

BACHELORARBEIT

IM STUDIENGANG AUDIOVISUELLE MEDIEN

BINAURALER TON FÜR VIRTUAL REALITY KONZERTE

AUFNAHME UND MISCHUNG VON POPULÄRE MUSIK FÜR 360°-VIDEOS

IM RAHMEN EINER INTERAKTIVEN VR-ANWENDUNG

VORGELEGT VON **JOHANNES MÄNNER** MATRIKEL-NR. 29366

AN DER **HOCHSCHULE DER MEDIEN STUTT GART**

EINGEREICHT AM 05.04.2019

ZUR ERLANGUNG DES AKADEMISCHEN GRADES

BACHELOR OF ENGINEERING

ERSTPRÜFER: **PROF. OLIVER CURDT**

ZWEITPRÜFER: **PROF. UWE SCHULZ**

Ehrenwörtliche Erklärung

„Hiermit versichere ich, Johannes Männer, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Binauraler Ton für Virtual Reality Konzerte – Aufnahme und Mischung von populärer Musik für 360°-Videos im Rahmen einer interaktiven VR-Anwendung“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.“

Johannes Männer

Stuttgart, den 5. April 2019

Kurzfassung

Die vorliegende Arbeit untersucht, anhand einer praktischen Umsetzung, Techniken in der Audioproduktion, welche bei der Entwicklung von Virtual Reality Konzerten auf Basis von 360°-Videos zum Einsatz kommen. Nach der theoretischen Aufarbeitung von Grundlagen, werden Besonderheiten und Beobachtungen bei der Aufnahme und Mischung zweier Elektro-Rock Konzerte der Stuttgarter Band „Eau Rouge“ dokumentiert. Der hierbei entstandene Prototyp einer interaktiven VR-Anwendung wird anschließend im Rahmen von Nutzertests evaluiert. Ziel ist es, den Mehrwert von binauralem, reaktivem Audio im Gegensatz zu herkömmlichem Stereo bei Virtual Reality Konzerten zu demonstrieren.

Abstract

This bachelor thesis examines the techniques of audio production used for virtual reality concerts in a 360°-video setting based on the reference of a practical implementation. Besides an explanation of theoretical foundations, the specialties and observations that occurred during the recording and mixing process of two Electronic Rock concerts of Stuttgart based band 'Eau Rouge' are documented and discussed. The prototype for an interactive VR-application which was developed in scope of this work will then be evaluated based on the results of multiple user tests. This bachelor thesis is aimed towards proving whether a considerable value is added to virtual reality concert experience by using omnidirectional and binaural sound, instead of conventional stereo.

Danksagung

Vielen Dank an Fabian Heidelberg für die Idee und die Leitung des Projektes. Weiterhin vielen Dank an Steffen Baalman für die Implementierung der Anwendung in der Unreal Spiele-Engine.

Ich möchte mich für den Beistand meiner Familie, sowie die Unterstützung meiner Kommilitonen Marcel Remy, Michael Dmoch und Jonathan Tenhunen bedanken.

Weiter danke ich den Mitgliedern der Band Eau Rouge, sowie den Verantwortlichen der beiden Konzertveranstaltungen im Kulturzentrum Merlin und im Club Schilli.

Besonderer Dank geht an Prof. Oliver Curdt für die Betreuung der Arbeit, sowie an Heiko Schulz und Matthias Adler für das Verleihen der notwendigen Technik.

Inhaltsverzeichnis

Ehrenwörtliche Erklärung	II
Kurzfassung	III
Abstract	III
Danksagung	IV
Abkürzungsverzeichnis	VIII
Abbildungsverzeichnis	IX
Formelverzeichnis	IX
1 Einleitung	1
2 Virtual Reality und 360°-Video	2
2.1 Definition: Virtual Reality	2
2.2 Definition: Immersion	3
2.3 Definition 360°-Video.....	4
2.4 Ein- und Ausgabetechniken	4
3 Räumliches Hören	6
3.1 Richtungswahrnehmung in horizontaler Ebene	7
3.2 Richtungswahrnehmung in vertikaler Ebene	8
3.3 Entfernungswahrnehmung	8
4 Binauraltechnik	9
4.1 Binaurale Aufnahme	10
4.2 Binaurale Synthese und Head-Tracking	10
5 Produktion von VR-Audio	12
5.1 Anforderungen.....	12
5.2 Audioformate für Virtual Reality und 360-Grad Video	12
5.2.1 Ambisonics.....	13
5.2.1.1 Grundlagen und Kodierung	14
5.2.1.2 Higher Order Ambisonics	15
5.2.1.3 Ambisonics-Formate	16
5.2.2 Objektbasiertes Audio	17
5.2.2.1 Idee und Funktionsweise	17
5.2.2.2 Vor- und Nachteile	18
5.3 Aufnahmetechniken für VR und 360-Grad Video	19
5.3.1 First-Order Ambisonics-Mikrofone	19
5.3.2 Higher-Order-Ambisonics-Mikrofone	20

5.3.3 Stereophone Mikrophonarrays	21
5.3.3.1 IRT-Kreuz	22
5.3.3.2 ORTF 3D.....	23
5.4 Softwaretools in der Postproduktion.....	24
5.4.1 Sennheiser Ambeo A-B Converter	24
5.4.2 Facebook 360 Spatial Workstation.....	25
5.4.3 IEM Plugin-Suite	27
6 Praktische Umsetzung der Virtual Reality Konzerte	29
6.1 Beschreibung der Anwendung.....	29
6.1.1 Anforderungen an den Ton	29
6.1.2 Die Band „Eau Rouge“	30
6.1.3 Der Song „Melt“.....	30
6.1.4 Kulturzentrum Merlin	31
6.1.5 Club Schilli.....	32
6.2 Besonderheiten bei der Aufnahme.....	32
6.2.1 Verwendete Technik.....	32
6.2.2 Mikrofonierung.....	32
6.2.3 Analyse der Aufnahmen	35
6.2.3.1 Hauptmikrofone im Publikum (Merlin).....	35
6.2.3.2 Band- und Bühnenmikrofone (Merlin):.....	36
6.2.3.3 Aufnahme im Club Schilli	37
6.3 Besonderheiten bei der Mischung.....	38
6.3.1 Produktionsformat	38
6.3.2 Stereo-Musikmischungen.....	38
6.3.2.1 Stereo-Musikmischung: Merlin	38
6.3.2.2 Stereo-Musikmischung: Schilli	40
6.3.2.3 Vergleich beider Stereomischungen	41
6.3.3 Ambisonics-Mischungen aller Kameraperspektiven	41
6.3.3.1 Merlin im Publikum	42
6.3.3.2 Merlin auf der Bühne	44
6.3.3.3 Schilli im Publikum und auf der Bühne	46
6.3.4 Dekodierung	46
6.3.5 Alternative Surround-Mischung	47
6.4 Implementierung des Tons in der VR-Anwendung	49
7 Benutzertests und Evaluation	50
7.1 Methode der Evaluation	50
7.1.1 Versuchsanordnung.....	50
7.1.2 Probanden	50
7.1.3 Verwendete Technik.....	51
7.1.4 Fragebogen.....	51
7.1.5 Mitsingoption	52
7.2 Auswertung und Diskussion des Fragebogens.....	53
7.2.1 Frage 12: Wie empfandest Du den Sound der Anwendung?	53
7.2.2 Frage 13: An welchem Ort hat Dir der Sound am besten gefallen?	53
7.2.3 Frage 15: Gib an, was Dir am besten gefallen hat.	54
7.2.4 Frage 16: Gib an, was Dir nicht gefallen hat.	54

7.2.5 Frage 17: Konntest Du einen Unterschied feststellen?	55
7.2.6 Frage 18: Welche Version hat Dir besser gefallen?	55
7.2.7 Frage 19a: Wieso hat dir Version A besser gefallen?	55
7.2.8 Frage 19b: Wieso hat dir Version B besser gefallen?	56
7.2.9 Zusammenfassung und mögliche Verbesserungen	56
8 Fazit und Ausblick.....	57
Literaturverzeichnis.....	X
Anhang A: Fragebogen	XIII
Anhang B: Evaluation des Fragebogens	XVI
Anhang C: Fotografien der Produktionen	XX
Anhang D: Datenträger.....	XXI

Abkürzungsverzeichnis

3DOF	3 Degrees of Freedom
6DOF	6 Degrees of Freedom
ACN	Ambisonic Channel Numbering
ADAT	Alesis Digital Audio Tape
AllRA	All-Round Ambisonics
AOÜF	Außenohrübertragungsfunktion
DAW	Digital Audio Workstation
dB	Dezibel
EBU	European Broadcasting Union
ESMA	Equal Segment Microphone Array
FOA	First Order Ambisonics
FuMa	Furse and Malham
HMD	Head-Mounted Display
HOA	Higher Order Ambisonics
HRIR	Head Related Impulse Response
HRTF	Head-Related Transfer Function
Hz	Hertz
IRT	Institut für Rundfunktechnik
ITU-R	International Telecommunication Union, Radiocommunication Sector
LUFS	Loudness Units Full Scale
ORTF	Office de Radiodiffusion Télévision Français
PA	Public Address
SN3D	Schmidt Semi-Normalisation
VR	Virtual Reality

Abbildungsverzeichnis

Abbildung 1: 6 Degrees of freedom. aus: Chaudhari.....	5
Abbildung 2: Kopfbezogenes Koordinatensystem nach Blauert	7
Abbildung 3: Richtungsbestimmende Bänder nach Blauert.....	8
Abbildung 4: Typische HRTFs für frontale bzw. seitliche Schalleinfallrichtung.....	11
Abbildung 5: Sphärische Harmonische 0. bis 4. Ordnung.....	16
Abbildung 6: Objektbasierte Audioproduktion und -wiedergabe im Überblick.....	17
Abbildung 7: Sennheiser Ambeo VR Mikrofon	20
Abbildung 8: mh acoustics Eigenmike®	21
Abbildung 9: "Rundum-Übertragung" mit IRT-Kreuz nach Wuttke.....	22
Abbildung 10 (links): ORTF-3D Mikrofonarray im Windkorb.....	23
Abbildung 11 (rechts): Orientierung eines X/Y Paares	23
Abbildung 12: Sennheiser Ambeo A-B Converter Plugin-Oberfläche.....	25
Abbildung 13: FB360 Spatializer Plugin	26
Abbildung 14: AllRADecoder Plugin.....	28
Abbildung 15: Spurplan der Konzerte	34
Abbildung 16: Saalplan des Konzertes im Merlin	34
Abbildung 17: Mixerkanäle in FL Studio	40
Abbildung 18: Matching EQ des Izotope Plugins auf dem Master-Kanal	41
Abbildung 19: Schema des Routings in der DAW	42
Abbildung 20: Einstellungen im Controll-Plugin	43
Abbildung 21: Mixertracks des Reaper-Projektes (Merlin Publikum)	44
Abbildung 22: räumliche Platzierung des Lead-Gesangs.....	45
Abbildung 23: Oktagon-Lautsprecher Layout im AllRADecoder.....	47
Abbildung 24: Surround-Panning des Lead-Gesangs in ReaSurround	48
Abbildung 25: Anordnung der Schallquellen in der Unreal-Engine.....	49
Abbildung 26: Ranking Ergebnisse aller Befragten.....	XVI
Abbildung 27: Ranking Ergebnisse (Audioprofis und -studierende).....	XVI
Abbildung 28: Positive Reaktionen auf den Ton der Anwendung (Version A)	XVII
Abbildung 29: Negative Reaktion auf den Ton der Anwendung (Version A)	XVIII
Abbildung 30: Warum hat Dir Version A besser gefallen?	XVIII
Abbildung 31: Warum hat Dir Version B besser gefallen?	XIX
Abbildung 32: Aufbau im Merlin	XX
Abbildung 33: Konzert in Ulm.....	XX

Formelverzeichnis

Formel 1.....	14
Formel 2.....	14
Formel 3.....	20

1 Einleitung

Die Idee, durch Einsatz modernster Technik Konzerte zu Hause im Wohnzimmer zu erleben und sich dabei so zu fühlen als stünde man inmitten einer feiernden Menge oder sogar neben der Band auf der Bühne, ist nicht neu und war Teil des großen Hypes, den Virtual Reality (kurz: VR) ca. 2016 auslöste. Dieser Hype scheint seinen Zenit längst überschritten zu haben. Einen vorläufigen Tiefpunkt erreichte die Technologie wohl im vergangenen Herbst, als das kalifornische Unternehmen *Jaunt*, welches einst das „Netflix des VR“ werden wollte, den Ausstieg aus dem Virtual Reality Geschäft bekannt gab.¹

Das erfolgreichste Produkt des Virtual Reality Pioniers war bis dahin ein Live-Konzert von Paul McCartney, welches mit stereoskopischen 360°-Kameras gefilmt wurde und mit unidirektionalem, reaktivem Ton ausgestattet war.^{2,3}

Die Telekom, sowie das Startup-Unternehmen *concertVR*, versuchen sich am Live-Streaming von 360°-Konzerten auf mobile Endgeräte. Allerdings funktioniert derzeit die VR-Funktion der *MagentaMusik 360* App schlichtweg nicht (laut zahlreichen Nutzerbewertungen auf Google Play⁴) und auch bei *concertVR* scheint es Schwierigkeiten in der Entwicklung zu geben, der letzte Facebook Post des Startups liegt schon neun Monate zurück.⁵

Das Angebot an 360°-Konzerten, das im Netz zu finden ist, ist momentan sehr überschaubar. Zudem leiden die meisten frei zugänglichen Aufnahmen an schwacher Bildqualität und verwenden einfachen Stereo-Sound.

Laut der Analysefirma IDC setzt sich der generelle Negativtrend im VR-Geschäft jedoch nicht länger fort. Der Absatz an VR-Brillen ist im dritten Quartal 2018 wieder angestiegen. Vor allem die auf externe Hardware angewiesenen Displays mit hoher Rechenleistung, wie die *Oculus Rift* oder *PlaystationVR*, konnten wieder zahlreich verkauft werden.⁶ Da diese Plattformen das Potential bieten, VR-Inhalte in exzellenter Qualität zu präsentieren, kann davon ausgegangen werden, dass auch die Produktion von VR-Konzerten wieder an Bedeutung gewinnt. Zudem unterstützen *Youtube* und *Facebook* mittlerweile 360°-Videos mit binauralem 3D-Sound, was ebenfalls darauf hoffen lässt, dass die Zukunft hochwertigere Konzert-Inhalte im Bereich 360°-Video und Virtual Reality bereithält.

Die Audioqualität spielt hierbei, wie auch in herkömmlichen Konzertproduktionen, eine sehr bedeutende Rolle. Die entscheidenden Stichworte sind 3D-Sound, Binauraltechnik und Head-Tracking. Remy untersucht in seiner Bachelorthesis⁷ bereits 3D-Audio für 360°-Videos im Rahmen klassischer Konzerte. Er betont dabei mehrfach, dass ein möglichst natürliches Klangbild angestrebt wird und bezeichnet den Einsatz von 3D-Audio als unverzichtbar.

¹ Vgl. Bastian (2018)

² Vgl. McDonald (2016) nach Lobeck (2017)

³ Vgl. Jaunt, Inc. (2014) nach James (2014)

⁴ Vgl. Google (2019)

⁵ Vgl. Concert VR (2018)

⁶ Vgl. IDC (2018)

⁷ Vgl. Remy (2018)

In dieser Arbeit wird die Produktion von Unterhaltungsmusik-Inhalten für Virtual Reality beleuchtet, welche mehr Freiheiten im Aufnahme- und Mixing-Prozess bietet. Dabei wurde der Ton für eine VR-Anwendung erstellt, die dem Nutzer die Möglichkeit gibt, sich zwei Konzerte einer Rockband anzusehen und dabei beliebig zwischen verschiedenen Kameraperspektiven zu wechseln.

Zunächst werden die Grundbegriffe Virtual Reality und 360°-Video erklärt, daraufhin folgt ein kurzer Überblick über das räumliche Hören und die Binauraltechnik. Anschließend werden gängige Audioformate und Aufnahmetechniken für 360°-Video vorgestellt, sowie eine engere Auswahl der Software, die zur Produktion für den Prototyp eingesetzt wurde. Im praktischen Teil dieser Arbeit werden alle wesentlichen Schritte der Aufnahme und Mischung dokumentiert, sowie ein Nutzertest ausgewertet. Es soll die Frage beantwortet werden, ob reaktiver, unidirektionaler Ton einen entscheidenden Mehrwert für Virtual Reality Konzerte im Bereich der Unterhaltungsmusik bieten und die Immersion solcher Anwendungen steigern kann.

2 Virtual Reality und 360°-Video

Die Begriffe Virtual Reality und 360°-Video sind teilweise nur schwer zu trennen, da es viele Gemeinsamkeiten, aber auch Unterschiede gibt.

Ein 360°-Video kann einerseits Bestandteil einer virtuellen Realität sein, muss andererseits aber nicht zwangsläufig im technischen Rahmen von Virtual Reality konsumiert werden. Eine virtuelle Realität kann wiederum, durch den Einsatz von computergenerierten Umgebungen, auch vollkommen ohne Nutzung von 360°-Videos umgesetzt werden.

2.1 Definition: Virtual Reality

Es sind etliche Definitionen und Charakterisierungen zu Thema Virtual Reality zu finden, die wenigsten davon decken alle Facetten von VR ab oder sind allgemeingültig. Der Brockhaus definiert Virtual Reality als:

eine von Computern und entsprechenden Programmen simulierte Welt, die dem Nutzer durch spezielle Techniken und Schnittstellen vermittelt wird.⁸

Lobeck unterscheidet außerdem zwei inhaltliche Ausrichtungen von VR. Neben der Neuerschaffung einer synthetischen, frei gestalteten Welt nennt er auch die Erzeugung des

⁸ Brockhaus Enzyklopädie

Abbilds der realen Welt mittels 360°-Videoaufnahmen als mögliche Darstellungsform der virtuellen Realität.⁹

Wenn der Nutzer nicht nur Beobachter in der simulierten Welt bleibt, sondern aktiv in das dortige Geschehen eingreifen kann, spricht man von interaktiver Virtuellen Realität. Dabei kommt es zwischen dem Nutzer und der virtuellen Welt zu einer direkten Interaktion.¹⁰ Im Idealfall, so beschreibt Brill, kann sich der Nutzer innerhalb der simulierten Realität wie in der bekannten realen Umgebung verhalten, vorausgesetzt die Reaktion des Systems auf Interaktion geschieht in Echtzeit.¹¹

Schlese¹² differenziert fünf Stufen von VR:

1. Der Nutzer befindet sich an einem realen Ort (z.B. auf dem Mars), welcher virtuell simuliert wird.
2. Der Nutzer befindet sich in einer rein erfundenen Welt, diese existiert nur in der Simulation.
3. Der Nutzer kann das Geschehen in der virtuellen Welt beeinflussen.
4. Reale Akteure, repräsentiert durch Avatare, interagieren miteinander in der virtuellen Welt.
5. Zusätzlich gibt es künstliche Intelligenzen, die als Avatare mit realen Akteuren handeln.

2.2 Definition: Immersion

Ein zentraler Begriff, der immer im direkten Zusammenhang mit Virtual Reality steht, ist die sog. Immersion. Brill bezeichnet als Immersion das Gefühl, welches eintritt, wenn sich der Nutzer in der virtuellen Umgebung befindet.¹³

Lobeck definiert Immersion als das Schlüsselwort in Virtual Reality:

*Immersive virtuelle Realität meint die Darstellung der virtuellen Welt, die so authentisch aussieht, sich anhört, sich anfühlt und riecht, dass unser Gehirn psychologisch getäuscht wird und es als real akzeptiert.*¹⁴

Laut Brill beruht Immersion auf Wahrnehmungen, die den visuellen, akustischen und manchmal auch den taktilen Eingabekanal des Menschen betreffen.¹⁵ In der Praxis wird dies durch moderne Ein- und Ausgabetechniken gewährleistet, deren Qualitätseigenschaften den Grad der Immersion mitbestimmen.

Die Intensität von Immersion kann allgemein stark variieren und hängt nicht alleinig von der eingesetzten Technik ab. Die Anzahl an Interaktionsmöglichkeiten, bzw. das Ausmaß der

⁹ Vgl. Lobeck (2017), S.17

¹⁰ Vgl. Engelmann (2018), S.22

¹¹ Vgl. Brill (2008), S.6-7

¹² Vgl. Schlese (2017), S.106

¹³ Brill (2008), S.6

¹⁴ Lobeck (2017), S.19

¹⁵ Brill (2008), S.13

Beeinflussbarkeit der virtuellen Realität durch den Nutzer, spielt dabei, neben dem Grad an Übereinstimmung zwischen virtueller und realer Umgebung, ebenfalls eine bedeutende Rolle.¹⁶

Bei einer Anwendung vollkommen ohne Interaktionsmöglichkeiten, beispielsweise bei einem Film oder linearen 360°-Video, ist somit die Intensität der Immersion eher gering, Engelmann spricht hier von einer passiven Immersion. Erlaubt die virtuelle Realität dem Nutzer jedoch eine direkte Interaktion mit der Umgebung, so wird eine höhere Intensität erreicht.¹⁷

2.3 Definition 360°-Video

Wie bei herkömmlichen Filmen und Videos handelt es sich auch bei 360°-Videos um lineare Medien. Das heißt der Nutzer ist nicht dazu in der Lage, in das gezeigte Geschehen einzugreifen, wie es z.B. bei Videospiele der Fall ist. Anders als im Kino jedoch befindet sich der Betrachter inmitten einer Szene und beobachtet diese nicht von außen. Dabei kann er selbst den Blickwinkel bestimmen, bzw. sich in alle erdenklichen Richtungen umsehen.

Im Gegensatz zur computergenerierten Welt, die in vielen Virtual Reality Anwendungen betrachtet wird, wird bei 360°-Videos das reale Abbild einer realen Situation mittels multipler Kameraobjektive erzeugt. Lobeck bezeichnet das 360°-Video als die einfachste Form von Virtual Reality.¹⁸

Allerdings können 360°-Videos auch auf gewöhnlichen Bildschirmen, wie z.B. Monitoren oder Handydisplays, abgespielt werden. Dadurch verlieren sie jedoch den Effekt der Immersion, der ein wesentlicher Bestandteil von Virtual Reality ist.

In Verbindung mit moderner Ausgabetechnik können 360°-Videos eine hohe Intensität an Immersion erreichen. Schließlich hat das Gezeigte einen hohen Grad an Übereinstimmung mit einer realen Umgebung (*siehe 2.2*), es ist ja ein fotorealistisches Abbild der realen Welt. Andererseits leidet die Immersion auch wiederum unter den fehlenden Interaktionsmöglichkeiten.

Die Frage „Sind 360°-Videos Virtual Reality?“ kann nur schwer beantwortet werden, hier unterscheiden sich auch die Meinungen in der Literatur. Zusammenfassend muss man sagen, dass der Begriff des 360°-Videos nicht eindeutig dem Bereich der virtuellen Realität zuzuordnen ist.

2.4 Ein- und Ausgabetechniken

Grundlage für eine gelungene Immersion sind, neben der Interaktivität, die eingesetzten Ein- und Ausgabegeräte.

¹⁶ Vgl. Brill (2008), S.6-7

¹⁷ Vgl. Engelmann (2018), S.22

¹⁸ Vgl. Lobeck (2017), S.17

Für etwa 70 Prozent unserer Wahrnehmungen ist das Auge verantwortlich, was die visuelle Wahrnehmung zum wichtigsten Eingabekanal den Menschen macht.¹⁹

Auf visueller Ebene werden zur Ausgabe in VR-Anwendungen meist Head-Mounted Displays (kurz: HMDs; umgangssprachlich: VR-Brillen) verwendet. Um die Bildgenerierung an den Blickpunkt des Nutzers in der virtuellen Umgebung anpassen zu können wird sog. Head-Tracking eingesetzt. Dabei werden die Position und Kopforientierung des Nutzers vom VR-System sensorisch erfasst. Die gewonnenen Daten werden verwendet um die virtuelle Umgebung aus der Nutzerperspektive darzustellen. Somit kann das Bild in Echtzeit auf Kopfbewegungen reagieren.²⁰

Bei der Registrierung von Bewegungen wird generell in zwei Fällen differenziert. Diese unterscheiden sich jeweils im Grad der Freiheit (DOF²¹). Bei einfachen 360°-Videos ist der Nutzer nur dazu in der Lage, seinen Kopf zu drehen („yaw“), zu neigen („pitch“) oder zu kippen („roll“). Dieser Grad an Bewegungsfreiheit wird als 3DOF bezeichnet. Bei 6DOF, einem Grad der Bewegungsfreiheit, der vor allem bei Virtual Reality Games angewendet wird, ist der Nutzer zusätzlich in der Lage, sich im dreidimensionalen Raum fortzubewegen. (links-rechts / vorne-hinten / oben-unten)²²

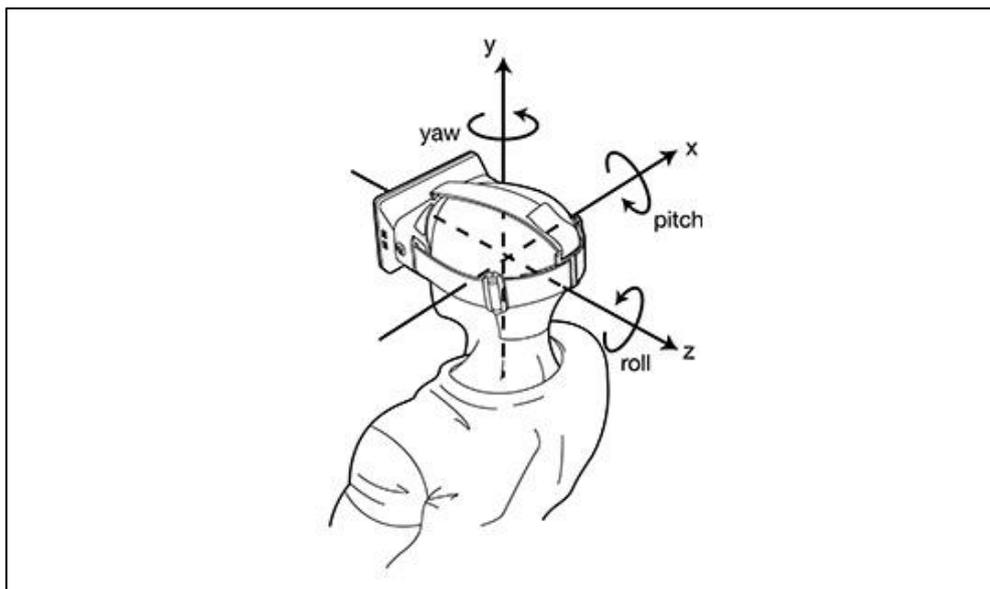


Abbildung 1: 6 Degrees of freedom. aus: Chaudhari²³

Derzeit ist ein breites Angebot an HMDs vorhanden. Die Vielfalt reicht dabei von einfachen „Low End“ Papphalterungen für Smartphones (z.B. das *Google Cardboard*) über mobile VR-Brillen (wie z.B. die *Oculus Go*) bis zu stationären „Full Feature-Systemen“ (z.B. die *HTC Vive*), welche an

¹⁹ Brill (2008), S. 13

²⁰ Vgl. Dörner, Jung et al. (2013), S.13-14

²¹ Degrees of Freedom

²² Vgl. Google Developers (2018)

²³ Bildquelle: Chaudhari (2007)

externe Hardware gekoppelt sind um eine höhere Grafikleistung zu bieten.²⁴ Je höher die Auflösung des Bildes und je Größer das Sichtfeld des jeweiligen HMDs ist, desto immersiver wirken die damit präsentierten virtuellen Welten.²⁵

Neben der sichtbaren Darstellung der Welt kann die Immersion von VR-Anwendungen durch akustische Reize erheblich verstärkt werden. Die Synchronisation von Bild und Ton ist dabei von besonders hoher Bedeutung.²⁶ In der Praxis werden für die Audiowiedergabe in VR-Anwendungen meist Kopfhörer oder Mehrkanalsysteme eingesetzt. Kopfhörer haben den zusätzlichen Vorteil, den Nutzer akustisch von der Außenwelt zu isolieren. Isolation von der realen Welt ist laut Schlese Voraussetzung für Immersion.²⁷

Für die Interaktion in der virtuellen Welt werden in der Regel Hilfstechniken eingesetzt, die Arme und Beine des Nutzers simulieren. Neben einfachen Handcontrollern, wie z.B. die der *HTC Vive*, deren Bewegungen ebenfalls getrackt und an das Programm gesendet werden, gibt es unzählige alternativen Techniken verschiedener Hersteller. Als Eingabegeräte dienen z.B. auch Armbänder, Handschuhe, komplette Anzüge oder am Körper verteilte Sensoren.²⁸

3 Räumliches Hören

Räumliches Hören spielt für Virtual Reality und 360°-Videos eine wichtige Rolle. Die visuelle Information bekommt der Nutzer nur über einen bestimmten Bildausschnitt, er kann aber nicht sehen, was sich hinter ihm befindet. Schall hingegen wird aus allen Richtungen gleichzeitig wahrgenommen und kann so die Erfahrung einer Welt, die den Nutzer umhüllt, noch verbessern. Schallereignisse können innerhalb der virtuellen Umgebung lokalisiert werden, zudem wird auch der virtuelle Raum durch Raumreflexionen akustisch beschrieben.

Richtungs- und Entfernungswahrnehmung bilden zusammen die räumliche Wahrnehmung des Gehörs, was auch die Wahrnehmung der räumlichen Ausdehnung einer Schallquelle miteinschließt. Dickreiter beschreibt die Räumlichkeit einer Schallquelle nicht als ihre tatsächliche Ausdehnung, sondern als die Empfindung, dass der Raum um die Schallquelle mit Schall erfüllt ist.²⁹

Ein kopfbezogenes Polarkoordinatensystem (siehe Abbildung 2) wird verwendet, um die Richtung eines Hörereignisses zu beschreiben. Dabei befindet sich dessen Ursprung genau zwischen beiden Gehörkanaleingängen. Die Horizontalebene wird durch die interaurale Achse und die Unterkanten der Augenhöhlen aufgespannt, orthogonal dazu verläuft die Frontalebene. Die vertikale- oder

²⁴ Vgl. Haug, Pagel (2015), S.80-84

²⁵ Vgl. Dörner et al. (2013), S.14

²⁶ Vgl. Brill (2008), S.25

²⁷ Vgl. Schlese (2015)

²⁸ Vgl. Haug, Pagel (2015), S.93

²⁹ Dickreiter; Goeres-Petri (2014), S.127

Medianebene trennt den Kopf in zwei Hälften. Mittels den Polarkoordinaten *Azimus*, *Elevation* und *Entfernung* kann eine Schallquelle im dreidimensionalen Raum dargestellt werden.³⁰

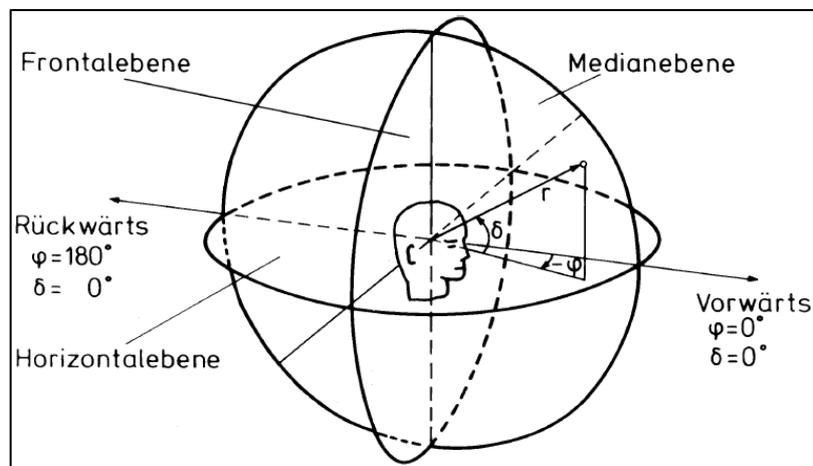


Abbildung 2: Kopfbezogenes Koordinatensystem nach Blauert³¹

3.1 Richtungswahrnehmung in horizontaler Ebene

Sobald eine Schallquelle in horizontaler Ebene aus der Mitte (0°) wandert, werden unterschiedliche Signale an beiden Ohren registriert. Es entstehen interaurale Laufzeitdifferenzen und frequenzabhängige interaurale Pegeldifferenzen. Die Lokalisierungsschärfe einer Schallquelle hängt dabei stark von der Art des Signals ab. Einzelne Sinusschwingungen sind deutlich schwerer zu lokalisieren als breitbandige, impulshaltige Signale.³²

Interaurale Laufzeitdifferenzen

Auf Grund des Abstands der Ohren von ca. 17cm kommt es zu einem winkelabhängigen Schallwegunterschied. Der kleinste wahrnehmbare Unterschied im Schallweg von nur 1cm entspricht einer interauralen Laufzeitdifferenz von etwa 0,03ms. In diesem Fall wird das Hörereignis ca. $3^\circ - 5^\circ$ außerhalb der Mitte wahrgenommen.³³

Interaurale Pegeldifferenzen

Für die Lokalisierung von Schallquellen spielen interaurale Pegeldifferenzen, welche durch Abschattung von Kopf und Ohrmuschel entstehen, eine untergeordnete Rolle. Sie sind frequenzabhängig und existieren nur oberhalb von etwa 300Hz.³⁴

³⁰ Vgl. Blauert; Braasch (2008), S. 87f.

³¹ Bildquelle: Blauert; Braasch (2008), S.88

³² Vgl. Dickreiter, Goeres-Petri (2014), S.129-131

³³ Vgl. Ebenda, S.129

³⁴ Vgl. Ebenda, S.130

3.2 Richtungswahrnehmung in vertikaler Ebene

In vertikaler Ebene gibt es keine interauralen Pegel- und Laufzeitdifferenzen. Entscheidend für eine Lokalisation in der Höhe sind Klangfarbenunterschiede, welche unterbewusst wahrgenommen werden und durch die Form und Beschaffenheit von Kopf und Ohren entstehen. Die Wahrnehmung solcher feinen Unterschiede funktioniert nur bei breitbandigen, dem Hörer bekannten Signalen. Unbekannte Signale oder Sinustöne sind in der Medianebene nicht lokalisierbar. Die Schärfe der Lokalisation ist allerdings auch bei gut bekannten Signalen mit nur 10° immer noch sehr gering.³⁵

Blauert untersuchte 1969 diese richtungsabhängigen Klangfarbenunterschiede und kam zu dem Ergebnis, dass bestimmte Frequenzbänder je nach vertikalem Einfallswinkel angehoben werden.³⁶

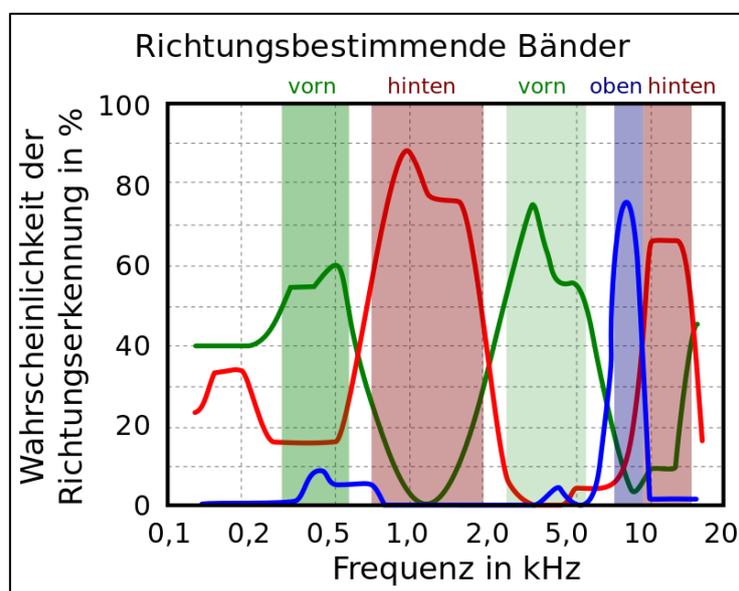


Abbildung 3: Richtungsbestimmende Bänder nach Blauert³⁷

3.3 Entfernungswahrnehmung

Je nach Entfernung ändern sich Lautstärke und Spektrum des Hörereignisses. Die Bekanntheit der Schallquelle ist dabei für deren Entfernungswahrnehmung essenziell. Nur wenn das Ohr die Schallquelle bereits in einer bestimmten Entfernung kennt lässt sich abschätzen, wie sich das Hörereignis bei zunehmender Entfernung verändert.³⁸

Blauert & Braasch unterscheiden Merkmale der Entfernungswahrnehmung in vier verschiedenen Bereichen:³⁹

³⁵ Vgl. Dickreiter; Goeres-Petri, S.131f.

³⁶ Vgl. Blauert, Braasch (2008), S.94f.

³⁷ Bildquelle: Wikipedia (2017)

³⁸ Vgl. Dickreiter; Goeres-Petri (2014), S.132

³⁹ Vgl. Blauert, Braasch (2008), S.98f

Sehr geringe Entfernungen (≤ 25 cm)

Ändert sich der Abstand einer sehr nah am Kopf auftretenden Schallquelle zu den Ohren, so ändern sich auch monaurale und interaurale Übertragungsfunktionen, d.h. Spektrum und Pegel des Hörereignis werden stark beeinflusst. Der Signalpegel wächst bei sinkendem Abstand.

Mittlere Entfernungen (ca. 25cm – 15 m)

Bei Entfernungsverdopplung nimmt der Lautstärkepegel um 6 dB ab (bei allseitiger Abstrahlung). Natürliche Schallquellen verändern bei zunehmender Entfernung ihre Klangfarbe, da tiefe Frequenzanteile schneller an Pegel verlieren, wie gerichtete hohe Frequenzanteile. Aufgrund der Kurven gleicher Lautstärke ändert sich der Klang ebenfalls bei abfallendem Pegel. Nicht zuletzt spielt das Verhältnis aus Direktschall zu reflektiertem Schall eine Rolle bei der Entfernungswahrnehmung (Je weiter eine Schallquelle entfernt ist, desto größer wird der Diffusschallanteil).

Große Entfernungen (> 15 m)

Bei größeren Entfernungen werden hohe Frequenzen in der Luft absorbiert, daher klingt z.B. entferntes Donnerrollen dumpf. Laut Blauert liegt der akustische Horizont allerdings bei etwa 15m, d.h. alle Schallereignisse aus größerer Entfernung werden aufgrund von Erfahrungswerten nur abgeschätzt.

Grenzfall: Im-Kopf-Lokalisierung

Kann ein Signal keiner natürlichen Schallquelle außerhalb des Kopfes zugeordnet werden, kommt es zur sog. Im-Kopf-Lokalisation. Das geschieht vor allem bei der Wiedergabe durch Kopfhörer, da die Signale hier direkt am Ohr anliegen und somit keine typischen Klangfarbenänderungen, welche durch Außenohrübertragungsfunktionen entstehen, vorweisen. In diesem Fall entsteht das Klangbild einer Szene im Kopf zwischen den Ohren.^{40,41}

4 Binauraltechnik

Die gängigste Art der Audiowiedergabe bei Virtual Reality Inhalten oder 360°-Videos ist die über Kopfhörer. Um Schallquellen dabei räumlich ortbar zu machen und die Im-Kopf-Lokalisierung zu verhindern, wird die Binauraltechnik angewandt. Hierbei wird der Vorteil ausgenutzt, dass beide Stereosignale L R unter Kopfhörern akustisch getrennt sind, d.h. jedes Ohr erhält nur ein diskretes Signal und es kommt nicht zu Übersprechen wie bei der Lautsprecherstereophonie. Entsprechen diese Signale den realen Ohrsignalen einer realen Szene, so entsteht beim Hörer die Illusion, genau an dieser Szene teilzunehmen.

⁴⁰ Vgl. Blauert, Braasch (2008), S.98f

⁴¹ Vgl. Dickreiter; Goeres-Petri (2014), S.133f.

„Da die Trommelfelle [...] unser einziges Tor zur auditiven Welt sind, leistet eine Synthese des Schallfeldes vor dem Trommelfell im Idealfall eine komplette Rekonstruktion unserer akustischen Umgebung.“⁴²

4.1 Binaurale Aufnahme

Bei der binauralen Aufnahme werden die Ohrsignale mittels zwei Druckempfänger-Kapseln aufgenommen. Die Mikrofone werden dabei im Gehörgang einer echten Person oder eines nachgebauten Kunstkopfes platziert. Das Schallsignal wird hierbei je nach Einfallrichtung durch die Außenohrübertragungsfunktion (AOÜF oder HRTF⁴³) spezifisch gefiltert (siehe Abbildung 4). Dabei haben Kopf- und Ohrmuscheln durch Abschattung, Beugung, Verzögerung, Resonanzen und Reflexionen den größten Einfluss auf die HRTF. HRTFs sind für jeden Menschen individuell, was zu Lokalisationsproblemen bei binauralen Aufnahmen, welche nicht am eigenen Gehörgang entstanden sind, führt.⁴⁴

Kopfbezogene Impulsantworten (HRIRs⁴⁵) werden im reflexionsarmen Raum binaural aufgenommen und können durch Fouriertransformation in den Frequenzbereich überführt werden, wodurch wiederum HRTFs entstehen. Werden die HRIRs in reflexionsreichen Räumen aufgezeichnet, spricht man von binauralen Raumimpulsantworten, welche später in der Synthese von binauralem Faltungshall gebrauch finden.⁴⁶

Da HRTFs stark richtungsabhängig sind, werden für die Erstellung von HRIR-Datenbanken HRIRs aus beliebig vielen Schalleinfallrichtungen gemessen, zwischen den Ergebnissen kann dann interpoliert werden, sodass für jede erdenkliche Schalleinfallrichtung eine HRTF bekannt ist. In der Praxis ist das aber nur schwer individuell umsetzbar, daher werden für die Erstellung solcher Datensätze „Durchschnittsköpfe“ verwendet.⁴⁷ Ein weiteres Problem, welches zu Ungenauigkeit führt, ist die Wahl der Lautsprecher, die zur Messung eingesetzt werden. Schließlich hat jeder Lautsprecher eine frequenzabhängige Richtcharakteristik und kann so nicht jedes natürliche Schallereignis simulieren.⁴⁸

4.2 Binaurale Synthese und Head-Tracking

Die binaurale Aufnahme mittels eines Kunstkopfes ist statisch an eine Perspektive gekoppelt. Für die Anwendung bei 360°-Videos oder interaktiven VR Anwendungen sind Kunstkopfaufnahmen ungeeignet, da sie keine Klangänderung bei Kopfbewegungen realisieren können. Viele Klangszene werden darüber hinaus künstlich erstellt. Dabei gibt es viele Einzelsignale, die virtuell im Raum verteilt werden, z.B. bei Videospiele. Auch in diesem Fall musste eine Alternative zur Kunstkopfaufnahme gefunden werden.

⁴² Slavik; Weinzierl (2008), S.671

⁴³ Head-related transfer function

⁴⁴ Vgl. Weinzierl (2008), S. 586f.

⁴⁵ Head-related impulse responses

⁴⁶ Vgl. Slavik; Weinzierl (2008), S.673

⁴⁷ Vgl. Weinzierl (2008), S.587

⁴⁸ Vgl. Slavik, Weinzierl (2008), S.674

Bei der binauralen Synthese kommen zuvor ermittelte HRTFs zum Einsatz, die zur Filterung beliebiger Signale angewandt werden. So erhalten einzelne Schallquellen durch Faltung eine kopfbezogene räumliche Information, was beim Hörer wiederum die Illusion einer realen Szene bewirkt. Die Simulation des virtuellen Raumes geschieht durch die zusätzliche Faltung mittels binauraler Raumimpulsantworten, welche ebenfalls zuvor gemessen oder mittels einer raumakustischen Computersimulation errechnet worden sind.⁴⁹

Durch Head-Tracking kann die aktuelle Kopfposition des Hörers ermittelt werden, was für die Erfahrung in der virtuellen Realität essenziell ist. Bei einer Kopfbewegung müssen nun in Echtzeit die zur Faltung gebrauchten Impulsantworten aktualisiert werden. Entscheidend für eine gelungene Binauralsynthese sind hier kurze Latenzzeiten von unter 60ms.⁵⁰

Neben der nun gegebenen Interaktivität hat die Anwendung von Head-Tracking den positiven Effekt, dass durch Peilbewegungen des Kopfes Vorne-Hinten-Vertauschungen verhindert werden. Solche Vertauschungen treten bei statischen Kunstkopfaufnahmen aufgrund der fehlenden interauralen Laufzeit- und Pegeldifferenzen bei Schalleinfall aus frontaler oder rückseitiger Richtung auf.⁵¹

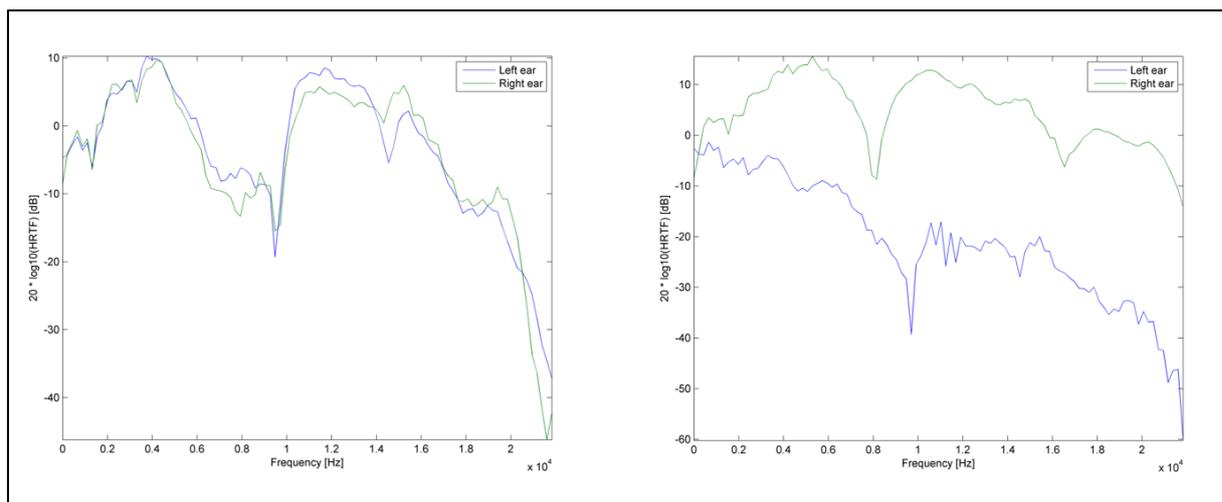


Abbildung 4: Typische HRTFs für frontale bzw. seitliche Schalleinfallrichtung⁵²

⁴⁹ Vgl. Slavik; Weinzierl (2008), S.674ff

⁵⁰ Vgl. Ebenda

⁵¹ Vgl. Slavik; Weinzierl (2008), S.674ff

⁵² Bildquelle: Woźniak (2015)

5 Produktion von VR-Audio

In diesem Kapitel werden zunächst die Anforderungen an VR-Audio erläutert. Anschließend werden gängige Audioformate und Aufnahmetechniken für VR-inhalte zusammengefasst und eine engere Auswahl an Möglichkeiten in der Audiopostproduktion vorgestellt.

5.1 Anforderungen

Eine hohe Audioqualität in Virtual Reality ist nach Schütze⁵³ nicht nur vorteilhaft, sondern absolut ausschlaggebend für eine Immersive Erfahrung. Er betont dabei, dass nicht der absolute Realismus in der simulierten Welt angestrebt wird, sondern eine unterhaltsame und packende Nutzererfahrung. Seinem Ideal nach folgt der Ton in den „New Realities“ dem Vorbild im Spielfilmton, bei dem die Realität signifikant übertrieben und manipuliert wird.

Eine Gemeinsamkeit fast aller VR-Anwendungen ist die binaurale Wiedergabe über Kopfhörer. Schallquellen in der virtuellen Welt sollen räumlich dargestellt werden, mit richtungsunabhängiger, gleichbleibend hoher Qualität. Wichtig dabei ist eine dynamische Anpassung des Klangbildes an die Kopfbewegungen des Nutzers, was eine nahezu latenzfreie Binauralsynthese mit präzisiertem Head-Tracking voraussetzt.⁵⁴

Die genaueren Anforderungen an VR-Audio variieren je nach Anwendung stark. Generell wird zwischen linearen und interaktiven Anwendungen, sowie zwischen den verschiedenen Graden an Bewegungsfreiheit unterschieden. Die Vertonung eines linearen 360-Grad Videos ähnelt der eines Spielfilm noch eher als die eines komplexen interaktiven VR-Games. Darüber hinaus muss beachtet werden, ob der Nutzer nur den Blickwinkel ändern (3DOF), oder sich in der virtuellen Welt fortbewegen kann (6DOF).

Remy beschreibt in seiner Bachelorarbeit die Wiedergabe von Hörsignalen als zwingend notwendig, um ein natürliches und realistisches Hörerlebnis zu schaffen.⁵⁵ Solche Hörsignale können in anderen Fällen, in welchen kein natürliches Klangbild angestrebt wird, wiederum nicht zweckdienlich sein. Beispielsweise können realitätsgetreue, diffuse Hallsignale laut Schütze für die meisten Hörer als irrelevant eingestuft werden.⁵⁶ Hier muss der/die Tonschaffende entscheiden, ob signifikante Signale in der Medianebene dargestellt werden sollen, oder ob auf den Einsatz von Höhenkanälen verzichtet wird.

5.2 Audioformate für Virtual Reality und 360-Grad Video

Das wichtigste Format für 360°-Videos ist das bereits in den 1970er Jahren entwickelte szenenbasierte Ambisonics. Es wird von den meisten gängigen Plattformen für 360-Grad Videoinhalte, wie z.B. *Youtube* oder *Facebook*, unterstützt. Auch bei Abwandlungen des

⁵³ Vgl. Schütze; Irwin-Schütze (2018), S.5f.

⁵⁴ Vgl. Shivappa u.a. (2016), S.3

⁵⁵ Vgl. Remy (2018), S.17

⁵⁶ Vgl. Schütze; Irwin-Schütze (2018), S.6

Ambisonics Formats, wie z.B. dem *Two Big Ears*-Format von *Facebook* handelt es sich um schallfeldbasiertes Audio.⁵⁷

Die meisten interaktiven Virtual Reality Anwendungen und Spiele basieren auf Spiele-Engines, wie z.B. *Unreal* oder *Unity*, welche objektbasiertes Audio einsetzen.

Sehr selten kommen kanalbasierte Formate zum Einsatz, da hier die Wiedergabe in der Regel über Lautsprecher erfolgt, was wiederum der Immersion schadet. Bei der Wiedergabe über Kopfhörer passt sich das Klangbild nicht an Kopfbewegungen an, es sei denn die Einzelnen Lautsprecher signale werden als Objekte in der virtuellen Welt aufgestellt und anschließend binauralisiert. Dann handelt es sich allerdings wieder um objektbasiertes Audio.

Das für die Zukunft eher unbedeutende, quad-binaurale Format ist eine Mischform aus schallfeld- und kanalbasiertem Audio. Dabei werden vier binaural aufgezeichnete Stereopaare verwendet, welche je nach Blickwinkel ineinander überblendet werden.⁵⁸

Im praktischen Teil dieser Arbeit wurden Ambisonics sowie objektorientiertes Audio eingesetzt. Da die anderen Formate in Virtual Reality Anwendungen heute nur noch selten Gebrauch finden, werden diese hier nicht weiter behandelt.

5.2.1 Ambisonics

Ambisonics ist ein Verfahren zur Übertragung und Synthese von dreidimensionalen Schallfeldern. Dabei hängt dessen Genauigkeit von der Anzahl der Übertragungskanäle bei der Aufnahme (bzw. Synthese) und der Anzahl der Wiedergabekanäle (Lautsprecher bzw. virtuelle Lautsprecherobjekte) ab.⁵⁹

Da Ambisonics Schall kugelförmig beschreibt, d.h. Klangquellen aus allen Richtungen projiziert, ist das Format hervorragend für 360°-Videos geeignet. Zudem ist die Wiedergabe kanalunabhängig und kann durch Dekodierung direkt in ein beliebiges Lautsprechersystem, oder sogar in ein dynamisches, binaurales Signal für Kopfhörer übersetzt werden.⁶⁰

Bereits in den 1970er Jahren von Michael Gerzon entwickelt, konnte sich das Format zunächst nicht am Tonträgermarkt behaupten. Mit der Einführung von Ambisonics höherer Ordnung (HOA⁶¹), welche eine höhere Genauigkeit ermöglicht, erlebt das Format aktuell eine Renaissance.⁶²

⁵⁷ Vgl. Rieger (2019a)

⁵⁸ Vgl. Ebenda

⁵⁹ Vgl. Slavik; Weinzierl (2008), S.659

⁶⁰ Vgl. Rieger (2019b)

⁶¹ Higher Order Ambisonics

⁶² Vgl. Slavik; Weinzierl (2008), S.659f.

5.2.1.1 Grundlagen und Kodierung

Ambisonics beschreibt Schallfeldverteilungen, gemessen am Ort des Hörers. Dieses Schallfeld kann durch Kugelflächenfunktionen zerlegt werden, ähnlich wie die Fourier-Transformation ein periodisches Signal zerlegt.⁶³

Es gilt die Fourier-Bessel-Reihe mit den Kugelkoordinaten der Hörerperspektive (Radius r , Azimuth ϕ und Elevation δ):⁶⁴

$$p(r) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\phi, \delta)$$

Formel 1

Das Schallfeld $p(r)$ ergibt sich aus der Überlagerung von sog. sphärischen Harmonischen der Ordnung m (mit dem Faktor der Komponenten B_{mn}^{σ} multiplizierte Funktionen Y_{mn}^{σ}).

Diese werden mit sphärischen Besselfunktionen $j_m^{(kr)}$ und einem Phasenfaktor i^m jeweils radial gewichtet.⁶⁵ Da die Kugelkoordinaten in dieser Rechnung an die Hörerperspektive angepasst werden können, ist eine einfache Rotation der gesamten Szene möglich, was einen wesentlichen Vorteil, im Gegensatz zu kanalbasierten Formaten, bietet.

Das Ambisonics-Verfahren überträgt die Komponenten B_{mn}^{σ} der Fourier-Bessel-Reihe zur Synthese, bzw. Resynthese eines Schallfeldes. Die Genauigkeit dieser Kodierung hängt dabei von der Anzahl der übertragenen Komponenten, also von der Länge der Reihe, ab.

Die ersten drei Komponenten beschreiben den Schalldruck im Ursprung (B_{00}^1) und die Druck- bzw. Schnellekomponenten in den drei Raumrichtungen (B_{11}^1 , B_{11}^{-1} und B_{10}^1).

In der Fachterminologie werden diese vier Signale als W, X, Y, Z bezeichnet und bilden das sog. B-Format.⁶⁶

Das B-Format (oder auch FOA⁶⁷) benötigt somit vier PCM-Übertragungskanäle. Die Anzahl der Übertragungskanäle N , die zur Resynthese eines Schallfeldes mit der Ordnungszahl m benötigt werden, lässt sich einfach errechnen.⁶⁸

Es gilt:

$$N = (m+1)^2$$

Formel 2

⁶³ Vgl. Gerzon (1973) nach Villegas (2015), S.2

⁶⁴ Slavik; Weinzierl (2008), S.660

⁶⁵ Vgl. Slavik; Weinzierl (2008), S.660f.

⁶⁶ Vgl. Slavik; Weinzierl (2008), S.660f.

⁶⁷ First Order Ambisonics

⁶⁸ Slavik; Weinzierl (2008), S.662

Bei der Dekodierung werden die gewonnenen Komponenten durch eine Überlagerung von Lautsprecher signalen für einen zentralen Hörerpunkt resynthetisiert, vorausgesetzt das aufgenommene Schallfeld ist eine ebene Welle und die Wiedergabelautsprecher strahlen ebene Wellen ab. Sind die Orte der Lautsprecher und die Ambisonics-Komponenten bekannt, so können in einem Gleichungssystem die gesuchten Signale der Lautsprecher ermittelt werden. Wenn die Lautsprecheranzahl nicht mit der Komponentenanzahl übereinstimmt, werden sog. pseudoinverse Matrizen zur Signalermittlung eingesetzt, die die beste Näherung für ein nicht exakt lösbares Gleichungssystem liefern.⁶⁹

Für die binaurale Wiedergabe von Ambisonics Schallfeldern wird zunächst ein beliebiges virtuelles Lautsprecher setup dekodiert. Daraus kann anschließend per Faltung mit HRIRs das zweikanalige Kopfhörersignal gerendert werden.⁷⁰

5.2.1.2 Higher Order Ambisonics

Ambisonics-Signale ab der dritten Ordnung werden Higher Order Ambisonics genannt. Höhere Ordnungen der sphärischen Harmonischen sind gleichbedeutend mit einer steigenden Anzahl an Komponenten und somit auch Übertragungskanälen. Ambisonics-Signale dritter Ordnung benötigen für die Kodierung 16 Übertragungskanäle, bei Signalen fünfter Ordnung sind es schon 36 (siehe *Formel 2*). Da es sich bei der Fourier-Bessel-Reihe um eine Reihe handelt, sind alle HOA⁷¹Formate, durch einfaches Nichtberücksichtigen der Komponenten höherer Ordnung, abwärts kompatibel.

Mit steigender Anzahl an Komponenten steigt sowohl die Genauigkeit der Annäherung an das reale Schallfeld, als auch die Größe des Sweetspots an der Hörposition.⁷² Während das Schallfeld in erster Ordnung noch sehr diffus ist, kommen bei höheren Ordnungen immer feinere Auflösungen zustande.

Das lässt sich in etwa mit der Anzahl an verwendeten Lautsprechern bei Surround Systemen vergleichen. In 360°-Videos steigt die Lokalisationsschärfe von Schallquellen mit dem Einsatz von HOA.

Die sphärischen Harmonischen in Abbildung 5 zeigen deren winkelabhängige Amplitude und können als Richtcharakteristiken von Mikrofonen verstanden werden.⁷³ Ambisonics nullter Ordnung enthält demnach keine Richtungsinformation, bei der ersten Ordnung kommen Signale von Mikrofonen mit Achtercharakteristik dazu, welche alle drei Raumrichtungen abdecken. Mit zunehmender Ordnung gibt es (im bildlichen Sinne) immer mehr stärker gerichtete Mikrofone und somit auch mehr Richtungsinformation.

⁶⁹ Vgl. Slavik; Weinzierl (2008), S.663f.

⁷⁰ Vgl. Dieterle u.a. (2018)

⁷¹ Higher Order Ambisonics

⁷² Vgl. Shivappa u.a. (2016), S. 4

⁷³ Vgl. Remy (2018), S. 24

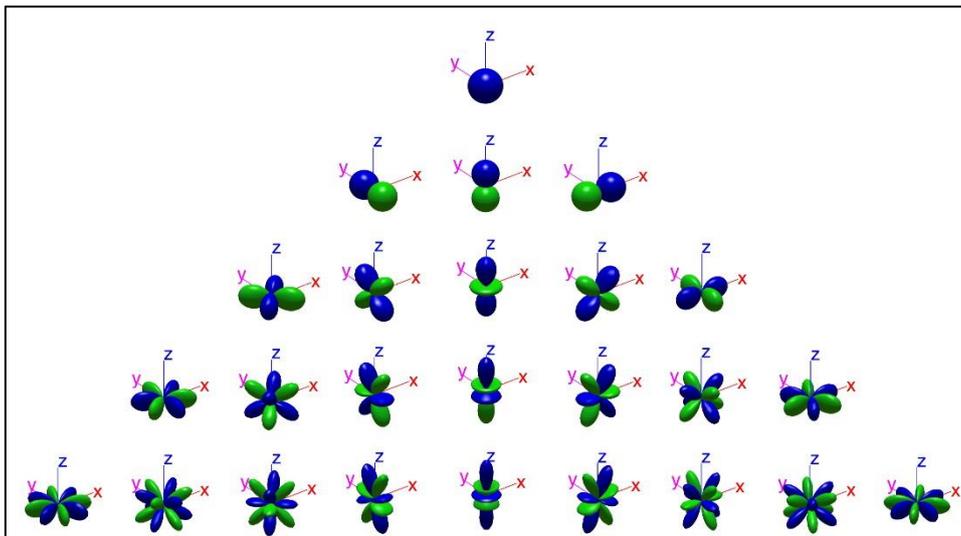


Abbildung 5: Sphärische Harmonische 0. bis 4. Ordnung⁷⁴

5.2.1.3 Ambisonics-Formate

Die verschiedenen Formate von Ambisonics unterscheiden sich in Reihenfolge und Gewichtung der Übertragungskanäle. Bei der Verwendung von Ambisonics Softwaretools und Wiedergabeplattformen muss genau darauf geachtet werden, welche Formate unterstützt und aktuell berechnet werden. Andernfalls kann es schnell passieren, dass Wiedergabefehler aufgrund von Inkompatibilität aufkommen.

Die rohen Aufzeichnungen der vier Kanäle eines FOA-Tetraedermikrofons bilden das sogenannte *A-Format*, welches dann in das für die Decodierung relevante *B-Format* konvertiert werden muss.⁷⁵

Derzeit gibt es verschiedene Ausrichtungen für B-Formate und Higher Order Ambisonics mit jeweils unterschiedlicher Kanalreihenfolge: Das traditionelle *FuMa* nach Furse und Malham, sowie das *ACN* Format sind die gängigsten.⁷⁶

Auch verschiedene Gewichtungen der Kanäle kommen zu Einsatz, die Normalisierungsmethoden heißen *maxN*, *SN3D* und *N3D*.⁷⁷

Das weit verbreitete Format *ambiX* verwendet eine Kanalreihenfolge nach *ACN* und eine *SN3D* Normalisierung.⁷⁸

⁷⁴ Bildquelle: Politis (2018)

⁷⁵ Vgl. Sennheiser (2019a)

⁷⁶ Vgl. Facebook (2019)

⁷⁷ Vgl. Malham (2003), S.2

⁷⁸ Vgl. Facebook (2019)

5.2.2 Objektbasiertes Audio

Aufgrund der Starrheit von kanalbasiertem Audio kommt dieses, wie bereits in 5.2 erwähnt, für professionelle Vertonung von VR-Anwendungen und 360-Grad Videos nicht in Frage.

Bei 6DOF VR-Anwendungen stößt allerdings auch szenenorientiertes Audio schnell an seine Grenzen. Der Sweetspot eines Ambisonics Systems beschränkt sich auf die Kopfposition des Hörers, diese kann zwar rotiert werden, jedoch reagiert das Tonsignal nicht auf eine Positionsänderung des Hörers, was vor allem bei interaktiven VR-Games verlangt wird.

Der objektorientierte Ansatz, der beispielsweise in Game-Engines, wie *Unreal* oder *Unity* verfolgt wird, antwortet auf dieses Problem.

5.2.2.1 Idee und Funktionsweise

Komplexe Klangszene werden als Gesamtheit einer Sammlung einzelner Elemente betrachtet. Bei diesen Objekten handelt es sich um monophone Audiosignale, die mittels in Metadaten verpackte Parameter beschrieben werden. Diese Metadaten beinhalten die künstlerische Intention des Tonschaffenden und ermöglichen eine Wiedergabe dessen, unabhängig vom Zielsystem.⁷⁹

Abbildung 6 veranschaulicht die einzelnen Schritte der objektbasierten Audioproduktion bis hin zur Wiedergabe auf der Seite des Konsumenten.

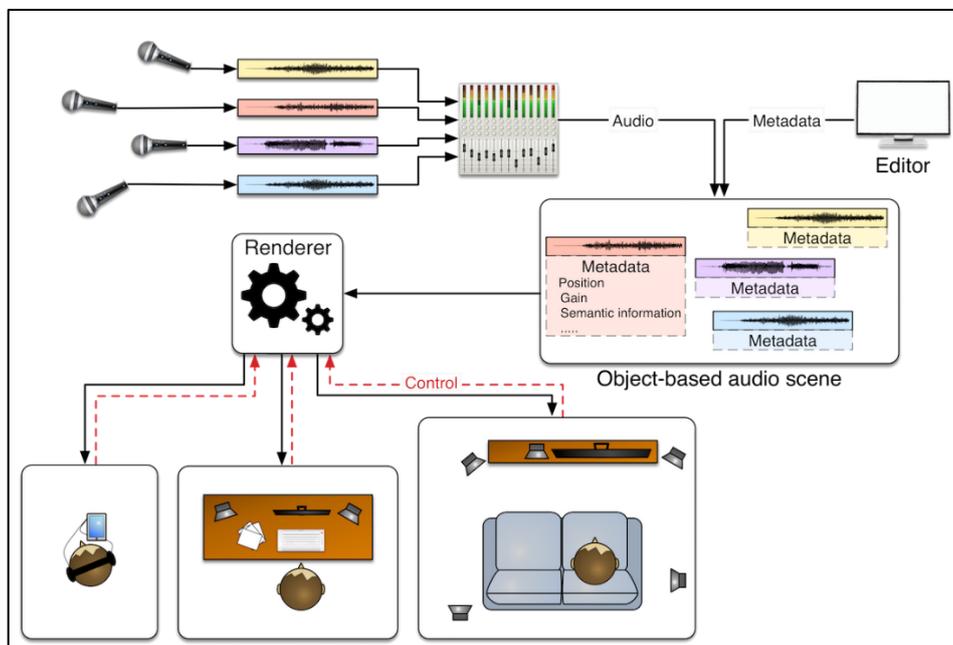


Abbildung 6: Objektbasierte Audioproduktion und -wiedergabe im Überblick⁸⁰

⁷⁹ Vgl. Altman u.a. (2016), S.3

⁸⁰ Bildquelle: Meier; Weitnauer (2018)

Monophone Audiosignale werden nach der Aufnahme bzw. Synthese zunächst auf herkömmliche Art klanglich nachbearbeitet. Parameter wie Lautstärke, Position und Abstrahlverhalten der Klangobjekte werden in den nun hinzukommenden Metadaten beschrieben. Im Gegensatz zum ursprünglichen Soundfile sind die Metadaten flexibel und reagieren auf das Programm. So ändern sich bei VR-Anwendungen durch Head-Tracking z.B. die Positionsparameter des Klangobjektes.⁸¹

Im Rendering-Prozess werden aus der entstandenen Klanglandschaft wiederum diskrete Lautsprecher-signale generiert. Dabei kann jedes beliebige Lautsprechersystem, das dem Renderer bekannt ist, verwendet werden. Dieser Prozess ist dynamisch, da sich Metadaten, vor allem durch Interaktion und Head-Tracking, fortlaufend verändern.⁸² Für die binaurale Wiedergabe kann ein virtuelles Lautsprechersystem zum Einsatz kommen. Die einzelnen Objekte können aber auch individuell mittels HRTF-Faltung binauralisiert werden.⁸³

5.2.2.2 Vor- und Nachteile

Objektbasiertes Audio hat den Vorteil, je nach Rendering Algorithmus, eine hohe Genauigkeit in der Lokalisationsschärfe der einzelnen Schallquellen zu erzielen. Außerdem ist es nicht an eine Hörerposition gebunden und somit sehr gut für 6DOF Anwendungen geeignet.⁸⁴

Allerdings geht die Menge der erforderlichen Daten zur Beschreibung einer Szene mit der Menge an Objekten einher. Damit verbunden sind ein hoher Rechenaufwand und eine hohe Anforderung an Bandbreite bei der Übertragung. Aus diesem Grund sind komplexe Szenen für Streaming oder die Anwendung auf Mobilgeräten noch schwer umsetzbar.⁸⁵

Ebenfalls problematisch bei der Produktion von objektbasierten Klangszenen ist akustisches Übersprechen.⁸⁶ Gerade bei Musikproduktionen führt das zu Schwierigkeiten, da einzelne Instrumente, die nicht vollständig isoliert abgenommen wurden, nicht mehr eindeutig lokalisierbar sind. Beispielsweise führt das Übersprechen einer Gitarre auf das Gesangsmikrofon zu einer Fehllokalisierung: Die Gitarre wird nun zusätzlich aus der Richtung des Sängers wahrgenommen.

⁸¹ Vgl. Meier; Weitnauer (2018)

⁸² Ebenda

⁸³ Vgl. Shivappa u.a. (2016), S.4

⁸⁴ Vgl. Altman u.a. (2016), S. 3f.

⁸⁵ Vgl. Shivappa u.a. (2016), S.4

⁸⁶ Ebenda

5.3 Aufnahmetechniken für VR und 360-Grad Video

Laut Dieterle besteht eine Szene für Virtual Reality (gemeint sind computergenerierte Anwendungen) aus Objektsounds, Effekten und Atmo.⁸⁷ Bei der Aufnahme dieser einzelnen Elemente muss, wie schon erwähnt, Übersprechen weitestgehend verhindert werden. In der Praxis werden viele Sounds computergeneriert oder stammen aus Foley-Aufnahmen im Studio. Hinzu kommen oft nicht-diegetische Elemente, wie z.B. Sprechertexte oder Hintergrundmusik, welche nicht auf Head-Tracking reagieren.⁸⁸

Bei szenenbasiertem Audio steht die Hörposition fest und kann nicht verändert werden (3DOF). Daher kann die komplette Szene mit einem Mikrofonsystem an dieser Position aufgenommen werden. Zu vergleichen ist diese Methode mit dem Einsatz der Hauptmikrofone bei Aufnahmen von klassischer Musik. Die einzelnen Objekt-Sounds kann man als Stützmikrofone verstehen.⁸⁹

Es gibt zahlreiche Arten dieser Hauptmikrofone, die sich in der Anzahl der Mikrofonkapseln, deren Charakteristik und Ausrichtung, sowie der Komplexität unterscheiden. Grundsätzlich sind der Kreativität bei der Entwicklung solcher Arrays keine Grenzen gesetzt. Im Folgenden soll eine engere Auswahl an Hauptmikrofonsystemen vorgestellt werden.

Wichtig ist dabei zu wissen, dass die verwendeten Systeme unabhängig vom Produktionsformat des Anwendungstons sind. Beispielsweise können die vier Signale eines IRT-Kreuzes problemlos in der Ambisonics Umgebung gemischt werden. Ebenso können Aufnahmen, welche mit einem Ambisonics Mikrofon Array durchgeführt worden sind, auch in jedes beliebige Zielformat konvertiert werden.

5.3.1 First-Order Ambisonics-Mikrofone

First-Order Ambisonics-Mikrofone (auch: Soundfield-Mikrofone) zielen darauf ab, das Schallfeld an einem möglichst kleinen Punkt im Ambisonics B-Format zu beschreiben. Das kann mittels der vier eingesetzten Mikrofonkapseln direkt, oder nach Summen- und Differenzbildung der Signale, ähnlich wie beim MS-Verfahren, geschehen.⁹⁰

Es wird zwischen B-Format-Mikrofonen (1x Kugel, 3x Acht), Triple-MS-Mikrofonen (2x Niere, 2x Acht) und Tetraedermikrofonen (4x Niere) unterschieden.⁹¹

Letztere sind in der Praxis am bedeutsamsten und erhielten ihren Namen aufgrund der vier Kapseln mit Nierencharakteristik, welche in Form eines Tetraeders angeordnet sind. Die 4 Signale *Left Front Up LFU*, *Right Front Down RFD*, *Left Back Down LBD* und *Right Back Up RBU* bilden das sog. A-Format. Um daraus die vier Komponenten *W*, *X*, *Y*, *Z* für das bei Ambisonics relevante B-Format zu erzeugen wird folgende Matrizierung angewandt.⁹²

⁸⁷ Vgl. Dieterle (2018)

⁸⁸ Vgl. Altman u. a. (2016), S.5

⁸⁹ Vgl. Dieterle (2018)

⁹⁰ Vgl. Weinzierl (2008), S.591-593

⁹¹ Vgl. Dieterle (2018)

⁹² Vgl. Weinzierl (2008), S. 592f.

Es gilt:

$$W = LFU + LBD + RFD + RBU$$

$$X = LFU - LBD + RFD - RBU$$

$$Y = LFU + LBD - RFD - RBU$$

$$Z = LFU - LBD - RFD + RBU$$

Formel 3

Bekannte Beispiele für FOA-Mikrofone sind das *Sennheiser Ambeo* oder das *Rode NT-SF1*. Sie sind aufgrund ihrer Kompaktheit gut für die Produktion von 360°-Videos geeignet, gleichzeitig sind sie jedoch sehr ungenau in der räumlichen Auflösung, was mit der hohen Kanalkohärenz zusammenhängt.⁹³ Im Bereich der Musikproduktion konnten sie sich trotz einer hohen Flexibilität, was die Richtcharakteristik angeht, nicht durchsetzen.⁹⁴



Abbildung 7: Sennheiser Ambeo VR Mikrofon⁹⁵

5.3.2 Higher-Order-Ambisonics-Mikrofone

Aufnahmen in HOA basieren auf Mikrofonarrays mit einer Vielzahl an Kapseln, welche auf der Oberfläche einer Kugel angebracht sind. Das *mhacoustics: Eigenmike* (siehe Abbildung 8) besitzt 32 multidirektionale Elektret-Kapseln, verteilt auf einer Kugel von 8.4cm Durchmesser. Das gemessene Schallfeld wird mittels Beamforming wiederum in sphärische Harmonische zerlegt. Mit dem *Eigenmike* sind Ambisonics-Aufnahmen der vierten Ordnung möglich.⁹⁶

Laut Bates stellt es dabei als eine große Herausforderung dar, ein hohes Level der Klangfarbe von HOA-Mikrofonarrays zu erreichen. In seinen Tests schneidet das *mhacoustics: Eigenmike* zwar in der räumlichen Auflösung besser als die getesteten FOA-Mikrofone ab, jedoch verliert es Qualitätspunkte im Klang.⁹⁷ Camerer erklärt, dass diese Einbußen mit der mangelnden Qualität der Elektret-Mikrofone zusammenhängen. Er betont außerdem, dass der mangelnde Abstand

⁹³ Vgl. Lee (2016), S. 2

⁹⁴ Vgl. Weinzierl (2008), S.593

⁹⁵ Bildquelle: Sennheiser (2019a)

⁹⁶ Vgl. mh acoustics LLC (a)

⁹⁷ Vgl. Bates (2017)

zwischen den Kapseln bei HOA-Mikrofonen zu hoher Signalkorrelation führt, vor allem im tieffrequenten Bereich, was sich wiederum negativ auf die Klangqualität auswirkt.⁹⁸

Weitere Beispiele für HOA-Mikrofone sind das achtkanalige *CoreSoundsOctomic*, welches Ambisonics in 2. Ordnung aufnimmt⁹⁹ und das mit einer 360°-Kamera kombinierte *VisiSonic 5/64*, welches über 64 Kapseln verfügt und damit Ambisonics in 7. Ordnung ermöglicht.¹⁰⁰



Abbildung 8: mh acoustics Eigenmike¹⁰¹

5.3.3 Stereophone Mikrofonarrays

Gerade bei Musikaufnahmen wird ein ausgewogener Klang, der frei von Artefakten ist, welche bei HOA-Aufnahmen durch zu hohe Korrelation der Signale auftreten¹⁰², verlangt. Um die erforderliche Klangqualität zu erreichen, werden hochwertige Mikrofonkapseln eingesetzt, welche durch unterschiedliche Ausrichtungen und Abstände zu stark individuellen Signalen führen. Diese Signale werden anschließend diskret auf feststehende Lautsprecher setups übertragen. Aufgrund von Pegel- und Laufzeitunterschieden entstehen hierbei Phantomschallquellen zwischen den Lautsprechern.¹⁰³ Die Lokalisationsschärfe dieser Phantomschallquellen ist von der Anzahl der diskreten Kanäle, also Lautsprechersignalen und der Signaltrennung abhängig. Bei VR-Anwendungen und 360°-Videos werden die einzelnen Lautsprechersignale wiederum als Objekte um die Hörerposition aufgestellt und anschließend binauralisiert. Voraussetzung hierbei ist, dass der Mittelpunkt des Mikrofonarrays bei der Aufnahme möglichst mit der Position der 360°Kamera übereinstimmt. In der Praxis werden solche Arrays meist oberhalb der Kamera platziert.

Typische Beispiele stereophoner Hauptmikrofone sind:

- A/B, X/Y, ORTF (für zweikanal Stereo)
- OCT Surround, ORTF Surround, IRT-Kreuz, Decca Tree (für Surround Sound)
- ORTF 3D, ESMA 3D, 2L-Cube (für Lautsprecher setups mit Höhenebene, z.B. Dolby Atmos oder Auro3D)

⁹⁸ Vgl. Camerer (2018)

⁹⁹ Vgl. Core Sound LLC (2018)

¹⁰⁰ Vgl. VisiSonics Corporation (2014)

¹⁰¹ Bildquelle: mh acoustics LLC (b)

¹⁰² Vgl. Wittek u.a. (2007) nach Wittek & Theile (2017)

¹⁰³ Vgl. Wittek; Theile (2017)

Stellvertretend für Mikrofonarrays, die auf dem Prinzip von Laufzeit- und Pegeldifferenzen basieren, werden hier zwei dieser Anordnungen exemplarisch vorgestellt.

5.3.3.1 IRT-Kreuz

Beim IRT-Kreuz handelt es sich um 4 Mikrofone mit Nierencharakteristik, die an den Eckpunkten eines Quadrates mit ca. 25cm Kantenlänge angeordnet sind. Dabei sind alle Mikrofone jeweils um 90° zueinander versetzt ausgerichtet, sprich sie zeigen vom Mittelpunkt des Kreuzes weg. In der Wiedergabe werden die vier resultierenden Signale auf 4 Lautsprecher, welche ebenfalls im Quadrat angeordnet sind, übertragen. Somit ergeben sich je nach horizontalem Einfallswinkel einer Schallquelle verschiedene Laufzeit- und Pegelunterschiede, wodurch die Lokalisierung der Schallquelle bei der Wiedergabe ermöglicht wird. Befindet sich der Hörer am Sweetspot, so stimmt seine Richtungswahrnehmung der Schallquellen mit der realen Szene überein.¹⁰⁴

Im Gegensatz zur herkömmlichen 5.1 Lautsprecheranordnung, ist die Lokalisationsschärfe beim IRT-Kreuz in allen Richtungen gleich, da kein Center-Lautsprecher zu Einsatz kommt und der Öffnungswinkel zwischen den Lautsprechern konstant 90° beträgt. Das macht diese Aufnahmetechnik interessant für 360°-Videos, da es hier keine bevorzugte Richtung gibt. Laut Wuttke führt das IRT-Kreuz bei Rundum-Übertragungen (siehe Abbildung 9) zu einem guten Ergebnis. Bei klassischen Musikaufnahmen findet es allerdings seltener Verwendung als Hauptmikrofon, da hier der wichtigste Schallanteil von vorne einfällt und ein zusätzlicher Centerkanal zu einem besseren Ergebnis führt. Das IRT-Kreuz wird bei Konzerten eher als Ergänzung für die Aufnahme einer den Hörer umhüllenden Publikums-Atmo eingesetzt.¹⁰⁵

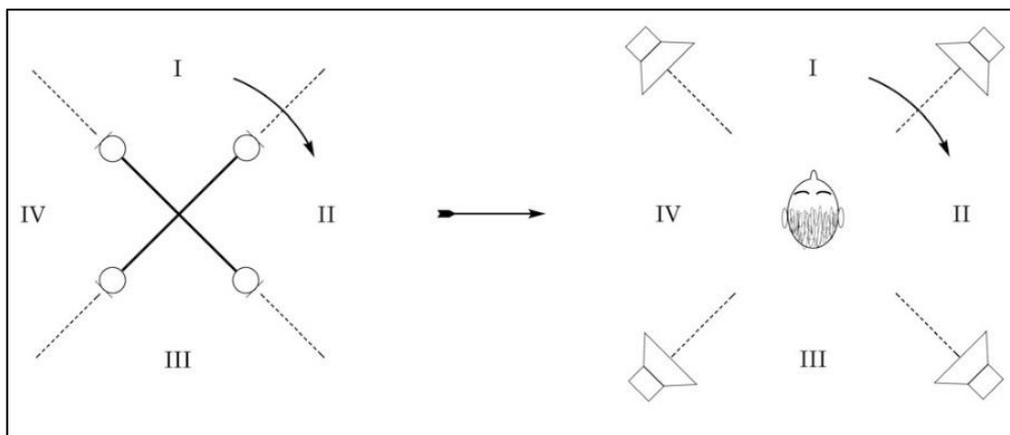


Abbildung 9: "Rundum-Übertragung" mit IRT-Kreuz nach Wuttke¹⁰⁶

¹⁰⁴ Vgl. Wuttke (2002)

¹⁰⁵ Vgl. Wuttke (2002)

¹⁰⁶ Bildquelle: Ebenda

5.3.3.2 ORTF 3D

Im Gegensatz zum IRT-Kreuz, gibt das von Wittek und Theile entwickelte ORTF-3D Array (Abbildung 10) eine Höhenebene wieder und dient somit als dreidimensionales Aufnahmeverfahren. Es besteht aus insgesamt 8 Mikrofonen mit Supernierencharakteristik, welche zu vier X/Y Paaren zusammengefasst werden. Die Paare sind in den Ecken eines Rechteckes mit Kantenlängen von 10cm x 20cm angeordnet.¹⁰⁷ Auf horizontaler Ebene funktioniert das System also ähnlich, wie das IRT-Kreuz, die Abstände und Öffnungswinkel der Stereopaare sind jedoch nicht alle gleich, sondern betragen 100° bzw 80°. Die X/Y Paare sind ebenfalls nicht symmetrisch zur horizontalen Ebene gekippt und haben einen Öffnungswinkel von 90° (siehe Abbildung 11).

Laut Wittek und Theile führt das Verfahren zu optimaler Diffusfeld-Dekorrelation mit hoher Signaltrennung und guter 3D-Abbildung.¹⁰⁸ Durch die kompakte Bauweise stellt es ein praktisches System zur Aufnahme von 360°-Inhalten da. Die acht Signale werden beim Einsatz in VR-Anwendungen auf virtuelle Lautsprecher in zwei Ebenen geroutet. Dabei ergeben die Lautsprecher die Eckpunkte eines Würfels und werden anschließend binauralisiert.¹⁰⁹ Je nach Anforderung kann eine 3D-Aufnahme der Szene für VR einen bedeutenden Mehrwert, im Gegensatz zu einem rein horizontalen Klangbild, führen. Lee kritisiert allerdings, dass der nicht identische Aufnahmewinkel zwischen den Stereopaaren beim ORTF-3D System, zu einer ungleichmäßigen Lokalisierung führt.¹¹⁰

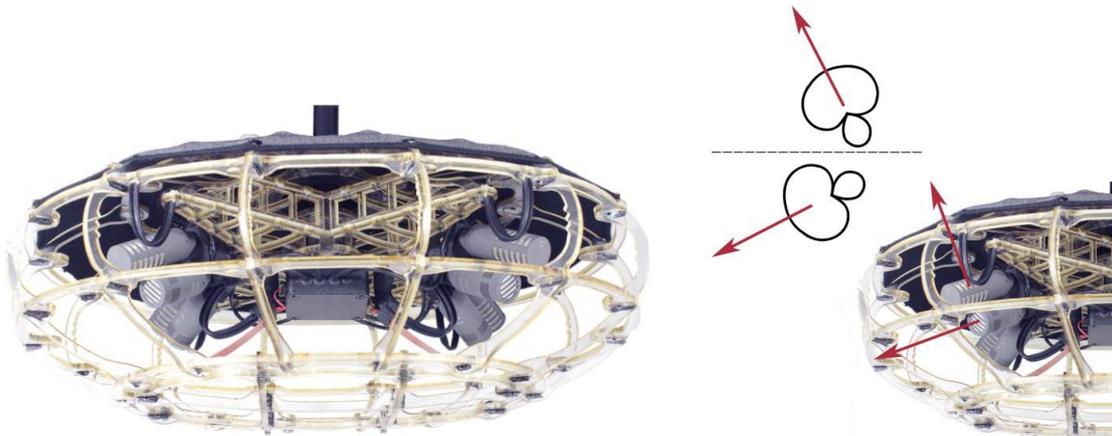


Abbildung 10 (links): ORTF-3D Mikrofonarray im Windkorb¹¹¹

Abbildung 11 (rechts): Orientierung eines X/Y Paares¹¹²

¹⁰⁷ Vgl. Wittek; Theile (2017), S.5f.

¹⁰⁸ Vgl. Wittek; Theile (2017), S. 6

¹⁰⁹ Vgl. Wittek; Theile (2017), S.5

¹¹⁰ Vgl. Lee (2016), S. 2

¹¹¹ Bildquelle: Wittek; Theile (2017)

¹¹² Bildquelle: Ebenda

5.4 Softwaretools in der Postproduktion

Wie auch bei der Vertonung von herkömmlichem, linearem Filmmaterial, empfiehlt sich bei der Mischung von 360°-Videos ein bildbezogenes Arbeiten. Die zuvor aufgenommenen Mono-, Stereo- und Ambisonic-Files werden in der Postproduktion nun zu einem gesamten, räumlichen Mix „gelayert“. In der Praxis geschieht dies mittels eines 3D-Panners, der die Signale im virtuellen Raum verteilt.¹¹³

Solche 3D-Panner haben, neben der Berücksichtigung der Medianebene, weitere wesentliche Vorteile gegenüber Surround-Pannern. Sie sind kanalunabhängig und daher sehr flexibel. Die resultierende Tonmischung im Ambisonics-Format kann in jedes gewünschte Zielformat überführt werden, je nachdem auf welcher Zielplattform das 360°-Video abgespielt werden soll. Außerdem kann neben der Richtung einer Schallquelle auch deren Entfernung definiert werden. Je näher sich die Schallquelle an der Hörerposition befindet, desto lauter wird sie dann abgespielt. Die Art der Positionierung einer Schallquelle erinnert stark an das objektbasierte Audio bei der Entwicklung von 3D-Games.¹¹⁴ Eine Rektangulärprojektion des 360°-Videos bietet dabei die Grundlage für die räumliche Verteilung der Signale.

Die Auswahl an Plugins zur Erstellung von Ambisonics-Mischungen ist groß und wächst stetig. Unterschiede gibt es nicht nur im Preis, sondern auch in der unterstützten Ambisonics-Ordnung und im Umfang der Einstellungsmöglichkeiten. Nicht alle Plugins bieten die Möglichkeit einer bildbezogenen Mischung, es können allerdings einzelne Tools der Workstations miteinander kombiniert werden. Knpfer stellt in seiner Masterarbeit¹¹⁵ eine große Auswahl verschiedener Software tabellarisch gegenüber. Im Folgenden sollen die Plugins vorgestellt werden, welche im Praktischen Teil dieser Arbeit zum Einsatz kamen.

5.4.1 Sennheiser Ambeo A-B Converter¹¹⁶

Aufnahmen von FOA-Mikrofonen, wie dem *Sennheiser Ambeo*, müssen zunächst in das B-Format konvertiert werden, welches direkt kompatibel zu den 360°-Videoplayern von *Facebook* und *Youtube* ist. Das *A-B Converter* Plugin unterstützt sowohl das FuMa- als auch das ambiX-Format und bietet Einstellungen zur Rotation der Szene, was praktisch ist, wenn die Mikrofonausrichtung nicht mit der Kameraausrichtung übereinstimmt. Außerdem kann ein Korrektur- sowie ein Hochpassfilter geschaltet werden. Das Plugin ist kostenlos und für alle gängigen Systeme verfügbar.

¹¹³ Vgl. Schütze; Irwin-Schütze (2018), S.182ff.

¹¹⁴ Vgl. Ebenda, S.187

¹¹⁵ Vg. Knpfer (2018), S.36f.

¹¹⁶ Vgl. Sennheiser (2019b)

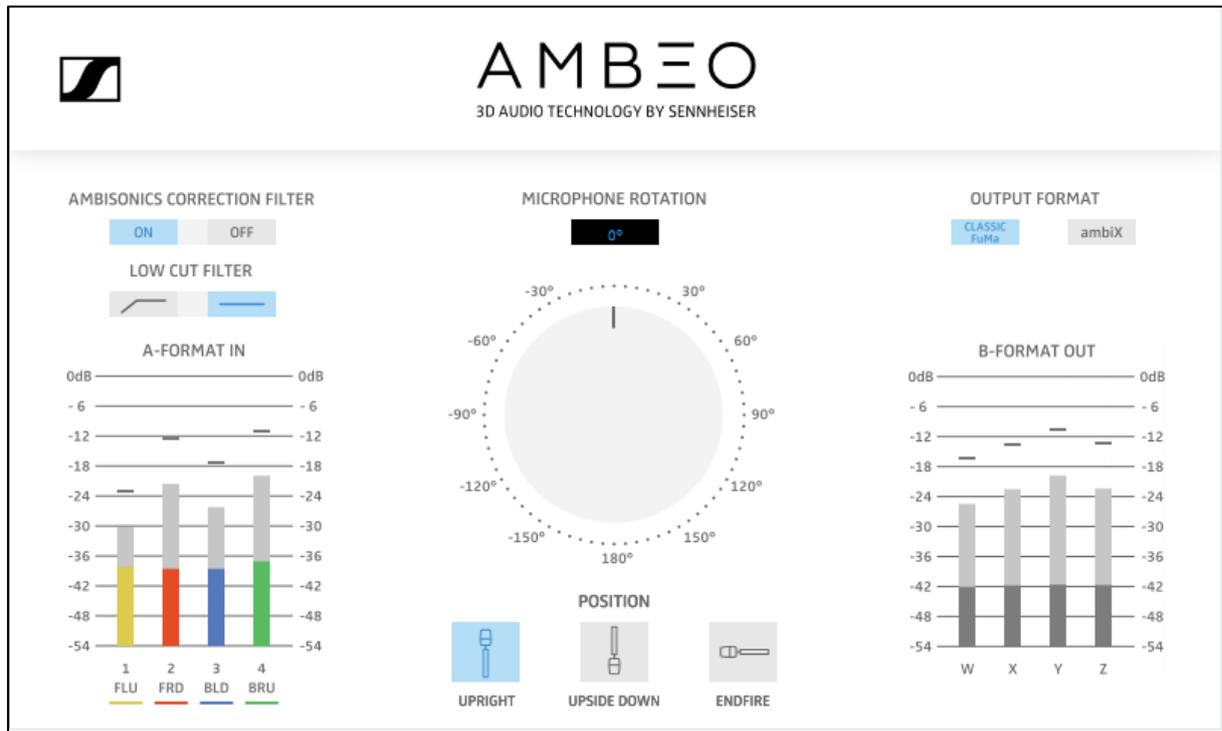


Abbildung 12: Sennheiser Ambeo A-B Converter Plugin-Oberfläche¹¹⁷

5.4.2 Facebook 360 Spatial Workstation¹¹⁸

Die *FB360 Spatial Workstation* besteht aus mehreren Plugins für DAWs¹¹⁹ zur Erstellung von Ambisonics-Mischungen, einem VR-Videooplayer und einer Engine zur Kodierung von 360°-Videos. Die Workstation unterstützt in der aktuellen Version 3.3.1 Ambisonics bis zur dritten Ordnung und ist frei erhältlich. Die Plugins sind mit den DAWs *Avid Pro Tools HD*, *Steinberg Nuendo* & *Cockos Reaper* kompatibel, bei der Mischung in der dritten Ordnung müssen die entsprechenden Ambisonics-Busse über 16 Kanäle verfügen.

Zentrales Element der *FB360 Workstation* ist der *FB360 Spatializer* (Abbildung 13), ein 3D-Panner mit Videoreferenz, welcher sowohl Mono-, als auch Mehrkanal- und Ambisonicsquellen als Eingangssignale akzeptiert. Mittels der Parameter *Position*, *Attenuation*, *Directionality* und *Doppler*, können die Eigenschaften jeder Schallquelle definiert werden. Dabei funktioniert die Positionierung direkt im Videofenster, welches die Rektangluar-Darstellung eines sphärischen Videos zeigt. Zusätzlich kann definiert werden, wie weit eine Schallquelle entfernt ist und wie sich die Entfernung auf die Dämpfung des Hörereignisses auswirkt. Es wird außerdem definiert, ob es sich um eine Punktschallquelle oder eine breite Fläche handelt und in welche Richtungen der Schall abgestrahlt wird. Durch hinzuschalten des Doppler-Effekts werden entfernte Schallereignisse zum Hörer zeitlich verzögert.

Eine weitere Besonderheit der *FB360 Workstation* ist die Simulation von Raumreflexionen. Im Einzelnen werden die Reflexionen für jedes Signal berechnet, nachdem deren Grad und Pegel im

¹¹⁷ Bildquelle: Sennheiser (2019a)

¹¹⁸ Vgl. Facebook (2019)

¹¹⁹ Digital Audio Workstation

Spatializer definiert wurde. Die Eigenschaften des virtuellen Raumes werden im separaten FB360 *Control* Plugin festgelegt. Dieses Plugin wird einmal im Projekt geladen und sorgt gleichzeitig für die Binauralisierung der Ambisonics-Mischung. So kann jederzeit der finale Mix auf Kopfhörern vorgehört werden.

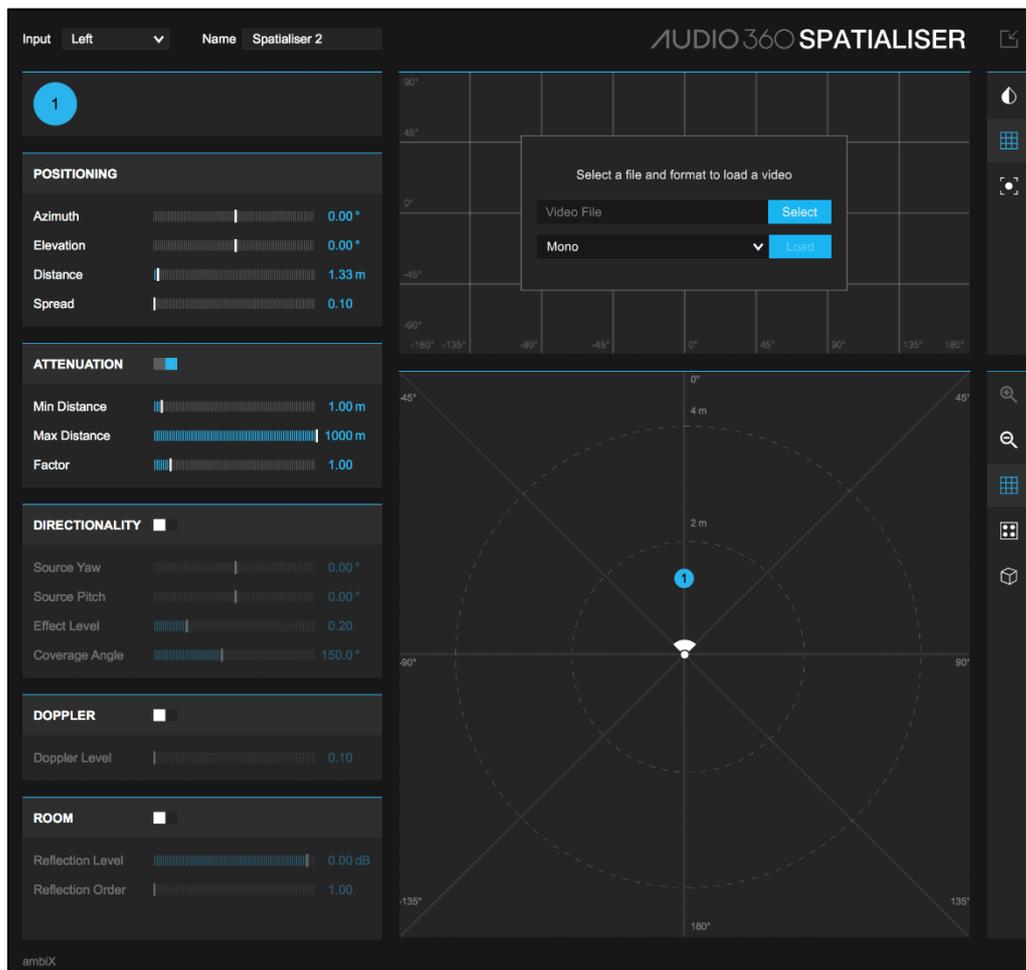


Abbildung 13: FB360 Spatializer Plugin

Dabei passt sich die Ausrichtung des Mixes an den Blickwinkel des Videos an. Das Video, welches im *Spatializer* integriert wird, öffnet sich automatisch im *FB360 Player* und kann über ein HMD mit Head-Tracking-System, oder auf dem Bildschirm per Mausbedienung abgespielt und kontrolliert werden. Objekte, die sich im Video bewegen, können mittels Automationen auch im Ambisonic-Mix bewegt werden. Dieser Vorgang kann völlig autonom per „Object Tracking“ erfolgen.

Das *FB360 Converter* Plugin dient dazu, vorhandene Ambisonics-Signale zu rotieren. Mit dem *FB360 Loudness Meter* lässt sich die Lautheit im Mix messen, davon ausgehend, dass sich der Nutzer immer in die lauteste Richtung orientiert. Zur Messung der LUFs kommt ein Logarithmus, wie ihn die EBU-128-Norm empfiehlt, zum Einsatz.

Der *FB360 Encoder* ist eine standalone App und dient zur Integration von Ambisonics-Files in die entsprechenden 360°-Videos. Als Zielformat kann *Facebook Video*, *Youtube Video* und *Rift:Oculus*

Video gewählt werden, *Youtube* funktioniert derzeit allerdings nur mit einem FOA-Mix. Nach Auswahl der Audio- und Videodaten werden die benötigten Metadaten erstellt, die zur Wiedergabe auf den entsprechenden Plattformen benötigt werden.

5.4.3 IEM Plugin-Suite¹²⁰

Die am *Institute of Electronic Music and Acoustics* entwickelte Plugin-Suite bietet eine sehr große Auswahl an Tools zur Erstellung und Mischung von Ambisonics-Signalen bis hin zur 7. Ordnung. In diesem Extremfall werden alle möglichen 64 Kanäle eines Audiobusses in *Reaper* benötigt (siehe *Formel 2*). Derzeit gibt es noch keine Möglichkeit ein Video mit in den Bearbeitungsprozess mit einzuschließen, es kann also nur ohne Bildbezug gemischt werden.

Mit dem *Stereo Encoder* lassen sich Mono- und Stereofiles im Raum verteilen, ähnlich wie beim *FB360 Spatializer*. Der *RoomEncoder* ist für die Berechnung der Reflexionen in Abhängigkeit von der Hörerposition zuständig, dabei ist der Doppler-Effekt bereits mit integriert. Wie auch die *FB360 Spatial Workstation* sind hier ein Binaural-Decoder und ein Tool zur Szenen-Rotation enthalten.

Die Suite bietet darüber hinaus noch andere Tools, die in der Form nicht in der *FB360 Spatial Workstation* vorkommen, welche aber im Bearbeitungsprozess mit der *Spatial Workstation* kombiniert werden können (dann allerdings nur bei Mischung in dritter Ordnung), hier wird nur eine Auswahl vorgestellt:

- Der *DirectionalCompressor* funktioniert als Kompressor bzw. Limiter in der Ambisonics-Umgebung. Dabei können verschiedene Regionen im Raum auch verschieden effektiert werden. Beispielsweise kann der Raum aus frontaler Richtung stärker komprimiert werden, als der Raum hinter der Hörerposition.
- Der *EnergyVisualizer* veranschaulicht die Schallleistung der virtuellen Lautsprecher im Ambisonic-Signal mittels Farbcodes. So kann auf einen Blick erfasst werden, aus welcher Richtung die meiste Energie einfällt.
- Mittels *MultiEQ* können Ambisonics-Signale, wie aus der klassischen Mischung gewohnt, gefiltert werden.
- *FdnReverb* errechnet einen Ambisonics-Hall, auch hier funktioniert die Bedienung wie in klassischen Hall-Plugins.

Ein besonders nützliches Tool der *IEM Plugin-Suite* ist der *AllRA*¹²¹ *Decoder*¹²² (*Abbildung 14*). Hier können Ambisonics-Mischungen bis zur 7. Ordnung in jedes gewünschte Lautsprecher-Layout dekodiert werden. Das Ergebnis ist dann wiederum ein kanalbasiertes, feststehendes Format, welches perfekt an das Ziel-Wiedergabesystem angepasst ist. Im Plugin werden nach Baukastenprinzip Lautsprecher aufgestellt. Jedem Lautsprecher werden Kanalnummer und Positionsdaten zugeteilt. Durch Hinzufügen von imaginären Lautsprechern werden Probleme verhindert, die auftreten falls keine eindeutigen Flächen zwischen den einzelnen Lautsprechern aufgespannt werden können. Beispielsweise müssen bei einem System ohne Lautsprecher in der

¹²⁰ Vgl. IEM (2019a)

¹²¹ Steht für: All-Round Ambisonics

¹²² Vgl. IEM (2019b)

Medianebene zwangsläufig imaginäre Lautsprecher definiert werden. Die fertigen Layouts können exportiert und für den erneuten Einsatz gesichert werden.

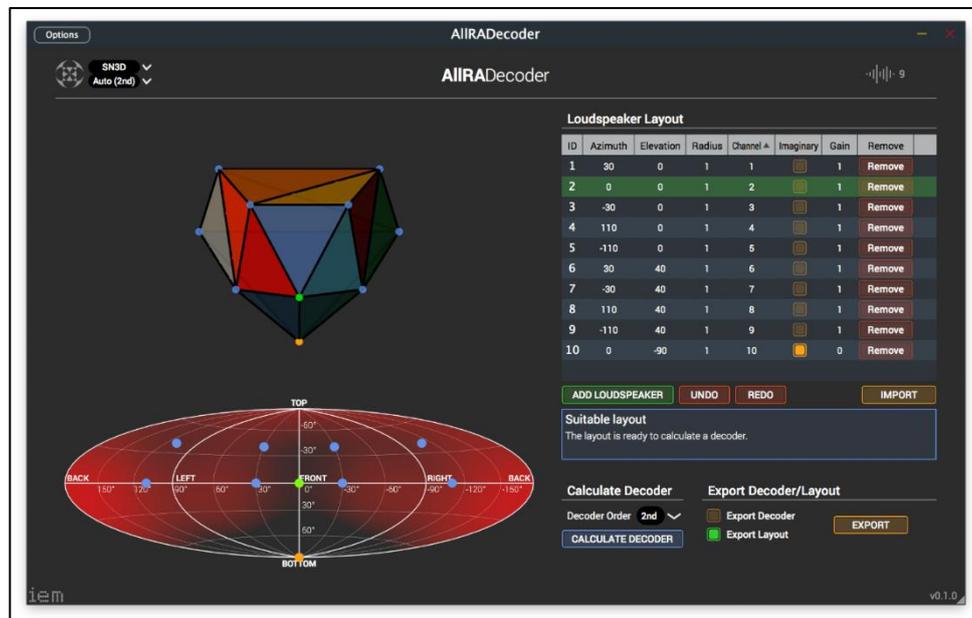


Abbildung 14: AIIRADecoder Plugin

6 Praktische Umsetzung der Virtual Reality Konzerte

6.1 Beschreibung der Anwendung

Bei der in dieser Arbeit behandelten VR-Anwendung handelt es sich um einen für die HTC Vive entwickelten Prototyp, der lineare 360-Grad Konzertvideos mit Interaktionsmöglichkeiten für den Nutzer ausstattet, um ein interessanteres VR-Erlebnis zu erreichen.

Grundidee der Anwendung ist es, dem Nutzer nicht nur die Möglichkeit zu geben sich Konzerte mit einem HMD anzusehen, sondern darüber hinaus über Kameraposition und sogar Location des Konzerts zu entscheiden.

Der Nutzer sieht sich einen Song der Band „Eau Rouge“ in deren Live-Performance an und kann während der Show, auf Knopfdruck, zwischen Publikums- und Bandperspektive wechseln. So wird der Eindruck erweckt, man stünde entweder als Bandmitglied mit auf der Bühne und beobachtet die Reaktionen des Publikums oder man genießt als Konzertgast die Veranstaltung. Außerdem kann der Nutzer sogar das gesamte Konzert wechseln, um sich eine andere Live-Show der Band anzusehen, ohne dass der dargebotene Song dabei unterbrochen wird. Ein weiteres interaktives Element ist eine Art Mitsing-Option, welche eintritt, wenn der Nutzer den Controller der HTC Vive vor seinen Mund hält und damit ein Mikrofon simuliert. In diesem Fall wird der Gesang der Aufnahme stark gedämpft, sodass der Nutzer selbst in die Rolle des Leadsängers schlüpfen kann.

Für den Prototypen wurde der Song „Melt“ an zwei verschiedenen Konzert-Terminen in Stuttgart und Ulm mit jeweils zwei 360°-Kameras aufgezeichnet. Die Anwendung läuft auf Basis der *Unreal Engine 4*.

6.1.1 Anforderungen an den Ton

Der Ton spielt grundsätzlich bei jedem musikalischen Medium eine bedeutende Rolle. Ziel bei dieser Anwendung war es jedoch nicht nur, die Darbietung der Band durch sauberes Abmischen hervorzuheben, sondern viel mehr ein Gefühl des „Dabei Seins“ an den Nutzer zu vermitteln.

Immersion war an dieser Stelle wichtiger als Realismus. Es sollte nicht versucht werden, das reale Klangbild der Konzerte unverändert wiederzugeben, wie es bei den meisten Klassik-Veranstaltungen der Fall ist, sondern einen glaubwürdigen, aber angenehm klingenden Kompromiss zwischen Realaufnahme und Kunstmischung zu finden.

Besonders anspruchsvoll war hierbei die Mischung für die Bühnenposition, da diese sehr von der Realität abweicht. Neben der reinen Musik waren außerdem die Reaktionen des Publikums und der Klang des Raums im Fokus, sowohl bei der Aufnahme als auch später in der Mischung.

Eine kleine Hürde war es, sicherzustellen, dass beide Konzert-Aufnahmen genau gleich lang sind und die Songs jederzeit überblendet werden können, ohne dass ein Zeitsprung im Song wahrnehmbar ist. Dafür musste die Band in beiden Konzerten auf einen Click-Track spielen, außerdem durften keine in der Länge voneinander abweichenden Soli oder ähnliches vorkommen. Die Struktur des Songs musste also in beiden Konzerten die gleiche sein.

Für die Wiedergabe auf Kopfhörern über die *Unreal Engine 4* musste ein geeignetes Tonformat gefunden werden. Es wurde entschieden, eine räumliche Mischung zu erstellen, die mittels Head Tracking auf die Kopfbewegungen des Users reagiert. Die vertikale Ebene wurde bewusst nicht berücksichtigt, da auf den hier gezeigten Rock-Konzerten keine signifikanten Schallereignisse aus vertikaler Richtung vorkamen.

In den später in dieser Arbeit folgenden Usertests sollte die Frage behandelt werden, ob der interaktive, binaurale Sound, wie er hier verwendet wird, zu einer besseren Nutzererfahrung führt, oder ob ein einfacher Stereomix für die Anwendung ausgereicht hätte.

6.1.2 Die Band „Eau Rouge“

Die dreiköpfige Stuttgarter Band „Eau Rouge“ beschreibt ihren musikalischen Stil selbst als *Dream Pop* oder *Noise Pop*.¹²³ Charakteristisch für ihre Musik ist der Einsatz von zahlreichen Gitarreneffekten sowie Synthesizern, was zu einem halb elektronischen und sehr halligen Klangbild führt.

*„Unsere Musik besteht aus viel Hall und Sounds in denen man sich verlieren kann, bei denen man die Augen schließt und dann glitzert es. Da passt der Begriff Dream Pop für uns einfach sehr gut, und wenn es dann von den Klängen etwas härter wird, besonders live, dann ist es massiv und deshalb der Noise –Aspekt.“*¹²⁴

Die beiden Gitarristen und Sänger Jonas und Bo sind mit ihren Gitarren durch ihre verschiedenen Bodeneffektgeräte sehr flexibel und können so selbst den gewünschten Klangcharakter ihrer Instrumente bestimmen. Außerdem werden ihre Gitarrensignale gesplittet, sodass beide Musiker, wenn es gewollt ist, parallel zwei unterschiedliche Signale erzeugen können. Beispielsweise können mit nur einem Instrument gleichzeitig sowohl eine Bassgitarre als auch eine Rhythmus-Gitarre gespielt werden.

Der Schlagzeuger Magnus spielt abwechselnd ein akustisches Schlagzeug oder ein Samplepad, welches neben Drum-Samples auch andere synthetische Klänge abspielt.

6.1.3 Der Song „Melt“

Die Wahl des Musikstücks, welches für die Anwendung aufgezeichnet werden sollte fiel auf den Song „Melt“.

Zum Zeitpunkt der beiden Konzerte war die Single (Release: 14.12.2018) noch nicht veröffentlicht worden. Das Publikum hörte das Musikstück also zum ersten Mal, und auch für die Band war es ein besonderer Moment, ihr neuestes Werk aufzuführen. Anzumerken ist, dass sich die Studioaufnahme, die aktuell auf allen gängigen Plattformen zu hören ist, deutlich von der Live-

¹²³ Vgl. Eau Rouge (2019)

¹²⁴ Vgl. Eau Rouge (2018)

Performance unterscheidet. In dieser Arbeit wird ausschließlich auf die „Live-Version“ der Single eingegangen.

Der Track beginnt zunächst rein elektronisch, mit Synthesizer und elektronischen Drums, dann wird das Drum Pad vom akustischen, aber harten Schlagzeug abgelöst und die Gitarren und der Bass setzen ein. Ein halliger Pluck-Synth zieht sich fast durch den gesamten Song.

Jonas Stimme funktioniert gleichzeitig als Lead-Melodie und wird von verzerrten Gitarrensounds begleitet, die sich für den instrumentalen Part vor dem dritten Chorus im Klang ändern und deutlich aggressiver werden.

Der gesamte Song wurde auch live auf Click gespielt, was für die interaktive VR-Anwendung später essenziell war. Durch den Einsatz des Click-Tracks konnte sichergestellt werden, dass die Aufnahmen der beiden Konzerte perfekt synchron abgespielt werden können, ohne auseinander zu laufen.

6.1.4 Kulturzentrum Merlin

Das Konzert im Kulturzentrum Merlin im Stuttgarter Westen fand am 30.11.2018 statt. Vor heimischem Publikum freute sich die Band über einen restlos ausverkauften Gig vor ca. 200 bestens gelaunten Gästen. Der Song „Melt“ wurde ca. in der Mitte des einstündigen Sets gespielt.

Der Saal im Merlin ist ca. 16m lang, 8m breit und nur ungefähr 4m hoch. Von der Bühne aus gesehen rechts wurde die Wand des Saals mit schwarzem Mollton abgedeckt. Aufgrund dieser Raumeigenschaften und der Tatsache, dass der gesamte Saal mit Menschen gefüllt war, gab es während des Konzertes sehr wenig Nachhall. Die Konzertbesucher konnten sich außerdem jederzeit an der Bar im hinteren Bereich des Saals mit Getränken versorgen und auch sonst ließen sich viele, trotz der lauten Musik, nicht davon abhalten Unterhaltungen zu führen, was zu einem insgesamt hohen Geräuschpegel im Publikum sorgte.

Aufgrund der kleinen Bühne (3m x 6m) konnte die Bühnenkamera nicht mittig aufgestellt werden, sondern wurde halblinks zwischen Drummer und dem in der Mitte der Bühne stehenden Leadsänger platziert.

Die Publikumskamera wurde ca. 6,5m von der Bühne entfernt in der Mitte des Saals aufgestellt um das Gedränge in Bühnennähe zu vermeiden.

Beschallt wurde der Saal durch ein PA-System von *JBL*, welches von einem *Midas Venice 320* Beschallungspult angesteuert wurde. Gemischt wurde allerdings separat durch den bandeigenen Mischer mit einem digitalen und aufzeichnungsfähigen *Alan & Heath dLive C3500* Beschallungspult.

6.1.5 Club Schilli

Das Konzert im Club Schilli in Ulm fand zwei Tage nach dem Gig im Merlin Kulturzentrum, am 02.12.2018 statt. Im Gegensatz zum vorherigen Event gab es hier nur ein sehr kleines Publikum (ca. 30 Personen), welches sich auch kaum unterhielt. Während der Aufführung gab es somit kaum Nebengeräusche, wie es noch in Stuttgart der Fall war. Der Club ist zwar deutlich kleiner als das Kulturzentrum Merlin (der Bereich vor der Bühne bietet gerademal ca. 30 Quadratmeter Tanzfläche), allerdings sind die Wände nicht mit Mollton verkleidet und die Decke ist über fünf Meter hoch, was in Kombination mit lauter Beschallung zu relativ viel Nachhall führte. Der Bereich neben der Bühne, wo sich Merchandise und Bar befinden, ist außerdem etwas größer als der eigentliche Tanzbereich. Einige Gäste hielten sich vorzugsweise dort auf, was den Effekt hatte, das Konzert allgemein sehr leer erscheinen zu lassen.

Zudem gab es technische Probleme am Beschallungspult, weswegen die Einzelsignale der Bandmikrofone nicht aufgezeichnet werden konnten.

6.2 Besonderheiten bei der Aufnahme

6.2.1 Verwendete Technik

Die Mittel zur Realisierung der Aufnahme waren begrenzt, da es sich bei den Konzerten um kleinere Veranstaltungen ohne Teilnahme von Sponsoren oder Rundfunkanstalten handelte. So musste bei der Umsetzung des Prototyps der Kostenpunkt berücksichtigt werden. Daher wurde bereits im Vorfeld genau überlegt, welche Mikrofonierung für das Projekt als sinnvoll erschien und auf hochpreisige Technik verzichtet.

Insgesamt wurden für die Atmo-Aufnahmen 11 Mikrofone im mittleren Preisbereich eingesetzt. Als Mikrofonpreamps und Wandler dienten zwei *RME Octamics* welche sowohl analog, als auch über die ADAT¹²⁵-Schnittstelle mit einem *RME Fireface UFX* USB-Interface verbunden waren. Das Interface wurde per USB an einen herkömmlichen Acer-Laptop angeschlossen. In der DAW *Reaper* von *Cockos* wurden, mit Ausnahme des *Ambeo* Mikrofons, alle Atmo-Spuren sowie der Stereo-Mastertrack der Beschallung aufgezeichnet.

Um Kabelwege im Publikum zu vermeiden wurden die vier Signale des *Sennheiser Ambeo* Mikrofons mittels eines *Zoom F8* Recorders aufgezeichnet, welcher sich direkt unter dem Mikrofonstativ befand.

6.2.2 Mikrofonierung

Die Grundlage für eine räumliche Aufnahme des Konzertes war ein IRT-Kreuz, welches ca. 1,5 Meter über der Publikumskamera an der Saaldecke aufgehängt wurde. Hierfür wurden vier *Audio-Technica AE3000* Kondensatormikrofone mit Nierencharakteristik auf einer Konstruktion aus drei

¹²⁵ Alesis Digital Audio Tape

Mikrofonschienen befestigt. Mittels dreier Fäden und einem Kunststoffrohr wurde sichergestellt, dass sich das IRT-Kreuz nicht drehen oder kippen konnte. Die vorderen Mikrofone wurden jeweils im 90° Winkel voneinander nach außen ausgerichtet und im 45° Winkel nach unten gebeugt. Ziel war es, mehr Schall aus dem Publikum als von der PA aufzunehmen. Die Basisbreite betrug etwa 30 Zentimeter, so sollte ein möglichst breites Klangbild erreicht werden.

Auf eine Höhenebene durch zusätzliche, nach oben ausgerichtete Mikrofone, wurde aus folgenden Gründen bewusst verzichtet: Für ein 3D Mikrofonarray wären mindestens acht baugleiche Mikrofone nötig gewesen, um alle Himmelsrichtungen in zwei Ebenen (horizontale und Höhenebene) abzubilden. Der Mehrwert, im Kontext eines Pop Rock Konzertes, zusätzlich diffusen Schall von oben aufzunehmen, war nicht ausreichend ersichtlich und konnte somit den Aufwand nicht rechtfertigen. In der VR Applikation hat der User zwar die Möglichkeit sich neben einer 360° Drehung auch nach oben bzw. unten umzuschauen, es wurde allerdings entschieden, dass diese Bewegungen keine Bedeutung für den Ton der Applikation haben sollten. Alle wesentlichen Geräuschquellen des Konzertes befinden sich für den Hörer auf der horizontalen Ebene und so wurde angenommen, dass zusätzliche Reflexionen vom Boden bzw. der Decke des Saals kaum positive Auswirkungen auf die Immersion der Anwendung hätten.

Mit der Absicht, die Reaktionen der Konzertbesucher in unmittelbarer Nähe des Zuschauers aufzunehmen, wurde ein *Sennheiser Ambeo* Mikrofon direkt unterhalb der 360° Kamera im Publikum in ca. 1,70m Höhe platziert. Die Wahl fiel auf das First Order Ambisonics Mikrofon, da es so möglich war, den räumlichen Klang an einem konkreten Punkt, nämlich direkt an der User-Position, aufzunehmen. Auf diese Weise konnte auf eine komplexere Mikrofon-Installation mittels mehrerer Einzelmikrofone im Publikum verzichtet werden.

Im hinteren Bereich des Saals im Merlin, über der Bar und über der Garderobe, wurden zusätzlich Mikrofone installiert. Der Gedanke dabei war es, das Publikum noch präsenter einzufangen und besser von der Beschallung zu trennen. Hierfür wurde das *Sennheiser mkh 50* Kondensatormikrofon mit Supernierencharakteristik und das *Sennheiser mkh 40* mit Nierencharakteristik verwendet, welche senkrecht, von der Decke hängend, nach unten zeigten.

Die gleichen Mikrofone kamen auch in Ulm zum Einsatz, ebenfalls über der Bar, bzw. im Backstage-Bereich.

Der Schall auf der Bühne wurde von zwei baugleichen *Beyerdynamic MCE 83* Kondensatormikrofonen abgenommen. Die Mikrofone mit Nierencharakteristik wurden hinter dem PA System von der Decke gehangen und zeigten senkrecht auf den Bühnenboden.

Im Abstand von ca. 6 Metern wurden zwei *Rode NT5* Kleinmembran Kondensatormikrofone an der vorderen Bühnenkante aufgestellt, um eine große Breite des Publikums in den vorderen Reihen aufzuzeichnen.

Aus Platzgründen musste auf Installationen, wie beispielsweise ein IRT-Kreuz, auf der Bühne verzichtet werden. Allerdings war dies nicht von Bedeutung, da der Bühnensound ohnehin künstlich so stark verfälscht werden sollte, dass die realen Aufnahmen hier nur eine stark untergeordnete Rolle spielten.

Alle Einzelsignale der Band wurden im Beschallungspult prefader aufgezeichnet, um eine eigene, unabhängige Mischung zu ermöglichen. Dabei gab es keine ungewöhnlichen Mikrofonierungen,

das Schlagzeug und der Gesang wurden herkömmlich abgenommen und die Bass- und Synthesignale führten direkt in die Stagebox des Beschallungspult. Die beiden Gitarrenamps wurden sehr nah mikrofoniert. Per Aux-Ausgang wurde der Summenmaster an das Interface gereicht und ebenfalls in *Reaper* aufgezeichnet.

Insgesamt wurden so im Kulturzentrum Merlin 31 Einzelspuren aufgezeichnet (das A-Format des *Ambeos* als vier Einzelsignale gezählt). Im Club Schilli musste, aufgrund der sehr kleinen Bühne, auf die vier Bühnenmikrofone verzichtet werden.

Band (von Beschallungspult)	Atmo (Aufnahme in Reaper)
Bassdrum innen	Hauptmikrofon VL
Bassdrum außen	Hauptmikrofon VR
Snare Oben	Hauptmikrofon HL
Snare Unten	Hauptmikrofon HR
Hightom	Bühnenkante L
Lowtom	Bühnenkante R
Overhead L	Bühnenmitte L
Overhead R	Bühnenmitte R
Lead Vocal	Raum L
Backing Vocal	Raum R
Guitar 1	PA-Master L
Guitar 2	PA-Master R
Guitar 3 (Bass)	
Drumpad / Synth L	Sennheiser Ambeo (Zoom Recorder)
Drumpad / Synth R	4-Kanal Ambisonics A-Format

Abbildung 15: Spurplan der Konzerte (Band- und Bühnenmikrofone entfallen im Schilli-Konzert)¹²⁶

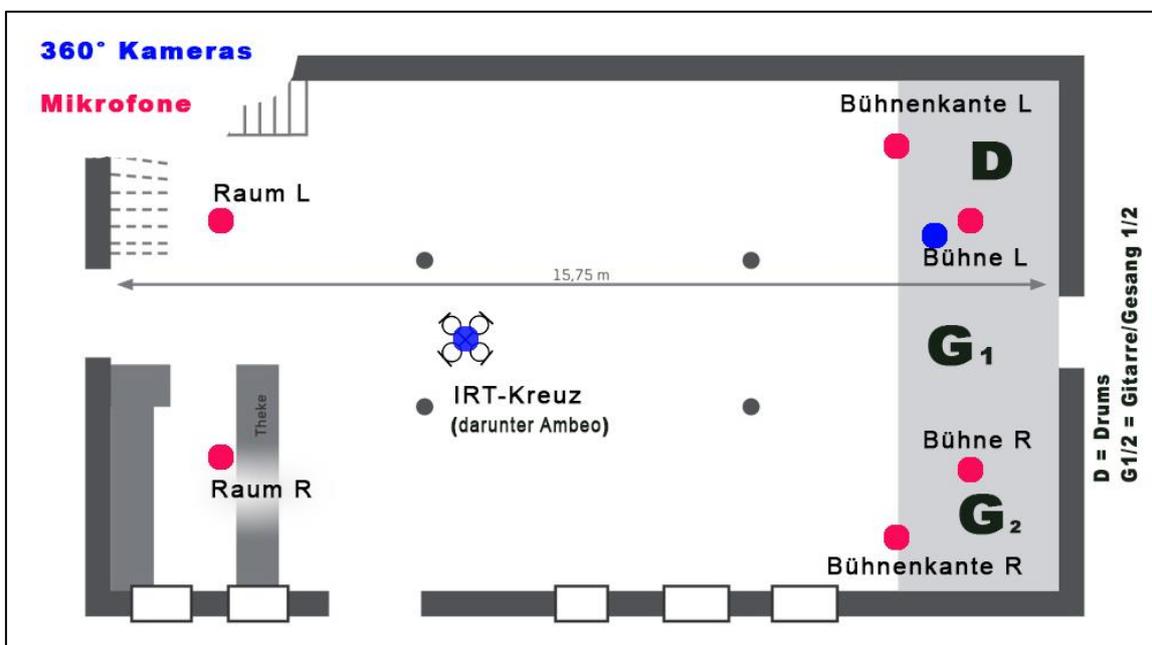


Abbildung 16: Saalplan des Konzertes im Merlin¹²⁷

¹²⁶ Bildquelle: Eigene Abbildung

¹²⁷ Bildquelle: Eigene Abbildung auf Grundlage von: Merlin Stuttgart (2019)

6.2.3 Analyse der Aufnahmen

Nach den beiden Konzertaufnahmen wurden alle Signale analysiert und bewertet. Die Ergebnisse sind in den folgenden Abschnitten festgehalten worden.

6.2.3.1 Hauptmikrofone im Publikum (Merlin)

Das Publikum ist überraschend präsent auf allen Atmo-Spuren zu hören. Das wurde im Vorfeld nicht erwartet, da die Beschallung an den jeweiligen Konzerten als sehr laut wahrgenommen wurde.

IRT-Kreuz:

Um das gesamte Frequenzspektrum der Hauptmikrofone zu nutzen, wurde absichtlich kein Lowcut-Filter bei der Aufnahme eingesetzt, jedoch kam es so zu einem sehr bassigen Signal, welches subjektiv als sehr unschön empfunden wurde. Eine Erklärung hierfür könnte im Bereich Raumakustik liegen. (Wahrscheinlich führten stehende Wellen zu der bassigen Aufnahme des IRT-Kreuzes.)

Generell war der Klang der PA-Anlage auf den Atmo-Spuren sehr unausgewogen und übertönte mit dröhnenden Bässen die Publikums-Atmo. Durch Abtrennung der tiefen Frequenzen mittels eines steilen Lowcut-Filters mit einer Grenzfrequenz von 180Hz konnte ein deutlich angenehmeres Ergebnis erzielt werden. Zwar wurden die Pegelspitzen noch von Snare-Drum und Lead-Gesang bestimmt, jedoch war das Publikum deutlich klarer zu hören. Durch Panning der vier Spuren auf vier Surround-Kanäle konnte eine große Breite der Aufnahme erreicht werden.

Sennheiser Ambeo:

Um die Aufnahme des *Ambeo*-Mikrofons beurteilen zu können, musste das Signal zunächst mittels *Ambeo A-B Converter* in das Ambisonics B-Format konvertiert werden. Anschließend konnte es binaural auf Kopfhörern abgehört werden.

Sofort auffallend war, dass der dröhnende Bass hier im Vergleich zu den Aufnahmen des IRT-Kreuzes deutlich weniger problematisch ausfiel. Trotzdem konnte durch einen Lowcut bei 120Hz ein klareres Ergebnis erzielt werden.

Vergleich IRT-Kreuz und Ambeo-Mikrofon:

Um beide Hauptmikrofon-Aufnahmen direkt vergleichen zu können, wurden die vier Spuren des IRT-Kreuzes nach der Filterung in der *FB360 Spatial Workstation* um die Hörposition ausgerichtet und binauralisiert.

Generell klingt die Aufnahme des IRT-Kreuzes etwas ausgewogener und klarer als die des Ambisonics-Mikrofons. Im direkten Vergleich klingt das *Ambeo-Mikrofon* dumpfer.

Der Vorteil der Ambisonics-Aufnahme liegt in der Präsenz einzelner Personen aus dem Publikum, nämlich denjenigen, die direkt neben der Kamera standen. Hier können einzelne Stimmen und sogar ganze Sätze wahrgenommen werden, während das Publikum in der Aufnahme des IRT-Kreuzes eher wie eine breite Fläche klingt.

Der Nachteil des *Ambeos* besteht ganz klar in dessen Räumlichkeit. Das Klangbild des IRT-Kreuzes ist viel breiter und räumlicher. Auch bei der simulierten Kopfbewegung des binauralen Mixes mittels der „Yaw-Funktion“ des *FB360* Plugins konnte bei der Mischung des IRT-Kreuzes deutlich mehr Bewegung des Klanges festgestellt werden. Die Lokalisation der Musik-Schallquellen, in diesem Fall der PA-Anlage, fällt beim IRT-Kreuz schärfer aus.

Raummikrofone:

Auf beiden Raummikrofon-Spuren ist das Publikum deutlich präsenter als die Beschallung, welche jeweils sehr dumpf klingt (Kulturzentrum Merlin). Dennoch konnte hier kein Mehrwert zum IRT-Kreuz festgestellt werden. Bei Hinzumischung der Raumsignale in den binauralen Mix verschlechtert sich nur die Lokalisationsschärfe der Band.

6.2.3.2 Band- und Bühnenmikrofone (Merlin):

Bühnen-Atmo:

Die beiden hängenden Mikrofone über der Bühne zeigen (nach Einsatz des Lowcut-Filters) ein sehr präsenten Bild des Schlagzeugs. Zu vergleichen ist der Sound direkt auf der Bühne mit der Aufnahme von Drum-Overheads, jedoch mit gedämpfter Publikums-Atmo und Beschallung im Hintergrund.

Die beiden Nieren, welche von der Bühne ins Publikum zeigten, erzielten nach Panning ein sehr breites Klangbild des Publikums. Allerdings ist die Beschallung ebenso präsent wie das Publikum. Die Idee, die Mikrofone in den Schallschatten der PA-Anlage zu platzieren, um weniger Pegel der Beschallung einzufangen, hat also eindeutig nicht funktioniert.

PA-Master:

Die Live-Mischung des Konzertes klingt insgesamt sehr dünn im Bassbereich, der Peak der Bassdrum befindet sich bei 70Hz, darunter fällt der Frequenzgang steil ab. Im Mittenbereich sind starke Verzerrungen der Gitarren zu hören, was der gesamten Mischung einen sehr aggressiven Charakter verleiht. Besonders auffallend ist die Bearbeitung der Snare, die ihren natürlichen Klang fast vollständig verliert und sehr Druckvoll den Rhythmus des Songs angibt, während die Bassdrum im Mix untergeht.

Insgesamt ist der Live-Mix auf Kopfhörern nicht ausgewogen und eher unangenehm zu hören, was jedoch Sinn ergibt, da dieser einzig und allein auf die Akustik im Club abgestimmt wurde.

Einzelsignale der Band:

Alle Schlagzeugspuren klingen „gewöhnlich“, die Bassdrum hat ihre Resonanz bei ca. 45Hz, auf den Overheads ist deutliches Übersprechen zu hören, insgesamt gibt es keine Auffälligkeiten.

Sowohl die Drumpad- als auch die Bass-Spur gingen direkt ins Mischpult und haben daher keinerlei Probleme mit Übersprechen. Die beiden Gitarrenspuren wurden direkt am Verstärker abgenommen, somit ist das Übersprechen hier nur minimal. Allerdings klingen die Gitarren nicht so aggressiv wie im PA-Mix, da dort zusätzliche Effekte hinzu gemischt wurden.

Der Leadgesang leidet enorm unter Übersprechen, hier ist das gesamte Schlagzeug deutlich zu hören.

6.2.3.3 Aufnahme im Club Schilli

Im Club Schilli gab es während des Songs keine Publikums-Atmo, da die wenigen Konzertbesucher nur nach dem Stück applaudierten. Im Gegensatz zum Merlin haben die Aufnahmen im Club Schilli deutlich mehr Reflexionen und somit einen halligeren Klang. Auch hier kann das IRT-Kreuz mehr überzeugen als das *Ambeo*.

Aufgrund von Platzmangel wurden keine Mikrofone auf der Bühne installiert.

Die Aufnahme der Einzelsignale der Band hat aus technischen Problemen im Beschallungspult nicht funktioniert.

Im Gegensatz zum Live-Mix im Kulturzentrum Merlin klingt der PA-Stereo-Master der Aufnahme im Club Schilli deutlich ausgereifter und voller. Die Bassdrum ist druckvoller und die Snare besser im Mix eingebettet. Allerdings sind die Gitarreneffekte noch aggressiver und die Spuren zudem nicht im Stereofeld platziert worden, was den Mix insgesamt sehr „eingengt“ klingen lässt. Zudem kommt es beim Leadgesang zu hörbaren Übersteuerungen, welche allerdings gut in das Gesamtgefühl der Mischung passen.

6.3 Besonderheiten bei der Mischung

Ziel der Mischungen war es, die Unterschiede in den Locations, sowie zwischen Bühne und Publikumsposition deutlich hörbar zu machen. Zugleich sollte jedoch eine ausreichend gute Klangqualität für jedes 360°-Video erreicht werden. Somit wurde im Endeffekt ein Kompromiss zwischen realer Aufnahme und Kunstmischung gefunden.

6.3.1 Produktionsformat

Alle Mischungen wurden im Ambisonics Format dritter Ordnung erstellt und gespeichert. Dies hatte den Vorteil maximaler Flexibilität, wodurch im Anschluss an die Mischung jede Zielplattform für die Wiedergabe der 360°-Videos genutzt werden konnte. Für die Wiedergabe in der *Unreal Engine* musste das Ausgabeformat später in ein kanalbasiertes Format konvertiert werden. Hierfür wurden die Ambisonics-Mischungen in jeweils acht Monosignale (virtuelle Lautsprecher) decodiert, die dann wiederum in der Game-Engine um den Spieler herum aufgestellt und für die Wiedergabe mittels Kopfhörern binauralisiert wurden. Weiterhin wurden Youtube-Videos erstellt, welche Ambisonics erster Ordnung als Audioformat unterstützen.

6.3.2 Stereo-Musikmischungen

Zuvor wurde beobachtet, dass die Atmo-Mikrofone den Bandsound nur sehr unausgewogen und subjektiv schlecht aufzeichnen konnten. Außerdem musste der Bassbereich aller Atmo-Signale herausgefiltert werden, um eine dröhnende Wiedergabe zu verhindern. Ziel war es aber, die Musik in angenehmer Klangqualität zu präsentieren, was nur durch Hinzumischung der Bandsignale möglich war.

Als Basis für die beiden Mischungen aus Publikums-Perspektive sollten daher Stereo-Mischungen der Band dienen. Allerdings waren auch die Stereo-Aufzeichnungen des Beschallungs-Masters für eine Wiedergabe auf Kopfhörern subjektiv nicht ausreichend gut geeignet.

6.3.2.1 Stereo-Musikmischung: Merlin

Von der Aufnahme im Kulturzentrum Merlin standen alle Einzelsignale der Band zur Verfügung. Diese wurden dann in der DAW *FL Studio 20* des belgischen Unternehmens *Image Line* zu einer neuen Stereo Mischung geformt (siehe Abbildung 17).

Zuerst wurden alle Instrumente, mit Ausnahme des Leadgesangs freigeschnitten, das heißt jede Spur wurde auf das entsprechende Nutzsignal reduziert. Im Falle der Bassdrum wurden beispielsweise alle Parts, in denen die Bassdrum selbst nicht spielt, herausgeschnitten usw.

Zusätzlich wurden Bassdrum, Snare und High-Hats mit einem Noise-Gate versehen, um Übersprechen zu minimalisieren. Durch diese Technik konnte das Schlagzeug aus der Aufnahme nahezu freigestellt werden und im anschließenden Mixing-Prozess wie ein Studio-Schlagzeug behandelt werden (mit einer kleinen Ausnahme: Ein zu stark eingestellter Kompressor führte

dazu, dass das Übersprechen von Gitarre und Gesang wieder präsenter wurde, daher wurde die Dynamik des Schlagzeugs nur wenig bearbeitet).

Die beiden Gitarrenamps waren nah genug mikrofoniert worden, ein Noise-Gate war hier nicht notwendig, und auf den Direktsignalen von Drumpad-Synthesizer und Bass gab es logischerweise gar kein Übersprechen anderer Instrumente.

Schwieriger war die Bearbeitung der Lead-Gesangsspur. Hier war das Übersprechen des Schlagzeugs so präsent, dass ein Noise-Gate zu keinem Ergebnis führte. Der Gesang konnte auch nicht freigeschnitten werden, da sich sonst der Gesamtklang der Band immer dann ändern würde, wenn der Gesang und somit auch das Übersprechen des Schlagzeugs fehlte. Als Kompromiss wurde die Gesangsspur mittels einer einfachen Fader-Automation in den Gesangspausen jeweils dynamisch im Pegel ca. 4dB zurückgenommen. Insgesamt konnte das Problem so etwas in Grenzen gehalten werden, dennoch wurde der gesamte Stereomix durch das Schlagzeugübersprechen auf dem Lead-Vocal negativ beeinflusst.

Nach dem Editing der Live Aufnahme wurde wie im Studioalltag üblich gemischt. Zunächst wurde das Schlagzeug durch Einstellung der Pegel und Filterkurven ausgewogen gestaltet. Um die noch sehr „brav“ klingende Bassdrum, sowie die Snare, kraftvoller zu machen wurde in beiden Mixer-Kanälen ein Multibandkompressor eingeschliffen, was entscheidend zum Gesamteindruck der Mischung beitragen konnte. Der Snare wurde außerdem noch ein Halleffekt hinzugefügt, wodurch sofort das gesamte Schlagzeug weniger trocken wirkte.

Danach wurden Bass und Synthesizer hinzugemischt. Die Spuren wurden nur wenig bearbeitet, unangenehme Frequenzen wurden aus der Drumpad-Spur gefiltert.

Um die Bassdrum präsenter im Mix zu platzieren, wurde mittels Sidechain-Kompression ein minimaler Ducking Effekt auf Bass, Synthesizer und Gitarren gebracht.

Die beiden Gitarrensignale wurden zwar direkt am Verstärker abgenommen und waren somit schon ausreichend verzerrt, jedoch fehlten sowohl Hall- als auch Delayeffekte. Wie auch in der Live Mischung wurden die Gitarren dann durch Panning links und rechts im Stereofeld platziert.

Die natürliche Dynamik des Lead-Gesangs war relativ groß und musste daher mittels eines Kompressors zurückgehalten werden. Aufgrund des vielen Übersprechens konnte der Kompressor allerdings nur vorsichtig eingesetzt werden. Nach Einstellung des Equalizers wurden auch hier Hall und Tap Delay hinzugemischt.

Der gesamte Mix wurde im Masterkanal durch Einsatz des Mastering-Plugins *IzotopeOzone 7* zuletzt etwas verdichtet (Multibandkompressor) und vorsichtig in die Breite gezogen.

Zuletzt wurde die Summe noch einmal ohne die Lead-Gesangsspur gebounced, da für die Anwendung auch eine Mitsing-Mischung ohne Gesang benötigt wurde. Diese unterschied sich zwar minimal im Klang von der Version mit Gesang, da die Effektkette auf dem Masterkanal hier etwas anders reagierte, jedoch war der Unterschied so minimal, dass das Umschalten zwischen den Versionen kaum hörbar war.



Abbildung 17: Mixerkanäle in FL Studio

6.3.2.2 Stereo-Musikmischung: Schilli

Subjektiv klang der Beschallungs-Master des Ulmer Konzerts schon deutlich besser als der des Kulturzentrums Merlin. Da hier keine Einzelsignale zur Verfügung standen, musste mit den zwei Spuren allein gearbeitet werden. Dennoch: Ohne eine weitere Bearbeitung war das Hören dieser Mischung über Kopfhörer sehr unangenehm, unausgewogen und dumpf. Das ist dadurch zu erklären, dass die Mischung vor Ort auf sehr unvorteilhafte Raumbedingungen angepasst wurde.

Hier wurde in der Nachmischung erneut mit dem Plugin *IzotopeOzone 7* gearbeitet. Um den Frequenzgang ausgewogen zu gestalten, wurde die Funktion *Matching EQ* verwendet (siehe Abbildung 18). Dabei werden zunächst die Spektren zweier Audiosignale analysiert und als *Snapshot* gespeichert, nämlich das Spektrum eines Referenzsongs und das Spektrum des zu bearbeitenden Songs. Im nächsten Schritt wird eine Filterkurve berechnet und auf den zu bearbeitenden Song angewendet, sodass im Endeffekt beide Songs ein annähernd gleiches Spektrum besitzen.¹²⁸ In diesem Fall wurde die zuvor erstellte Mischung des Songs aus dem Kulturzentrum Merlin als Referenztrack ausgewählt. Die resultierende Filterkurve des *Matching EQs* ist in Abbildung 18 zu sehen.

Durch Anwenden der Kurve änderte sich das Klangbild des Ulmer Songs deutlich zum positiven, alle Instrumente klangen sofort präsenter, die Snare deutlich schärfer, und im Bassbereich konnte mehr Druck erzeugt werden.

Allerdings wurde hinter die Dynamik-Bearbeitung noch der *Ozone Imager* in die Masterchain eingeschliffen um das Stereobild zu verbreitern. Ohne das Plugin klang der Mix deutlich zu flach, im Gegensatz zum Merlin wurden die Gitarren auch nur sehr vorsichtig nach außen gelegt.

¹²⁸ Vgl. Izotope (2011)



Abbildung 18: Matching EQ des Izotope Plugins auf dem Master-Kanal

6.3.2.3 Vergleich beider Stereomischungen

Obwohl der Frequenzgang beider Mixe sehr ähnlich ist, gibt es doch deutliche Unterschiede im Klang. Die Gitarren sind im Ulmer Konzert deutlich aggressiver als noch im Merlin. Der gesamte Mix aus dem Club Schilli hat weniger Dynamik und klingt daher sehr dicht und auch etwas verzerrt. Die einzelnen Spuren sind im Merlin deutlicher herauszuhören, alles klingt aufgeräumter und auch etwas braver. Die Performance der Musiker unterscheidet sich zudem auch, der Solo-Gitarrenpart transportiert im Schilli deutlich mehr Energie.

Das Lautheitsanalyse-Tool aus *Adobe Audition* gibt für die Merlin-Mischung einen Wert von $-10,02$ *LUFS*¹²⁹ an, die Mischung des Club Schillis kommt auf einen Wert von $-13,04$ *LUFS*. (Nach ITU-R BS 1770-3-Algorithmus).

6.3.3 Ambisonics-Mischungen aller Kameraperspektiven

Im nächsten Schritt der Postproduktion sollten die neu erstellten Musikmischungen zusammen mit den Atmoaufnahmen in die Form einer für 360°-Videos funktionierenden 3rd Order Ambisonics-Mischung gebracht werden. Dazu wurde in *Reaper* mit der *FB 360 Spatial Workstation* gearbeitet. Abgehört wurde der binaurale Ton abwechselnd mit *Beyerdynamic DT770* und *Shure SRH940* Kopfhörern, deren geschlossene Bauweise sich für VR-Anwendungen eignet, da Umgebungsgeräusche unterdrückt werden. Auf den Einsatz eines Subwoofers zur Erzeugung von Körperschall wurde verzichtet. Alle resultierenden 3rd Order Ambisonics-Mischungen im ambiX-Format sind auf der beigefügten DVD zu finden.

Zunächst wurden vier Projekte in *Reaper* angelegt (für jede Kameraperspektive ein Projekt), die alle das gleiche Channel-Routing besaßen. Die Mikrofonsignale, sowie die jeweilige Stereo-Musikmischung wurden auf Audiospuren mit je 16 Kanälen gezogen. (Ambisonics dritter Ordnung benötigt 16 Übertragungskanäle). Durch Insertion des *FB 360 Spatializer* Plugins auf jeder

¹²⁹ Loudness Units Full Scale

Audiospur, wurden die Signale räumlich verteilt und in das Ambisonics-Format konvertiert. Jede dieser Spuren wurde auf eine gemeinsame Masterspur geroutet, die ebenfalls 16 Kanäle übertrug und mit dem *FB 360 Mix Loudness* Plugins ausgestattet war, womit die Aussteuerung kontrolliert werden konnte. Dieser Master konnte später dann als fertige Mischung gerendert werden. Außerdem wurde er auf eine separate Spur gesendet, auf der das *FB 360 Control* Plugin platziert war, welches für die Simulation des Raums, die Drehung der Szene und die Binauralisierung für Kopfhörer verwendet wurde. Die ersten zwei Kanäle dieser „Kontroll-Spur“ wurden an die physikalischen Ausgänge der Soundkarte geschickt, was die direkte Abhöre der Ambisonics-Szene durch Kopfhörer ermöglichte. Die in *Reaper* angelegten Spuren sind in Abbildung 21 dargestellt.)

Die gebounceten Ambisonics-Mischungen dienen dann als Speicherform und konnten in weiteren Schritten für eine beliebige Zielplattform angepasst werden.

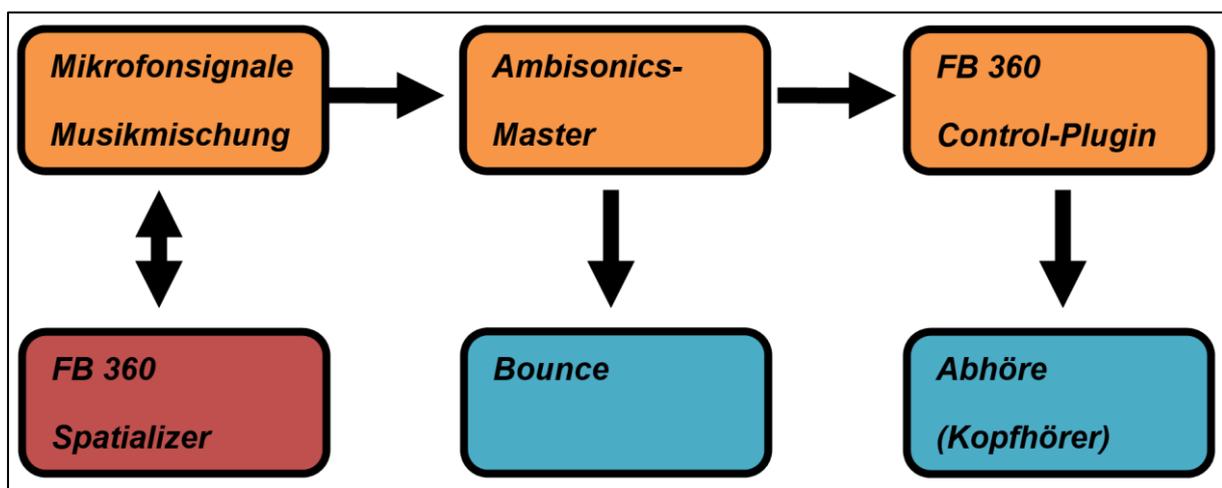


Abbildung 19: Schema des Routings in der DAW¹³⁰

6.3.3.1 Merlin im Publikum

Für die Ambisonics-Mischung der Publikumsposition im Merlin wurden insgesamt 9 Spuren verwendet: Neben Kontroll-Spur und 3D Master gab es 7 Signalspuren, darunter die vier Signale des IRT-Kreuzes, das B-Format-Signal des *Ambeo*-Mikrofons und die beiden Fassungen des Stereo-Musikmixes (Normal oder Karaoke-Version). Da Details, wie z.B. die PA-Anlage auf dem Video nicht zu erkennen waren, wurde beschlossen das Panning der jeweiligen Spuren ohne Berücksichtigung der Videoprojektion durchzuführen. Das Video wurde ich Nachhinein so ausgerichtet, dass die 0°-Achse in Blickrichtung Bühne zeigt.

Zuerst wurde die Stereo-Musikmischung angelegt und mit dem Video synchronisiert. Im *Spatializer* Plugin wurden der linke und rechte Kanal jeweils im 35°-Winkel zur Mittelachse positioniert. Gleichzeitig wurden im *Control* Plugin (Abbildung 20) die Maße des Raums (16m x 9m x 4m) eingestellt, um die Reflexionssimulation zu ermöglichen. Nachdem *room* im *Spatializer* Plugin aktiviert wurde, fühlte sich das Klangbild schon sehr räumlich an, außerdem konnte durch Einstellung des *yaw* Reglers im *Control* Plugin die Szene gedreht werden, was eine spätere

¹³⁰ Bildquelle: Eigene Abbildung

Kopfdrehung vorab simulierte. Die Live-Berechnung der Reflexionen funktionierte gut und nach kurzer Zeit des Ausprobierens konnten ideale Einstellungen für den Hall gefunden werden. Dabei galt es das passende Maß zwischen Anzahl, Verstärkung und Dämpfung der Reflexionen zu finden. Das Einschalten der *Attenuation* führte zu einem sehr diffusen und etwas dumpfen Gesamtklang, daher wurde dieser Effekt bewusst ausgelassen. Ziel war schließlich ein satter Klang, dem zu viel Realismus subjektiv nicht nur Gutes tut.

Die Einstellungen des *Spatializer* Plugins wurden auf die Spur der Karaoke-Mischung kopiert. Beim Rendern der beiden Versionen konnte später einfach die jeweilige Stereomischung, welche nicht gebraucht wurde, gemuted werden.

Die vier Spuren des IRT-Kreuzes wurden ebenfalls auf das Video geschnitten, dann von den tiefen Frequenzen getrennt (siehe 6.2.3.1) und schließlich im *Spatializer* um die Hörposition herum angeordnet. Da hier der Raumklang schon Teil der Aufnahme war, wurden keine Reflexionseffekte hinzugefügt. Dasselbe geschah mit dem B-Format-Signal des *Ambeo* Mikrofons. Dieses musste im *Spatializer* nur als solches definiert werden.

Da die zusätzlichen Atmo-Mikrofone nicht zu einem besseren Klang führten, wurden diese in der Mischung nicht berücksichtigt.

Nach der räumlichen Verteilung der Signale war der Gesamtklang des Konzertes schon sehr glaubwürdig, allerdings wurde die Musik Stellenweise von der Publikums-Atmo übertönt, was den Musikgenuss negativ beeinflusste. Im nächsten Schritt wurde daher eine Lautstärkeautomation für alle Atmosignale geschrieben. Der Anfangs- und Schlussapplaus blieb unverändert, während des Songs wurden -10db Pegel aus der Atmo genommen. An den Stellen, in denen es während des Stückes zu Applaus kam, wurde die Atmo wiederum kurzzeitig

hinzugemischt. Dabei wurde stets darauf geachtet, dass die Sprünge in der Automation nicht hörbar waren. Insgesamt konnte so ein dynamischer Wechsel zwischen Applaus und purem Musikgenuss gefunden werden. Im Solopart des Stückes, welcher sich langsam aufbaut, wurde das Publikum parallel zur musikalischen Steigerung immer weiter zurückgenommen, wodurch der Höhepunkt im dritten Chorus stärker hervorgehoben werden konnte.

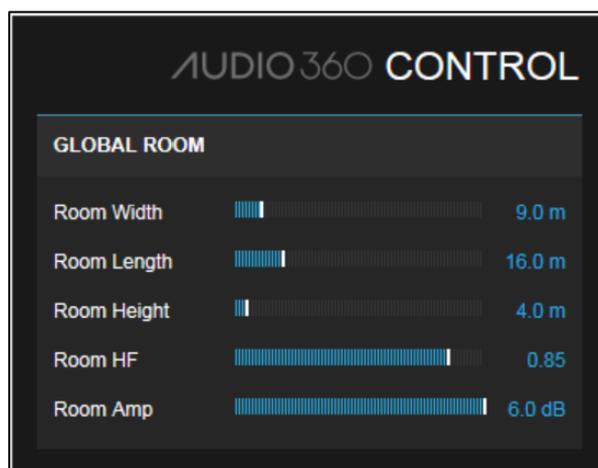


Abbildung 20: Einstellungen im Controll-Plugin



Abbildung 21: Mixertracks des Reaper-Projektes (Merlin Publikum)

6.3.3.2 Merlin auf der Bühne

Die Bühnenmischung stellte eine besondere Herausforderung dar. Zunächst stand hierbei die Frage im Raum, welcher Höreindruck dem Nutzer in der Applikation vermittelt werden sollte. Sollte der Bühnensound realistisch übertragen werden, oder sollte eine Art „Kunstmischung“ präsentiert werden. Die Entscheidung lehnte eine naturgetreue Abbildung des Bühnensounds ab, da bei der Anwendung ganz klar der Unterhaltungswert und die Immersion für den Zuhörer im Vordergrund standen. Konkret bedeutete das, dass die einzelnen Schallquellen (Instrumente) akustisch und optisch überlagert werden sollten, sprich der Gesang sollte aus der Richtung des Sängers wahrgenommen werden und nicht etwa aus der Beschallungsanlage, wie es eigentlich zu erwarten wäre.

Um dieses Ziel zu erreichen, konnte nicht mit der Stereomusikmischung gearbeitet werden, es wurden die Einzelspuren jedes Instruments benötigt. Da aber die Mischung im Vergleich zur Publikumsposition nicht völlig anders klingen sollte, wurden die bereits bearbeiteten Spuren aus *FL Studio* in Mono- (Lead Vocal, Backing Vocal, Bass) und Stereofiles (Drums, Synthesizer, Gitarre 1 & 2) gerendert und in das *Reaper* Projekt eingefügt.

Insgesamt ergaben sich so für die Darstellung der Band 7 Spuren, welche jeweils mit dem *Spatializer* Plugin im 360°-Raum ausgerichtet werden konnten (siehe Abbildung 22). Dafür wurde das 360°-Video in das Vorschaufenster des Plugins geladen, die jeweiligen Instrumente mussten dann nur noch in der Rektangular-Projektion des Videos platziert werden. Die Raumreflexionen wurden, wie bereits in 6.3.3.1 angepasst. Abbildung 19 zeigt exemplarisch die Einstellungen für den Lead-Gesang, alle anderen Spuren wurden auf dieselbe Weise hinzugemischt.

Eine Schwierigkeit bei dieser Mischung entstand durch das akustische Übersprechen des Schlagzeugs auf das Gesangsmikrofon. Bei aktivierter Gesangsspur wanderte das Schlagzeug im 3D-Panorama in Richtung des Sängers. Dieses Problem konnte nicht vollständig gelöst, jedoch in Grenzen gehalten werden. Dazu wurde die Gesangsspur im Pegel automatisiert: Große Textpausen wurden um -7dB reduziert, so war das ungewollte Schlagzeugübersprechen kaum noch wahrnehmbar. Vollständig stummgeschaltet konnte die Spur allerdings nicht werden, denn bei weiterer Reduktion des Pegels war eine deutliche Klangverfärbung im Gesamtmix an den

jeweiligen Stellen hörbar. Kurze Textpausen konnten sogar nur um -3dB reduziert werden, da sonst eine kurze Ab- und wieder Anschwellung des Schlagzeug-Sounds zu hören wäre.

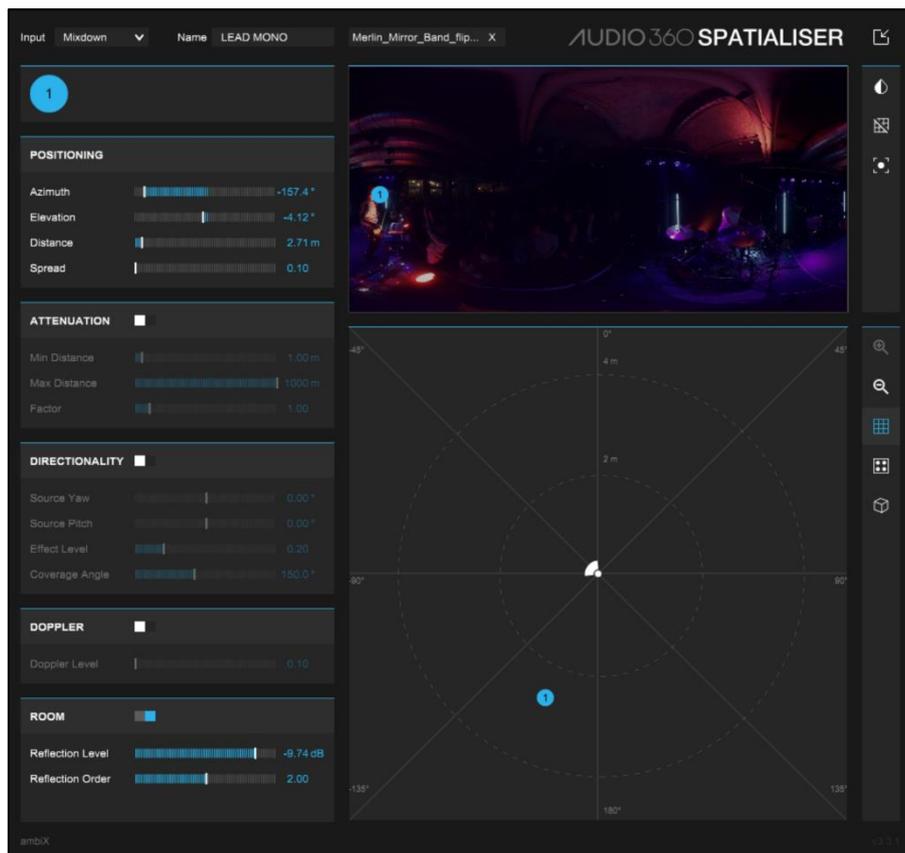


Abbildung 22: räumliche Platzierung des Lead-Gesangs

Nachdem alle Bandspuren gemischt waren, stand ein Zwischenergebnis fest, welches eine hohe räumliche Auflösung der einzelnen Instrumente vorweisen konnte. Allerdings klang die Mischung insgesamt sehr steril, da es noch keinerlei Publikumsatmo gab. Im nächsten Schritt wurden die vier Atmo-Signale, welche auf der Bühne aufgenommen worden sind, geschnitten und hinzu gemischt. Wie auch zuvor die einzelnen Instrumente, wurden die Bühnenmikrofone entsprechend ihrer Position im Panning-Fenster des *Spatializers* platziert.

Die beiden Signale der senkrecht auf die Bühne zeigenden *Beyerdynamic* Mikrofone spiegelten den realen Bühnensound wider, welcher im Wesentlichen das Schlagzeug abbildete. Das Publikum war nur wenig präsent. Bei starkem Hinzumischen der beiden Signale kam es zu einer Überbetonung des Schlagzeugs, daher wurden die Mikrofone nur vorsichtig gepegelt.

In den Signalen der beiden *Rode* Mikrofone, welche ins Publikum zeigten, war zwar reichlich Atmo enthalten, allerdings stellte hier das Übersprechen der PA ein Problem dar. Bei aktiven Atmo-Spuren fiel der zuvor so präzise Mix der Band in sich zusammen. Aus diesem Grund konnte die Atmo nur in der Pause vor dem Solopart und während des Schlussapplauses hinzugemischt werden. Eine Möglichkeit, dieses Problem zu umgehen, besteht darin, eine künstliche Atmo über die Band zu legen, die separat aufgenommen wurde. In diesem konkreten Fall wurde allerdings auf solch einen Trick verzichtet.

6.3.3.3 Schilli im Publikum und auf der Bühne

Die Mischungen für die Videos im Club Schilli sind prinzipiell sehr ähnlich und können daher in einem Punkt zusammengefasst werden. Da hier die Einzelsignale der Band nicht zur Verfügung standen, musste für beide Mischungen das Stereosignal der Musik verwendet werden. Darüber hinaus gab es, bis auf den Schlussapplaus, keine Publikums-Atmo, was das Arbeiten insgesamt sehr übersichtlich machte.

Zunächst wurden die beiden Kanäle (L; R) der Musikmischung auf die beiden Lautsprechersysteme der PA-Anlage platziert. Dadurch ergab sich für die Publikumsposition ein gleichseitiges Dreieck zwischen Hörposition und Stereosignal, was klanglich optimal war. Für die Bühnenmischung geschah dasselbe, allerdings wurden die Schallquellen hier ca. einen Meter hinter der PA platziert, da die Stereobreite auf Kopfhörern sonst unnatürlich extrem klang und es bei Kopfdrehung zu Artefakten kam, wenn die Hörposition mit beiden Signalen auf einer Linie stand.

Die Publikumssignale, sowie das Ambisonics B-Format-Signal, wurden entsprechend ihrer Aufnahmepositionen verteilt und im Pegel automatisiert. Sie geben den Raumklang wieder und werden für den Schlussapplaus nochmal um ca. 18dB angehoben.

6.3.4 Dekodierung

Die finalen Mischungen mussten für die Implementierung in der Spiele-Engine in ein kanal- bzw. objektbasiertes Tonformat umgewandelt werden. Beim herkömmlichen 5.1 oder 7.1 Standard wird die frontale Wiedergabe allerdings, im Gegensatz zur rückseitigen, bevorzugt, was bei 360°-Anwendungen, in denen sich der Hörer in alle Richtungen orientieren kann, ungelegen kommt. Daher wurde beschlossen ein Lautsprechersetup aus 8 Lautsprechern mit jeweils gleichem Öffnungswinkel von 45° zueinander zu erstellen. Aufgrund der gleichbleibenden Abstände zwischen den Stereopaaren, werden Phantomschallquellen aus allen Richtungen gleich hoch aufgelöst. Das resultierende Oktagon wurde im *AllRA Decoder* Plugin implementiert (Abbildung 23), dabei bildet der erste Kanal den aus 45° vorne linksstehenden Lautsprecher. Danach wird schrittweise im Uhrzeigersinn hochgezählt, Kanal 2 bildet den Center, Kanal 3 den Lautsprecher vorne rechts, usw. Das Plugin verlangt zwingend die Definition von Höhenkanälen, daher wurden dem Layout zwei imaginäre Lautsprecher hinzugefügt.

Für den Einsatz in der Spiele-Engine wurden Mono-Signale benötigt, daher wurden alle Kanäle der 8.0 Oktagon-Mischung einzeln gebounced. Diese Files sind ebenfalls auf der beiliegenden DVD enthalten.

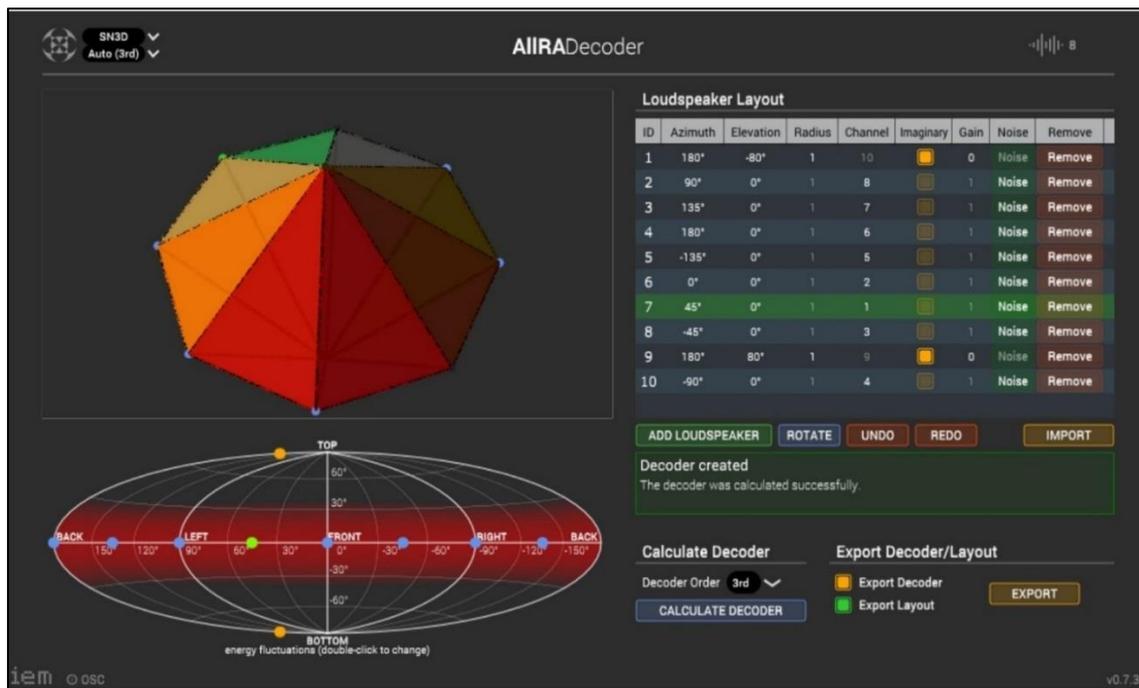


Abbildung 23: Oktagon-Lautsprecher Layout im AIRRDecoder

6.3.5 Alternative Surround-Mischung

Es musste eine Möglichkeit gefunden werden, die Funktionalität des Ambisonics-Workflows zu prüfen. Um das Arbeiten mit Ambisonics zu rechtfertigen, sollte das Resultat mindestens die Qualität einer herkömmlichen 7.1 Surround Mischung erreichen. Außerdem musste geprüft werden, ob die Werkzeuge, die im Rahmen der Ambisonics-Mischung verwendet wurden, auch fehlerfrei funktionierten. Um dies zu testen, wurde eine Vergleichsmischung im herkömmlichen 7.1 Format erstellt.

Dazu wurden alle *Spatializer* Plugins auf den Spuren gelöscht und durch den *ReaSurround Panner* von *Cockos* (Abbildung 24) ersetzt. Dort wurde zunächst das gängige Surround-Setup *7.1 ITU-R* eingestellt, was einem Öffnungswinkel von 30° zwischen den drei vorderen Lautsprechern entspricht. Die seitlichen Lautsprecher sind mit 110° bzw. -110° zum Center geöffnet, die hinteren Lautsprecher mit 150° bzw. -150°. Jedes Instrument wurde dann wiederum einzeln gepannt. Da keine Bildreferenz für den Surroundpanner zur Verfügung stand, wurden die jeweiligen Azimuth-Winkel aus dem Ambisonics-Panner in den Surround-Panner übertragen.

Die Hallerzeugung gelang mittels des *Altverb* Plugins von *Audio Ease*. Zuvor wurden Raumimpulsantworten im leeren Veranstaltungssaal aufgenommen, welche die Software zur Erzeugung eines Faltungshalls benötigt. Auf diese Weise sollte der Hall den tatsächlichen Gegebenheiten vor Ort sehr nahekommen. Als Input des Halleffekts diente die Stereo-Musikmischung, welche selbst nicht Teil des Surround-Mixes war. Der gerenderte Stereohall wurde dann auf alle Kanäle verteilt, sodass eine ausreichende Umhüllung der Hörerposition gegeben war.

Auf dem 8-kanaligen Masterbus wurde zur Abhörfunktion eine Instanz des *Waves Nx Virtual Room Mix* geladen. Bei diesem Plugin handelt es sich um einen Binaural-Decoder, der es möglich macht, Surround-Mischungen auf Kopfhörern zu simulieren.¹³¹

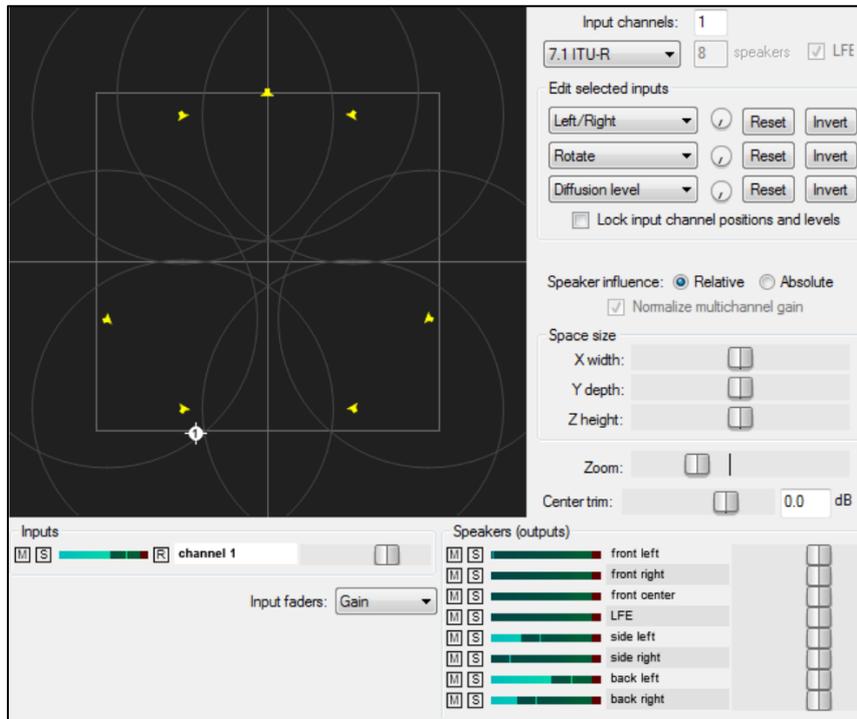


Abbildung 24: Surround-Panning des Lead-Gesangs in ReaSurround (Azimuth = -157°)

Anschließend wurde das Ergebnis im direkten Vergleich mit der Ambisonics-Mischung gegenübergestellt (hierfür wurde die Ambisonics-Mischung in dasselbe 7.1 Format dekodiert). In der Hörprobe zeigt sich kein großer Unterschied zwischen den Versionen, was zeigt, dass im Ambisonics-Workflow kein Fehler unterlaufen ist. Alle Instrumente sind gleich im Raum verteilt und auch klanglich in beiden Versionen sehr ähnlich. Der Hauptunterschied liegt im Raumklang: Während der Ambisonics-Mix etwas brav und sehr klar klingt, wirkt der 7.1 Mix halliger, vor allem der Bassbereich ist wesentlich diffuser. Diese feinen Unterschiede sind allerdings vernachlässigbar und finden ihre Ursache in den entsprechenden Hall-Effekten der beiden Mixing-Methoden. Der Test sollte primär grobe Fehler im Resultat ausschließen, was gelungen ist. Eine echte Alternative zur Ambisonics-Mischung bietet Surround-Sound nicht, da zum einen nicht alle Richtungen gleich aufgelöst werden und zum anderen ohne eine bildliche Darstellung des 360°-Videos gearbeitet werden muss.

¹³¹ Vgl. Waves Audio Ltd. (2019)

6.4 Implementierung des Tons in der VR-Anwendung

Die Monosignale der einzelnen 8.0 Mischungen wurden von Steffen Baalman in die *Unreal* Spiele-Engine integriert. Die 6 verschiedenen Mischungen ergaben also insgesamt 48 Monofiles, die in *Unreal* als Objekte, bzw. virtuelle Lautsprecher um die Spielerposition aufgestellt wurden (Abbildung 25). Dabei musste wiederum auf die Ausrichtung der Videos und deren Übereinstimmung mit der Lautsprecherausrichtung geachtet werden. Auch die Binauralisierung der Szenen musste in *Unreal* erneut programmiert werden. Dafür wurde das *SteamAudio* Plugin verwendet, eine Software zur Einbindung von 3D-Sound in Videospiele.¹³²

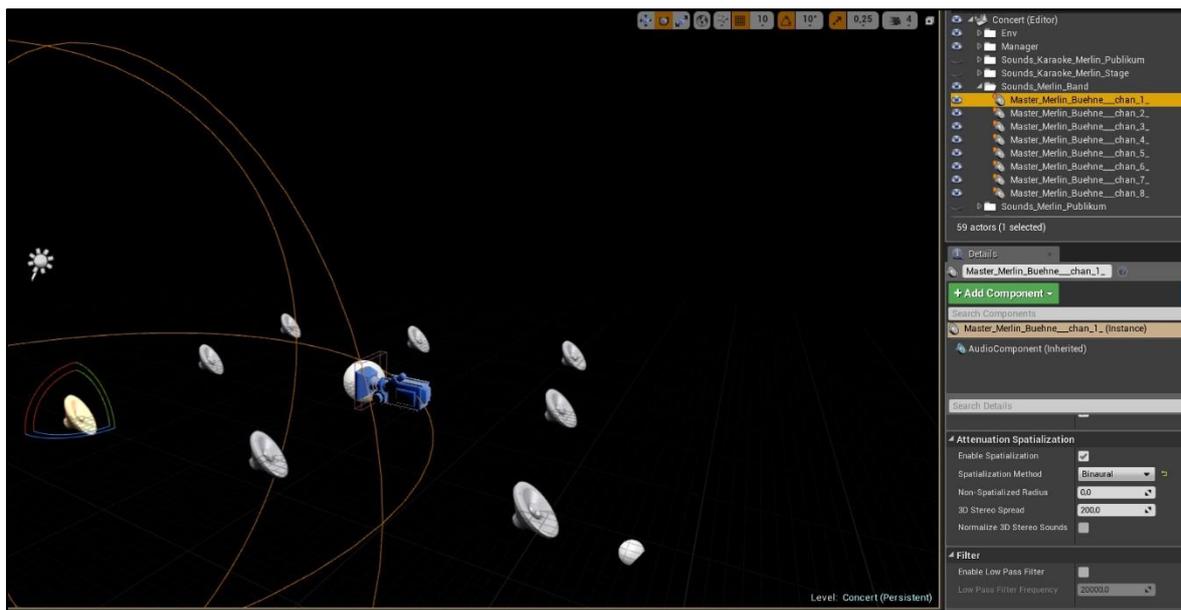


Abbildung 25: Anordnung der Schallquellen in der Unreal-Engine

Die Abbildung zeigt die Blickrichtung des Users (Kamera in der Mitte) und die einzelnen Monosignale, die darum angeordnet sind. Diese Anordnung hat im Vergleich zu einer Anordnung im Kreis keine Klanglichen Unterschiede, da lediglich der Winkel zur Hörerposition entscheidend ist, nicht aber der Abstand. Es befinden sich jeweils 6 Signale am gleichen Ort, hier sind aktuell nur die Signale der Bühnen-Mischung im Merlin aktiv.

Wechselt der User die Location so werden automatisch alle Soundfiles, bis auf die der aktuell gefragten Mischung, gemuted. In der Praxis kam es beim Wechsel der Videos zu Latenzproblemen, da dieser Rechenvorgang nicht ganz ohne Unterbrechung im Ton funktionierte. Bei aktiver Mitsingoption wurden wiederum die entsprechenden Karaoke-Mischungen aktiv.

¹³² Vgl Valve Corp. (2018)

7 Benutzertests und Evaluation

Die VR-Anwendung wurde insgesamt 24 Versuchspersonen vorgeführt und deren Eindrücke mithilfe von Fragebögen abgefragt. Ziel der Evaluation war es, erste Meinungen bezüglich des Prototyps zu sammeln. Die Ergebnisse der Evaluation sind aufgrund der kleinen Versuchsgruppe nur begrenzt aussagekräftig, lassen dennoch allgemeine Tendenzen erkennen.

Bei den Tests handelte es sich um keine reine Evaluation des Audios, sondern um eine übergreifende, allgemeine Untersuchung der Benutzererfahrung. Da sich diese Arbeit aber ausschließlich mit dem Ton der Anwendung befasst, wird hier nicht auf das Userfeedback bezüglich des Bildes oder der Bedienung der Applikation eingegangen.

7.1 Methode der Evaluation

Grundsätzlich handelt es sich bei der Evaluation um eine einfache Erhebung an Nutzerfeedback. Dabei gab es keinen konkreten Schwerpunkt, auf welchen die Aufmerksamkeit der Probanden gelenkt wurde. Mittels eines A/B-Vergleichs sollte der Erfolg der räumlichen Tonmischung nachgewiesen werden.

7.1.1 Versuchsanordnung

Für die Tests wurden zwei verschiedene Versionen des Prototyps vorbereitet. Version A mit binauralem, reaktivem Sound, Version B mit starrem Stereo-Ton. Jeder Nutzertest bestand aus insgesamt drei Durchläufen.

Im ersten Durchgang des Tests wurde Version A abgespielt. Die Probanden wurden jeweils gebeten sich alle Locations in beiden Konzerten anzusehen, waren jedoch frei, den Test jederzeit abzubrechen. Anschließend sollte der Fragebogen bis einschließlich Frage 16 ausgefüllt werden.

Im zweiten Durchgang wurde die alternative Version B abgespielt. Den VP wurde nicht gesagt, ob und wie sich die Version von der ursprünglichen unterscheidet. Nach Beenden des Durchlaufs wurde darum gebeten, Brille und Kopfhörer aufgesetzt zu lassen.

Im dritten Durchlauf wurde nochmals Version A gezeigt, um einen direkten Vergleich zu ermöglichen. Zwischen den Durchläufen 2 und 3 gab es technisch bedingt je ca. eine halbe Minute Pause, in der die Probanden auf den Start der Anwendung warten mussten. Nachdem alle Tests abgeschlossen waren, wurden die letzten drei Fragen beantwortet.

7.1.2 Probanden

Die Probanden sollten einen Teil einer möglichen Zielgruppe der Anwendung repräsentieren. Es sollte ein breites Spektrum an Männern und Frauen in verschiedenen Altersgruppen getestet

werden, darunter sowohl Personen mit als auch ohne Hintergrund im professionellen Audiobereich.

An fünf aufeinanderfolgenden Tagen nahmen insgesamt 24 Personen an den Nutzertests teil, davon war nur eine Person älter als 35 und 9 Personen jünger als 25. 15 Personen beschrieben sich selbst als audiophil, davon waren wiederum 7 beruflich oder im Studium im Bereich Audio tätig. Von allen Befragten gaben nur 8 Personen an, mindestens einmal im Monat VR-Anwendungen zu nutzen, der überwiegende Teil der VP war also in Sachen VR noch relativ unerfahren. Dahingegen gab die Hälfte aller Probanden an, mindestens einmal die Woche ein PC-Spiel zu spielen. Nur fünf der Getesteten gingen regelmäßig (mindestens einmal im Monat) zu Live-Konzerten, der Großteil gab an ca. ein- bis viermal im Jahr ein Konzert zu besuchen.

7.1.3 Verwendete Technik

Die Benutzertests fanden im CA-Labor der HdM während der vorlesungsfreien Zeit statt. Dies hatte den Vorteil, dass die Tests in der leisen Umgebung ohne Störungen von außen durchgeführt werden konnten, was für VR Anwendungen essenziell ist.

Beide Versionen des Prototyps A und B wurden direkt aus der *Unreal Engine 4* gestartet.

Neben der *HTC Vive* und deren beiden Controllern kamen *Beyerdynamic DT 770 Pro* als Wiedergabekopfhörer zum Einsatz. Deren geschlossene Bauweise ist optimal für VR Anwendungen geeignet, da Umgebungsgeräusche weitestgehend unterdrückt werden. Außerdem war es naheliegend, dasselbe Modell für die Usertests zu verwenden, das zuvor in der Mischung zum Einsatz kam. Die Kopfhörer wurden von einem *FocusriteSaffire 6 USB* Audiointerface angesteuert, dessen Wiedergabelautstärke wurde vor den Versuchsdurchläufen von den Probanden selbst reguliert.

Die Erstellung des Fragebogens und die Erhebung der Daten wurde mittels der Online Software *Survey Monkey* realisiert. In der kostenpflichtigen, browserbasierten Anwendung war es sehr einfach, simple Multiple Choice Fragen oder Freitext Fragen zu stellen. Außerdem war es schnell möglich bestimmte Attribute der Anwendung von den VP mittels Schieberegler auf einer Skala von 0 (negativ) und 100 (positiv) bewerten zu lassen oder sogar Rankings einzelner Attribute aufzustellen.

7.1.4 Fragebogen

Der Fragebogen sollte in wenigen Minuten ausfüllbar sein. Alle Fragen sollten einfach verständlich und allgemein gehalten werden. Aufgrund einer besseren Übersichtlichkeit werden an dieser Stelle nicht alle Fragen einzeln aufgeführt. Der gesamte Fragebogen ist im Anhang¹³³ dieser Arbeit zu finden.

¹³³ Siehe Anhang A

Um eine Art Stimmungsbild der Probanden festzuhalten wurden neben Freitext-Fragen viele Schieberegler eingesetzt, die Attribute der Anwendung ohne dazugehörige Begründung bewerten sollten.

Von den Insgesamt 19 Fragen dienten die ersten sechs zur Einordnung der jeweiligen Testperson in Alters- und Interessengruppen, welche in 9.1.2 beschrieben worden sind.

Die Fragen 7 bis 16 bezogen sich auf Version A der Anwendung. Dabei wurden sowohl Fragen zum Thema Audio, als auch zum Bild und der Bedienung gestellt.

In Punkt 13 wurde untersucht, wie der Sound in den vier unterschiedlichen Perspektiven empfunden wurde. Dazu sollten die Testpersonen per Drag and Drop ein Ranking der einzelnen Locations aufstellen.

Die Fragen 15 und 16 waren Freitextfelder, in denen die VP beschreiben sollten, was ihnen an der Anwendung gut bzw. schlecht gefallen hat.

Der zweite Teil des Fragebogens (Frage 17 – 19) wurde von den VP nach dem zweiten und dritten Testdurchlauf der Anwendung ausgefüllt. Es wurde gefragt, ob zwischen den beiden Versionen ein Unterschied erkannt wurde und, wenn ja, welche Version besser gefällt. In einem Freitextfeld sollte dies begründet werden.

7.1.5 Mitsingoption

Ein etwas problematischer Teil des Prototyps war die noch sehr hakelig funktionierende Mitsingoption. Geplant war eine sofortige Ausschaltung der Gesangsspur bei der Registrierung der Controller-Geste.

In der Praxis führte diese Funktion leider gelegentlich zu Ton-Aussetzern oder zur Zerstörung der Binauralität. Die Gründe hierfür müssen im Programmcode des Unreal-Projektes liegen, da alle Tonmischungen mehrmals auf solche Fehler überprüft worden sind. Es wird vermutet, dass die Mitsingoption oftmals versehentlich durch eine fehlerhafte Erkennung der Geste getriggert wird. Beim Sprung zur Mitsingoption setzt der Ton für ca. eine halbe Sekunde aus, was die Immersion der Anwendung negativ beeinflusst.

Im Laufe der Tests kam es zudem manchmal zu einem Verlust der Binauralität in der Publikumsposition im Merlin, was mit dem Auslösen der Mitsingoption einher ging. In diesem Fall konnte der Sound keine glaubwürdige, räumliche Szene darstellen. Der gesamte Mix verrutschte in den Kopfhörern auf eine Seite und stand so im Widerspruch zum Bild.

Während der Tests wurde dieses Problem von manchen Probanden beklagt, andere wiederum hatten keinerlei Probleme.

7.2 Auswertung und Diskussion des Fragebogens

In dieser Arbeit liegt der Fokus auf der Evaluation des Anwendungs-Tons, daher wurden Fragen, die sich auf das Bild oder die Bedienung des Prototyps beziehen, in der Auswertung nicht berücksichtigt. (D.h. die Fragen 7 – 11, sowie Frage 14 wurden ausgeklammert)

Es wurde weiterhin nicht versucht, die persönlichen Angaben (Fragen 1-6) mit den Ergebnissen zu korrelieren. Eine Ausnahme hierbei ist Frage 5: „Hast Du beruflich oder im Studium mit dem Thema Ton zu tun?“. Die Gruppe der „Ton-Profis“ wird in der Auswertung zusätzlich separat betrachtet.

Im Folgenden werden die wesentlichen Aussagen zu jeder einzelnen Ton-Frage kurz aufgeführt und direkt diskutiert. Die dazugehörigen Schaubilder und Einzelantworten sind weiterhin im Anhang beigefügt.

7.2.1 Frage 12: Wie empfandest Du den Sound der Anwendung?

Die Probanden sollten sich bei dieser Frage entscheiden, ob sie den Sound allgemein als eher flach (0 Punkte) oder räumlich (100 Punkte) empfanden.

Im Durchschnitt aller VP ergab sich ein Ergebnis von 74 Punkten, damit wurde der Ton insgesamt als eher räumlich bewertet. Die VP, die aus dem Audio Bereich kamen, bewerteten den Ton im Durchschnitt mit 64 Punkten als etwas flacher, im Gegensatz zu den Testpersonen ohne einen beruflichen Schwerpunkt auf Audio, welche im Durchschnitt 78 Punkte vergaben.

Generell sollte die Frage eher ein allgemeines Stimmungsbild als präzise Antworten liefern. Es war zu erwarten, dass professionelle Hörer höhere Erwartungen an den Ton haben als Laien, und somit den Gesamtklang etwas kritischer bewerten, was tendenziell bestätigt worden ist.

7.2.2 Frage 13: An welchem Ort hat Dir der Sound am besten gefallen?

Jeder Proband sollte eine Reihenfolge der verschiedenen Tonmischungen erstellen, beginnend mit der Mischung, die subjektiv am besten gefiel an erster, und der am wenigsten ansprechenden Mischung an vierter Stelle. Auf diese Weise sollte untersucht werden, wie die grundlegend unterschiedlichen Mischungen bei den Nutzern ankommen.

Im Anhang sind die aus dieser Erhebung resultierenden Balkendiagramme beigefügt.¹³⁴ Nicht jede Testperson ordnete jeder Mischung einen Platz im Ranking zu, daher kommt es bei der Auswertung nicht für jede Mischung auf 24 Gesamtstimmen.

In einer separaten Nebenrechnung¹³⁵ wurde die Durchschnittliche „Bewertungsnote“ für jeden Mix ermittelt. Insgesamt schneidet der Mix *Schili Publikum* am besten ab, gefolgt von *Merlin Stage* und *Merlin Publikum*, die schlechteste Bewertung erhält *Schili Stage*.

¹³⁴ Siehe Anhang B: Abbildungen 26 & 27

¹³⁵ Siehe Anhang B: Nebenrechnung 1

Ein eindeutiger Trend zu einer favorisierten Mischung ist allerdings nicht zu erkennen, die Durchschnittsplatzierungen liegen dafür zu nahe beieinander. Die Ergebnisse wirken generell ziemlich zufällig und willkürlich. Keine Mischung kann sich weder positiv noch negativ hervorheben. Interessant ist jedoch, dass die Mischung für die Publikumsposition im Kulturzentrum Merlin sowohl die meisten „Nummer 1“, als auch die meisten „Nummer 4“-Stimmen bekommen hat. Man könnte vorsichtig annehmen, dass die beiden Mischungen im Merlin die Meinungen deutlicher spalten, als die Mischungen für das Konzert im Club Schili.

Obwohl sich die beiden Tonmischungen im Club Schili kaum unterscheiden, wurde die Mischung für das Publikum hier insgesamt deutlich positiver bewertet. Eine mögliche Erklärung hierfür könnte sein, dass auch der visuelle Eindruck hier eine bedeutende Rolle spielt. Steht man im Publikum in der ersten Reihe und sieht die Musiker, wirkt sich das möglicherweise auch auf die Hörempfindung aus. Von der Stage aus sieht man zunächst den leeren Publikumsbereich, die Musik kommt von hinten und ist somit beim Zuschauer nicht im Fokus.

Im Ganzen betrachtet kann man hier nur schwer von einem aussagekräftigen Ergebnis sprechen. Die Versuchsgruppe war für ein Ranking zu klein und die Auswirkungen der visuellen Eindrücke auf den Höreindruck sind nicht messbar und vermutlich ergebnisverfälschend.

Es war anzunehmen, dass sich die Profis im Tonbereich intensiver auf die verschiedenen Mischungen konzentrierten. Außerdem sind Personen mit geschultem Gehör vermutlich schwerer durch visuelle Eindrücke ablenkbar. Doch auch hier kann man nicht von einem eindeutigen Ergebnis sprechen, es gibt keine besonderen Auffälligkeiten. Selbst Audio-Profis scheinen sich hier nicht einig zu sein. Allerdings ist die Versuchsgruppe hier mit sieben Personen auch zu klein für ein aussagekräftiges Ergebnis.

7.2.3 Frage 15: Gib an, was Dir am besten gefallen hat.

In diesem Freitextfeld äußerten sich 12 von insgesamt 24 Probanden positiv zum Ton der Anwendung. Fünf Personen gaben sogar direkt den räumlichen Klang, der auf Kopfbewegungen reagiert, als ein Highlight der Anwendung an (Es wurde keine VP im Vorhinein darauf hingewiesen, auch wurde nicht darum gebeten, auf den Ton der Anwendung zu achten). Die Tatsache, dass der Sound sich von Location zu Location ändert, wurde ebenfalls erwähnt. Viele Kommentare beschreiben das Hören auf der Bühne oder in der ersten Reihe als besonderes Erlebnis. Alle positiven Kommentare sind im Einzelnen im Anhang aufgeführt.¹³⁶

7.2.4 Frage 16: Gib an, was Dir nicht gefallen hat.

Drei Personen gaben an, dass der Sound gelegentlich springt oder ausfällt. Außerdem wurde zweimal ein fehlerhafter Sound erwähnt, was vermutlich der noch schlecht implementierten Mitsing-Option zu schulde kommt (siehe 9.1.5).

Es wird sowohl erwähnt, dass die Lokalisation nicht funktioniert, als auch dass die Reaktion auf Kopfbewegungen zu stark sei. Da für den Prototypen weder personalisierte HRTFs noch kalibrierte

¹³⁶ Siehe Anhang B, Abbildung 28

Kopfhörer zum Einsatz kamen, war zu erwarten, dass die Genauigkeit der Richtungsortung nicht für jede Testperson gleich gut funktioniert.

Kritisiert wurde auch, dass der Sänger auf der Bühne im Merlin zu präsent und linkslastig ist. Das ist durchaus nachvollziehbar, da der Sound auf der Bühne im Merlin tatsächlich nie ausgewogen ist und immer von der Blickrichtung abhängt.

Ein Kommentar wünscht sich noch mehr Raumklang und ein präserteres Publikum. Höchstwahrscheinlich bezieht sich der Kommentar auf die Bühnenposition im Merlin. Aufgrund des Übersprechens während des Songs mussten hier die Atmo-Mikrofone stark zurückgenommen werden. Eine künstliche Atmo und mehr Hall auf den Bandspuren wären hier vermutlich besser für die Immersion gewesen.

Alle negativen Kommentare zum Ton der Anwendung sind im Anhang aufgeführt.¹³⁷

7.2.5 Frage 17: Konntest Du einen Unterschied feststellen?

Frage 17 wurde nach den beiden weiteren Testdurchläufen gestellt. Hier galt es die kopfgebundene Stereo-Version (Version B) und die ursprüngliche Head-Tracking-Version (Version A) gegenüberzustellen.

Insgesamt gaben 20 Personen an, einen Unterschied festgestellt zu haben. Von diesen 20 konnten 17 wiederum die Änderung im Klang der Anwendung identifizieren, was rund 70% der Befragten entspricht. Diese 17 Fragebögen wurden als „geltend“ markiert und für die weitere Untersuchung berücksichtigt.

Aus der Gruppe der Audio-Profis und -Studenten konnten fünf von 7 Personen den Unterschied erkennen.

7.2.6 Frage 18: Welche Version hat Dir besser gefallen?

Von den nun noch 17 Testpersonen, die den Unterschied im Ton ausmachen konnten, empfanden 14 die Version mit binauralem Ton als besser, darunter vier der fünf Audio-Profis bzw. -Studenten. Die drei übrigen entschieden sich wiederum für den von Kopfbewegungen unabhängigen Stereoton. Das zeigt deutlich den Mehrwert, den die binaurale Surround Mischung für die Anwendung erreicht, selbst bei einer relativ kleinen Versuchsgruppe.

7.2.7 Frage 19a: Wieso hat dir Version A besser gefallen?

Alle Probanden, die Version A präferierten, gaben an, dass der räumliche Klang ausschlaggebend für eine bessere Erfahrung der Anwendung ist. Selbst wenn der Ton nicht als richtungsabhängig

¹³⁷ Siehe Anhang B, Abbildung 29

identifiziert wurde, so wurde er dennoch als „räumlich“, „dreidimensional“ oder „voller“ beschrieben. Die einzelnen Antworten sind im Anhang aufgelistet.¹³⁸

7.2.8 Frage 19b: Wieso hat dir Version B besser gefallen?

Interessant ist, dass die Qualität des Bildes zweimal als besser beschrieben wurde, obwohl es hier keinen Unterschied gab. Der Ton wurde von zwei Personen als räumlicher empfunden, eine Person beschreibt den Sound als „klarer“.¹³⁹

Der Stereomix bildet den Sound der Konzerte in einem optimalen Blickwinkel ab, das heißt alle Mixe klingen immer ausgewogen und nie links- oder rechtslastig. Vermutlich hat das auch Auswirkungen auf das Empfinden der Musik. Es ist nachvollziehbar, dass manche Hörer einen ausgewogenen Mix gegenüber einem räumlichen bevorzugen, diese sind jedoch stark in der Minderheit.

7.2.9 Zusammenfassung und mögliche Verbesserungen

Der Nutzertest hat gezeigt, dass das Feedback zum Ton des Prototyps insgesamt positiv ausgefallen ist. Trotz der kleinen Versuchsgruppe kann man davon ausgehen, dass der Multidirektionale Ton einen Mehrwert für VR-Konzerte darstellt. Weiterhin wird vermutet, dass der Anteil an Negativen Nutzererfahrungen durch eine Weiterentwicklung des Programms noch weiter reduziert werden kann. Bezüglich der Qualität der verschiedenen Mischungen gab es kein eindeutiges Ergebnis, was zeigt, dass Musik in erster Linie Geschmackssache bleibt.

Kinderkrankheiten der Anwendung, wie Tonausfälle und -Ruckler, die bei der Eingabe von Nutzerbefehlen auftreten, müssen behoben werden. Es sollte außerdem in Erwägung gezogen werden, dem Nutzer eine statische Stereomischung als Alternative anzubieten um die Personen, denen der räumliche Klang weniger gefällt, zufrieden zu stellen. Durch das Einbeziehen eines Subwoofers zur Erzeugung von Körperschall kann die Immersion der Anwendung vermutlich weiter gesteigert werden, da hier der größte Unterschied zwischen der akustischen Realität auf den Live-Konzerten und deren Simulation liegt.

¹³⁸ Siehe Anhang B, Abbildung 30

¹³⁹ Siehe Anhang B, Abbildung 31

8 Fazit und Ausblick

Ziel dieser Arbeit war es, einen möglichen Weg zur Erstellung von multidirektionalem, reaktivem Ton für populäre VR-Konzerte auszuarbeiten, sowie dessen Mehrwert im Rahmen eines Nutzertests zu demonstrieren.

Zunächst wurden die Grundbegriffe Virtual Reality und 360°-Video erklärt und eingeordnet. Nach einer kurzen Erläuterung der Grundlagen im Bereich des räumlichen Hörens, sowie der Binauraltechnik, wurden wichtige Werkzeuge für die Produktion von 360°-Video Sound vorgestellt. Es lässt sich festhalten, dass Ambisonics das derzeit flexibelste Format zur Erstellung von multidirektionalem Ton darstellt. Durch höhere Ordnungen kann eine beliebig hohe Genauigkeit erzielt werden, vorausgesetzt das Wiedergabesystem unterstützt diese. Ambisonics-Mikrofonsysteme sind zwar klein und praktisch für 360°-Videos, leiden allerdings an mangelnder Qualität und sind daher für musikalische Inhalte eher ungeeignet. Zur Aufnahme von Atmo empfiehlt sich der Einsatz von stereophonen Hauptmikrofonen, wie z.B. einem IRT-Kreuz oder, falls eine Wiedergabe in der Medianebene gewünscht ist, einem ORTF-3D Mikrofonarray.

Im praktischen Teil dieser Arbeit konnten weitere Erkenntnisse gewonnen werden. Bei der Aufnahme von populärer Live-Musik muss in zwei Teilen kategorisiert werden: Hier wird die Atmo von der reinen Musik getrennt. Darin besteht der Hauptunterschied zur Aufnahme von ernster Musik, bei der die Hauptmikrofone alle Instrumente naturgetreu abbilden sollen und klanglich nicht weiter gestaltet wird. In der populären Musik muss hingegen ein von der Beschallung unabhängiger Mix erstellt werden. Hierzu werden die Instrumente einzeln abgenommen und gemischt. Die fertige Mischung stellt die Basis für den VR-Ton dar.

Im zweiten Schritt wird die Atmo dazu genommen. In der Praxis hat sich gezeigt, dass gerade der Bassbereich der Atmomikrofone sehr verwaschen abgebildet wird und deshalb gefiltert werden muss. Bei der Aufnahme musste darauf geachtet werden, einen Pegelunterschied zwischen Publikum und Beschallung zu gewährleisten, was durch entsprechende Ausrichtung der Mikrofone realisiert werden konnte. Vollständig ausgeblendet werden konnte die Beschallung jedoch nicht.

Die Mischung in der Ambisonics-Umgebung hat sich als intuitiv und übersichtlich erwiesen. Hier wurden alle Schallquellen im Video räumlich verteilt. Anschließend wurde der Konzertsaal durch Hall- und Reflexionseffekte simuliert. Zuletzt wurde die Publikumsatmo dynamisch automatisiert, sodass Publikum und Musik abwechselnd im Fokus lagen und nicht ineinander verschwommen sind.

Die Ergebnisse der Mischung, 6 verschiedene Ambisonics-Mischungen dritter Ordnung, konnten nach der Dekodierung in ein achtkanaliges Format, in die Spiele-Engine integriert werden. Der fertige Prototyp wurde zum Großteil positiv evaluiert, leidet jedoch noch an kleineren Fehlern, die ihre Ursache in der Implementierung des Programms finden.

In zukünftigen Produktionen ähnlicher Natur gilt es, die hier gezeigten Aufnahme- und Produktionstechniken zu verbessern. Eine deutlichere Kanaltrennung und die Verhinderung von Übersprechen bei der Aufnahme der Band wären von großem Vorteil für die Bühnenperspektive. Das Experimentieren mit 3D-Hauptmikrofonen, welche die Medianebene abbilden, wäre ebenfalls interessant. Dabei sollte das entsprechende Konzert allerdings auch Klänge auf verschiedenen

Höhenebenen anbieten. Auch das Hinzufügen eines separaten Subwoofers wäre ein mögliches Forschungsobjekt.

In dieser Produktion wurde ein Prototyp für die *HTC Vive*, eines der aktuell Leistungsstärksten VR-Systeme, erstellt. Interessant wäre außerdem eine Implementierung für mobile Endgeräte mit weniger Leistung. Darüber hinaus ist die Live-Übertragung solcher VR-Konzerte ein Thema, was es zu realisieren gilt. Es stellt sich die Frage, inwiefern eine unidirektionale Mischung in Echtzeit kodiert, übertragen und für die binaurale, auf Kopfbewegungen reagierende, Abhöre wieder enkodiert werden kann.

Die Rolle, die Virtual Reality Konzerte in der Zukunft einnehmen werden, ist nur schwer vorhersehbar und wohl stark an den Verkauf von VR-Brillen gekoppelt. Es ist allerdings zu erwarten, dass die Geschichte der Entwicklung dieser Medienform noch lange nicht zu Ende erzählt ist.

Literaturverzeichnis

- Altman, Marcus; Krauss, Kurt; Susal, Joel; Tsingos, Nicolas (2016):** *Immersive Audio for VR*. AES Conference Paper. URL: <http://www.aes.org/e-lib/browse.cfm?elib=18512> [Zugriff am 06.03.2019]
- Bastian, Matthias (2018):** *So sieht es aus, wenn ein VR-Startup aus Virtual Reality aussteigt*. Mixed.de Artikel. URL: <https://mixed.de/so-sieht-es-aus-wenn-ein-vr-startup-aus-virtual-reality-aussteigt/> [Zugriff am 09.03.2019]
- Bates, Enda (2017):** *Comparing Ambisonic Microphones*. URL: <https://endabates.wordpress.com/2017/06/19/comparing-ambisonic-microphones/> [Zugriff am 07.03.2019]
- Blauert, Jens; Braasch, Jonas (2008):** *Räumliches Hören*. In: Weinzierl, Stefan: *Handbuch der Audiotechnik*. Berlin; Heidelberg: Springer, S.87-122
- Brill, Manfred (2009):** *Virtuelle Realität*. Berlin; Heidelberg: Springer
- Brockhaus Enzyklopädie:** *virtuelle Realität*. URL: <http://brockhaus.de/ecs/enzy/article/virtuelle-realität> [Zugriff am 28.02.2019]
- Camerer, Florian (2018):** *Die Kirche, das Dorf und 3D-Audio – was gehört wohin?*. Schoeps MikroForum. URL: <http://mikroforum.schoeps.de/index.php/info/14-2018personen/38-camerer> [Zugriff am 07.30.2019]
- Chaudhari, Mayur (2017):** *Do People View All 360°?*. URL: <https://blog.vrtigo.io/do-people-view-all-360-f60b858059fe> [Zugriff am 03.03.2019]
- Concert VR (2018):** Concert VR auf Facebook. URL: <https://www.facebook.com/concertvr/> [Zugriff am 11.03.2019]
- Core Sound, LLC (2018):** *Introduction: OctoMic 2nd Order Ambisonic Microphone for VR and Surround*. URL: <http://www.core-sound.com/OctoMic/1.php> [Zugriff am 07.03.2019]
- Dickreiter, Michael; Goeres-Petri, Jürgen (2014):** *Schallwahrnehmung*. In: Dickreiter, Michael; Dittel, Volker; Hoeg, Wolfgang, Wöhr, Martin (Hrsg.): *Handbuch der Tonstudientechnik Band 1, 8. Auflage*. Berlin, Boston: DE GRUYTER, S. 115–135
- Dieterle, Hannes; Sagnowski, Kasper; Wittek, Helmut (2018):** *Mikrofontechnik für VR*. Schoeps Mikroforum 2018. Karlsruhe, 06.04.2018. URL: <https://www.youtube.com/watch?v=ut9rvTsxeyEY> [Zugriff am 05.03.2019]
- Dörner, R., Jung, B., Grimm, P., Broll, W. & Göbel, M. (2013).** *Einleitung*. In: R. Dörner, W. Broll, P. Grimm & B. Jung (Hrsg.): *Virtual und Augmented Reality (VR / AR) (Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. Berlin; Heidelberg: Springer Vieweg, S.1-31
- Eau Rouge (2019):** *Instagram: eaurougemusic*. URL: <https://www.instagram.com/eaurougemusic/?hl=de> [Zugriff am 17.03.2019]
- Eau Rouge (2018):** *Eau Rouge im Interview*. URL: <https://indiesfinest.blogspot.com/2018/11/und-dann-glitzert-es-eau-rouge-im.html?fbclid=IwAR28VeNz1ABDJpegigf1p1w7yvNQPxh6xLqU9n9AB02zGx8lx5YUWYEO1K4> [Zugriff am 05.02.2019]
- Engelmann, Nikolayi (2018):** *Virtual Reality Gaming: Potential der Technologie für die Welt der digitalen Spiele*. Baden-Baden: Tectum Verlag
- Facebook (2019):** *Facebook 360 Spatial Workstation Documentation*. URL: <https://facebookincubator.github.io/facebook-360-spatial-workstation/Documentation/SpatialWorkstation/SpatialWorkstation.html> [Zugriff am 06.03.2019]
- Gerzon, Michael A. (1973):** *Periphony: With-Height Sound Reproduction*. AES E-Library. URL: <http://www.aes.org/e-lib/browse.cfm?elib=2012> [Zugriff am 05.03.2019]
- Google (2019):** *MagentaMusik 360*. Google Play Store. URL: <https://play.google.com/store/apps/details?id=de.telekom.magentamusik&hl=de> [Zugriff am 11.03.2019]
- Google Developers (2018):** *Degrees of freedom*. URL: https://developers.google.com/vr/discover/degrees-of-freedom#top_of_page [Zugriff am 03.03.2019]

Haug, Marcel; Pagel, Sven (2015): *Augmented und Virtual Reality – Eine Marktanalyse der Ein- und Ausgabetechniken*. In: Lobeck, Ralf (Hrsg.): *VIRTUAL REALITY: Impulse zu einem Phänomen der erweiterten Wahrnehmung*. Hamburg; Düsseldorf; München; Berlin: AMD Akademie Mode & Design, S.79-99

IDC (2018): *VR Headset Market Rebounds as Standalone Products Gain Traction While AR Headset Market Also Saw Positive Movement, According to IDC*. FRAMINGHAM, Mass., December 4, 2018. URL: <https://www.idc.com/getdoc.jsp?containerId=prUS44509518> [Zugriff am 09.03.2019]

IEM (2019a): *Plugin-Descriptions*. In: *IEM Plugin-Suite Documentation*. URL: <https://plugins.iem.at/docs/pluginDescriptions/> [Zugriff am 23.03.2019]

IEM (2019b): *AllRADecoder Guide*. In: *IEM Plugin-Suite Documentation*. URL: <https://plugins.iem.at/docs/allradecoder/> [Zugriff am 23.03.2019]

Izotope (2011): *Izotope Ozone 5 help, Matching EQ*. URL: http://help.izotope.com/docs/ozone/index.html?page=pages%2Fmodules_matching_eq.htm [Zugriff am 27.02.2019]

James, Paul (2014): *Paul McCartney's 'Live and Let Die' Performance is Jaunt VR's First Publically Available Cinematic VR Content*. Roadtovr.com Artikel. 20.11.2014. URL: <https://www.roadtovr.com/jaunt-vr-releases-paul-mccartney-live-let-die-concert-performance-3d-360/> [Zugriff am 09.03.2019]

Knupfer, Pablo (2018): *Binauraler Ton für einen interaktiven VR-Film*. Masterarbeit. URL: <https://curdt.home.hdm-stuttgart.de/PDF/Knupfer.pdf> [Zugriff am 22.03.2019]

Lee, Hyunkook (2016): *Capturing and Rendering 360° VR Audio Using Cardioid Microphones*. AES Conference Paper. URL: <http://www.aes.org/e-lib/browse.cfm?elib=18511> [Zugriff am 14.03.2019]

Lobeck, Ralf (2017): *VIRTUAL REALITY: Impulse zu einem Phänomen der erweiterten Wahrnehmung*. Hamburg; Düsseldorf; München; Berlin: AMD Akademie Mode & Design

Malham, Dave (2003): *Higher order Ambisonicssystems*. URL: https://www.york.ac.uk/inst/mustech/3d_audio/higher_order_ambisonics.pdf [Zugriff am 06.03.2019]

McDonald, A. (2016): *Virtually real*. In Digital TV Europe, Ausgabe August/ September 2016, S.16-20

Meier, Michael; Weitnauer, Michael (2018): *Object-based audio – an overview*. IRT 2018. URL: <https://lab.irt.de/demos/object-based-audio/> [Zugriff am 06.03.2019]

Merlin Stuttgart (2019): *Räume im Merlin: Saal*. URL: <https://www.merlinstuttgart.de/index.php?id=21> [Zugriff am 02.04.2019]

mh acoustics LLC (a): *Eigenmike® microphone*. URL: <https://mhacoustics.com/products> [Zugriff am 07.03.2019]

mh acoustics LLC (b): *Home*. URL: <https://mhacoustics.com/home> [Zugriff am 07.03.2019]

Politis, Archontis (2018): *Real/Complex Spherical Harmonic Transform, Gaunt Coefficients and Rotations*. URL: <https://www.mathworks.com/matlabcentral/fileexchange/43856> [Zugriff: 05.03.2018]

Remy, Marcel (2018): *Audio für 360°-Video - Qualitätskriterien und Produktionsworkflow im Kontext musikalischer Aufführungspraxis am Beispiel des Hessischen Rundfunks*. Bachelorarbeit. URL: <https://curdt.home.hdm-stuttgart.de/PDF/Remy.pdf> [Zugriff am 04.03.2019]

Rieger, Martin (2019a): *Formate für 360° Sound*. URL: <https://www.vrtonung.de/virtual-reality-audio-file-format-pro-contra/> [Zugriff am 04.03.2019]

Rieger, Martin (2019b): *Ambisonics für Virtual Reality Videos*. URL: <https://www.vrtonung.de/ambisonics/> [Zugriff am 05.03.2019]

Schlese, Michael (2015): *Alice Welt. Grafische Konstruktion virtueller Kommunikationsräume*. In: Baumgartner, Ekkehart, Zitzlsperger, Phillip (Hgg.): *Wandel visueller Kommunikation. Grafik und Design zwischen Illustration und Gestaltung, Visuelle Kulturen, Band 2*. Hamburg etc.: AMD Akademie Mode & Design, S.97 – 117

Schlese, Michael (2017): *“Dir glaub’ ich nicht mit dem Ohr, dir glaub’ ich nur mit dem Aug“*. In: Lobeck, Ralf (Hrsg.): *VIRTUAL REALITY: Impulse zu einem Phänomen der erweiterten Wahrnehmung*. Hamburg; Düsseldorf; München; Berlin: AMD Akademie Mode & Design, S.103-126

Sennheiser (2019a): *Ambeo VR Mic – Produkt Highlights*. URL: <https://de-de.sennheiser.com/mikrofon-3d-audio-ambeo-vr-mic> [Zugriff am 06.03.2019]

Sennheiser (2019b): *Ambeo Plugins*. URL: <https://en-us.sennheiser.com/ambeo-blueprints-downloads#> [Zugriff am 22.03.2019]

Shivappa, Shankar; Morrell, Martin; Sen, Deep; Peters, Nils & Salehin, S. M. A. (Hrsg.) (2016): *Efficient, Compelling, and Immersive VR Audio Experience Using Scene Based Audio/Higher Order Ambisonics*. AES Conference Paper. URL: <http://www.aes.org/e-lib/browse.cfm?elib=18493> [Zugriff am 16.12.2017]

Slavik, Karl M.; Weinzierl, Stefan (2008): *Wiedergabeverfahren*. In: Weinzierl, Stefan: *Handbuch der Audiotechnik*. Berlin; Heidelberg: Springer, S.609-685

Villegas, Julián (2015): *Locating virtual sound sources at arbitrary distances in real-time binaural reproduction*. URL: https://www.researchgate.net/publication/283532765_Locating_virtual_sound_sources_at_arbitrary_distances_in_real-time_binaural_reproduction [Zugriff am 05.03.2019]

VisiSonics Corporation (2014): *Technology Overview*. URL: <https://visisonics.com/tech/> [Zugriff am 07.03.2019]

Waves Audio Ltd. (2019): *Nx – Virtual Mix Room over Headphones*. URL: <https://www.waves.com/plugins/nx#introducing-nx-virtual-mix-room> [Zugriff am 24.03.2019]

Weinzierl, Stefan (2008): *Aufnahmeverfahren*. In: Weinzierl, Stefan: *Handbuch der Audiotechnik*. Berlin; Heidelberg: Springer, S.551-607

Wikipedia (2017): *Blauertsche Bänder*. URL: https://de.wikipedia.org/w/index.php?title=Blauertsche_B%C3%A4nder&oldid=172310304 [Zugriff am 02.03.2019]

Wittek, Helmut; Rumsey, Francis; Theile, Günther (2007): *Perceptual Enhancement of Wavefield Synthesis by Stereophonic Means*. URL: http://hauptmikrofon.de/theile/2007-2_WFS-enhancement-by-Stereophonic-means_JAES_V55_9_2007.pdf [Zugriff am 14.03.2019]

Wittek, Helmut; Theile, Günther (2017): *Development and Application of a Stereophonic Multichannel Recording Technique for 3D Audio and VR*. AES Convention Paper. URL: <http://www.aes.org/e-lib/browse.cfm?elib=19266> [Zugriff am 14.03.2019]

Woźniak, Tomasz (2015): *Creating binaural sound: Head Related Transfer Functions*. URL: <https://codeandsound.wordpress.com/2015/03/11/creating-binaural-sound-head-related-transfer-functions/> [Zugriff am 04.03.2019]

Wuttke, Jörg (2002): *16 Gedanken zu Surround-Aufnahmen*. Vortrag auf der 22. Tonmeistertagung 2002. URL: <http://www.ingwu.de/index.php/mikrofontechnik/mikrofontechnik/mikrofontechnik/30-16-gedanken-zu-surround-aufnahmen> [Zugriff am 14.02.2019]

Anhang A: Fragebogen

VR-Konzert

1. Wie alt bist Du?

- 18 - 24
 25 - 36
 älter als 36

2. Wie häufig nutzt Du VR-Anwendungen?

- Mindestens einmal die Woche. Nur auf Messen oder ähnlichen Anwendungen.
 Höchstens einmal im Monat. Ich hatte noch nie ein VR-Headset auf.
 Weniger als ein mal Monat.

3. Wie häufig spielst Du Computerspiele?

- Mindestens einmal die Woche. Nur auf Messen oder ähnlichen Anwendungen.
 Höchstens einmal im Monat. Ich habe noch nie Computerspiele gespielt.
 Weniger als ein mal Monat.

4. Würdest du dich als Audiophil beschreiben?

- Ja
 Nein

5. Hast Du beruflich oder im Studium mit dem Thema Ton zu tun?

- Ja
 Nein

6. Wie oft gehst du auf Konzerte/Festivals?

- Häufig. (mehrmals im Monat) Selten. (einmal im Jahr)
 Regelmäßig. (mindestens einmal im Monat) Ich war noch nie auf einem Konzert.
 Gelegentlich. (ca. alle drei Monate)

7. Wie empfandest Du die Immersion zum Konzert?

Niedrig Hoch

8. War ein Unterschied im Gefühl zwischen den Locations zu spüren?

kein Unterschied spürbar

9. Wie gefällt dir die Idee der Mitsingoption?

langweilig originell

10. Wie empfandest Du die Steuerung?

kopliert intuitiv

11. Wie empfandest Du das Bild?

minderwertig hochwertig

12. Wie empfandest Du den Sound der Anwendung?

flach räumlich

13. An welchem Ort hat Dir der Sound am besten gefallen?

Merlin im Publikum

Merlin auf der Bühne

Schili im Publikum

Schili auf der Bühne

14. Hättest Du Interesse an weiteren VR-Konzerten?

- Ja
- Nein

15. Bitte gib mit Deinen eigenen Worten an, was Dir an diesem neuen VR-Konzert am besten gefallen hat.

16. Bitte gib mit Deinen eigenen Worten an, was Dir an diesem neuen VR-Konzert NICHT gefallen hat.

VR-Konzert

17. Konntest du einen Unterschied feststellen zwischen den beiden Versionen?

Ja

Nein

18. Welche Version hat dir besser gefallen?

Version 1

Version 2

19. Wieso hat dir Version 1 / Version 2 besser gefallen?

Anhang B: Evaluation des Fragebogens

Frage 13: An welchem Ort hat Dir der Sound am besten gefallen?

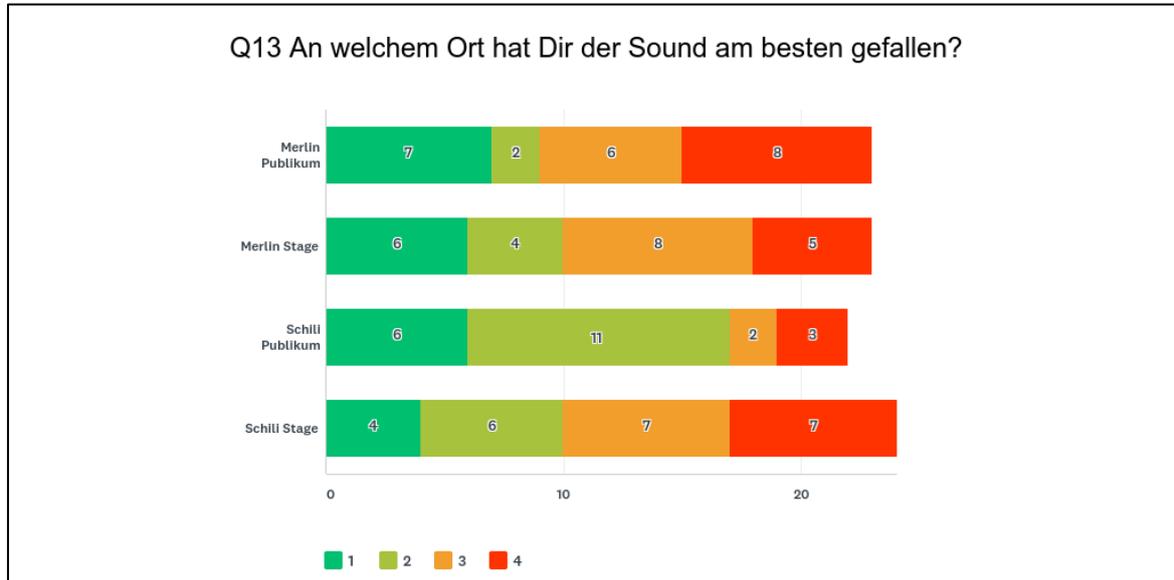


Abbildung 26: Ranking Ergebnisse aller Befragten

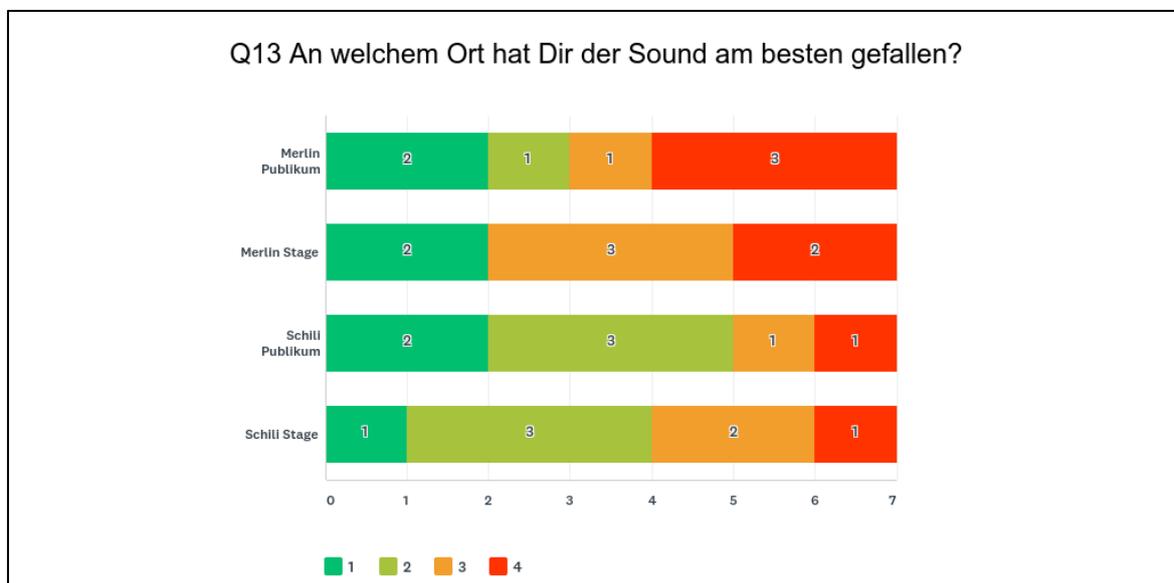


Abbildung 27: Ranking Ergebnisse (Audioprofis und -studierende)

Nebenrechnung: Ermittlung der Durchschnittsnoten

$$\text{Merlin Publikum: } \frac{7*1+2*2+6*3+8*4}{23} = 2,65$$

$$\text{Merlin Stage: } \frac{6*1+4*2+8*3+5*4}{23} = 2,52$$

$$\text{Schili Publikum: } \frac{6*1+11*2+2*3+3*4}{22} = 2,09$$

$$\text{Schili Stage: } \frac{4*1+6*2+7*3+7*4}{24} = 2,70$$

Frage 15: Positive Reaktionen auf den Ton

NR.	BEANTWORTUNGEN	DATUM
1	Wechseln der Positionen in den Clubs (man befindet sich ja sonst eher selten mit auf der Bühne), unterschiedliche Hörerlebnisse bei den verschiedenen Locations,	15.02.2019 15:00
2	Es war sehr interessant zu sehen, wie die verschiedenen Locations von der Bühne und vom Publikum aus aussahen und wie der Raum den Klang beeinflusst. Es ist sicherlich ein interessantes Konzept, gerade auch für Bands, um einen Konzertraum im Vorab zu sehen und zu hören. Somit kann man sich evtl. - vorausgesetzt natürlich, die Qualität von Video und Audio sind sehr gut - ein Bild darüber machen, wie viel Platz man für Equipment o.Ä. zur Verfügung hat und wie ein Raum klingt (Nachhallzeit, evtl. stehende Wellen usw.). Und es ist natürlich interessant zu sehen, welchen Blickwinkel die Band auf das Publikum hat, so eine Perspektive kennt man als Nicht-Bandmusiker ja meist nicht.	15.02.2019 14:30
3	Es wirkte sehr real, dass Bild war scharf, der Sound sehr gut und man bekam das Gefühl, als sei man live auf einem Konzert! Kein Schwindelgefühl.	15.02.2019 12:10
4	Die Immersion war super, man war total drin und hat vergessen, dass man eigtl in einem stillen Raum ist. Auf der Bühne stand ich auch um ersten mal ;D	15.02.2019 11:04
5	Immersion ist gegeben. Das Konzept hat Potential, v.a. mit großen bekannten Bands. Sound folgt Kopfbewegungen	14.02.2019 15:34
6	gewisse Immersion, auch im Audio, besser als bei den meisten youtube VR Konzerten	14.02.2019 14:28
7	- Einfache Bedienung - Guter Sound	13.02.2019 15:00
8	3D Sound und das Gefühl "mitten" in der Crowd	13.02.2019 14:41
9	Der räumliche Klang. Man konnte die Bühne gut verorten, und hatte dadurch das Gefühl, wirklich dort zu sein.	12.02.2019 15:32
10	Der Räumliche klang beim drehen des Kopfs.	12.02.2019 15:01
11	Die Möglichkeit beim Musik hören in eine Konzert Atmosphäre einzutauchen	12.02.2019 14:27
12	Ich muss nicht zu einer gewissen Zeit Vorort sein. Kann ich mir wiederholt anschauen. Bin immer in der ersten Reihe und habe den Sänger/Band direkt vor mir und muss es nicht aus der letzten Reihe betrachten. Rotieren vom Kopf passt den Ton auf die Ohren an	12.02.2019 13:55

Abbildung 28: Positive Reaktionen auf den Ton der Anwendung (Version A)

Frage 16: Negative Reaktionen auf den Ton

NR.	BEANTWORTUNGEN	DATUM
1	Das Bild war leider nicht besonders schön, es war sehr dunkel und verrauscht. Meiner Meinung nach sind GoPros aber gerade für den Einsatz bei wenig Licht nicht die beste Wahl, somit könnte da der Umstieg auf eine andere Kamera Besserung bringen. Der Sound war sehr räumlich und die Binauralisierung hat für mich persönlich nicht wirklich funktioniert. Wenn man auf der Bühne den Kopf gedreht hat, dann sind die einzelnen Instrumente nicht so wirklich auf den Positionen zu hören gewesen, wo man sie hätte hören müssen.	15.02.2019 14:30
2	3D Sound fehlerhaft im Merlin. Controls umständlich. UI sollte Feedback geben (Hover, Vibrate, etc.). Merlin Publikum zu dunkel. Musik nicht mein Geschmack. Tracking beim Umdrehen geht verloren. Verzögerung im Video?	15.02.2019 13:04
3	Als Merlin im Publikum war die sound Abmischung nicht richtig. Wenn ich rechts auf das "Lichtpult" geschaut habe, war der Sound auf beiden Ohren gleich laut, wenn ich auf die Bühne geschaut habe, war links auf dem Ohr wenig zu hören. Das Bild könnte etwas besser sein - beim zweiten mal der 1. Version war es nicht mehr so. lol	15.02.2019 11:04
4	Video leider zu dunkel. Sound springt. zu weit weg von der Bühne (im Merlin). man schwebt leicht über dem Boden. Im Club in Ulm spiegelt sich das Bild sobald man die Position wechselt.	14.02.2019 15:34
5	Audio-Abhängigkeit von der Kopposition ist übertrieben, es gibt Stellen, an denen der Klang hart umschaltet	14.02.2019 14:28
6	Es waren hin und wieder kurze Ausfälle im Ton zu hören, was etwas zu Lasten der Immersion war	13.02.2019 14:14
7	zu wenig Raumklang und Publikums-Atmo. Es klingt noch recht brav (Energie, Dynamik). Bildqualität zu schlecht, Perspektive ist unnatürlich	13.02.2019 12:50
8	Der Ton auf der Merlin-Bühne war deutlich links-lastig. Der Sänger/Gitarrist war durch die räumliche Nähe zu dominant.	12.02.2019 15:32
9	Der Sound war blechern. Ich hatte das Gefühl als wäre es eine schlechter MP3 Rip gewesen.	12.02.2019 15:01

Abbildung 29: Negative Reaktion auf den Ton der Anwendung (Version A)

Frage 19a: Warum hat Dir Version A besser gefallen?

NR.	BEANTWORTUNGEN	DATUM
1	geltend akustische Perspektiven	16.02.2019 17:58
2	geltend Schwerpunkt Ton Bei Version 2 gab es keine Binauralisierung des Audios, d.h. es wurde einfach ein statischer Mix abgespielt, der immer gleich geblieben ist, wenn man den Kopf gedreht hat. Aber wenn es für den User schon die Möglichkeit gibt, auf der Bühne zu stehen, dann ist es auf alle Fälle sinnvoll, wenn sich die Instrumente der Kopfbewegung des Users anpassen. Auch wenn die Binauralisierung bei Version 1 noch nicht super funktioniert hat, war es vom Gesamterlebnis dennoch deutlich besser.	15.02.2019 14:31
3	geltend Die erste Version klingt weitaus räumlicher. Die zweite Version hörte sich für mich an, als fehlte die räumliche Komponente komplett.	15.02.2019 12:35
4	geltend Räumlicher Klang hat das Erlebnis authentischer gemacht	15.02.2019 11:05
5	geltend Schwerpunkt Ton Sound kommt gefühlt "mehr dreidimensional"	14.02.2019 15:34
6	geltend Schwerpunkt Ton Weil in Version 2 das Audio praktisch Kopfrichtungsunabhängig war.	14.02.2019 12:43
7	geltend Raumklang	14.02.2019 12:16
8	geltend Räumlicher Ton	14.02.2019 10:44
9	geltend Der Sound klingt räumlicher	13.02.2019 15:11
10	geltend Fand den Sound besser, da er auch räumlich wahrnehmbar war.	13.02.2019 14:15
11	geltend Durch den 3D sound bekommt man des Gefühl wirklich auf der Bühne zu stehen.	13.02.2019 13:47
12	geltend Schwerpunkt Ton Der Klang wirkte räumlicher. Man konnte einzelne Elemente nur durchs Gehör orten.	12.02.2019 15:33
13	geltend Der 3D Effekt war bei der 2ten Version gar nicht vorhanden. Ich habe ihn richtig vermisst.	12.02.2019 15:02
14	geltend Der Ton klang irgendwie voller	12.02.2019 13:15

Abbildung 30: Warum hat Dir Version A besser gefallen?

Frage 19b: Warum hat Dir Version B besser gefallen?

NR.	BEANTWORTUNGEN	DATUM
1	geltend Als man die erste Version am Anfang gesehen hat, war das schon sehr gut, aber als Version 2 danach gezeigt wurde, konnte ich eine bessere Qualität feststellen. Auch der Sound bei Version 2 war im Vergleich zu Version 1 besser, klarer und räumlicher.	15.02.2019 12:13
2	geltend Audio / Bild gefühlt klarer	15.02.2019 10:28
3	geltend Schwerpunkt Ton Fühlte sich etwas räumlicher an	13.02.2019 13:04

Abbildung 31: Warum hat Dir Version B besser gefallen?

Anhang C: Fotografien der Produktionen



Abbildung 32: Aufbau im Merlin (Unten das 360°-Kamera-Rig. Oben das IRT-Kreuz)



Abbildung 33: Konzert in Ulm

Anhang D: Datenträger

Im beigefügten Datenträger sind die Videos im Youtube-Format mit Ambisonics-Ton erster Ordnung enthalten. Außerdem sind alle 3rd Order Ambisonics-Mischungen, die Monosignale des 8.0 Lautsprecher-Setups, die Stereo-Downmixe, die Ergebnisse des Fragebogens, sowie eine elektronische Version dieser Arbeit beigefügt.