

# Audio für 360°-Video

**Qualitätskriterien und Produktionsworkflow im Kontext musikalischer  
Aufführungspraxis am Beispiel des Hessischen Rundfunks**

Bachelorarbeit im Studiengang Audiovisuelle Medien

vorgelegt von

Marcel Remy

Matr.-Nr.: 30421

am 31. August 2018

an der Hochschule der Medien Stuttgart zur Erlangung des akademischen

Grades

Bachelor of Engineering

Erstprüfer: Prof. Oliver Curdt

Zweitprüfer: Dipl.-Ing. Christoph Niesar

Praxisbetreuer: M.A. Simon Binnenmarsch

---

## Ehrenwörtliche Erklärung

„Hiermit versichere ich, Marcel Remy, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: *„Audio für 360°-Video - Qualitätskriterien und Produktionsworkflow im Kontext musikalischer Aufführungspraxis am Beispiel des Hessischen Rundfunks“* 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.“

---

Marcel Remy

Stuttgart, den 31.08.2018

---

## Kurzfassung

Die vorliegende Arbeit befasst sich mit den Besonderheiten der Audioproduktion für 360°-Videos im Kontext musikalischer Aufführungspraxis. Neben den theoretischen Grundlagen werden die Anforderungen aufgezeigt, die das Medium des 360°-Videos an die Audioproduktion stellt und was dies speziell für musikalische Inhalte bedeutet. Es wird der gesamte Produktionsprozess von der Aufnahme bis zur Distribution erörtert und praktisch erprobt. Verschiedene Werkzeuge der Aufnahmetechnik und Postproduktion werden aufgezeigt. Zusätzlich wird die Entwicklung und Anwendung eines stereofonen 3D-Mikrofonaufnahmeverfahrens vorgestellt. Ziel dieser Bachelorthesis ist die Erarbeitung eines praktikablen Workflows, für die Produktion von 360°-Videos mit hochwertigem 3D-Audio im Rahmen von Live-Konzerten.

## Abstract

This work deals with the specialties of audio-production for 360°-Video in the context of musical live performances. Among the theoretical foundations, the requirements of the 360°-Video medium concerning the audio-production of musical content is pointed out. The complete production process, including recording, postproduction and distribution, is discussed and field-tested. A variety of microphone-techniques und postproduction-tools are presented. Furthermore, the development and application of a stereophonic 3D microphone recording technique is revealed. This bachelor thesis aims for the development of a practical workflow to produce 360°-Video with high quality 3D-Audio of musical live performances.

---

## Danksagung

An dieser Stelle möchte ich all jenen danken, die durch ihre fachliche und persönliche Unterstützung zum Gelingen dieser Bachelorarbeit beigetragen haben. Besonderer Dank gilt Professor Oliver Curdt, Christoph Niesar und Simon Binnenmarsch für die Betreuung meiner Arbeit.

Ich danke den Kollegen der Hörfunkproduktion des Hessischen Rundfunks Robin Bös, Lisa Harnest, Alexander Kolb, Frank Wagner und Freya Hartwig, ohne die die Testproduktion mit dem hr Sinfonieorchester nicht hätte entstehen können.

Herzlich bedanken möchte ich mich bei meiner Familie und meinen Freunden Bernd Rischner, Jörg Seebold, Silvana Moers, Jan Bielezki und Pablo Knupfer für ihren Beistand und ihre Unterstützung.

Des Weiteren danke ich meinen Sponsoren Audio Ease, Avid, Noisemakers, Sennheiser und Waves für den Support der Produktionen mit hochwertigen Produkten.

# Inhaltsverzeichnis

|  |      |
|--|------|
| Ehrenwörtliche Erklärung .....                             | II   |
| Kurzfassung .....  | III  |
| Abstract .....   | III  |
| Danksagung.....  | IV   |
| Inhaltsverzeichnis .....                                   | V    |
| Abbildungsverzeichnis.....                                 | VIII |
| Abkürzungsverzeichnis.....                                 | IX   |
| Formelverzeichnis.....                                     | IX   |
| 1. Einleitung .....  | 1    |
| 2. 360°-Video – Definition und Abgrenzung.....             | 2    |
| 2.1 Definition 360°-Video .....                            | 2    |
| 2.2 Definition Virtual Reality.....                        | 3    |
| 2.3 Gemeinsamkeiten von VR und 360°-Video .....            | 3    |
| 2.4 Unterschiede von VR und 360°-Video.....                | 3    |
| 2.5 Vorteile des 360°-Video-Formats .....                  | 4    |
| 3. Räumliches Hören .....                                  | 5    |
| 3.1 Richtungswahrnehmung.....                              | 6    |
| 3.1.1 Horizontale Ebene.....                               | 6    |
| 3.1.2 Vertikale Ebene.....                                 | 7    |
| 3.2 Entfernungswahrnehmung.....                            | 9    |
| 4. Binauraltechnik .....                                   | 10   |
| 4.1 HRTF/HRIR .....  | 10   |
| 4.2 Binauralaufnahmetechnik.....                           | 11   |
| 4.3 Binauralsynthese und Head-Tracking .....               | 13   |
| 5. Immersion und Präsenz.....                              | 14   |
| 5.1 Immersion.....   | 14   |
| 5.2 Präsenz.....   | 15   |
| 5.3 Erzeugen von Immersion.....                            | 15   |
| 6. Anforderungen an die Tongestaltung für 360°-Video ..... | 17   |
| 7. 3D-Audioformate für 360°-Video.....                     | 20   |
| 7.1 Next Generation Audio.....                             | 20   |
| 7.2 Ambisonics .....                                       | 22   |
| 7.2.1 Grundkonzept.....                                    | 22   |
| 7.2.2 Funktionsweise.....                                  | 23   |

---

|        |   |    |
|--------|---|----|
| 7.2.3  | Ordnungszahl .....  | 24 |
| 7.2.4  | Ambisonics-Formate .....  | 26 |
| 7.2.5  | Vorteile.....   | 27 |
| 7.2.6  | Nachteile .....   | 28 |
| 7.2.7  | Ambisonics für 360°-Video .....   | 28 |
| 8.     | Ausspielwege und Plattformen für 360°-Videos.....   | 29 |
| 9.     | Mikrofonaufnahmetechnik für 360°-Videoproduktionen .....  | 31 |
| 9.1    | Ambisonics-Mikrofone .....  | 31 |
| 9.1.1  | FOA-Mikrofone .....   | 31 |
| 9.1.2  | HOA-Mikrofone.....  | 34 |
| 9.2    | Stereofone 3D-Mikrofonarrays .....  | 34 |
| 9.2.1  | 2L-Cube .....   | 35 |
| 9.2.2  | Twins Square/Cube .....   | 36 |
| 9.2.3  | Coincident Z-Microphone Technique.....  | 37 |
| 9.2.4  | ESMA 3D.....  | 38 |
| 9.2.5  | ORTF 3D.....  | 39 |
| 9.3    | Polymikrofonie .....  | 41 |
| 9.3.1  | Einzelmikrofonie.....   | 41 |
| 9.3.2  | Stützmikrofonie.....  | 42 |
| 9.4    | Kamerastandpunkt und Perspektive .....  | 44 |
| 9.5    | Zusammenfassung .....   | 44 |
| 10.    | Entwicklung und Umsetzung eines stereofonen 3D-Audioaufnahmeverfahrens für 360°-<br>Videoproduktionen ..... | 45 |
| 10.1.1 | Verfahren .....   | 46 |
| 10.1.2 | Praktische Umsetzung .....  | 52 |
| 11.    | Audiopostproduktion für 360°-Video.....   | 55 |
| 11.1   | Produktionsformat .....   | 55 |
| 11.2   | DAW-Anforderungen.....  | 56 |
| 11.3   | Mischungs-Konzept für 360°-Video .....  | 56 |
| 11.4   | Softwaretools.....  | 58 |
| 11.4.1 | Sennheiser AMBEO A-B Format Converter.....  | 58 |
| 11.4.2 | Audio Ease 360 Pan Suite.....   | 59 |
| 11.4.3 | Noisemakers Ambi HD Bundle .....  | 61 |
| 11.4.4 | Facebook 360 Workstation.....   | 62 |
| 11.4.5 | Weitere Tools.....  | 63 |
| 12.    | Testproduktionen .....  | 64 |

---

|        |  |      |
|--------|--|------|
| 12.1   | Teil 1 – Studio-Session Akustik Duo .....            | 64   |
| 12.1.1 | Aufnahme .....                                       | 64   |
| 12.1.2 | Nachbearbeitung .....                                | 65   |
| 12.1.3 | Beobachtungen .....                                  | 66   |
| 12.2   | Teil 2 – Livekonzert hr Sinfonieorchester .....      | 66   |
| 12.2.1 | Aufnahme .....                                       | 69   |
| 12.2.2 | Nachbearbeitung .....                                | 70   |
| 12.2.3 | Beobachtungen .....                                  | 73   |
| 13.    | Workflowempfehlung für den Hessischen Rundfunk ..... | 74   |
| 14.    | Fazit und Ausblick .....                             | 77   |
|        | Literaturverzeichnis .....                           | X    |
|        | Anhang .....   | XVII |

## Abbildungsverzeichnis

|   |    |
|---|----|
| Abbildung 1 - Die drei Freiheitsgrade bei 360°-Video.....                                     | 4  |
| Abbildung 2 - Kopfbezogenes Koordinatensystem zur Richtungsbeschreibung .....                 | 5  |
| Abbildung 3 - Richtungsbestimmende Bänder nach Blauert.....                                   | 8  |
| Abbildung 4 - Informationsgewinnung durch Kopfdrehungen .....                                 | 9  |
| Abbildung 5 - Omni Binaural Mikrofon von 3Dio .....   | 12 |
| Abbildung 6 - Kanalbasierter Workflow .....   | 20 |
| Abbildung 7 – Szenenbasierter Workflow.....   | 21 |
| Abbildung 8 – Objektbasierter Workflow .....  | 22 |
| Abbildung 9 - Sphärische Harmonische bis zur dritten Ordnung.....                             | 25 |
| Abbildung 10 - Signale eines Ambisonics-A-Format Mikrofons.....                               | 32 |
| Abbildung 11 - Sennheiser AMBEO VR Mic mit (links) und ohne Korb (rechts).....                | 33 |
| Abbildung 12 - 2L Mikrofonarray .....   | 36 |
| Abbildung 13 - Twins Cube Mikrofonarray .....   | 37 |
| Abbildung 14 - 3D-Mikrofonarray nach der Coincident Z-Microphone Technique.....               | 38 |
| Abbildung 15 - ESMA 3D Mikrofonarray.....   | 39 |
| Abbildung 16 - Koinzidente Mikrofonanordnung in der Vertikalebene beim Schoeps ORTF 3D.....   | 40 |
| Abbildung 17 - Schoeps ORTF 3D Mikrofonarray im Windkorb.....                                 | 41 |
| Abbildung 18 - Laufzeit- und Pegeldifferenzen für Nierenmikrofone 24 cm/90° .....             | 48 |
| Abbildung 19 - Kohärenzfunktion für Nierenmikrofone 24 cm/90° .....                           | 49 |
| Abbildung 20 - Vertikales Signalübersprechen bei Aufnahme und Wiedergabe .....                | 50 |
| Abbildung 21 - Reduktion des vertikalen Signalübersprechens .....                             | 50 |
| Abbildung 22 - Virtuelle Lautsprecher Anordnung.....  | 51 |
| Abbildung 23 - Signaldekorrelation bei Nierenmikrofonen mit einem Versatzwinkel von 53° ..... | 52 |
| Abbildung 24 - Grundaufbau der Mikrofonschienen.....  | 53 |
| Abbildung 25 - Kontrolle des Öffnungswinkels mit einem Zollstock.....                         | 54 |
| Abbildung 26 - Der komplette Aufbau des Mikrofonarrays.....                                   | 55 |
| Abbildung 27 – Ambisonics-Monitoring via Spatializer.....                                     | 57 |
| Abbildung 28 - Sennheiser AMBEO A - B Format Converter.....                                   | 59 |
| Abbildung 29 - Darstellung von Schallquellen als farbige Kreise .....                         | 60 |
| Abbildung 30 - Sphärische Aufsicht im Ambisonics-Panner "Ambi Pan" von Noisemakers.....       | 61 |
| Abbildung 31 - Produktionsaufbau "Akustik Duo" im Studio U54 der HdM.....                     | 64 |
| Abbildung 32 - Hallenplan des Eröffnungskonzertes.....  | 68 |
| Abbildung 33 - Orchesteraufbau für das Eröffnungskonzert.....                                 | 70 |
| Abbildung 34 - Verteilung der Signale des 3D-Mikrofonarrays .....                             | 71 |
| Abbildung 35 - Videobasiertes Panning der Stützmikrofone .....                                | 72 |

## Abkürzungsverzeichnis

|        |   |
|--------|---|
| AAC    | Advanced Audio Coding                         |
| ACN    | Ambisonics Channel Numbering                  |
| ADM    | Audio Definition Model                        |
| AR     | Augmented Reality                             |
| BRIR   | Binaural Room Impulse Response                |
| DAW    | Digital Audio Workstation                     |
| DOF    | Degrees Of Freedom                            |
| DSP    | Digital Signal Processing                     |
| ESMA   | Equal Segement Microphone Array               |
| FOA    | First Order Ambisonics                        |
| FuMa   | Furse and Malham                              |
| HMD    | Head Mounted Display                          |
| HOA    | Higher Order Ambisonics                       |
| hr     | Hessischer Rundfunk                           |
| HRIR   | Head Related Impulse Response                 |
| HRTF   | Head Related Tranfer Function                 |
| IRT    | Institut für Rundfunktechnik                  |
| ITU    | International Telecommunication Union         |
| MADI   | Multi Channael Audio Digital Interface        |
| MPEG-H | Moving Picture Experts Group Standard H       |
| MR     | Mixed Reality                                 |
| NOS    | Nederlandse Omroep Stichting                  |
| ORTF   | Office de Radiodiffusion Télévision Française |
| SID    | Single Index Designation                      |
| VR     | Virtual Reality                               |

## Formelverzeichnis

|   |    |
|---|----|
| Formel (1) - Fourier-Besselreihe                | 23 |
| Formel (2) - Ambisonics-Komponenten             | 24 |
| Formel (3) - Übertragungskanäle für Ambisonics  | 25 |
| Formel (4) - Ambisonics A/B-Format Matrizierung | 32 |

# 1. Einleitung

Neue Medienformen, wie Virtual Reality und 360°-Video, gewinnen zunehmend an Bedeutung und haben laut Specht „das Potenzial, nach PCs und Mobilgeräten die dritte große Computing-Plattform zu werden“<sup>1</sup>. Im Jahr 2015 wurden weltweit 10 Millionen Virtual Reality Headsets verkauft und im Jahr 2020 sollen es schon 75 Millionen sein<sup>2</sup>. Inzwischen wird 360°-Video mit 3D-Audio von großen Internetplattformen, wie *Youtube* und *Facebook*, unterstützt, wodurch sich potentiell eine enorme Reichweite ergibt. Darüber hinaus sind weitere Distributionsplattformen entstanden, die sich auf die Verbreitung von Virtual Reality- (VR) und 360°-Medien spezialisiert haben. Diese ermöglichen die Wiedergabe auf Smartphones und VR-Brillen mit Hilfe spezieller Apps. Vieles spricht also dafür, dass 360°-Videos eine immer größere Rolle in der Mediennutzung spielen werden. Auch Gershon ist sich dessen sicher und hält fest: „It ultimately will change entertainment“<sup>3</sup>.

Diese neue Form der Mediennutzung stellt gleichzeitig Chance und Risiko für den Hessischen Rundfunk und den Rundfunk im Allgemeinen dar. Durch den Einstieg in die Produktion von hochwertigen 360°-Videoinhalten kann der Hessische Rundfunk auf die sich verändernde Mediennutzung reagieren und seinen Programmauftrag auch im Bereich der neuen Medien nachkommen. Zudem besteht die Chance, sich neue Zielgruppen zu erschließen und Mediennutzer zu erreichen, die über das bisherige Fernseh-, Radio- und Onlineangebot nicht erreicht werden können.

Musikalische Aufführungen erfreuen sich großer Beliebtheit und die Besucherzahlen von Live-Konzerten steigen seit einigen Jahren kontinuierlich<sup>4</sup>. Zudem wird das Publikum von Klassik-Konzerten in Deutschland zunehmend jünger<sup>5</sup>. Das Medium des 360°-Video bietet die Möglichkeit das Konzerterlebnis in das Wohnzimmer zu bringen und musikalische Aufführungen aus einer völlig neuen Perspektive zu erleben. Mit der hr-Bigband und dem hr-Sinfonieorchester besitzt der Hessische Rundfunk zwei renommierte Klangkörper. Pro Saison werden rund 50 Konzerte des hr-Sinfonieorchesters und 25 Projekte der hr-Bigband aufgezeichnet. Dies stellt eine vielversprechende Grundlage für die Produktion von 360°-Videos im Kontext musikalischer Aufführungen mit den Klangkörpern des Hessischen Rundfunks dar.

Der Hessische Rundfunk besitzt langjährige Erfahrung in der Produktion technisch hochwertiger audiovisueller Medien mit Stereo- und Surround-Ton, doch die Produktion hochwertiger

---

<sup>1</sup> Specht (2018)

<sup>2</sup> vgl. Ebenda

<sup>3</sup> Gershon zitiert nach Grover (2017)

<sup>4</sup> vgl. Deutsches Musikinformationszentrum (2016); Deutsches Musikinformationszentrum (2018)

<sup>5</sup> vgl. Berg (2017)

360°-Videoinhalte erfordert den Einsatz von 3D-Audio und bedarf eines anderen tongestalterischen Ansatzes. Die vorliegende Arbeit beschreibt deshalb die Grundlagen und Anforderungen der Produktion von Audio für 360°-Video im Kontext musikalischer Aufführungspraxis und soll dabei die folgenden Fragen beantworten: Welche technischen und gestalterischen Anforderungen müssen bei der Produktion von Audio für 360°-Video beachtet werden und was sind die qualitätsbestimmenden Faktoren? Welche Distributionswege stehen für 360°-Video zur Verfügung und was muss hinsichtlich des Produktionsformates beachtet werden? Wie kann eine geeignete Mikrofon-Aufnahmetechnik für 360°-Video aussehen und welche Werkzeuge kommen innerhalb der Postproduktion zum Einsatz? Abschließend soll die Frage beantwortet werden, wie ein empfehlenswerter Produktionsworkflow für den Hessischen Rundfunk gestaltet sein sollte.

Dazu wird zunächst 360°-Video als Medienform definiert. Anschließend werden die Grundlagen des räumlichen Hörens und der Binauraltechnik erläutert. Es folgt eine Analyse der wichtigsten Gestaltungsmerkmale von 360°-Video, aus denen Anforderungen an die Tongestaltung abgeleitet werden. Im nächsten Schritt wird das Audioformat, welches für 360°-Videoproduktionen zum Einsatz kommt, detailliert vorgestellt und verschiedene Distributionsmöglichkeiten aufgezeigt. Für die praktische Umsetzung erfolgt eine Untersuchung und Entwicklung verschiedener Mikrofonierungstechniken für 3D-Audio. Zusätzlich werden die Besonderheiten und Werkzeuge der Postproduktion vorgestellt. Im Anschluss daran werden zwei Beispielproduktionen beschrieben, in deren Rahmen verschiedene Mikrofonierungstechniken und Nachbearbeitungsschritte erprobt wurden. Die vorliegende Arbeit schließt ab mit einer Workflowempfehlung für den Hessischen Rundfunk zur Audioproduktion für 360°-Videos im Kontext musikalischer Aufführungspraxis.

## **2. 360°-Video – Definition und Abgrenzung**

360°-Videos und VR-Inhalte erfreuen sich zunehmender Beliebtheit und werden vielfach in den Medien diskutiert. Auch wenn 360°-Video und VR viele Gemeinsamkeiten haben, so gibt es doch einige grundlegende Unterschiede, die beachtet werden müssen, wenn man solche Medien produzieren will. Dieses Kapitel gibt daher Definitionen für 360°-Video und VR und zeigt deren Unterschiede auf.

### **2.1 Definition 360°-Video**

Im Gegensatz zur klassischen Videotechnik ist bei 360°-Video das Blickfeld der Kamera erweitert und erfasst eine Szene mit einem Rundumblick von 360°. Bei der Reproduktion von 360°-Video-Inhalten befindet sich der Zuschauer im Zentrum des Mediums und kann durch seine Blickrichtung den Bildausschnitt jederzeit selbst festlegen. 360°-Videos sind lineare Medien. Der Betrachter kann zwar den Bildausschnitt, jedoch nicht den zeitlichen Ablauf beeinflussen.

## 2.2 Definition Virtual Reality

Das Medium VR besteht aus einer computergenerierten Welt, die sich interaktiv erfahren lässt. Diese Welt spricht dabei möglichst viele Sinne des Nutzers an und umgibt ihn vollständig. Mittels geeigneter Technologien wie Head-Mounted Displays (HMD)<sup>6</sup>, Head-Tracking und Kopfhörern wird die reale Welt weitestgehend ausgeblendet, um ein möglichst intensives Erfahren der virtuellen Welt zu ermöglichen. In diesem Zusammenhang ist der Begriff der „Immersion“<sup>7</sup> von zentraler Bedeutung.

Berger und Kiani definieren VR wie folgt: „Virtual Reality ist ein Medium, welches aus einer computergenerierten, interaktiven Welt besteht, die den Nutzer vollständig umgibt und durch die Ansprache ein oder mehrerer Sinne mittels geeigneter Systeme besonders immersiv erlebt werden kann.“<sup>8</sup>

Neben der VR bestehen weitere Mischformen wie Augmented Reality (AR) und Mixed Reality (MR), die sich zumindest durch eine Teil-Virtualität des Mediums auszeichnen. Schütze prägt in diesem Zusammenhang den Begriff der „New Realities“ um die Medien VR/AR/MR/360°-Video gleichermaßen zu erfassen<sup>9</sup>. Dieterle et. al stellt fest, dass alle Medien der „New Realities“ in Bezug auf Audio dasselbe Zielformat anstreben - binaurales<sup>10</sup> 3D-Audio mit Headtracking<sup>11</sup>.

## 2.3 Gemeinsamkeiten von VR und 360°-Video

Die Medien 360°-Video und VR versuchen beide einen möglichst hohen Grad an Immersion zu erreichen. Zur Wiedergabe der Inhalte bedienen sie sich HMDs und dreidimensionalem Audio, das zumeist binauralisiert über Kopfhörer wiedergegeben wird. Im Gegensatz zu konventionellen Bildschirmen kann durch HMDs die reale Welt vollständig ausgeblendet werden und erhöht so den Grad an Immersion. Durch den Einsatz von Head-Tracking reagiert der Inhalt auf die Kopfbewegung des Nutzers und erzeugt so ein hohes Maß an Realismus. Dazu sollte sich durch das Head-Tracking nicht nur die visuelle, sondern auch die akustische Perspektive ändern.

## 2.4 Unterschiede von VR und 360°-Video

360°-Video kann als Unterkategorie von VR betrachtet werden, nutzt dabei zum Teil dieselben Technologien und verfolgt ähnliche Absichten. VR-Inhalte stammen eher aus dem Bereich der Computergrafik, während 360°-Videos ihren Ursprung meist in der Videotechnik haben.

---

<sup>6</sup> Umgangssprachlich werden HMDs als VR-Brille bezeichnet

<sup>7</sup> Eintauchen in die virtuelle Umgebung, vgl. Kapitel 5.Immersion und Präsenz, S. 14

<sup>8</sup> Berger u.a. (2017)

<sup>9</sup> vgl. Schütze; Irwin-Schütze (2018), S. 1

<sup>10</sup> Zweikanaliges Format für Kopfhörer, vgl. 4

<sup>11</sup> vgl. Dieterle u.a. (2018)

Im Gegensatz zu VR-Inhalten, kann 360°-Video nicht nur über HMDs wiedergegeben werden. Es ist ebenso eine Wiedergabe über konventionelle Bildschirme (PCs, Laptops) oder Handheld-Systeme (Smartphones, Tablet Computer) möglich, obgleich dies den Grad an Immersion reduziert.

Des Weiteren unterscheiden sich die Medien durch den Grad an Interaktivität. Nach Susal, Krauss, Tsingos & Altman lassen sich VR-Inhalte in interaktive und lineare Anwendungen unterteilen<sup>12</sup>. Klassische VR-Inhalte sind interaktiv, wodurch der Nutzer das Geschehen und den Ablauf aktiv steuert. Sämtliche Inhalte müssen daher in Echtzeit berechnet werden. 360°-Video kann der linearen VR zugeordnet werden. Bei der linearen VR ist der zeitliche Ablauf festgelegt, der Inhalt selbst kann nicht verändert werden, was auch durch den Begriff „Passive VR“ zum Ausdruck kommt. Bei der Wiedergabe werden die Blickwinkel des Nutzers auf bereits aufgezeichnetes Video- und Audiomaterial berechnet. Der Grad an Freiheit (DOF)<sup>13</sup> ist im Gegensatz zu VR geringer. Wie in Abbildung 1 dargestellt ist, weist 360°-Video nur drei DOF auf, da lediglich Rotationen des Kopfes möglich sind. Bei interaktiven VR Inhalten sind dagegen zusätzlich Translationen möglich, wodurch sechs DOF erreicht werden.

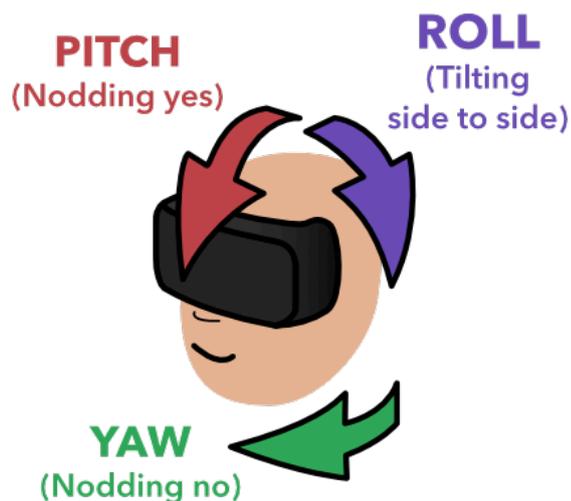


Abbildung 1 - Die drei Freiheitsgrade bei 360°-Video<sup>14</sup>

## 2.5 Vorteile des 360°-Video-Formats

Laut Schütze hat das 360°-Video-Format das größte Potenzial für eine schnell wachsende Zahl an Nutzern, da dieses Format mit einem einfachen Headset-Adapter für Smartphones konsumiert werden kann<sup>15</sup>.

<sup>12</sup> vgl. Susal u.a. (2016), S. 2

<sup>13</sup> Degrees of freedom

<sup>14</sup> Bildquelle: Alistair (2018)

<sup>15</sup> vgl. Schütze; Irwin-Schütze (2018), S. 26

Da 360°-Video seinen Ursprung in der Videotechnik hat, stellt dieses Medium einen geeigneten Einstiegspunkt in die Produktion immersiver VR-Medien für die Rundfunkanstalten dar, die auf ihrem großen Erfahrungsschatz in der Produktion linearer Video- und Audioinhalte aufbauen können.

360°-Videos können im Gegensatz zu interaktiven VR-Inhalten mit relativ geringem Aufwand produziert werden. Insbesondere im Kontext musikalischer Aufführungen im Rundfunk, die ohnehin aufgezeichnet werden, können mit geringem zusätzlichen Zeit- und Kostenaufwand 360°-Videoinhalte produziert werden.

Des Weiteren lässt sich mit 360°-Video eine wesentlich größere Reichweite erzielen, als mit interaktiven VR-Inhalten. Das liegt zum einen daran, dass für die Wiedergabe nicht zwingend HMDs erforderlich sind, welche noch nicht in der breiten Masse angekommen sind. Zum anderen ist es möglich 360°-Videoinhalte über große Online-Plattformen wie Facebook und YouTube zu verbreiten.

### 3. Räumliches Hören

Räumliches Hören beschreibt die Wahrnehmung der Richtung und der Entfernung eines Schallereignisses. Dies schließt auch die Wahrnehmung der räumlichen Ausdehnung eines Schallereignisses mit ein<sup>16</sup>. Bei korrekter Lokalisierung stimmen Hörereignisrichtung und Schallereignisrichtung überein. Zur Beschreibung der Lokalisation wird ein kopfbezogenes Koordinatensystem eingesetzt, welches in Abbildung 2 dargestellt ist. Dabei werden drei Dimensionen unterschieden: Die horizontale Ebene, die vertikale Ebene und die frontale Ebene. Dieses Kapitel zeigt die grundsätzlichen Mechanismen auf, nach denen das räumliche Hören funktioniert. Das Hören außerhalb der frontalen Ebene weist Besonderheiten auf, die bei der Produktion von dreidimensionalen Audioinhalten beachtet werden müssen.

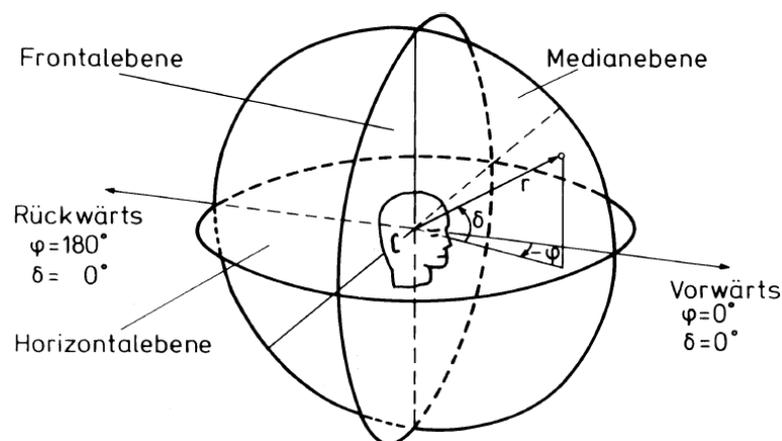


Abbildung 2 - Kopfbezogenes Koordinatensystem zur Richtungsbeschreibung

<sup>16</sup> vgl. Dickreiter; Goeres-Petri (2013), S. 127

## 3.1 Richtungswahrnehmung

Die Wahrnehmung der Richtung wird bestimmt von interauralen Laufzeit- und Pegeldifferenzen. Abhängig von der Ebene, in der ein Schallereignis auftritt und der Beschaffenheit des Signals, sind verschiedene Faktoren für die Lokalisation verantwortlich. Dabei ist die Güte der Wahrnehmung nicht auf allen Ebenen gleich und vor allem auf der horizontalen Ebene im frontalen Bereich stärker ausgeprägt<sup>17</sup>.

### 3.1.1 Horizontale Ebene

Die Lokalisation auf horizontaler Ebene erfolgt durch interaurale Laufzeit- und frequenzabhängige Pegeldifferenzen. Befindet sich eine Schallquelle außerhalb der 0°-Achse, kommt es so zu unterschiedlichen Signalen am linken und rechten Ohr, da eines der Ohren früher erreicht wird und durch den Kopf Verdeckungseffekte auftreten, die frequenzabhängige Pegeldifferenzen hervorrufen<sup>18</sup>.

Der Abstand der Ohren zueinander beträgt ca. 17 cm, was zu einer maximalen Laufzeitdifferenz von 0,63 ms führt. Laut Dickreiter beträgt die geringste noch wahrnehmbare Laufzeitdifferenz 0,03 ms und bedingt die Lokalisation einer Schallquelle bei 3° bis 5° außerhalb der Mittelachse<sup>19</sup>. Die Lokalisation ist jedoch nur dann eindeutig, wenn die Wellenlänge des Schalls größer ist, als die Wegdifferenz zu beiden Ohren<sup>20</sup>. Befindet sich die Schallquelle einer reinen Sinusschwingung auf horizontaler Ebene unter 90°, so erreicht dieses Signal eines der Ohren mit der maximalen Laufzeitdifferenz von 0,63 ms. In dieser Zeit durchläuft eine Schwingung von 1,59 kHz eine volle Periode, womit beide Ohrsignale phasengleich sind und es dadurch zu einer Fehllokalisierung im Frontbereich kommen kann. Bis zu dieser Frequenz kann das Gehör Phasendifferenzen beider Ohrsignale direkt erkennen<sup>21</sup>. Für höhere Frequenzen müssen vom Ohr die Hüllkurven beider Ohrsignale verglichen werden. Dabei können breitbandige, transientenhaltige Signale wie Sprache oder Musik, einfacher lokalisiert werden<sup>22</sup>.

Neben den interauralen Laufzeitdifferenzen treten zusätzlich interaurale Pegeldifferenzen auf. Diese werden durch Abschattungseffekte von Kopf und Ohrmuschel hervorgerufen<sup>23</sup>. Dies tritt jedoch nur für Frequenzen auf, deren Wellenlänge klein gegenüber den Abmessungen des Kopfes sind, andernfalls treten Beugungserscheinungen auf. Eine größere Pegeldifferenz führt dabei zu einer Lokalisierung der Schallquelle weiter außerhalb der Mittelachse. Diese Abhängigkeit von Pegeldifferenz und Schallereignisrichtung ist jedoch unregelmäßig und in starkem Maße von der spektralen

---

<sup>17</sup> vgl. Görne (2015), S. 126

<sup>18</sup> vgl. Wenzel u.a. (2018), S. 11-16

<sup>19</sup> vgl. Dickreiter; Goeres-Petri (2013), S. 129

<sup>20</sup> vgl. Görne (2015), S. 126-127

<sup>21</sup> vgl. Dickreiter; Goeres-Petri (2013), S. 129

<sup>22</sup> vgl. Görne (2015), S. 127-128

<sup>23</sup> vgl. Howard; Angus (2009), S. 111

Zusammensetzung des Signals abhängig. Bei breitbandigen Signalen ergeben sich frequenzabhängige Pegelunterschiede. Diese unterstützen ebenfalls die Lokalisation, auch wenn sie nur unterbewusst wahrgenommen werden. Die Auswertung dieser Klangfarbenunterschiede basiert auf Erfahrungswerten und Lernprozessen, weshalb uns bekannte Stimmen beispielweise deutlich einfacher korrekt lokalisiert werden können<sup>24</sup>.

Für die Lokalisierung tiefer Frequenzen sind also vornehmlich interaurale Laufzeitdifferenzen verantwortlich und für die Lokalisierung hoher Frequenzen interaurale Pegeldifferenzen<sup>25</sup>. Beim natürlichen Hören werden Laufzeit- und Pegeldifferenzen simultan ausgewertet. Die Zusammenhänge von Klangfarbe, Pegel und Einfallswinkel sind dabei in starkem Maße von der Signalbeschaffenheit abhängig und komplex<sup>26</sup>. Es lässt sich festhalten, dass für die Lokalisation auf der horizontalen Ebene, Laufzeit- oder Pegeldifferenzen allein nicht zu einer optimalen Lokalisation führen.

Die Genauigkeit der Lokalisation im 360°-Raum variiert stark und ist für den Bereich der Blickrichtung in der horizontalen Ebene optimiert. Dabei ist die zeitliche und örtliche Auflösung begrenzt. Die Lokalisierungsschärfe beträgt örtlich maximal 1° bei frontaler Schalleinfallrichtung<sup>27</sup> und nimmt bei seitlicher Schalleinfallrichtung ab. Durch die begrenzte zeitliche Auflösung wird ein Wechsel der Schalleinfallrichtung von links/rechts nach 150 ms wahrgenommen, ein Wechsel von vorne/hinten erst nach 250 ms.

Zwei gleiche Signale, welche mit kurzer Verzögerung beim Hörer ankommen, werden nicht als getrennte Schallsignale wahrgenommen. Es stellt sich der Präzedenz-Effekt ein, wonach der zuerst eintreffende Schall entscheidend ist für die Lokalisation, unabhängig davon, aus welcher Richtung der später eintreffende Schall tatsächlich kommt<sup>28</sup>. Der Präzedenzeffekt tritt bis zu einer Verzögerung von etwa 40 ms auf. Bei einer größeren Verzögerung kommt es zur Echobildung, wodurch zwei getrennte Schallereignisse wahrgenommen werden. Haas fand heraus, dass der Präzedenzeffekt selbst dann auftritt, wenn das zeitlich nachfolgende Signal einen bis zu 10 dB höheren Pegel aufweist<sup>29</sup>.

### 3.1.2 Vertikale Ebene

Die Lokalisation auf vertikaler Ebene beruht auf einem anderen Mechanismus und ist weniger genau. Die Lokalisierungsunschärfe beträgt für einen unbekanntem Sprecher 15° und verschlechtert sich weiter mit zunehmendem Erhebungswinkel<sup>30</sup>.

---

<sup>24</sup> vgl. Dickreiter; Goeres-Petri (2013), S. 132

<sup>25</sup> vgl. Howard; Angus (2009), S. 112

<sup>26</sup> vgl. Dickreiter; Goeres-Petri (2013), S. 130

<sup>27</sup> vgl. Görne (2015), S. 126

<sup>28</sup> vgl. Blauert; Braasch (2008), S. 103-105

<sup>29</sup> vgl. Dickreiter; Goeres-Petri (2013), S. 131

<sup>30</sup> vgl. Ebenda, S. 132

Schalleinfall aus der vertikalen Ebene führt nicht zu unterschiedlichen Ohrsignalen, da ein Schallsignal beide Ohren mit derselben Laufzeit und Klangfarbe erreicht. Demnach kann die Lokalisation auf der vertikalen Ebene nicht durch die Auswertung der Signalunterschiede erreicht werden. Stattdessen werden Klangfarbenunterschiede zu einem Schalleinfall aus Blickrichtung ausgewertet. Voraussetzung für eine solche Auswertung ist, dass der Klang aus Blickrichtung zuvor erfahren wurde<sup>31</sup>. Breitbandige und bekannte Signale, die häufig erfahren wurden, lassen sich somit einfacher lokalisieren. Untersuchungen von Blauert zeigen, dass in Abhängigkeit vom Einfallswinkel auf vertikaler Ebene spezifische Frequenzbänder angehoben werden<sup>32</sup>. Diese Bänder werden richtungsbestimmende oder „Blauertsche“ Bänder genannt. Abbildung 3 zeigt, welche Frequenzbereiche entscheidend sind, um eine Schallquelle in der vertikalen Ebene vorne, oben oder hinten zu lokalisieren.

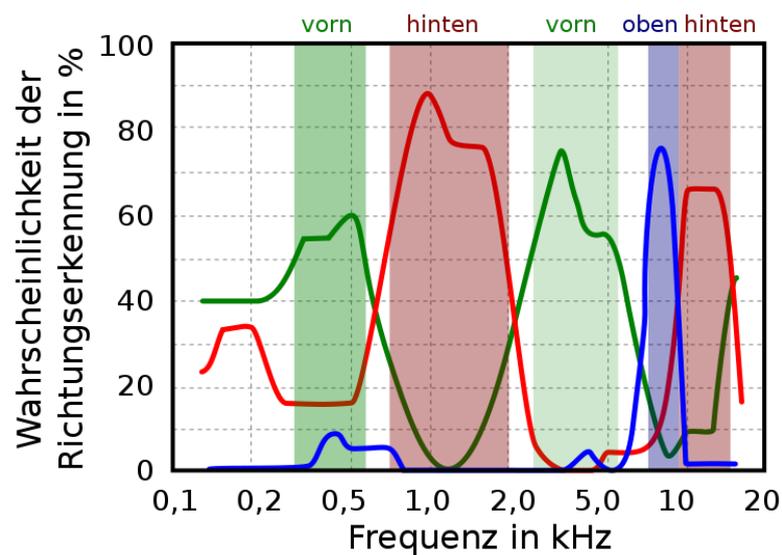


Abbildung 3 - Richtungsbestimmende Bänder nach Blauert<sup>33</sup>

Des Weiteren helfen Kopfbewegungen bei der korrekten Lokalisation von Schallquellen in der vertikalen Ebene. Abbildung 4 zeigt, dass kleinste Drehungen als „Peilbewegungen“ fungieren, wodurch das Gehör zusätzliche Informationen gewinnen kann<sup>34</sup>. Es entstehen interaurale Laufzeit- und Pegeldifferenzen<sup>35</sup>, welche die Bestimmung einer Schallquellenrichtung mit hoher Sicherheit ermöglichen<sup>36</sup>.

<sup>31</sup> vgl. Dickreiter (1995), S. 100

<sup>32</sup> vgl. Blauert (1974), S. 89 ff.

<sup>33</sup> Bildquelle: Wikipedia (2017)

<sup>34</sup> vgl. Blauert; Braasch (2008), S. 88

<sup>35</sup> vgl. Blauert (1974), S. 143 ff.

<sup>36</sup> vgl. Ebenda, S. 153

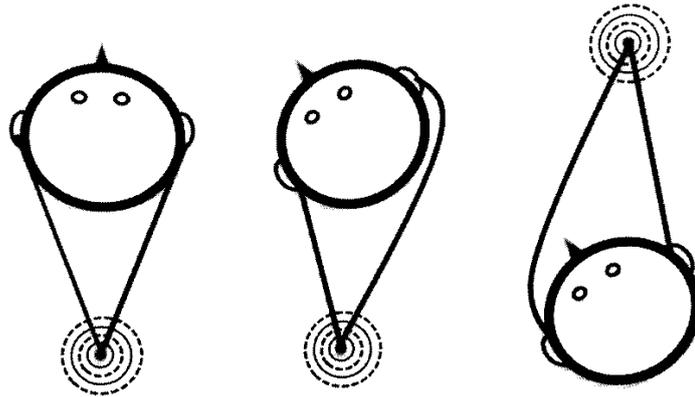


Abbildung 4 - Informationsgewinnung durch Kopfdrehungen<sup>37</sup>

### 3.2 Entfernungswahrnehmung

Für die Wahrnehmung der Entfernung einer Schallquelle werden vornehmlich die Lautstärke und Veränderungen im Klangspektrum eines Hörereignisses ausgewertet. Hörerfahrung und Bekanntheit der Schallquelle sind in diesem Zusammenhang von großer Bedeutung.

Allgemein gilt, dass eine Schallquelle umso leiser wahrgenommen wird, desto weiter sie entfernt ist. Bei kugelförmiger Abstrahlung nimmt der Schalldruckpegel mit jeder Entfernungsverdopplung um 6 dB ab. Bei gerichteter Abstrahlung ist der Verlust geringer<sup>38</sup>, näherungsweise kann von 3 dB ausgegangen werden. Aufgrund der Tatsache, dass natürliche Schallquellen wie Musikinstrumente bestimmte Frequenzbänder kugelförmig und andere gerichtet abstrahlen, ergeben sich mit wachsender Entfernung zur Schallquelle neben Pegelverminderungen auch Klangfarbenänderungen. Untersuchungen von Fletcher und Munson zeigen, dass das Gehör unterschiedliche Empfindlichkeiten für verschiedene Frequenzbereiche aufweist und dass dies zusätzlich vom Schalldruckpegel abhängt<sup>39</sup>. So ist die Empfindlichkeit für tiefe Frequenzen grundsätzlich niedriger als für mittlere Frequenzen und nimmt mit sinkendem Schallpegel noch weiter ab. Diese Eigenschaft des Gehörs verstärkt somit eine Klangfarbenänderung mit sich veränderndem Abstand zu Schallquelle.

Innerhalb geschlossener Räume wird die Entfernungswahrnehmung unterstützt, indem das Verhältnis von Direkt- und Diffusschall ausgewertet wird<sup>40</sup>. Mit zunehmender Entfernung zur Schallquelle nimmt der Direktschallpegel ab, wohingegen der Diffusschallpegel gleich bleibt. In diesem Zusammenhang haben auch das Raumvolumen und die Nachhallzeit Einfluss auf das Verhältnis von Direkt- und

<sup>37</sup> Bildquelle: Roginska (2018), s.103

<sup>38</sup> vgl. Dickreiter; Goeres-Petri (2013), S.132

<sup>39</sup> vgl. Fletcher und Munson zitiert nach Ellermeier; Hellbrück (2008), S. 55

<sup>40</sup> vgl. Dickreiter (1995), S. 100

Diffusschall, weshalb auch hier Hörerfahrung zur Auswertung der Informationen notwendig ist. Neben der Entfernung zur Schallquelle wird auch die Beschaffenheit des Raumes erfasst.

Die Wahrnehmung der Entfernung ist grundsätzlich weniger präzise als die Wahrnehmung der Richtung einer Schallquelle. Das Gehör ist weniger gefordert Hörerfahrungen zur präzisen Entfernungswahrnehmung zu sammeln, da die Einschätzung der Entfernung schneller und präziser über das Auge erfolgen kann<sup>41</sup>. Im Zusammenhang mit der Produktion von 360°-Videos wird die Lokalisation einer Schallquelle im Hinblick auf Entfernung und Richtung durch den optischen Bezug zur Schallquelle unterstützt.

## 4. Binauraltechnik

Wie im vorangegangenen Kapitel erläutert wurde, ist der Mensch in der Lage mit zwei Ohren dreidimensional zu hören und Schall aus allen Richtungen wahrzunehmen. Die Binauraltechnik, die in diesem Kapitel erläutert wird, macht sich diesen Umstand zu Nutze und beruht auf der folgenden Annahme:

„Wenn die Schallsignale an den Trommelfellen bei der Wiedergabe genau denjenigen entsprechen, die sich bei der Aufnahmesituation an den Trommelfellen ergäben, dann werden die Zuhörer auch genau das gleiche hören, was sie in der Aufnahmesituation hören würden.“<sup>42</sup>

Werden einem Schallsignal während der Produktionskette die typischen Merkmale aufgeprägt, die für das Richtungshören notwendig sind, so kann bei der Wiedergabe eine realitätsnahe, dreidimensionale Reproduktion erfolgen. Dabei muss die Wiedergabe zwingend über Kopfhörer erfolgen, da das Signal bereits die interauralen Signaldifferenzen enthält und diese nicht doppelt aufgeprägt werden dürfen. Bei der Binauraltechnik handelt es sich folglich um kopfbezogene Stereophonie oder Kopfhörer-Stereophonie<sup>43</sup>

### 4.1 HRTF/HRIR

Die für das räumliche Hören notwendigen Ohrsignale heißen auch interaurale Signaldifferenzen<sup>44</sup>. Es handelt sich um komplexe Laufzeit und Spektraldifferenzen, die aufgrund von „akustischer Abschattung, Beugung, Verzögerung, Resonanzen und Reflexionen durch Torso, Schulter, Kopf, Ohrmuscheln (pinnae), den Eingang in den Ohrkanal (cavum conchae) und den Ohrkanal selbst“<sup>45</sup> entstehen.

---

<sup>41</sup> vgl. Dickreiter; Goeres-Petri (2013), S. 133

<sup>42</sup> Blauert (2013)

<sup>43</sup> vgl. Sengpiel

<sup>44</sup> vgl. Ebenda

<sup>45</sup> vgl. Weinzierl (2008), S. 588

Mit Hilfe einer kopfbezogenen Impulsantwort (Head related Impulse Response, HRIR) lassen sich diese Verzerrungen messen. HRIRs werden in reflexionsarmen Räumen gemessen und beschreiben somit nur den akustischen Einfluss von Torso, Kopf und Ohrmuschel<sup>46</sup>. Mittels Fouriertransformation lässt sich eine Impulsantwort aus der Zeit- in die Frequenzdomäne überführen, wodurch eine kopfbezogene Übertragungsfunktion (Head related transfer function, HRTF) entsteht. Für die Messungen von HRIR werden Sondenmikrofone in den Hörkanal einer Versuchsperson oder eines Kunstkopfes eingebracht. Für jede Schalleinfallrichtung und jedes Ohr ändert sich die HRIR, weshalb mehrere Messungen erfolgen müssen, zwischen denen interpoliert werden kann. Bei einer HRIR handelt es sich daher um einen Datensatz mehrerer Messungen. Werden die Messungen in akustisch wirksamen Räumen durchgeführt, enthalten die HRIR zusätzlich die Reflexions- und Nachhallmuster des Raumes, weshalb die Impulsantworten binaurale Raumimpulsantworten (binaural room impulse response, BRIR) genannt werden.

Die Messung eines HRIR-Datensatzes ist sehr aufwendig und spezifisch für die Anatomie der vermessenen Person. Daher bedient man sich in der Praxis gemittelten HRIRs/HRTFs oder nutzt zur Messung einen Kunstkopf, der einen „Durchschnittskopf“ repräsentiert<sup>47</sup>.

## 4.2 Binauralaufnahmetechnik

Die binaurale Aufnahmetechnik ist eine zweikanalige Technik, bei der Mikrofone an beiden Ohrkanaleingängen eines Menschen oder eines Kunstkopfes platziert werden<sup>48</sup>. So wird den Signalen bereits bei der Aufnahme eine kopfbezogene Übertragungsfunktion aufgeprägt. Stereofone Aufnahmetechniken, wie das Kugelflächenmikrofon und die Trennkörperstereofonie, bilden ebenfalls den menschlichen Hörapparat annähernd nach und können daher zur binauralen Aufnahmetechnik gezählt werden<sup>49</sup>.

Ein Problem der binauralen Aufnahmetechnik ist die Tatsache, dass die HRTFs nicht individuell sind und somit wiedergabeseitig nicht für alle Menschen passend sind, was zwangsweise zu unterschiedlichen Qualitäten des Hörerlebnisses führt. Weiterhin tritt bei nicht individualisierten HRTFs oftmals das Problem der Vorne-Hinten-Vertauschung auf<sup>50</sup>, da frontale und rückwärtige Schalleinfallrichtungen dieselbe interaurale Laufzeitdifferenz hervorrufen und nur schwer unterschieden werden können.

Zwar ist es möglich, mit einer Kunstkopfaufnahme authentische, dreidimensionale „Schallfelder mit den komplexen Spektraldifferenzen und den Laufzeitdifferenzen allein über Kopfhörer eindrucksvoll

---

<sup>46</sup> vgl. Slavik; Weinzierl (2008), S. 673

<sup>47</sup> Weinzierl (2008), S. 587

<sup>48</sup> Sengpiel

<sup>49</sup> Sengpiel

<sup>50</sup> Møller u.a. (1996), S. 463

[...]“<sup>51</sup> wiederzugeben, aber die akustische Perspektive ist durch die Mikrofonanordnung fest vorgegeben und lässt sich nicht nachträglich ändern, womit Kopfdrehungen und Interaktivität ausgeschlossen sind. Daher ist diese Aufnahmetechnik für die Anwendung bei 360°-Videoproduktionen nicht geeignet.

Das Problem der festen Perspektive versuchen Algazi et al. mit ihrer Motion-Tracked Binaural Aufnahmetechnik zu umgehen. Dazu wird ein Mikrofonarray verwendet, welches entlang einer kopfähnlichen Kugelfläche angebracht wird. Bei der Wiedergabe wird mit einem Head-Tracker die Kopfposition des Hörers erfasst und diejenigen Mikrofonsignale des Arrays wiedergeben, die den Ohren des Hörers am nächsten sind. Bei Zwischenpositionen ist es möglich eine Interpolation zwischen zwei benachbarten Mikrofonen durchzuführen<sup>52</sup>. Abbildung 5 zeigt das Omni Binaural Microphone der Firma 3Dio, welches einen ähnlichen Ansatz verfolgt. Hier sind mehrere künstliche Ohrmuscheln mit Mikrofonen horizontal über eine rechteckige Fläche angebracht. Somit werden binaurale Aufnahmen für vier Richtungen erzeugt, zwischen denen nach Erfassung der Blickrichtung des Hörers interpoliert werden kann. Dies lässt jedoch nur Drehbewegungen auf der horizontalen Achse zu. Ein Neigen und Kippen des Kopfes lässt sich mit dieser Aufnahmetechnik nicht darstellen. Folglich ist der Nutzen für 360°-Videoproduktionen beschränkt, zumal es sich bei dynamischen Binauralaufnahmen um ein Sonderformat handelt, das von den wenigsten Plattformen für 360°-Medien unterstützt wird (vgl. Kapitel 8).



Abbildung 5 - Omni Binaural Mikrofon von 3Dio<sup>53</sup>

---

<sup>51</sup> Sengpiel

<sup>52</sup> Algazi u.a. (2004), S. 1146 ff.

<sup>53</sup> Bildquelle: 3Dio (2018)

### 4.3 Binauralsynthese und Head-Tracking

Die Binauralsynthese kann dazu genutzt werden, ein binaurales Signal einer Schallquelle zu erzeugen, sodass der Eindruck entsteht, ein Hörer befände sich am Ort der Schallquelle<sup>54</sup>. Die Binauralsynthese kann somit die Kunstkopfaufnahme einer konventionell aufgenommenen Schallquelle simulieren und jede beliebige Perspektive darstellen.

„Die Synthese basiert auf der Tatsache, dass die Schallübertragung von einem Punkt im Raum bis vor das menschliche Trommelfell ein weitgehend lineares, zeit-invariantes System ist und somit im Zeitbereich durch eine Impulsantwort bzw. im Frequenzbereich durch eine Übertragungsfunktion beschrieben werden kann“<sup>55</sup>.

Bei der statischen Binauralsynthese wird das Mono-Signal einer Schallquelle mit den Impulsantworten gefaltet, die die HRIRs des linken und rechten Ohrs für die gewünschte Schallereignisrichtung enthalten<sup>56</sup>. Um die benötigte Rechenleistung zu reduzieren, kann vor der Faltung eine Fouriertransformation durchgeführt werden, sodass die Faltung in der Frequenzdomäne ausgeführt wird und aus einer Multiplikation besteht.

Bei der dynamischen Binauralsynthese wird zusätzlich die Orientierung des Hörers mit einem Head-Tracker erfasst und ausgewertet, um die akustische Szene in Echtzeit anzupassen. Dazu werden anhand der Head-Tracker Daten die HRTFs bei jeder Veränderung der Hörerorientierung ausgetauscht<sup>57</sup>, damit eine Schallquelle ihre relative Position im Raum beibehält und sich nicht mit dem Hörer dreht.

Da der Austausch der HRTF-Datensätze in Echtzeit ein rechenintensives Verfahren ist und sich nicht für eine akustische Szene als Ganzes durchführen lässt, kommen bei 360°-Video virtuelle Lautsprecher zum Einsatz<sup>58</sup>. Dabei wird ein beliebiges Lautsprecher-Setup um eine zentrale Hörposition angeordnet und für jede Lautsprecherposition die BRIRs gemessen. Für die Wiedergabe einer akustischen Szene werden zunächst die Lautsprechersignale für das verwendete Setup erzeugt und anschließend mit den entsprechenden BRIRs gefaltet und über Kopfhörer ausgegeben. Für die Anpassung der akustischen Szene an die Orientierung des Hörers, die wieder mit einem Head-Tracker erfasst wird, kann die akustische Szene rotiert werden, bevor die virtuellen Lautsprechersignale berechnet werden<sup>59</sup>. Dies ist im Ambisonics-Format sehr einfach möglich. Daher kann bei der Binauralsynthese über virtuelle

---

<sup>54</sup> Roginska (2018), S. 101

<sup>55</sup> Slavik; Weinzierl (2008), S. 671

<sup>56</sup> Roginska (2018), S. 101

<sup>57</sup> Ebenda, S. 101-103

<sup>58</sup> Dieterle u.a. (2018)

<sup>59</sup> Roginska (2018), S. 105

Lautsprecher mit statischen HRTFs gearbeitet werden, was weniger rechenintensiv ist und dadurch die Binauralsynthese auf Mobile Devices und in Browsern ermöglicht.

Für eine überzeugende Wirkung von 360°-Medien spielt die Auswertung der Orientierungsdaten des Hörers eine entscheidende Rolle und ermöglicht erst das Erzeugen einer immersiven, akustischen Umgebung. Darüber hinaus kann Head-Tracking die Lokalisierung von Schallquellen verbessern, da dem Mediennutzer ermöglicht wird kurze Kopfbewegungen auszuführen, die zusätzliche interaurale Laufzeit- und Pegeldifferenzen generieren. Diese sind bei der Lokalisation von Schallquellen in der vertikalen Ebene hilfreich. Die Orientierungsdaten können dazu über Head-Tracker in HMDs, Bewegungssensoren in Smartphones oder die Perspektive eines Bildausschnitts in Videobildern erzeugt werden.

## 5. Immersion und Präsenz

Im Zusammenhang mit 360°-Video und VR-Medien tauchen häufig Begriffe wie „Immersion“, „immersive sound“ oder „immerersive Medien“ auf. In diesem Kapitel wird daher zunächst eine Definition dieser Begrifflichkeiten gegeben. Anschließend werden daraus Kriterien für die Tongestaltung zu 360°-Video Produktionen abgeleitet.

### 5.1 Immersion

In der Medienwissenschaft bestehen verschiedene Definitionen von „Immersion“. Murrays Beschreibung ist breit angelegt und zunächst unspezifisch:

"Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically immersive experience that we do from a plunge in the ocean or swimming pool: the sensation of being surrounded by a completely other reality."<sup>60</sup>

Nach Nohr bedeutet Immersion „das Eintauchen oder “Eingetaucht-Werden” des Mediennutzers in das Medium (lat. immergere = verschmelzen/eintauchen), was den Effekt des „Unsichtbar-Werdens“ der Medienapparate mit sich bringt<sup>61</sup>. Der Mediennutzer betrachtet die Szene nicht von außen, sondern ist vollständig von ihr umgeben, er befindet sich „mitten drin“<sup>62</sup>. Für Specht bedeutet Immersion das authentische Erleben der medialen Simulation<sup>63</sup>.

---

<sup>60</sup> Murray (1998), S. 98 ff.

<sup>61</sup> Neitzel; Nohr (2006), S. 223

<sup>62</sup> Kuhlen (2014), S. 134

<sup>63</sup> Specht (2018), S. 189

Darüber hinaus finden sich in der Literatur spezifischere Definitionen, die Immersion entweder als technische Eigenschaft eines Mediums in den Vordergrund stellen oder den psychologischen Zustand, den Immersion beim Mediennutzer auslöst. So ist Immersion nach Slater und Wilbur eine Eigenschaft von Technologie, die beschreibt in welchem Ausmaß Ausgabegeräte für Medien in der Lage sind, eine umfassende, einhüllende und lebendige Illusion für den Mediennutzer zu schaffen<sup>64</sup>. Witmer und Singer hingegen beschreiben Immersion als psychischen Zustand, der sich dadurch auszeichnet, dass man sich in einer Umwelt, die fortlaufend Reize aussendet, anwesend fühlt, von ihr eingeschlossen wird und mit ihr interagieren kann<sup>65</sup>.

## 5.2 Präsenz

Im Zusammenhang mit Immersion, bezeichnet der Begriff Präsenz das subjektive Gefühl des Nutzers in der virtuellen Umgebung anwesend zu sein<sup>66</sup>. Präsenz weist also auf einen psychischen Zustand des Medienkonsumenten hin, wie ihn auch Witmer und Singer in ihrer Definition von Immersion beschreiben. Bei Verwendung des Begriffs Präsenz zur Beschreibung des psychischen Zustands des Nutzers, ergibt es daher Sinn, den Begriff Immersion als technologische Eigenschaft von Medien zu verstehen. Ist die Immersion der künstlichen Umgebung in einem Medium ausreichend hoch, so funktioniert die Illusion und der Nutzer hat das Gefühl, sich in der künstlichen Welt zu befinden. Im Englischen wird dies oft mit „the sense of being there“ beschrieben<sup>67</sup>.

Präsenz kann somit als Wirkung und Ziel von Immersion als Ursache verstanden werden. Medien können mehr oder weniger immersiv sein, was Immersion als graduelle Eigenschaft auszeichnet.

## 5.3 Erzeugen von Immersion

Eine wichtige Voraussetzung für das Erzeugen von Immersion ist das Ansprechen mehrerer Sinne des Mediennutzers<sup>68</sup>. Um diese Multimodalität zu erreichen, ist es nötig die Sinne des Nutzers mit virtuellen Sinneseindrücken zu überlagern<sup>69</sup> und die reale Umgebung möglichst auszublenden.

Nach Slater und Wilbur lässt sich Immersion auf die vier folgenden technischen Eigenschaften der Ausgabegeräte zurückführen<sup>70</sup>:

---

<sup>64</sup> Slater; Wilbur (1997), S. 605

<sup>65</sup> vgl. Witmer; Singer (1998), S. 227

<sup>66</sup> vgl. Dörner; Steinicke (2013), S. 46

<sup>67</sup> vgl. Kuhlen (2014), S. 134

<sup>68</sup> vgl. Ebenda, S. 134

<sup>69</sup> vgl. Mühlberger (2014), S. 152

<sup>70</sup> vgl. Slater; Wilbur (1997), S. 605

1. Die Sinneseindrücke des Menschen sollen möglichst ausschließlich durch den Computer generiert werden, d. h. der Nutzer soll weitestgehend von der realen Umgebung isoliert werden;
2. möglichst viele Sinne sollen angesprochen werden;
3. die Ausgabegeräte sollen den Nutzer vollständig umgeben, anstatt nur ein enges Sichtfeld zu bieten;
4. zudem sollen die Ausgabegeräte eine „lebendige“ Darstellung bieten

Bezogen auf die Produktion von Audio für 360°-Video ist die Isolation des Nutzers durch die Verwendung von Kopfhörern gegeben. Als Quasi-Standard für 360°-Video steht als Audio-Ausgabeformat ein kopfbezogenes Stereosignal. Im Rahmen von 360°-Videos lassen sich derzeit nur Hör- und Sehsinn ansprechen, weshalb hohe Bild- und Tonqualitäten für die Immersion relevant sind. Ein vollständiges Umgeben des Nutzers kann audioseitig durch ein 3D-Audioformat erreicht werden. Das 360°-Video-Material lässt sich sowohl auf konventionellen Bildschirmen und Smartphones wiedergeben, als auch über HMDs, wobei ein vollständiges Umgeben des Nutzers nur über HMDs ermöglicht wird. Die „lebendige“ Darstellung der Inhalte beruht einerseits auf der Qualität der Wiedergabetechnik, andererseits aber auch zu großen Teilen auf der inhaltlichen Gestaltung der Medien.

Für das Erzeugen von Immersion kommt der Verfolgung der Orientierung des Nutzers (Head-Tracking) eine große Bedeutung zu. Idealerweise wird dazu die visuelle und akustische Ausgabe in Echtzeit an die Orientierung angepasst, denn „die gemeinsame Verwendung visueller und akustischer Reize verstärkt das Gefühl der Immersion ungemein“<sup>71</sup>. Dörner stellt ebenfalls fest, dass aktiviertes Head-Tracking das Potenzial besitzt die Präsenz des Mediennutzers zu erhöhen<sup>72</sup>. Steve Bryson hat die grundlegende Bedeutung dieser Eigenschaft prägnant auf den Punkt gebracht: „*If I turn my head and nothing happens, it ain't VR!*“<sup>73</sup>

Für die Immersion ist es nach Brill entscheidend, dass virtuelle und reale Umgebung in hohem Maße übereinstimmen<sup>74</sup>. Obwohl eine perfekte Illusion der Realität mit den heute verfügbaren Mitteln nicht herzustellen ist, muss dieses Ziel kritisch betrachtet werden. Laut Neitzel würde die perfekte Simulation der Realität zu einer totalen Immersion führen und hätte den kompletten Realitätsverlust zu Folge<sup>75</sup>.

Schütze macht deutlich, dass wir zwar heute in der Lage sind räumlichen 3D-Klang zu produzieren und mit ihm immersive Medienerfahrungen generieren können, aber dennoch viele Jahre davon entfernt sind

---

<sup>71</sup> vgl. Brill (2009), S. 25

<sup>72</sup> vgl. Dörner; Steinicke (2013), S. 46

<sup>73</sup> Bryson zitiert nach Dörner u.a. (2013), S. 13

<sup>74</sup> vgl. Brill (2009), S. 6

<sup>75</sup> vgl. Neitzel; Nohr (2006), S. 102

---

die reale Welt präzise zu simulieren. Daher postuliert Schütze, dass es nicht die Aufgabe des Tonschaffenden ist, die Realität möglichst exakt zu simulieren<sup>76</sup>.

Böcking zeigt auf, dass Mediennutzer sich gerne und bewusst auf eine Illusion, welche nicht den Regeln der Realität folgt, einlassen, sofern sie gut unterhalten werden<sup>77</sup>. Dieser Zusammenhang wird in der Literatur- bzw. Theaterwissenschaft als „willing suspension of disbelief“ bezeichnet. Demnach muss Fiktion nicht realistisch sein, sondern dem entsprechen was für den Mediennutzer plausibel erscheint. Im Kontext der Audioproduktion ist folglich der genregerechten Umsetzung und der Beachtung der Hörerfahrung der Mediennutzer höhere Priorität einzuräumen als einer möglichst realitätsnahen Umsetzung einer Audioszene.

Fink fasst die Bedeutung der Immersion wie folgt zusammen:

“Presence and agency in a photorealistic simulated world is the holy grail of immersion. In the meantime, suspension of disbelief allows us experiences that obscure the line between what’s real and what’s virtual.”<sup>78</sup>

## **6. Anforderungen an die Tongestaltung für 360°- Video**

Schütze stellt heraus, dass für alle Medien der „New Realities“ eine hohe Audioqualität nicht nur dienlich, sondern absolut notwendig ist<sup>79</sup>. Für die Bestimmung der Audioqualität muss die gesamte Bearbeitungskette beachtet werden. In diesem Kapitel sollen diejenigen Qualitätsanforderungen dargestellt werden, welche speziell für das Erzeugen von Immersion und letztlich dem Gefühl der Präsenz des Mediennutzers im Rahmen der Tongestaltung für 360°-Videoproduktionen von großer Bedeutung sind.

Beim natürlichen Hören wird Schall aus allen Richtungen wahrgenommen. Um ein möglichst natürliches und realistisches Hörerlebnis zu schaffen ist es daher zwingend notwendig, alle Raumdimensionen akustisch zu bedienen und ein dreidimensionales Audioformat einzusetzen, welches es ermöglicht, Hörsignale wiederzugeben. Bisherige Mehrkanalformate beschränkten sich auf die Abbildung der horizontalen Ebene, bei 3D-Audio wird die vertikale Ebene miteinbezogen. Um die

---

<sup>76</sup> vgl. Schütze; Irwin-Schütze (2018), S. 6

<sup>77</sup> vgl. Böcking (2008), S. 59 ff.

<sup>78</sup> Fink (2018), S. 34

<sup>79</sup> Schütze; Irwin-Schütze (2018), S. 5

---

Möglichkeiten der akustischen Höherebene nutzen zu können, müssen dezidierte Hörsignale aufgenommen oder in der Postproduktion erstellt werden.

Dieterle erklärt, dass für die naturgetreue Simulation der Umgebung der Ton „extrem wichtig“ ist, aber dass neben einem hohen Maß an Realismus genauso auf einen „schönen“ Klang ankommt<sup>80</sup>. Ein wichtiger Faktor, welcher vor allem für eine realistische Abbildung sorgt, ist die Lokalisation der Schallquellen. Für eine korrekte Wahrnehmung der Hörereignisrichtung müssen die im Bild sichtbaren Schallquellen mit den akustischen Lokalisationspunkten übereinstimmen<sup>81</sup>. Für die Tonaufnahme bedeutet dies, dass der Aufnahmewinkel möglichst dem Wiedergabewinkel entspricht. Zu beachten ist in diesem Zusammenhang, dass es im Gegensatz zu konventionellen bildbezogenen Medien keine bevorzugte Richtung gibt. Daher ist auf eine gleichmäßige Abbildung des gesamten 360°-Raumes zu achten<sup>82</sup>. Sofern mit Stützmikrofonen, Einzelmikrofonierung oder Mono-Signalen gearbeitet wird, ist ein am Bildmaterial orientiertes Panning der Signale im 360°-Schallfeld nötig.

Durch die Möglichkeit, die Blickrichtung im 360°-Video jederzeit ändern zu können, ergeben sich neben optischen auch akustische Perspektivwechsel, die möglichst natürlich wirken sollen<sup>83</sup>. Vor allem bei nebeneinander angeordneten Schallquellen sind in diesem Zusammenhang ihre Räumlichkeit und ihre Lautstärke relevant. Die Anpassung der Hörereignisrichtung geschieht erst bei Wiedergabe des 360°-Videos durch Auswertung der Head-Tracking Daten und entsprechende Rotation des Schallfeldes und die anschließende Binauralsynthese. Da diese Operationen erst auf dem Wiedergabegerät des Nutzers erfolgen, ist eine unmittelbare Einflussnahme des Tonschaffenden hierauf nicht möglich.

Ferner ist die korrekte Abbildung der Distanz vom Kamerastandpunkt zu einer Schallquelle von Bedeutung. Wie in Kapitel 3.2 aufgezeigt, ist die Entfernungswahrnehmung weniger kritisch als die Richtungswahrnehmung. Im Kontext von 360°-Video wird die Wahrnehmung der Entfernung zu großen Teilen durch die im Bild sichtbare Entfernung beeinflusst, aber dennoch muss die Tiefenstaffelung zum optischen Eindruck passen. Die Distanz wird durch die Mikrofonierung und den Einsatz von Hallgeräten festgelegt<sup>84</sup>. Aus diesem Grund ist es wichtig Direktsignale der Schallquellen aufzuzeichnen, vor allem von denjenigen, die sich sehr nahe an der Kamera befinden.

Die Gestaltung des Raumes ist im Zusammenhang mit der Produktion von 360°-Aufnahmen von großer Bedeutung. Der Raum als Ort für eine musikalische Darbietung hat stets Einfluss auf den Klang und wird gemeinsam mit der Musik erlebt und verschmilzt mit ihr zu einer Einheit. Darüber hinaus hat der

---

<sup>80</sup> Dieterle u.a. (2018)

<sup>81</sup> vgl. Maempel (2013), S. 375

<sup>82</sup> vgl. Dieterle u.a. (2018)

<sup>83</sup> vgl. Schütze; Irwin-Schütze (2018), S. 9

<sup>84</sup> vgl. Theile u.a. (2013), S. 334-335

Mensch durch Erfahrungen gelernt, welchen Klang ein bestimmter Raum besitzt. Da der Raum im 360°-Video stets sichtbar ist, muss dieser einen nachvollziehbaren, logisch richtigen Klang besitzen, um die Immersion des 360°-Video-Erlebnisses nicht zu stören. Bei der Tonaufnahme ist daher auf eine Erfassung des Raumklangs besonders zu achten. Durch die Gestaltung des Raumklangs kann die Immersion verstärkt werden. Ein wichtiges Kriterium in diesem Zusammenhang ist die Einhüllung<sup>85</sup>. Einhüllung funktioniert vor allem dann gut, wenn das Diffusfeld dekorreliert ist<sup>86</sup>. Bei der Verwendung eines Mikrofonarrays ist daher auf ausreichende Kanaltrennung zu achten. Weiterhin sollte der Raumklang realistisch gestaltet werden, was bedeutet, dass akustischer und optischer Raumeindruck zueinander passen. Dieser Umstand muss beim Einsatz von künstlichem Hall in der Nachbearbeitung beachtet werden. Je nachdem wie deutlich der Handlungsort des Geschehens im Videomaterial erkennbar ist, wird die Gestaltungsfreiheit in Bezug auf künstlichen Hall eingeschränkt. Besonderes Augenmerk sollte auf die frühen Reflexionen gelegt werden, da diese zum einen eine korrekte Lokalisation unterstützen können und zum anderen für die Wahrnehmung und Charakterisierung des Raumes von zentraler Bedeutung sind<sup>87</sup>.

Abschließend stellt sich die Frage nach einem medienästhetischen Ideal für die Aufnahme und Bearbeitung von Musik im Rahmen von 360°-Videoproduktionen. Nach Stolla sind zwischen musikalischer Live-Darbietung und Medienprodukt drei prinzipielle Verhältnisse voneinander abzugrenzen, welche jeweils ein medienästhetisches Ideal beschreiben<sup>88</sup>:

1. *Positivistisches Ideal*: Das Schallfeld der Live-Darbietung soll möglichst naturgetreu reproduziert werden. Einem streng dokumentarischen Anspruch folgend soll auf Manipulationen weitgehend verzichtet werden.
2. *Illusionistisches Ideal*: Dieser Ansatz verfolgt das Ziel, das psychische Erleben einer musikalischen Darbietung zu reproduzieren. Zur Schaffung einer perfekten Illusion werden klangliche Merkmale manipuliert.
3. *Medial-Autonomes Ideal*: Dieser Ansatz lässt eine freie und weitreichende Manipulation des Audiomaterials zu. Es erfolgt eine freie Interpretation mit Mitteln der Produktionstechnik, bei der eine Orientierung an einer Live-Darbietung nicht erforderlich ist.

Ogleich die meisten Kunstmusikproduktionen seit den 1970er Jahren dem illusionistischen Ideal folgen<sup>89</sup>, scheint dieser Ansatz in Reinform nicht für das Medium des 360°-Videos geeignet zu sein. Um die geforderte Wirkung von Immersion und Präsenz sowie das nötige Maß an Realismus zu erreichen,

---

<sup>85</sup> Im englischen ist der Begriff „Envelopment“ verbreitet

<sup>86</sup> vgl. Dieterle u.a. (2018)

<sup>87</sup> Ebenda

<sup>88</sup> Stolla zitiert nach Maempel (2013), S. 415-416

<sup>89</sup> Ebenda, S. 415

ist teilweise die Verfolgung des positivistischen Ideals nötig. Für das noch recht junge Format des 360°-Videos hat sich bisher noch kein allgemeingültiges, medienästhetisches Ideal ausgebildet, doch ein sinnvoll erscheinender Ansatz für musikalische Produktionen im Rahmen von 360°-Video liegt zwischen dem positivistischen und dem illusionistischen Ideal.

## 7. 3D-Audioformate für 360°-Video

In diesem Kapitel werden die für 360°-Video relevanten Audioformate vorgestellt. Einleitend werden zunächst verschiedene Grundkonzepte von 3D-Audioformaten vorgestellt. Anschließend werden die Formate des Ambisoncs-Systems näher vorgestellt, da dies die Standard-Audioformate für VR und 360°-Videoproduktionen sind.

### 7.1 Next Generation Audio

Unter dem Begriff Next Generation Audio werden die verschiedenen 3D-Audioformate zusammengefasst<sup>90</sup>. Um das Audioformat welches für 360°-Videoproduktionen relevant ist einordnen zu können, wird ein kurzer Überblick über die verschiedenen 3D-Audioformate gegeben und ihr grundlegendes Konzept vorgestellt.

Die heute verbreiteten 3D-Audioformate lassen sich in drei Kategorien einteilen, die jeweils ihren eigenen Ansatz besitzen:

1. Kanalbasiert: Kanalbasierte Formate beruhen auf den Prinzipien der Psychoakustik mit der Möglichkeit Phantomschallquellen zwischen Lautsprechern abbilden zu können. Das Wiedergabesystem muss im Vorfeld bekannt sein und eine Mischung wird speziell für dieses Wiedergabesystem angefertigt. Bei der Mischung werden diskrete Lautsprechersignale erzeugt. Diese werden übertragen und bei der Wiedergabe den entsprechenden Lautsprechern zugeordnet. Abbildung 6 zeigt den Prozess einer kanalbasierten Produktion. Beispiele für kanalbasierte Formate sind 5.1 (Dolby Digital) oder Auro 3D.

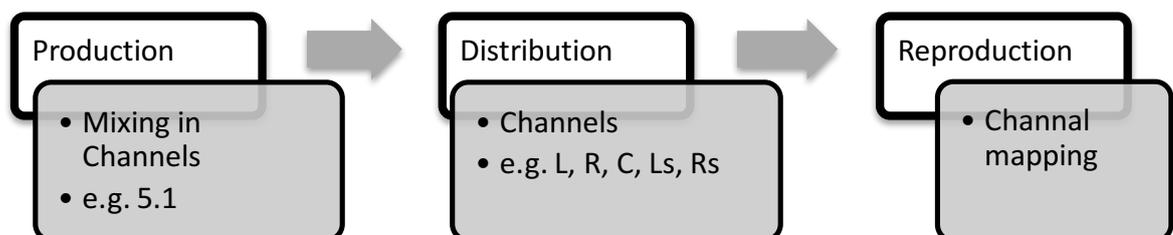


Abbildung 6 - Kanalbasierter Workflow<sup>91</sup>

<sup>90</sup> Camerer (2018)

<sup>91</sup> Bildquelle: Eigene Abbildung

2. Szenenbasiert: Szenenbasierte Formate verfolgen einen physikalischen Ansatz. Hierbei soll ein Schallfeld möglichst exakt am Wiedergabeort rekonstruiert werden. Einzelsignale werden gemischt und in einer Schallfeldbeschreibung dekodiert. Es wird eine akustische Szene konstruiert/aufgenommen, übertragen und rekonstruiert. Die Produktion ist dabei unabhängig vom Wiedergabesystem. Erst bei der Wiedergabe wird die akustische Szene für das Wiedergabesystem dekodiert. Bei der Wiedergabe ist es möglich das Schallfeld zu manipulieren und beispielsweise Rotationen umzusetzen. Abbildung 7 stellt den schematischen Ablauf für eine szenenbasierte Produktion dar. Ambisonics und Wellenfeldsynthese sind Vertreter der szenenbasierten Formate.

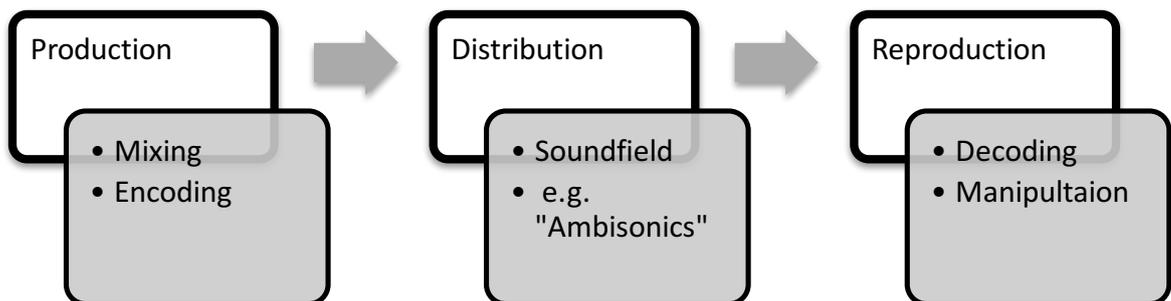


Abbildung 7 – Szenenbasierter Workflow<sup>92</sup>

3. Objektbasiert: Bei objektbasierten Formaten geht es um die genaue Beschreibung von einzelnen Sounds einer Mischung. Diese Sounds stellen die Objekte dar und werden während der Mischung nicht zusammengefasst, sondern mit Metadaten verknüpft. Diese beschreiben beispielsweise Panning, Level oder Filtereinstellungen eines Sounds. Es werden bei der Distribution folglich Einzelsounds und die zugehörigen Metadaten übertragen. Wiedergabeseitig wird ein Renderer benötigt, welcher genaue Informationen über das zur Verfügung stehende Wiedergabesystem besitzt. Der Renderer erzeugt dann zum Zeitpunkt der Wiedergabe die benötigten Lautsprechersignale, indem er die Metadaten der Einzelsounds auswertet und eine an das Wiedergabesystem angepasste Mischung erzeugt. Zusätzlich gibt es die Möglichkeit für den Nutzer das Rendering zu beeinflussen. So könnte beispielsweise ein Kommentator in der Lautstärke nachträglich angepasst werden. Der schematische Ablauf einer objektbasierten Produktion ist in Abbildung 8 dargestellt. Beispiele sind MPEG-H oder ADM.

<sup>92</sup> Bildquelle: Eigene Abbildung

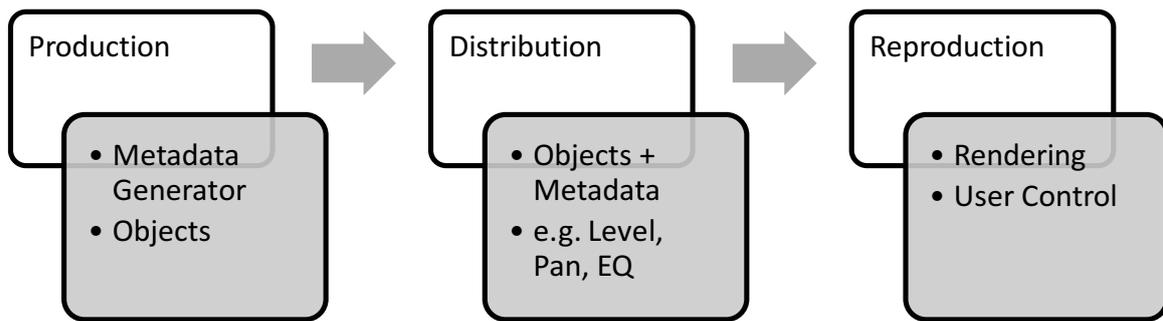


Abbildung 8 – Objektbasierter Workflow<sup>93</sup>

## 7.2 Ambisonics

Es ist anzumerken, dass es sich bei Ambisonics nicht um ein neues Format handelt, welches durch das vermehrte Aufkommen von 3D-Audio-Anwendungen in den letzten Jahren entstanden ist. Ambisonics wurde bereits in den 70er Jahren entwickelt, vornehmlich von Peter Fellget und Michael Gerzon<sup>94</sup>, konnte sich seinerzeit jedoch nicht etablieren und erlebt aktuell eine Renaissance. Es handelt sich um ein szenenbasiertes Audioformat, was bedeutet, dass es eine Repräsentation eines Schallfeldes darstellt und unabhängig von einem bestimmten Wiedergabesystem arbeitet. Erst bei der Reproduktion werden die Lautsprecher signale erzeugt, um das Schallfeld auf einem gegebenen Wiedergabesetup abspielen zu können.

### 7.2.1 Grundkonzept

Ambisonics zielt darauf ab, einen vollständigen, hierarchischen Ansatz für die Aufnahme, Speicherung, Verbreitung und Wiedergabe einer Audioszene zu bieten. Dabei ist die Größe des Wiedergabesystems von mono und Stereo über Surround bis hin zur vollen 3D-Wiedergabe inklusive Höheninformationen skalierbar. Je nach Kanalanzahl des Wiedergabesystems können mehr oder weniger Raumdimensionen des Schallfeldes reproduziert werden.

Ein Schallfeld besitzt mehrere spezifische Eigenschaften und besteht aus Direkt- und Diffusschall, die jeweils eigene zeitliche, räumliche und frequenzielle Merkmale aufweisen. Des Weiteren hat jedes Schallfeld eine räumliche Ausdehnung, durch die es definiert wird. Bei Ambisonics wird versucht, das Schallfeld in einem Punkt zu erfassen und für diesen Punkt bei der Wiedergabe eine möglichst genaue Reproduktion zu ermöglichen. Das Schallfeld kann in verschiedenen Auflösungen erfasst werden. Ähnlich dem Sampling innerhalb eines Analog-Digital Wandlers werden dem Schallfeld räumliche Richtungs-Proben entnommen. Je feiner die Abstufungen zwischen den Richtungen, desto besser die Auflösung. Dabei ist es zunächst irrelevant, ob es sich um ein reales Schallfeld handelt oder ob die akustische Szene erst in der Postproduktion erstellt wird – Ausgangsseitig entsteht ein generisches

<sup>93</sup> Bildquelle: Eigene Abbildung

<sup>94</sup> vgl. Elen (2001), S. 2

Format, durch das die Szene räumlich beschrieben wird<sup>95</sup>. In diesem generischen Raumformat ist die Mischung enkodiert. Für die Wiedergabe der akustischen Szene muss dieses Format für das jeweilige Wiedergabesystem dekodiert werden. Ambisonics arbeitet demnach unabhängig von einem Wiedergabesystem.

### 7.2.2 Funktionsweise

Die Beschreibung eines Schallfeldes in einem Punkt kann über die Zerlegung in Kugelflächenfunktionen erfolgen. Dies funktioniert ähnlich wie die Fourierreihe einer periodischen Schwingung<sup>96</sup>. Über Kugelflächenfunktionen wird beschrieben wie sich das Schallfeld in Abhängigkeit von Zeit und Richtung verändert<sup>97</sup>. Ambisonics beschreibt das dreidimensionale Schallfeld anhand einer Fourier-Bessel-Reihe. Die Richtungsbeschreibung erfolgt ausgehend von einem zentralen Punkt durch Kugelkoordinaten (Radius  $r$ , Azimut  $\varphi$  und Elevation  $\delta$ )<sup>98</sup>.

Für die Fourier-Besselreihe gilt:

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

Das Schallfeld  $p(r)$  ergibt sich in dieser Reihe aus Überlagerung der Funktionen  $Y_{mn}^{\sigma}$ , die mit den Faktoren  $B_{mn}^{\sigma}$  gewichtet werden. Die Faktoren  $B_{mn}^{\sigma}$  werden auch Komponenten genannt. Die gewichteten Funktionen werden sphärische Harmonische genannt. Durch diese winkelabhängigen Funktionen bildet sich ein orthogonales Basissystem, durch das sich ein beliebiges Schallfeld als Überlagerung von sphärischen Harmonischen der Ordnung  $m$  darstellen lässt. Dabei werden die sphärischen Harmonischen jeweils mit sphärischen Besselfunktionen  $j_m(kr)$  und einem Phasenfaktor  $i^m$  radial gewichtet<sup>99</sup>.

Voraussetzung für die Kodierung des Schallfeldes ist, dass das Schallfeld eine ebene Welle (Referenzwelle) darstellt und jeder Lautsprecher bei der Wiedergabe ebenfalls eine ebene Welle abstrahlt<sup>100</sup>.

Durch die Gewinnung der Komponenten  $B_{mn}^{\sigma}$  ist nach Formel (1) das Schallfeld vollständig beschrieben. Die Gewinnung der Komponenten  $B_{mn}^{\sigma}$  ist die Enkodierung des Schallfeldes. Die Enkodierung erfolgt für virtuelle Schallfelder auf Basis eines Schallfeldmodells oder für reale Schallfelder durch spezielle Ambisonics-Mikrofone (vgl. Kapitel 9.1). Ausgehend von Formel (1) ergibt

<sup>95</sup> vgl. Camerer (2018)

<sup>96</sup> vgl. Villegas (2015), S. 202

<sup>97</sup> vgl. Shivappa u.a. (2016), S. 4

<sup>98</sup> vgl. Slavik; Weinzierl (2008), S. 660

<sup>99</sup> vgl. Ebenda, S. 661

<sup>100</sup> vgl. Sontacchi; Höldrich, S. 2

sich für eine ebene Welle  $S_{Ref}$  mit der Wellenzahl  $k$  und dem Einfallswinkel  $(\theta_R, \delta_R)$  die Ambisonics-Komponente nach:

$$B_{mn}^\sigma = S_{Ref} \cdot Y_{mn}^\sigma(\varphi_R, \delta_R) = \hat{S} e^{jkr} \cdot Y_{mn}^\sigma(\varphi_R, \delta_R) \quad \text{Formel (2)}$$

Aus der Einfallrichtung einer ebenen Welle ergeben sich die reellen Gewichtungsfaktoren  $Y_{mn}^\sigma(\varphi_R, \delta_R)$  mit der sie enkodiert wird<sup>101</sup>.

Zur Resynthese eines Schallfeldes werden die Komponenten  $B_{mn}^\sigma$  der Fourier-Besselreihe übertragen. Je später die Reihe abgebrochen wird, desto mehr Komponenten werden übertragen und desto präziser ist die Rekonstruktion des Schallfeldes. Bei der Resynthese des Schallfeldes wird das Ambisonics-Signal dekodiert. Bei der Dekodierung werden die Ambisonics-Komponenten mit Lautsprecher Signalen für einen zentralen Hörpunkt überlagert<sup>102</sup>. Die Konfiguration des Wiedergabesetups muss für die Dekodierung bekannt sein. Originales und rekonstruiertes Schallfeld stimmen dann überein, wenn ihre Komponenten  $B_{mn}^\sigma$  übereinstimmen. Die Schallfeldkomponenten  $B_{mn}^\sigma$ , die Lautsprecherpositionen und die gesuchten Lautsprecher Signale ergeben ein lösbares Gleichungssystem. Dabei ist es unerheblich, ob das Ambisonics-Signal für die Wiedergabe auf realen Lautsprechern dekodiert wird oder für die Wiedergabe auf virtuellen Lautsprechern, um ein anschließendes binaurales Rendering durchzuführen. Das Lautsprecher-Setup kann grundsätzlich frei gestaltet werden, allerdings muss die Anzahl der Lautsprecher mindestens so hoch sein wie die Anzahl der Wiedergabekanäle damit das Gleichungssystem lösbar ist<sup>103</sup>. Darüber hinaus empfiehlt Gerzon eine äquidistantes Wiedergabe-Setup und mehr Lautsprecher als Wiedergabekanäle zu benutzen<sup>104</sup>.

### 7.2.3 Ordnungszahl

Die Ordnungszahl der Ambisonics-Szene ist ein Maß dafür, wie präzise die Richtungsabbildung innerhalb des Schallfeldes ist<sup>105</sup>. Je höher die Ordnungszahl, desto mehr sphärische Harmonische (räumliche Harmonische) werden genutzt, um das Schallfeld zu beschreiben.

Abbildung 9 zeigt die sphärischen Harmonischen von der 0. Bis zur 3. Ordnung. Die sphärischen Harmonischen der 0. und 1. Ordnung können als Richtcharakteristiken von Mikrofonen verstanden werden. So enthält das Signal eines reinen Druckempfängers die gesamte Schalldruckinformation eines Schallfeldes, jedoch keinerlei Richtungsinformationen. Die sphärischen Harmonischen 1. Ordnung gleichen einer Achtercharakteristik und erfassen so die Druckgradienten der Richtungen Oben – Unten, Links – Rechts und Vorne – Hinten. Die Komponente der 0. Ordnung wird W Komponente genannt,

<sup>101</sup> vgl. Slavik; Weinzierl (2008), S. 663

<sup>102</sup> vgl. Ebenda, S. 662

<sup>103</sup> vgl. Ebenda, S. 663

<sup>104</sup> vgl. Gerzon (1985), S. 2-3

<sup>105</sup> vgl. Camerer (2018)

die Komponenten der 1. Ordnung X, Y, und Z-Komponente. Ein Ambisonics-Signal welches die Komponenten W, X, Y, Z enthält wird First Order Ambisonics (FOA) genannt. Ambisonics-Signale höherer Ordnung werden Higher Order Ambisonics (HOA) genannt. Dabei schließt ein Ambisonics-Signal höherer Ordnung stets sämtliche Komponenten der nächst kleineren Ordnung mit ein. So ist durch ein Weglassen bestimmter Komponenten eine Down-Konvertierung sehr einfach möglich.

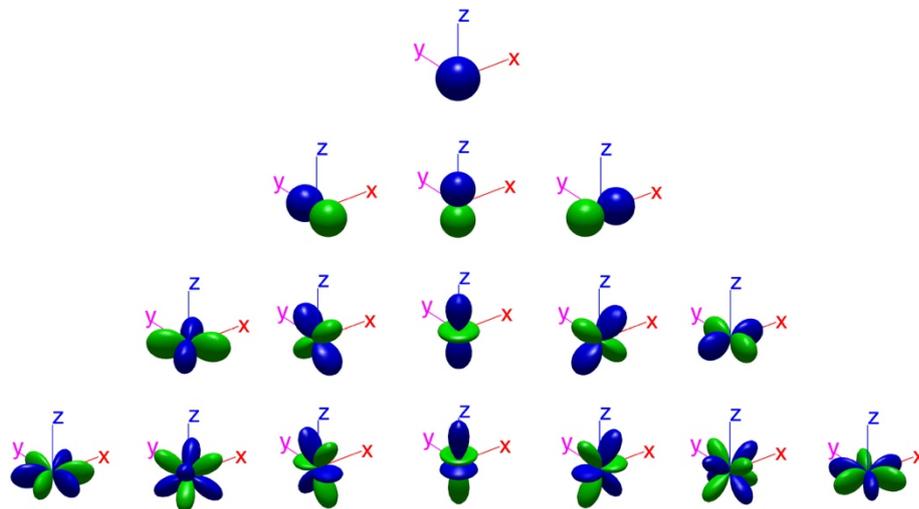


Abbildung 9 - Sphärische Harmonische bis zur dritten Ordnung<sup>106</sup>

Ambisonics wird in konventionellen Tonspuren enkodiert. Je höher die Ordnungszahl ist, desto mehr Übertragungskanäle werden benötigt um das Ambisonics-Signal zu speichern, zu verbreiten und wiederzugeben. Für ein Ambisonics-Schallfeld der Ordnungszahl  $M$  werden  $N$  Übertragungskanäle benötigt. Es gilt<sup>107</sup>:

$$N = (M + 1)^2 \quad \text{Formel (3)}$$

Die Anzahl der Übertragungskanäle hat unmittelbaren Einfluss auf die Größe des korrekt rekonstruierten Schallfeldes. Der korrekt rekonstruierte Bereich wird auch Sweetspot genannt und hat bei FOA etwa die Größe eines Tennisballs<sup>108</sup>. Da bei 360°-Video ein virtuelles Lautsprecher-setup binaural gerendert wird, sollte der Sweetspot optimalerweise so groß sein, dass zumindest ein (Kunst-)Kopf hineinpasst. Dies ist auch ein Grund dafür, weshalb FOA keine ausreichende Qualität für 360°-Videoproduktionen liefert.

<sup>106</sup> Bildquelle: Politis (2018)

<sup>107</sup> vgl. Hollerweger (2005), S. 6-7

<sup>108</sup> vgl. Wittek; Theile (2016), S. 2

## 7.2.4 Ambisonics-Formate

Innerhalb des Ambisonics-Systems existieren verschiedene Formate. Dabei wird nicht nur zwischen Ordnungszahlen unterschieden, sondern auch nach Kodierung, Kanalordnung und Normalisierung.

Das A-Format kann als Aufnahmeformat betrachtet werden und bezieht sich auf FOA. Um die sphärischen Harmonischen der 0. und 1. Ordnung aufzuzeichnen, werden nicht die diskreten Richtcharakteristiken Kugel und Acht verwendet, sondern eine spezielle Anordnung von Nieren. Ähnlich wie bei der MS-Stereofonie müssen diese Signale über Summen- und Differenzbildung matriziert werden, wodurch sie in das B-Format gewandelt werden<sup>109</sup>

Das B-Format besteht aus den Komponenten der 0. und 1. Ordnung: W, X, Y und Z. Dieses Format kommt zum Einsatz, wenn ein Ambisonics-Schallfeld gespeichert oder übertragen werden soll. Für jede der Komponenten ist ein Audiokanal zu reservieren. Es ist unabhängig vom Aufnahme- und Wiedergabe-Setup<sup>110</sup>.

Das C-Format ist ein Übertragungsformat, das auch unter dem Namen UHJ-Format bekannt ist. Gerzon entwickelte dieses Format für die Rundfunkausstrahlung und die Verbreitung über physische Medien. Das C steht dabei für Consumer<sup>111</sup>. Die Idee hinter dem Format ist, Ambisonics-Signale zu übertragen, die direkt ohne Dekodierung kompatibel zu konventionellen monophonen und stereofonen Wiedergabesystemen sind, für den Fall, dass der Konsument nicht über einen Decoder verfügt. Beim C-Format werden die Kanäle L, R, Q, und T übertragen. L und R repräsentieren das Stereosignal, Q enthält die Höheninformation und T erhöht die räumliche Auflösung auf horizontaler Ebene. Ist ein Decoder vorhanden, können diese Signale wieder ins B-Format gewandelt werden<sup>112</sup>. Das C-Format ist heute kaum noch gebräuchlich.

Das D-Format ist ein Wiedergabe-Format. Es handelt sich um die Lautsprechsignale, welche berechnet werden, wenn ein B-Format für ein bestimmtes Wiedergabe-Setup dekodiert wird<sup>113</sup>. Daneben findet sich in der Literatur auch noch das G-Format, welches als spezieller Fall eines D-Formats betrachtet werden kann. Da ursprünglich die Dekodierung des B-Formates nur für äquidistante Lautsprecher Systeme vorgesehen war, erhielt die Dekodierung für das 5.1 Lautsprecher Setup nach ITU-Standard den Namen G-Format. Die hierfür eingesetzten Decoder werden Vienna-Decoder<sup>114</sup> genannt<sup>115</sup>.

---

<sup>109</sup> vgl. Nicol (2018), S. 278-280

<sup>110</sup> vgl. Ebenda, S. 299

<sup>111</sup> vgl. Gerzon (1985), S. 862-865

<sup>112</sup> vgl. Nicol (2018), S. 299-300

<sup>113</sup> vgl. Rumsey (2001), S. 111

<sup>114</sup> Angelehnt an den Veranstaltungsort der AES-Convention auf der dieses Verfahren erstmals vorgestellt wurde

<sup>115</sup> vgl. Elen (2001), S. 3-4

Ein weiteres Unterscheidungsmerkmal von Ambisonics-Formaten ist die Reihenfolge, in der die Komponenten angeordnet und übertragen werden. Die Kanalordnung nach Furse und Malham (FuMa) bis zur dritten Ordnung lautet: (W), (X, Y, Z), (R, S, T, U, V), (K, L, M, N, O, P, Q)<sup>116</sup>. Die Anordnung nach Ambisonics Channel Number (ACN) ist dagegen: (W), (Y, Z, X), (V, T, R, S, U), (Q, O, M, K, L, N, P)<sup>117</sup>. Daniel postulierte eine weitere Anordnung der Komponenten – Single Index Designation (SID) mit: (W), (X, Y, Z), (U, V, S, T, R), (P, Q, N, O, L, M, K)<sup>118</sup>. Die ACN Kanalordnung ist aktuell am weitesten verbreitet, die SID Anordnung findet derweil nur in der Forschung Anwendung.

Ferner unterscheiden sich die Ambisonics-Formate durch ihre Normalisierung. Die verschiedenen Methoden heißen maxN, SN3D und N3D<sup>119</sup>. Sie unterscheiden sich durch die Gewichtung der Ambisonics-Komponenten. SN3D ist am weitesten verbreitet.

Je nach Kombination von Komponentenanzahl und Normalisierungsmethode existieren verschiedene Bezeichnungen für ein solches Format. Am weitesten verbreitet ist aktuell das Format ambix, welches die Komponentenanzahl nach ACN und die Normalisierungsmethode SN3D nutzt. Ein weiteres gängiges Format ist FuMa, welches die Kanalordnung nach Furse und Malham und eine leicht abgewandelte Form der maxN Normalisierungsmethode nutzt, bei der die W-Komponenten um -3 dB korrigiert wird. Ohne diese Korrektur spricht man von FuMa mit maxN Normalisierung, welches zum Teil auch von Plattformen als Format genutzt wird<sup>120</sup>.

Die verschiedenen Ambisonics-Formate sind mit entsprechenden Tools zwar ineinander überführbar, jedoch nicht automatisch miteinander kompatibel. Somit gibt es eine große Anzahl technischer Parameter, die beachtet werden müssen, wenn von Ambisonics die Rede ist. Je nach Zielplattform muss genau geprüft werden, welches Format gefordert ist, um Inkompatibilität zu vermeiden.

### 7.2.5 Vorteile

Ein großer Vorteil von Ambisonics ist, dass das Schallfeld sehr leicht nachträglich manipuliert werden kann. Da es keine Vorzugsrichtung gibt, ist es sehr einfach die komplette Szene zu drehen oder zu neigen<sup>121</sup>. Dadurch eignet sich Ambisonics optimal für den Einsatz bei 360°-Video, wo der Standpunkt fest ist, aber in jede Richtung des Raumes geblickt werden kann. Durch die einfache Rotation kann bei der Binauralisierung mit statischen HRTFs gearbeitet werden, wodurch die benötigte Rechenleistung

---

<sup>116</sup> vgl. Carpentier (2017), S. 2

<sup>117</sup> vgl. Chapman u.a. (2009), S. 3

<sup>118</sup> vgl. Daniel (2003), S. 13

<sup>119</sup> vgl. Malham (2003), S. 3-4

<sup>120</sup> vgl. Audiokinetic Inc. (2018)

<sup>121</sup> vgl. Camerer (2018)

sinkt – wichtig für Anwendung auf mobile devices, wo nur begrenzte Rechenleistung zur Verfügung steht<sup>122</sup>.

Ebenfalls von Vorteil ist, dass Ambisonics unabhängig von einem definierten Wiedergabesystem funktioniert. Eine in Ambisonics angelegte Mischung lässt sich durch entsprechende Decoder an jedes Wiedergabesystem anpassen. Aus diesem Grund eignet sich Ambisonics auch sehr gut als Speicherformat, da es sich auch problemlos an zukünftige Wiedergabesetups anpassen lässt. Eine Mischung muss somit nur einmal erfolgen und kann über verschiedene Plattformen und Wiedergabesysteme ausgewertet werden, wodurch der Aufwand in der Postproduktion reduziert werden kann.

Im Gegensatz zu vielen kanalbasierten Formaten, bildet Ambisonics den gesamten 360°-Raum um eine zentrale Hörposition ab. Dadurch ist es möglich, auch Schallquellen unterhalb der Hörposition darstellen zu können.

### 7.2.6 Nachteile

Ein Nachteil von Ambisonics ist, dass für eine räumlich hochauflösende Reproduktion des Schallfeldes eine große Anzahl an Übertragungskanälen erforderlich ist. Für Ambisonics der dritten Ordnung werden bereits 16 Kanäle benötigt.

Zwar liefert FOA mit nur vier Kanälen bereits eine vollständige Beschreibung des Schallfeldes, dabei kommt es jedoch laut Wittek zu starken Qualitätsverlusten: „Als Speicherformat für beliebige räumliche Signale ist Ambisonics sehr gut geeignet, aber wiederum nur, wenn die Ordnung groß genug ist. Ein Speicherformat mit nur vier Kanälen [...] erzeugt aus jeder 3D-Aufnahme einen Brei, denn die vormals gute Signaltrennung im 3D-Setup wird durch die Mischung auf 4 Kanäle zerstört.“<sup>123</sup>

### 7.2.7 Ambisonics für 360°-Video

Insgesamt stellt Ambisonics ein sehr geeignetes Format für 360°-Videoproduktionen dar und die Vorteile überwiegen. Ambisonics wird grundsätzlich bereits von vielen Plattformen unterstützt und weitere Verbreitungswege werden zukünftig die Wiedergabe von Ambisonics-Inhalten unterstützen. Aktuell gibt es noch keinen einheitlichen Standard innerhalb der Ambisonics-Formate für 360°-Video. Daher ist es empfehlenswert, zunächst in höchstmöglicher Qualität und Ordnungszahl zu produzieren. Eine nachträgliche Umwandlung in niedriger aufgelöste Ambisonics-Formate für eine spezifische Plattform ist einfach möglich. Dies sichert ein Stück weit die Wiederverwendbarkeit von Produktionen in der Zukunft, wenn Ambisonics höherer Ordnung von weiteren Plattformen unterstützt wird.

---

<sup>122</sup> vgl. Shivappa u.a. (2016), S. 4

<sup>123</sup> Wittek (2015)

Der Tonschaffende im Kontext von 360°-Videoproduktionen muss sich zunächst an das neue Produktionsformat Ambisonics gewöhnen. Die Philosophie bei der Mischung einer akustischen Szene unterscheidet sich von der einer kanalbasierten Mischung. Eine in der Digital Audio Workstation (DAW) erstellte Mischung kann nicht direkt abgehört werden und muss zunächst für ein Wiedergabegerät, in der Regel Kopfhörer, dekodiert werden. Die Dekodierung der Ambisonics-Signale ist aber nicht Bestandteil der Mischung, da sie erst nutzerseitig bei der Wiedergabe geschieht und sich damit außerhalb des Einflussbereichs des Tonschaffenden befindet. Je nach Ausspielplattform setzen Anbieter leicht unterschiedliche Dekodier-Verfahren und andere virtuelle Lautsprecheranordnungen und HRFT-Datensätze bei der Binauralsynthese ein. Das während der Mischung abgehörte Signal entspricht daher nicht unbedingt genau dem Signal bei Wiedergabe über eine bestimmte Plattform, weshalb eine Endkontrolle über den Verbreitungsweg empfehlenswert scheint.

## 8. Ausspielwege und Plattformen für 360°-Videos

*Youtube* unterstützt 360°-Video mit 3D-Audio im Ambisonics Format. Aktuell wird nur FOA im Ambix Format akzeptiert. Der Support für HOA ist jedoch angekündigt<sup>124</sup>. Außer dem Ambisonics-Audiotrack dürfen keine weiteren Audiodaten im Videocontainer enthalten sein. Als Videocontainer sind mp4 und mov möglich, wobei mp4 nur Audiodaten im AAC-Format unterstützt. Die Audiosamplerate sollte 48 kHz betragen<sup>125</sup>. 360°-Videos auf Youtube lassen sich sowohl über aktuelle Browser auf Desktop Computern betrachten, als auch über die Youtube-App auf Smartphones und Tablet PCs mit Android oder iOS Betriebssystem<sup>126</sup>. Für eine korrekte Wiedergabe eines 360°-Videos muss die Videodatei spezifische 360°-Metadaten enthalten<sup>127</sup>. Diese stellen auch die richtige Interpretation der Ambisonics-Audiodaten sicher. Metadaten können innerhalb der Postproduktion beispielsweise über *Adobe Premiere Pro*, *FB360 Encoder* oder den *Youtube Social Media Metadata Injector*<sup>128</sup> generiert werden.

Mit *Facebook* unterstützt eine weitere große Online-Plattform 360°-Videos. Neben FOA unterstützt *Facebook* auch ein hybrides HOA-Format, welches eine bessere Qualität bietet. Dieses Format ist auch unter dem Namen *two big ears*<sup>129</sup> (Dateiendung .tbe) bekannt. Das tbe Format beinhaltet ein achtkanaliges Ambisonics-Signal, das durch Weglassen einer Komponente aus dem neunkanaligen Ambisonics-Signal zweiter Ordnung generiert wird<sup>130</sup>. Optional ist es möglich zusätzlich zwei Audiokanäle zu integrieren, welche nicht zur 3D-Audioszene gehören. Diese werden bei der Binauralsynthese nicht berücksichtigt, wodurch sie unabhängig von der Orientierung des Nutzers im

---

<sup>124</sup> vgl. Dieterle u.a. (2018)

<sup>125</sup> vgl. Google; Youtube (2018a)

<sup>126</sup> vgl. Rieger (2017)

<sup>127</sup> vgl. Google; Youtube (2018b)

<sup>128</sup> vgl. Google; Youtube (2018c)

<sup>129</sup> Benannt nach der Entwicklerfirma, die von Facebook übernommen wurde

<sup>130</sup> vgl. Farina (2017)

Kopf lokalisiert werden<sup>131</sup>. Die Audiosamplerate beträgt 48 kHz. Für die korrekte Interpretation der Daten müssen ebenfalls entsprechende Metadaten vorhanden sein. Die Wiedergabe ist durch aktuelle Browser und die *Facebook App* möglich, welche für die meisten mobilen Geräte verfügbar ist<sup>132</sup>.

Die Videoplattform *Vimeo* unterstützt ebenfalls 360°-Videos. Diese lassen sich über aktuelle Browser und die *Vimeo App* auf vielen mobilen Geräten wiedergeben<sup>133</sup>. Aktuell wird nur Stereoton unterstützt, zukünftig soll auch 3D-Audio möglich sein<sup>134</sup>.

Das *Samsung VR-Portal* bietet eine weitere Möglichkeit zu Verbreitung von 360°-Medien. Die Inhalte können wahlweise über aktuelle Internetbrowser, die *Samsung VR-Android-App* oder bei Verwendung einer *Samsung VR-Brille*, über die *GearVR Oculus App* genutzt werden<sup>135</sup>. Audioseitig wird neben Mono und Stereo auch 5.1 Surround, Quad-Binaural und FOA im Ambix Format unterstützt<sup>136</sup>.

*Jaunt VR* hat sich auf die Verbreitung exklusiver 360°-Video und VR-Medien spezialisiert und veröffentlicht nur ausgewählte Beiträge nach einer Prüfung<sup>137</sup>. Neben aktuellen Internetbrowsern können ebenfalls Apps für *iOS*, *Android*, *Gear VR*, *Oculus Rift*, *Vive* und *PlayStation VR* genutzt werden<sup>138</sup>. *Jaunt* unterstützt 3D-Audio in Ambisonics und *Dolby Atmos VR*<sup>139</sup>.

Es lässt sich festhalten, dass aktuell mehrere Plattformen für die Verbreitung von 360°-Videos zur Verfügung stehen. Die Plattformen *Youtube* und *Facebook* sind vor allem deshalb interessant, weil sich über sie eine große Reichweite erzielen lässt und eventuell auf eine größere Anzahl bereits vorhandener Abonnenten und Follower zurückgegriffen werden kann. Die vielfältigen Wiedergabemöglichkeiten bedingen laut Schütze ein großes Wachstumspotenzial<sup>140</sup>. Aber auch Plattformen wie *Samsung VR* und *Jaunt VR* bieten Möglichkeiten für den Rundfunk. Auf diesen Plattformen sind vornehmlich Nutzer mit speziellem Interesse an 360°- und VR-Medien zu finden, weshalb sich durch hochwertige Produktionen eine neue Zielgruppe für den Rundfunk erschließen lässt. Eine Auswertung von 360°-Videoproduktionen über mehrere Plattformen ergibt daher Sinn, zumal die Anpassung einer Ambisonics-Mischung an die jeweiligen Verbreitungsplattformen in der Postproduktion einfach möglich ist.

---

<sup>131</sup> vgl. Facebook (2018)

<sup>132</sup> vgl. Rieger (2017)

<sup>133</sup> vgl. Vimeo (2018)

<sup>134</sup> vgl. Vimeo (2017)

<sup>135</sup> vgl. Rieger (2017)

<sup>136</sup> vgl. Samsung (2018)

<sup>137</sup> vgl. Jaunt VR (2018a)

<sup>138</sup> vgl. Jaunt VR (2018b)

<sup>139</sup> vgl. Rieger (2017)

<sup>140</sup> vgl. Schütze; Irwin-Schütze (2018), S. 26

---

## 9. Mikrofonaufnahmetechnik für 360°- Videoproduktionen

Die Anforderungen an die Mikrofonaufnahmetechnik sind im Allgemeinen sehr hoch, besonders im Kontext von Musikproduktionen. Als wichtigste Punkte nennt Camerer in diesem Zusammenhang eine gute Klangfarbe, gute Tiefenwiedergabe, eine hohe Dekorrelation des Diffusfeldes sowie ein möglichst großer Sweetspot bei der Wiedergabe<sup>141</sup>. Für 360°-Videoproduktionen kommen weitere Anforderungen hinzu. Es soll ein möglichst realistischer, dreidimensionaler Raumeindruck mit Höhenkanälen aufgezeichnet werden und der Mikrofonaufbau soll möglichst kompakt und praktikabel bleiben.

Bei vielen 360°-Videoproduktion wird aktuell ausschließlich ein FOA-Mikrofon eingesetzt, um direkt am Kamerastandort ein dreidimensionales Schallfeld aufzuzeichnen. Neben diesem Ansatz stellt dieses Kapitel weitere mögliche Mikrofonaufnahmetechniken für 360°-Videoproduktionen vor und untersucht diese hinsichtlich ihrer Eignung für Musikaufnahmen. Anschließend wird die Entwicklung und Umsetzung eines Mikrofonsetups für diesen Einsatzzweck vorgestellt.

### 9.1 Ambisonics-Mikrofone

Ambisonics-Mikrofone lassen sich in FOA- und HOA-Mikrofone unterteilen. Beiden Typen ist gemein, dass es sich um ein koinzidentes Aufnahmeverfahren handelt, welches „auf einer mathematischen Theorie der Schallfeldabtastung auf einer kugelsymmetrischen Oberfläche“ basiert<sup>142</sup>. Die mit ihnen aufgezeichneten Signale sind nicht unmittelbar für eine Lautsprecherwiedergabe geeignet und müssen zunächst in das Ambisonics-B-Format überführt und anschließend in Lautsprechersignale dekodiert werden.

#### 9.1.1 FOA-Mikrofone

Mit dem Einsatz von FOA-Mikrofonen zielt man darauf ab, direkt eine Schallfeldbeschreibung in FOA aufzuzeichnen. Der nächstliegende Ansatz ist, die Ambisonics-Komponenten der 0. und 1. Ordnung gemäß Abbildung 9 mit Mikrofonen der Richtcharakteristiken Kugel und Acht aufzuzeichnen. Wobei das Kugelmikrofon die Schalldruckinformation des Schallfeldes erfasst und die Achter-Mikrofone die richtungsabhängige Schallschnelle<sup>143</sup>. Bei diesem Verfahren werden direkt die Komponenten W, X, Y, Z generiert, wobei es sich um das Ambisonics-B-Format handelt. Ein solcher Mikrofonaufbau wird daher auch B-Format-Mikrofon genannt. Es ist ein koinzidentes Mikrofonaufnahmeverfahren und kann als 3D-Erweiterung der M/S-Stereofonie aufgefasst werden, bei dem alle Richtungen gleichwertig behandelt werden.

---

<sup>141</sup> vgl. Camerer (2018)

<sup>142</sup> Weinzierl (2008), S. 593

<sup>143</sup> vgl. Nicol (2018), S. 278-279

Da sich ein koinzidenter Aufbau von vier konventionellen Mikrofonen in der Praxis nur mühsam umsetzen lässt, entwickelten Craven und Gerzon einen weiteren Ansatz<sup>144</sup>. Dabei werden gemäß Abbildung 10 vier Nierenmikrofone auf einem Tetraeder angeordnet, wodurch alle vier Kapseln in einem Winkel von 60° zu einander stehen<sup>145</sup>. Durch elektrische Kompensation wird der bautechnisch bedingte Abstand der Kapseln ausgeglichen<sup>146</sup>. Wie in Abbildung 10 dargestellt, liefert ein solches Mikrofon die Signale Left Front Up (LFU), Right Front Down (RFD), Left Back Down (LBD), Right Back Up (RBU). Bei diesen Signalen handelt es sich um das Ambisonics-A-Format, weshalb ein solches Mikrofon auch A-Format-Mikrofon genannt wird. Zur Weiterverarbeitung werden die Signale des A-Formats durch folgende Summen- und Differenzbildung in das B-Format überführt:

$$\begin{aligned}
 W &= LFU + LBD + RFD + RBU && \text{Formel (4)} \\
 X &= LFU - LBD + RFD - RBU \\
 Y &= LFU + LBD - RFD - RBU \\
 Z &= LFU - LBD - RFD + RBU
 \end{aligned}$$

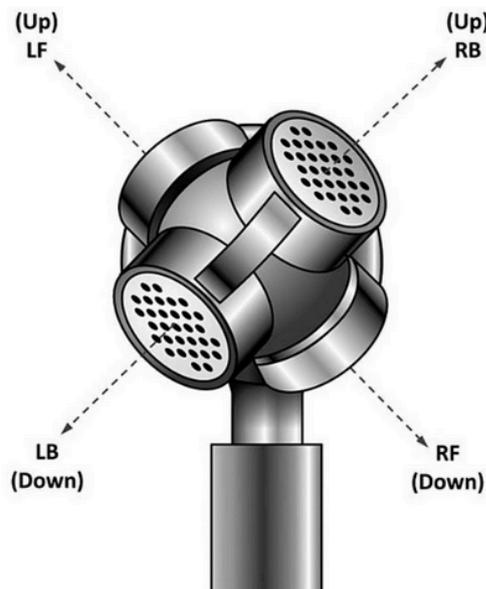


Abbildung 10 - Signale eines Ambisonics-A-Format Mikrofons<sup>147</sup>

Bekannte Vertreter für FOA-Mikrofone sind: *Soundfield MKV*, *Core Sound Tetra Mic*, *Sennheiser AMBEO VR Mic* (siehe Abbildung 11) oder das *Rode NT-SFI*. Bates et al., untersuchten in Hörvergleichen die Klangqualitäten einiger FOA-Mikrofone. Demnach lieferte das *Soundfield MKV* die

<sup>144</sup> vgl. Craven; Gerzon (1977)

<sup>145</sup> vgl. Friesecke (2014)

<sup>146</sup> vgl. Gerzon (1975), S. 1-4

<sup>147</sup> Bildquelle: Nicol (2018), s. 280

beste allgemeine Klangqualität vor dem *Sennheiser AMBEO* und dem *Core Sound Tetra Mic*. In Bezug auf die Richtungsabbildung lag das *Sennheiser AMBEO* vor dem *Soundfield MKV* und dem *Core Sound Tetra Mic*<sup>148</sup>.



Abbildung 11 - Sennheiser AMBEO VR Mic mit (links) und ohne Korb (rechts)<sup>149</sup>

Es lässt sich festhalten, dass FOA-Mikrofone eine praktikable Lösung darstellen, um mit einem koinzidenten Mikrofonaufbau ein Ambisonics-Schallfeld aufzuzeichnen. Ein Vorteil ist der geringe Platzbedarf, wodurch sich ein solches Mikrofon sehr einfach an einer 360°-Videokamera anbringen lässt. Wie bei allen rein auf Intensitätsstereofonie basierten Mikrofonsystemen ist auch bei FOA-Mikrofonen die räumliche Abbildung aufgrund fehlender Laufzeitdifferenzen und zu geringer Dekorrelation von minderer Qualität<sup>150</sup>. Zudem kommt es durch den nicht genau koinzidenten Aufbau, trotz elektronischer Kompensation, zu Klangfarbenänderungen, die als Phasing wahrgenommen werden<sup>151</sup>. Der Einsatz von FOA-Mikrofonen bei Musikaufnahmen wird demnach nicht zu optimalen Ergebnissen führen.

<sup>148</sup> vgl. Bates u.a. (2016), S. 7-8; Bates u.a. (2017), S. 9

<sup>149</sup> Bildquelle: Eigene Abbildung

<sup>150</sup> vgl. Wittek (2018); Lee (2018)

<sup>151</sup> vgl. Wittek (2018)

### 9.1.2 HOA-Mikrofone

Mit HOA-Mikrofonen wird versucht direkt ein Ambisonics-Schallfeld höherer Ordnung aufzuzeichnen. Dabei wird eine hohe Anzahl an Mikrofonkapseln mit Nierencharakteristik auf der Oberfläche einer Kugel angebracht. Mittels Beamforming-Verfahren werden aus den Signalen der Mikrofonkapseln Ambisonics-Komponenten höherer Ordnung berechnet<sup>152</sup>. Je mehr Mikrofonkapseln zum Einsatz kommen, desto höher ist die Ordnung des Ambisonics-Schallfeldes. Das OctoMic von CoreSounds beispielsweise verwendet acht Mikrofonkapseln<sup>153</sup>, womit Ambisonics 2. Ordnung möglich ist. Das *EigenMic* von *MH Acoustics* besitzt 32 Mikrofonkapseln<sup>154</sup> und das *VisiSonics 5/64* kombiniert 64 Mikrofonkapseln und fünf Kameras in einem kompakten Gehäuse<sup>155</sup>.

In den Vergleichstests von Bates et al. wurde das *EigenMic* von *MH Acoustics* als einziges HOA-Mikrofon mit den anderen FOA-Mikrofonen verglichen. Dabei fanden sie heraus, dass das *EigenMic* die beste Richtungsabbildung im Testfeld besaß, bei den Kriterien Klangfarbe und allgemeine Qualität aber hinter den anderen Mikrofonen lag<sup>156</sup>. Eine Erklärung für dieses Ergebnis ist bei Camerer zu finden. Er stellt fest, dass durch die hohe Anzahl an Mikrofonkapseln und -verstärkern auf kleinstem Raum viele Kompromisse auf Seiten der Klangfarbe in Kauf genommen werden<sup>157</sup>. Anstatt hochwertiger Kondensatorkapseln wird auf günstige Elektret-Kapseln zurückgegriffen, was sich negativ im Klang bemerkbar macht<sup>158</sup>. Rumsey hat in seinen Untersuchungen zur allgemeinen Audioqualität herausgefunden, dass die wahrgenommene Audioqualität zum größten Teil von der Klangfarbe abhängt<sup>159</sup>.

Es lässt sich festhalten, dass sich mit HOA-Mikrofonen Ambisonics-Schallfelder höherer Ordnung praktikabel aufzeichnen lassen und diese dabei eine präzise Richtungsabbildung bieten. Allerdings sind HOA-Mikrofone relativ teuer und auf Seiten der Klangfarbe stark kompromissbehaftet, was zu einer allgemein ungenügenden Audioqualität führt. Besonders im Kontext von Musikaufnahmen spielt die Klangfarbe eine entscheidende Rolle, weshalb HOA-Mikrofone zur Aufnahme musikalischer Darbietungen nicht optimal geeignet sind.

## 9.2 Stereofone 3D-Mikrofonarrays

Durch die logische Fortsetzung der Entwicklung von stereofonen Mikrofon-Setup und Surround-Setups sind bisher viele verschiedene 3D-Mikrofon-Setups entwickelt worden. Die Grundidee ist die

---

<sup>152</sup> vgl. Meyer; Elko (2016), S. 3-8

<sup>153</sup> vgl. Core Sound, LLC (2017)

<sup>154</sup> vgl. mh Acoustics LLC (2018)

<sup>155</sup> vgl. VisiSonics Corporation (2015)

<sup>156</sup> vgl. Bates u.a. (2016), S. 7-8; Bates u.a. (2017), S. 9

<sup>157</sup> vgl. Camerer (2018)

<sup>158</sup> vgl. Ebenda

<sup>159</sup> Rumsey zitiert nach Wittek (2018)

---

Erweiterung erprobter Surround-Setups um Höhenkanäle. Dabei lassen sich die verschiedenen 3D-Mikrofon-Setups in zwei Kategorien einteilen, je nachdem ob bei der Erfassung der Höherebene auf Laufzeit- oder auf Intensitätsstereofonie gesetzt wird.

So, wie die meisten Surround-Setups, besitzen auch ihre 3D-Weiterentwicklungen eine Vorzugsrichtung. Es kommen dezidierte Mikrofone zum Einsatz, welche für die Wiedergabe in einem Center-Kanal ausgelegt sind. Es liegt somit hohe Priorität auf einer frontalen Klangbühne. Bei 360°-Videoproduktionen ist jede Raumrichtung als gleichwertig zu betrachten, weshalb 3D-Mikrofonsetups mit deutlicher Vorzugsrichtung für den Einsatz bei 360°-Videoproduktionen weniger geeignet sind.

Durch stereofone 3D-Mikrofonarrays lässt sich eine akustische Szene dreidimensional aufzeichnen. Im Gegensatz zu Ambisonics-Mikrofonen können sehr hochwertige Schallwandler zum Aufbau des Mikrofon-Setups verwendet werden. Dadurch kann sichergestellt werden, dass für das im musikalischen Kontext wichtige Kriterium der Klangfarbe keine Kompromisse eingegangen werden<sup>160</sup>. Für eine hochwertige Aufnahme ist neben der Qualität der Schallwandler auch der Aufbau des 3D-Mikrofonarrays entscheidend. Die grundsätzlichen Anforderungen an einen solchen Aufbau beschreiben Wittek und Theile wie folgt<sup>161</sup>:

- Hohe Signaltrennung zwischen allen Kanälen, um Kammfilter zu vermeiden. Es darf kein Signal mit signifikantem Pegel auf mehr als zwei Kanälen anliegen.
- Die angestrebten Abbildungseigenschaften werden durch Laufzeit- und Pegeldifferenzen zwischen benachbarten Kanälen erreicht.
- Eine hohe Dekorrelation des Diffusfeldes ist wichtig für die Umhüllung und allgemeine Klangqualität.

Durch unterschiedliche ästhetische und praktische Anforderungen wurden viele 3D-Mikrofon-Setups entworfen. Im Folgenden sollen einige Mikrofon-Setups beschrieben werden, die sich dadurch auszeichnen, dass sie keine Vorzugsrichtung besitzen und sich damit grundsätzlich für den Einsatz bei 360°-Videoproduktionen eignen:

### **9.2.1 2L-Cube**

Das 2L-Cube Array geht auf Morten Lindberg vom norwegischen Label 2L zurück. Ausgehend vom Decca-Tree Array für Zweikanalstereofonie, werden beim 2L-Cube Array acht Mikrofone mit Kugelcharakteristik eingesetzt. Die Mikrofone werden, wie in Abbildung 12 dargestellt, in Würfelform

---

<sup>160</sup> Camerer (2018)

<sup>161</sup> vgl. Wittek; Theile (2016), S. 2-3

angeordnet<sup>162</sup>. Der Kapselabstand kann laut Lindberg zwischen 40 cm und 150 cm variieren, je nachdem ob das Array im Rahmen einer großen Orchester- oder Kammermusik-Aufführung eingesetzt wird<sup>163</sup>. Der Aufbau zielt darauf ab, auf einem System mit acht äquidistanten Lautsprechern in Würfelform wiedergegeben zu werden. Dieser Ansatz findet auch bei der Binauralisierung eines Ambisonics-Schallfeldes Anwendung, weshalb das 2L-Cube Array für den Einsatz bei 360°-Videoproduktionen interessant ist.

Es ist anzumerken, dass kleine Kapselabstände beim 2L-Cube für den Einsatz bei 360°-Videoproduktionen weniger geeignet scheinen. Erst bei Frequenzen über 500 Hz wird bei 40 cm Kapselabstand eine hohe Dekorrelation erreicht<sup>164</sup>, die für die gewünschte räumliche Abbildung des Diffusfeldes notwendig ist. Allerdings sind Arrays mit großen Mikrofonabständen grundsätzlich weniger praktikabel und oftmals ein Störfaktor im Videobild.

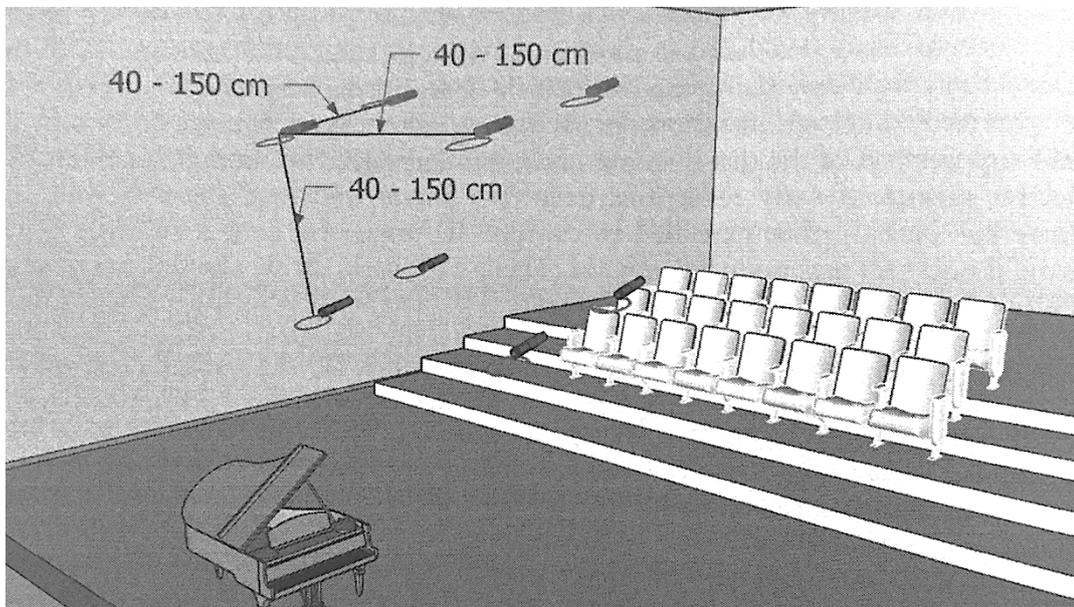


Abbildung 12 - 2L Mikrofonarray<sup>165</sup>

### 9.2.2 Twins Square/Cube

Das Twins Square und Twins Cube Array wurde von Gregor Zielinsky entwickelt und ist auch unter dem Namen *Ambeo* Square/Cube bekannt<sup>166</sup>. Bei diesen Mikrofonarrays kommen spezielle Twin-Mikrofone von Sennheiser zum Einsatz<sup>167</sup>. Bei diesem Mikrofontyp sind zwei Kapseln mit Nierencharakteristik verbaut, die rückwertig zueinander angeordnet sind und jeweils einen diskreten

<sup>162</sup> vgl. Kim (2018), S. 235-236

<sup>163</sup> vgl. Lindberg (2015), S. 15

<sup>164</sup> Weiss; Wittek (2018)

<sup>165</sup> Bildquelle: Kim (2018)

<sup>166</sup> Sennheiser Electronic GmbH und Co. KG (2018a)

<sup>167</sup> Das erste Twin Mikrofon wurde ursprünglich von der Firma Pearl Microphones entwickelt

Ausgang besitzen<sup>168</sup>. Beim Twins Square Mikrofonarray bilden vier Twin-Mikrofone ein aufrecht stehendes Quadrat mit einer Seitenlänge von etwa 2 m. Somit bildet sich frontal sowie rückseitig in horizontaler und vertikaler Ebene jeweils ein Laufzeitbasiertes Stereopaar aus. Seitlich handelt es sich somit um ein koinzidentes Stereopaar mit einem Öffnungswinkel von 180°. Die vorderen Kapseln der Mikrofone liefern die Signale der frontalen unteren und oberen Ebene und die hinteren Kapseln die Signale der rückwärtigen unteren und oberen Ebene. Mit nur vier Mikrofonen kann so eine 3D-Aufnahme erfolgen.

Abbildung 13 zeigt das Twins Cube Array. Bei diesem Aufbau werden zwei Twins Square Arrays hintereinander angeordnet, sodass sich Laufzeitunterschiede zwischen Front- und Rückseite ausbilden. In diesem Array dienen die Doppelkapselsysteme folglich nicht mehr dazu, die Signale der rückseitigen Kanäle zu generieren, sondern zur nachträglichen Anpassung der Richtcharakteristik mittels elektronischer Kombination. Wird diese Flexibilität nicht benötigt, kann der Aufbau auch mit Nieren oder breiten Nieren erfolgen<sup>169</sup>.

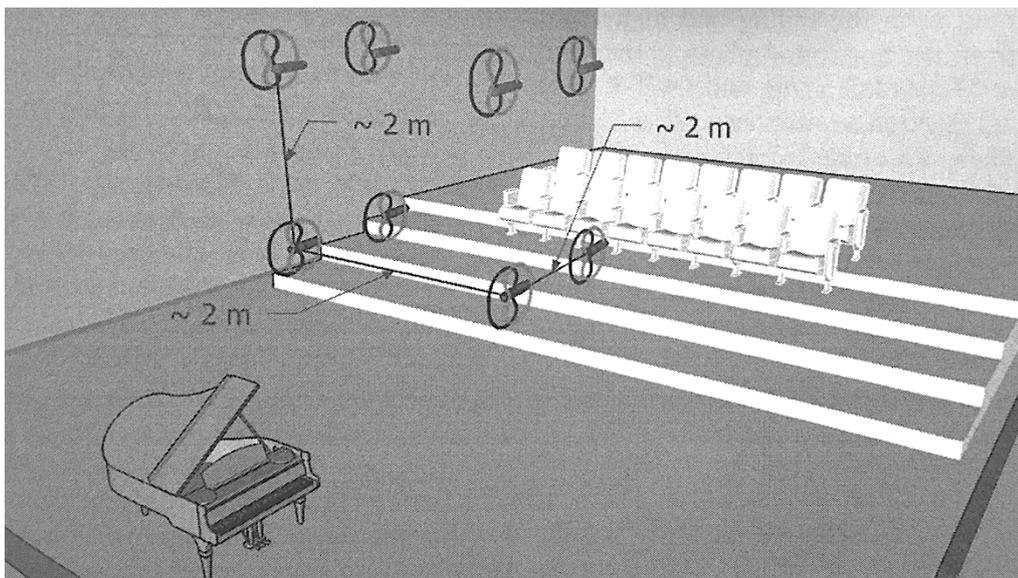


Abbildung 13 - Twins Cube Mikrofonarray<sup>170</sup>

### 9.2.3 Coincident Z-Microphone Technique

Die von Paul Geluso entwickelte Coincident Z-Microphone Technique basiert auf der Ambisonics- und M/S-Aufnahmetechnik. Dabei wird ein vertikal ausgerichtetes Mikrofon mit Achtercharakteristik mit einem horizontal ausgerichtetem Mikrofon kombiniert<sup>171</sup>. Somit wird durch die Acht nicht das Seitensignal S wie bei der M/S-Aufnahmetechnik, sondern das Höhsignal Z erfasst. Dieses Prinzip

<sup>168</sup> vgl. Kim (2018), S. 236-237

<sup>169</sup> vgl. Zielinsky zitiert nach Ebenda, S. 237

<sup>170</sup> Bildquelle: Ebenda

<sup>171</sup> vgl. Ebenda, S. 237

lässt sich auf viele Surround Mikrofonarrays übertragen und ist nicht auf einen bestimmten Aufbau beschränkt<sup>172</sup>. Abbildung 14 zeigt ein mögliches Array, das diese M/Z Aufnahmetechnik verwendet.

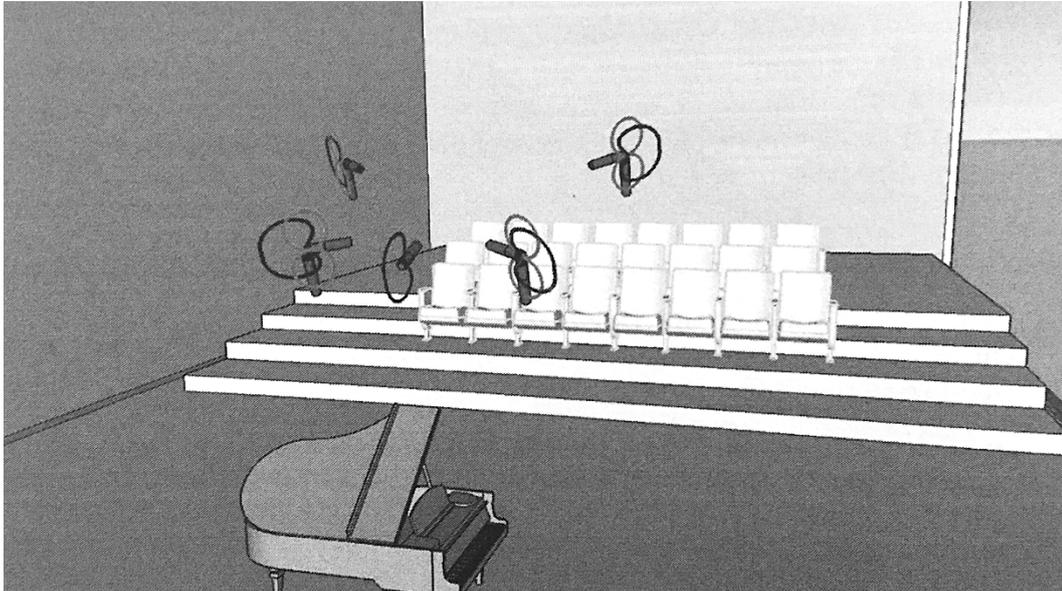


Abbildung 14 - 3D-Mikrofonarray nach der Coincident Z-Microphone Technique<sup>173</sup>

#### 9.2.4 ESMA 3D

Dieses Mikrofonarray basiert auf dem Konzept des Equal Segment Microphone Array (ESMA) von Williams. Dabei besitzen alle nebeneinander liegenden Mikrofone einen gleich großen Öffnungswinkel, sodass die Aufnahmewinkel aller Stereo-Segmente gleich sind<sup>174</sup>. Dieser Aufbau ermöglicht die gleichmäßige Abbildung des 360°-Raumes auf einer Ebene. Je mehr Stereo-Segmente im Mikrofonarray enthalten sind, desto feiner ist die Auflösung. Aufnahmen mit einem ESMA-Array werden idealerweise über ein Lautsprecher-Setup wiedergegeben, das dieselbe Winkelanordnung besitzt wie das Mikrofonarray<sup>175</sup>.

Im Rahmen von Untersuchungen zu VR-Mikrofontechniken experimentierte Lee mit verschiedenen ESMA-Mikrofonsetups. Dabei verglich er ESMA-Mikrofonarrays mit 4 Nierenmikrofonen und Kapselabständen von 24, 30 und 50 cm. Er fand heraus, dass der Aufbau mit 50 cm Kapselabstand die gleichmäßigste Abbildung und höchste Lokalisierungsgenauigkeit im Testfeld aufwies, die Setups mit 24 cm und 30 cm Kapselabstand aber ebenfalls gute Ergebnisse erzielten.

Zur 3D-Erweiterung des vierkanaligen ESMA-Array setzt Lee auf ein koinzidentes Verfahren zur Abbildung der Höheninformation<sup>176</sup>. Dazu werden die Nierenmikrofone um ein nach oben gerichtetes

<sup>172</sup> vgl. Geluso (2012), S. 5

<sup>173</sup> Bildquelle: Kim (2018)

<sup>174</sup> vgl. Lee (2018)

<sup>175</sup> vgl. Lee (2016), S. 4

<sup>176</sup> vgl. Ebenda, S- 8-10

Mikrofon mit Achtercharakteristik ergänzt, wodurch sich nach der Coincident Z-Microphone Technique ein vertikales MS-Stereosystem (M/Z) bildet. Durch diese Ausrichtung wird eine sehr gute Kanaltrennung erreicht, da horizontaler Direktschall weitgehend ausgeblendet wird. Die Signale der Achtermikrofone können entweder direkt auf Höhenlautsprecher gegeben werden, wenn lediglich eine Erweiterung des Höreindrucks gewünscht wird oder zusammen mit den Signalen der Nierenmikrofone einem MS-Decoder zugeführt werden, um eine vertikale Lokalisation zu ermöglichen. Ein alternativerer Aufbau verwendet zur Erfassung der Höheninformationen Mikrofone mit Supernierencharakteristik anstatt einer Achtercharakteristik<sup>177</sup>. Dazu werden die Mikrofone nach X/Y-Technik angewinkelt, was in Abbildung 15 gezeigt wird.



Abbildung 15 - ESMA 3D Mikrofonarray<sup>178</sup>

### 9.2.5 ORTF 3D

Das ORTF-3D Verfahren ist eine Entwicklung von Wittek und Theile. Das Array besteht aus zwei Ebenen und auf jeder Ebene kommen vier Mikrofone mit Supernierencharakteristik zum Einsatz. Die Mikrofone einer Ebene sind in einem Rechteck mit Kantenlängen von zehn mal zwanzig Zentimetern angeordnet<sup>179</sup>. Beide Ebenen des Mikrofonarrays sind ohne Abstand zueinander angebracht, sodass sich vertikal ein koinzidentes und horizontal ein gemischtes Stereoverfahren bildet.

Die Mikrofone beider Ebenen sind jeweils nach oben oder unten geneigt, wodurch sich die Signaltrennung in der vertikalen Ebene ergibt. So bildet sich ein gekipptes XY-Stereopaar, dessen Öffnungswinkel 90° beträgt. Diese Anordnung ist in Abbildung 16 dargestellt. Laut Wittek sorgt die hohe Richtwirkung der Supernierenmikrofone trotz der koinzidenten Anordnung für eine gute Abbildung und Diffusfeld-Dekorrelation<sup>180</sup>.

<sup>177</sup> vgl. Lee (2018)

<sup>178</sup> Bildquelle: Ebenda

<sup>179</sup> vgl. Wittek; Theile (2016), S. 3

<sup>180</sup> vgl. Wittek (2015)

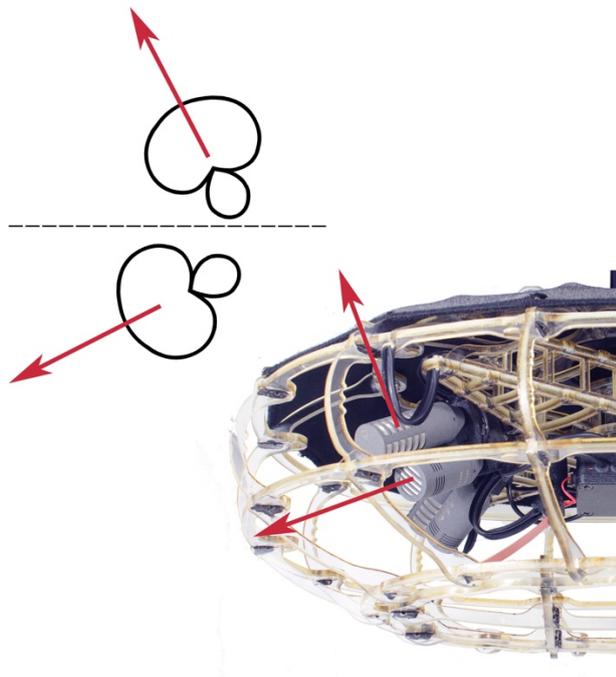


Abbildung 16 - Koinzidente Mikrofonanordnung in der Vertikalebene beim Schoeps ORTF 3D<sup>181</sup>

In der unteren Ebene des Arrays kommen kompakte Kleinmembran-Kondensatormikrofone mit frontaler Besprechungsrichtung zum Einsatz (Typ: *Schoeps CCM 41*). In der oberen Ebene werden Mikrofone mit seitlicher Besprechungsrichtung eingesetzt (Typ: *Schoeps CCM 41V*). Dadurch ist es möglich, die Mikrofongehäuse direkt übereinander anzuordnen. So ergibt sich, wie in Abbildung 17 gezeigt, eine kompakte Bauform des Mikrofonarrays, welches in einem Windkorb mit elastischer Aufhängung untergebracht ist. Die geringen Abmessungen sind praktisch für den Einsatz bei 360°-Videoproduktionen, wenn große Mikrofonarrays im Bild unerwünscht sind.

Laut Wittek und Theile entsteht durch das ORTF-3D Verfahren ein „8-Kanal-Array mit hoher Signaltrennung, optimaler Diffusfeld-Dekorrelation und guter 3D-Abbildung bei hoher Stabilität im Wiedergaberaum.“<sup>182</sup>

<sup>181</sup> Bildquelle: Wittek (2018)

<sup>182</sup> vgl. Wittek; Theile (2016), S. 4



Abbildung 17 - Schoeps ORTF 3D Mikrofonarray im Windkorb<sup>183</sup>

### 9.3 Polymikrofonie

Mikrofonaufnahmetechniken, die nach dem Prinzip der Polymikrofonie arbeiten, setzen mehrere Mikrofone in geringen Abständen ein, um einzelne Schallquellen einer Musikdarbietung getrennt voneinander zu erfassen. Die Balance des Klangbildes entsteht in der Nachbearbeitung, bei der nach dem Prinzip der Intensitätstereofonie die Schallquellen durch Panning innerhalb der Stereobasis verteilt werden. Bei der Polymikrofonie ist zwischen den Techniken der Einzelmikrofonie und der Stützmikrofonie zu unterscheiden.

#### 9.3.1 Einzelmikrofonie

Bei der Technik der Einzelmikrofonie werden ausschließlich einzelne Schallquellen eines Ensembles erfasst. Dieses Verfahren besitzt die größten Gestaltungsmöglichkeiten in der Nachbearbeitung<sup>184</sup>. Hauptanforderung dieses Verfahrens ist eine große Dämpfung einer Schallquelle zu ihren Nachbarn. Laut Dickreiter sollte die akustische Trennung einer Schallquelle mindestens 6 bis 10 dB betragen<sup>185</sup>, weshalb vornehmlich Richtmikrofone zum Einsatz kommen.

Weiterhin kann eine akustische Trennung der Schallquellen über die Anordnung der Musiker erreicht werden. Hierbei sollten möglichst gleich laute Instrumente zusammen angeordnet werden. In der E-Musik ist meist eine Standardsitzposition vorgegeben, die sich traditionell bewährt hat, einen guten akustischen Gesamtklang erzeugt und gute Sichtverbindung zwischen Dirigent und Musikern erlaubt<sup>186</sup>. Hinsichtlich der akustischen Trennung ist eine solche Anordnung mit Kompromissen behaftet.

<sup>183</sup> Bildquelle: Wittek (2018)

<sup>184</sup> vgl. Dickreiter (1995), S. 118

<sup>185</sup> vgl. Ebenda, S. 118-121

<sup>186</sup> vgl. Ebenda, S. 48-51

Durch den Einsatz von akustischen Trennwänden oder der Overdub-Aufnahmetechnik kann das akustische Übersprechen ebenfalls reduziert werden. Im Rahmen von 360°-Videoproduktionen sind sie jedoch nicht geeignet, da akustische Trennwände das Bild zu sehr beeinträchtigen und die Overdub-Aufnahmetechnik ein echtes Zusammenspiel der Musiker verhindert.

Die Einzelmikrofonierung von Schallquellen mit kurzen Mikrofonabständen sorgt für ein präsenten Klangbild. Nach Dickreiter gewinnen die Signale so an Deutlichkeit und Brillanz, ihre Kontur tritt deutlicher hervor und erzeugen ein Gefühl von Nähe<sup>187</sup>. Der Eindruck von Nähe kommt dadurch zustande, dass durch die Einzelmikrofonierung Diffusschall weitgehend ausgeblendet wird und die Einschwingvorgänge der Instrumente und die Geräuschanteile ihres Klangs in den Vordergrund treten<sup>188</sup>. Diese klanglichen Eigenschaften sind sehr hilfreich um in einem 360°-Video das Gefühl zu vermitteln, sich direkt neben einem Musiker zu befinden. Darüber hinaus ist es mit Einzelsignalen einfach möglich, die akustische Perspektive an den Kamerastandpunkt im 360°-Video anzupassen. Teile gibt jedoch zu bedenken, dass geringe Mikrofonabstände für die Präsenz zwar sehr förderlich seien, aber keinerlei Räumlichkeit übermitteln<sup>189</sup>. Wie in Kapitel 5.3 aufgezeigt, kommt der Wiedergabe der Räumlichkeit eine große Bedeutung für die Immersion des Mediennutzers zu. Demnach liefert der alleinige Einsatz der Einzelmikrofonie kein zufriedenstellendes Ergebnis für 360°-Videoproduktionen im Kontext musikalischer Aufführungspraxis.

### 9.3.2 Stützmikrofonie

Bei der Stützmikrofonie wird ein (3D-) Hauptmikrofon mit der Einzelmikrofonie kombiniert. Ziel ist es, durch das Hauptmikrofon einen räumlichen Gesamteindruck aufzuzeichnen und durch Einzelmikrofone die benötigte Nähe und Präsenz zu vermitteln. Je nach Gewichtung können die Stützmikrofone das Hauptmikrofonverfahren optimieren oder es dominieren, wobei der Übergang fließend ist<sup>190</sup>. Dickreiter merkt an, dass der Pegelunterschied zwischen Haupt- und Stützmikrofon ausreichend groß sein sollte, da es aufgrund unterschiedlicher Laufzeiten sonst zu störenden Kammfiltereffekten käme<sup>191</sup>. Nach Weinzierl kann ein Stützmikrofon mehrere Funktionen erfüllen<sup>192</sup>:

- Eine Schallquelle lauter erscheinen lassen.
- Den Entfernungseindruck einer Schallquelle vermindern.
- Erzeugen eines Lokalisationsreizes durch den Präzedenzeffekt.

---

<sup>187</sup> vgl. Ebenda, S. 138

<sup>188</sup> vgl. Ebenda, S. 138

<sup>189</sup> vgl. Theile u.a. (2013), S. 274-275

<sup>190</sup> vgl. Ebenda, S. 269

<sup>191</sup> vgl. Dickreiter (1995), S. 141

<sup>192</sup> vgl. Weinzierl (2008), S. 603

- Den Anteil hoher Frequenzen einer Schallquelle betonen.

Weinzierl merkt jedoch an, dass durch die vorauseilenden Signale der Stützmikrofone das räumliche Gesamtbild des Hauptmikrofons beeinträchtigt wird<sup>193</sup>. Laut Dickreiter bedingt die durch Stützmikrofone verbesserte Lokalisierbarkeit eine Reduktion der Tiefenstaffelung<sup>194</sup>. Dieser Mangel kann laut Theile durch eine zeitverzögerte und verhallte Zumischung der Stützmikrofone ausgeglichen werden<sup>195</sup>. Eine leichte Höhenbedämpfung kann ebenfalls dazu beitragen, eine größere Entfernung einer Schallquelle zu suggerieren<sup>196</sup>.

Theile empfiehlt zunächst das Hauptmikrofon optimal aufzustellen und im Anschluss das Klangbild durch Zumischung der Stützmikrofone zu optimieren, denn „grundsätzlich sollte das Stützmikrofonverfahren nicht dazu dienen, ein unkorrekt eingesetztes Hauptmikrofon zu korrigieren“<sup>197</sup>. Hierbei ist es wichtig darauf zu achten, die Perspektive der Stützmikrofone an die des Hauptmikrofons anzugleichen, damit es nicht zur Mehrfachabbildung einer Schallquelle kommt. Im Rahmen von 360°-Videoproduktionen ist zusätzlich dafür zu sorgen, dass akustische und optische Perspektive übereinstimmen. Kommen verschiedene Kamerapositionen zum Einsatz, so ist das Klangbild für jede Perspektive abzustimmen<sup>198</sup>.

Es können sowohl Mono- als auch Stereostützen eingesetzt werden. Stereostützen bieten sich dann an, wenn ganze Instrumentengruppen gestützt werden sollen. Hierbei ist es wichtig die Abbildungsbreite der Stereostützen auf die des Hauptmikrofons anzupassen<sup>199</sup>. Die erzeugte Nähe ist bei Stereostützen jedoch nicht so groß wie bei Monostützen, weshalb Monostützen bei 360°-Videoproduktionen zu bevorzugen sind.

Deboy betont, dass Stützmikrofone im Zusammenhang mit 360°-Videoproduktionen von großer Bedeutung sind und dabei helfen, verschiedene Perspektiven glaubhaft darzustellen. Neben dem Einsatz eines 3D-Hauptmikrofons und Stützmikrofonen empfiehlt er den Einsatz weiterer Raummikrofone<sup>200</sup>.

---

<sup>193</sup> vgl. Ebenda, S. 603

<sup>194</sup> vgl. Dickreiter (1995), S. 138

<sup>195</sup> vgl. Theile u.a. (2013), S. 270

<sup>196</sup> vgl. Dickreiter (1995), S. 138-141

<sup>197</sup> Theile u.a. (2013), S. 270

<sup>198</sup> vgl. Deboy (2018)

<sup>199</sup> vgl. Theile u.a. (2013), S. 271

<sup>200</sup> vgl. Deboy (2018)

---

## 9.4 Kamerastandpunkt und Perspektive

Der Standpunkt der 360°-Kamera hat ebenfalls Auswirkungen auf die Mikrofonierung, da dieser die Perspektive vorgibt, zu der die gesamte akustische Szene passen soll. Es ist ebenfalls möglich, dass mehrere Kameras eingesetzt werden, um von unterschiedlichen Standpunkten aus das Geschehen zu betrachten. Grundsätzlich sollten solche Kamerapositionen gewählt werden, die das 360°-Medium gut bedienen können. Im Idealfall befindet sich die Kamera mitten im Geschehen, sodass der Nutzer angeregt wird, sich in alle Richtungen umzuschauen und umzuhören. Kamerapositionen auf der Bühne oder mitten im Orchester bieten sich bei musikalischen Aufführungen an. Kamerapositionen an einem Zuhörerplatz eignen sich weniger, da hier kaum Anreize geschaffen werden, sich umzuschauen und die interessanten Inhalte nur in einem kleinen Winkel dargestellt werden.

Eine Möglichkeit um der Kameraperspektive akustisch zu entsprechen ist es, ein Hauptmikrofon unmittelbar am Kamerastandpunkt aufzustellen und die Ausrichtung auf die der Kamera abzustimmen. Bei dieser Herangehensweise müsste jedoch für jede Kameraposition ein Hauptmikrofon aufgebaut werden. Zum einen vergrößert dies den zeitlichen und materiellen Aufwand und zum anderen ist es aus Platzgründen oftmals nicht möglich, mehrere Hauptmikrofone mitten auf einer Bühne oder im Orchester aufzubauen, wenn die Kamera solche Positionen besetzt.

Eine weitere Möglichkeit die akustische Szene an die Kameraperspektive anzupassen ist die Rotation in der Nachbearbeitung. Dazu wird die musikalische Darbietung von einem zentralen Punkt aus erfasst. Dieser wird nach tonmeisterlichen Gesichtspunkten ausgewählt und muss sich nicht mit der Kameraperspektive decken. Ziel ist es, mit einem Hauptmikrofon eine möglichst gleichmäßige und ausgewogene Abbildung des gesamten Klangkörpers zu erreichen. Nur sollte darauf geachtet werden, das Hauptmikrofon nicht zu weit entfernt aufzubauen, da ein höherer Direktschallpegel nützlich ist, um die Nähe zu den Musikern auch akustisch abzubilden. In der Nachbearbeitung werden die Signale des Hauptmikrofons gemäß ihrer Ausrichtung zunächst in ein Ambisonics-Schallfeld gemischt, welches anschließend so weit rotiert wird, bis es mit der Kameraperspektive übereinstimmt. Der Einsatz von Stützmikrofonen kann sehr dabei helfen, Nähe darzustellen und kann durch korrektes Panning der Signale im 360°-Schallfeld dafür sorgen, dass die akustische Perspektive mit der optischen übereinstimmt.

## 9.5 Zusammenfassung

Insgesamt bieten sich vielfältige Möglichkeiten der Mikrofontechnik, um Audio für 360°-Videoproduktionen aufzunehmen. Soll ein hochwertiges, immersives Hörerlebnis geschaffen werden, so sind die Anforderungen an die Aufnahmetechnik hoch, vor allem im Kontext von Musikaufnahmen. Neben der musikalischen Darbietung selbst spielt der Konzertsaal und die Erfassung seines Raumklangs eine wichtige Rolle. Die häufig praktizierte, einfache Lösung des FOA-Mikrofons wird nicht zu einem

überzeugenden Ergebnis führen können. 3D-Audio kann als Erweiterung der tonmeisterlichen Arbeit begriffen werden<sup>201</sup>. Die Anforderungen von 360°-Video unterscheiden sich zwar von denen einer stereofonen Musikmischung, doch haben die alten Gesetze der Aufnahmetechnik nach wie vor Gültigkeit und müssen im Zuge eines neuen medienästhetischen Ideals anders angewandt werden. Wittek hält fest, dass das stereofone Array „für alle 3D-Formate das Mittel der Wahl“<sup>202</sup> ist. Wird es zusätzlich mit Stützmikrofonen ergänzt, lässt sich auch große Nähe zu einer Schallquelle darstellen, und es bieten sich weitreichende Gestaltungsmöglichkeiten in der Nachbearbeitung. Die Kombination aus stereofonem 3D-Hauptmikrofon, mono-Stützmikrofonen und weiteren Raummikrofonen scheint somit die vielversprechendste Mikrofonaufnahmetechnik zur Produktion von hochwertigen 360°-Videoproduktionen im Kontext musikalischer Aufführungspraxis zu sein.

## 10. Entwicklung und Umsetzung eines stereofonen 3D-Audioaufnahmeverfahrens für 360°-Videoproduktionen

Im Rahmen dieser Arbeit wurde ein stereophones 3D-Mikrofonarray entwickelt und einem Praxistest unterzogen. Da das Mikrofonarray im Hinblick auf einen Einsatz bei 360°-Videoproduktionen im Rundfunk entworfen werden sollte, galt es folgende Anforderungen zu erfüllen:

- Möglichst kleine Abmessungen, um nicht einen störenden Mikrofonaufbau im 360°-Videobild festzuhalten und hohe Praktikabilität im Live-Recording zu gewährleisten.
- Möglichkeit das Mikrofonarray mit nur einem Stativ aufstellen zu können, um zügige Aufbauzeiten zu realisieren und auch bei aufwendig mikrofonierten Bühnen noch einen Stellplatz zu finden.
- Einsatz von bereits im Rundfunk vorhandenem Material, um ohne Investitionskosten im Audiobereich in die Produktion von hochwertigen 360°-Videos einsteigen zu können.

Gemäß dieser Anforderung kommen Kleinmembran-Kondensatormikrofone zum Einsatz. Diese sind meist in ausreichender Stückzahl, in gängigen Richtcharakteristiken im Rundfunk verfügbar. In hochwertiger Ausführung verfügen sie über einen sehr natürlichen und transparenten Klang bei kleinen Abmessungen<sup>203</sup>. Das Mikrofonarray wurde exemplarisch mit den Kleinmembran-

---

<sup>201</sup> vgl. Wittek (2018)

<sup>202</sup> vgl. Wittek; Theile (2016), S. 2

<sup>203</sup> vgl. Henle (2001), S. 143

---

Kondensatormikrofonen *Neumann KM84* (Nierencharakteristik), *Schoeps CMC6/Mk4* (Nierencharakteristik) und *CMC6/Mk41* (Supernierencharakteristik) aufgebaut.

### 10.1.1 Verfahren

Beim 360°-Video gibt es keine Vorzugsrichtung und jede Achse, speziell auf der horizontalen Ebene, ist als gleichwertig zu betrachten. Daher ist es sinnvoll einen Mikrofonaufbau zu wählen, der ebenfalls eine gleichwertige Abbildung des 360°-Raumes gewährleistet. Aus dem Bereich der Surround-Aufnahmeverfahren werden solche Arrays vornehmlich zur Aufnahme von Atmosphären herangezogen<sup>204</sup>. Ein typisches Konzept ist das ESMA. Das IRT-Kreuz von Theile arbeitet ebenfalls nach dem ESMA-Prinzip<sup>205</sup>. Dabei werden vier Mikrofone mit Nierencharakteristik mit einem Öffnungswinkel von 90° zueinander angeordnet, der Kapselabstand beträgt etwa 20 bis 25 cm. Demnach handelt es sich um ein gemischtes stereofones Verfahren, das sowohl Laufzeit als auch Pegeldifferenzen erzeugt. Es kann letztlich als Surround-Erweiterung der Stereosysteme ORTF<sup>206</sup> oder NOS<sup>207</sup> betrachtet werden und verfügt demnach über vergleichbare Abbildungseigenschaften, wie einer guten Darstellung der Räumlichkeit bei gleichzeitig guter Lokalisation<sup>208</sup>.

Diese Abbildungseigenschaften, die kompakten Abmessungen und der einfach umzusetzende Aufbau machen das IRT-Kreuz zu einer geeigneten Ausgangsbasis für ein stereofones 3D-Mikrofonarray für 360°-Videoproduktionen. Es bieten sich mehrere Optionen an, um das IRT-Kreuz für eine Aufnahme von Höhenkanälen zu erweitern. Ein möglicher Ansatz ist es, jeweils ein IRT-Kreuz für die Horizontalebene und eines für die Höheebene einzusetzen und diese mit passendem Abstand zueinander zu positionieren. Hierdurch werden sich jedoch die Abmessungen des Arrays stark vergrößern. Wie Wallis und Lee in ihren Untersuchungen zeigen konnte,<sup>209</sup> sind Laufzeitunterschiede für die Lokalisation in der vertikalen Ebene weniger geeignet und führen nicht zu einer stabilen Lokalisation in der Höheebenen. Die Wirksamkeit des Präzedenz-Effektes für breitbandige Signale in vertikaler Ebene konnte durch seine Untersuchungen nicht gezeigt werden<sup>210</sup>. Dagegen macht Lee deutlich, dass vornehmlich Pegeldifferenzen eingesetzt werden sollten, um eine Lokalisation in Richtung der Höheebene zu ermöglichen. Aus diesem Grund wird für dieses Mikrofonarray in der vertikalen Ebene auf ein koinzidentes Verfahren gesetzt. Die Mikrofone werden in XY-Technik

---

<sup>204</sup> vgl. Theile u.a. (2013), S: 305-306

<sup>205</sup> vgl. Lee (2018)

<sup>206</sup> Nierenmikrofone, Kapselabstand 17 cm, Versatzwinkel 55°. Entwicklung des französischen Rundfunks

<sup>207</sup> Nierenmikrofone, Kapselabstand 30 cm, Versatzwinkel 45°. Entwicklung des niederländischen Rundfunks

<sup>208</sup> vgl. Dickreiter (1995), S. 130

<sup>209</sup> vgl. Wallis; Lee (2015), S.767-776

<sup>210</sup> vgl. Lee (2018)

angeordnet, sodass zwei IRT-Kreuze in einer Ebene zusammen liegen, wobei eine Ebene nach oben und die andere nach unten ausgerichtet ist.

Wittek und Lee betonen die Bedeutung einer guten Signaltrennung im Rahmen einer stereofonen Mehrkanalaufnahme<sup>211</sup>. Eine hohe Dekorrelation ist wichtig für eine verfärbungsfreie und breite Darstellung der Räumlichkeit. Zudem wird durch Dekorrelation die Durchhörbarkeit verbessert. Diese Faktoren sind besonders wichtig um die gewünschte Immersion zu erzeugen (vgl. Kapitel 6).

Um die zur Planung des Mikrofonarrays nötigen Versatzwinkel in Kombination mit verschiedenen Richtcharakteristiken abschätzen zu können, wurde das Softwaretool „Schoeps Image Assist“ verwendet. Nach der Eingabe von Richtcharakteristik, Mikrofonabstand und Versatzwinkel für eine zweikanalige Stereoanordnung, generiert das Tool eine grafische Darstellung der Abbildung des Aufnahmewinkels zwischen den Lautsprechern, der frequenzabhängigen Korrelation und der Laufzeit- und Pegeldifferenzen.

Auf horizontaler Ebene ergibt sich durch die Mikrofonanordnung ein gemischtes Aufnahmeverfahren, bei dem sowohl Laufzeit als auch Pegeldifferenzen ausgenutzt werden. Wittek betont, dass ein konsistenter Zusammenhang von Laufzeit- und Pegeldifferenzen wichtig ist<sup>212</sup>. Abbildung 18 zeigt ein ausgewogenes Verhältnis für jeweils ein Stereopaar des Arrays mit Nierencharakteristik. Ferner ist ersichtlich, dass die Lokalisationskurve, die optimalerweise streng linear verläuft<sup>213</sup>, eine gleichmäßige Verteilung aufweist, was zu einer maßstabsgerechten Abbildung führt. Der Aufnahmewinkel beträgt etwa 94°, wodurch sich eine gleichmäßige Rundumabbildung des 360°-Raumes ergibt. Anhand der Kohärenzfunktion in Abbildung 19 ist abzulesen, dass ab ca. 1 kHz eine sehr gute Dekorrelation der Diffusfeldsignale erreicht wird, was für eine gute räumliche Abbildung relevant ist<sup>214</sup>. Kommen dagegen Mikrofone mit Supernierencharakteristik zum Einsatz, verringert sich der Aufnahmewinkel auf etwa 80°. Die Lokalisationskurve flacht leicht nach außen ab, wodurch in der Mitte verhältnismäßig kleinere Teile des Aufnahmebereichs dargestellt werden<sup>215</sup>. Zusätzlich treten beim Einsatz von Supernierenmikrofonen höhere Pegeldifferenzen auf, was zum positiven Ergebnis einer größeren Dekorrelation des Diffusfeldes führt.

---

<sup>211</sup> vgl. Wittek (2018); Lee (2018)

<sup>212</sup> vgl. Wittek (2000), S. 15

<sup>213</sup> vgl. Ebenda, S. 15

<sup>214</sup> vgl. Wittek (2018)

<sup>215</sup> vgl. Wittek (2000), S. 18-19

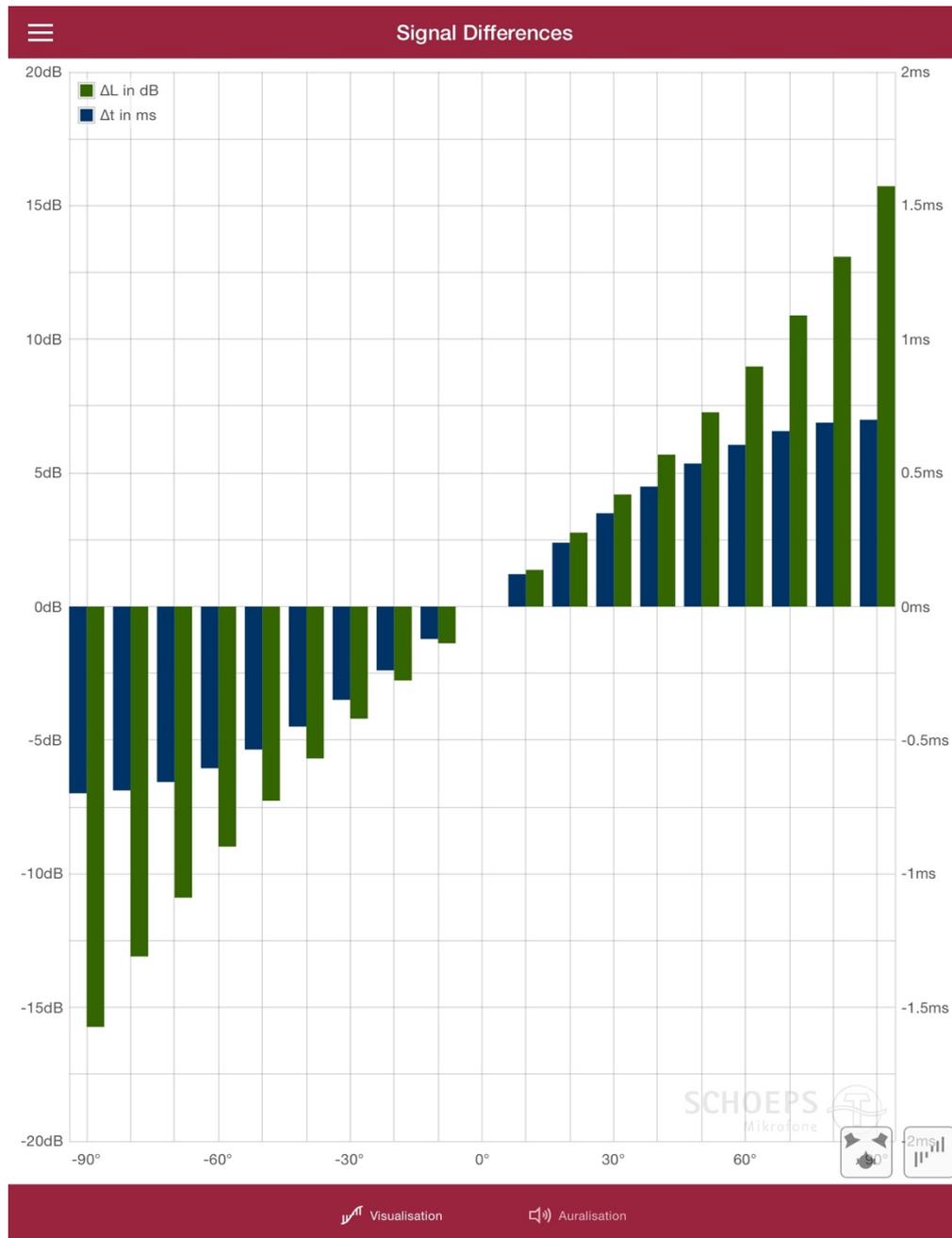


Abbildung 18 - Laufzeit- und Pegeldifferenzen für Nierenmikrofone mit 24 cm Kapselabstand und einem Öffnungswinkel von 90°<sup>216</sup>

<sup>216</sup> Bildquelle: Weiss; Wittek (2018)



Abbildung 19 - Kohärenzfunktion für Nierenmikrofone mit 24 cm Kapselabstand und einem Öffnungswinkel von 90°<sup>217</sup>

Da sich das räumliche Hören in vertikaler Ebene von dem auf horizontaler Ebene unterscheidet, müssen bei der Planung der Mikrofonaufnahmetechnik zur Erfassung der Hörsignale diese Unterschiede mit einbezogen werden. Besonderes Augenmerk ist daher auf das vertikale Kanalübersprechen zu legen. Mikrofone für Höhenkanäle sollen vornehmlich diffuse Atmo (Ambience) aufzeichnen, da Schallquellen deutlich oberhalb der Hörposition bei musikalischen Aufführungen in der Regel nicht vorkommen. Schallquellen in der unteren Ebene werden jedoch auch von den Mikrofonen der Höhenkanäle erfasst, wodurch es zu Kanalübersprechen kommt und die Signaltrennung leidet. Dies hat zur Folge, dass die

<sup>217</sup> Bildquelle: Ebenda

Lokalisation der Schallquellen nach oben driftet, es zu Klangverfärbungen kommt und die Schallquelle räumlich leicht verbreitert wird<sup>218</sup>. Dieser Effekt ist in Abbildung 20 angedeutet.

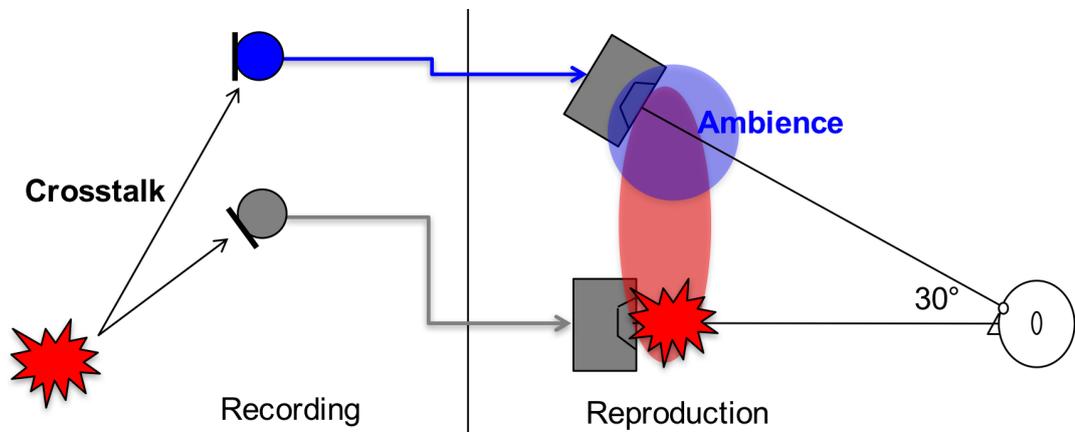


Abbildung 20 - Vertikales Signalübersprechen bei Aufnahme und Wiedergabe<sup>219</sup>

In ihren Versuchsreihen fanden Wallis und Lee heraus, dass eine Kanaltrennung von 6 bis 9 dB notwendig ist, um eine stabile Lokalisation einer Schallquelle in der unteren Ebene zu gewährleisten<sup>220</sup>. Um die Effekte des vertikalen Kanalübersprechens gänzlich unhörbar zu machen, ist jedoch eine Reduktion des Direktschallpegels um 10 dB im Mikrofon für die Höheninformation notwendig<sup>221</sup>. Dieses Prinzip ist in Abbildung 21 dargestellt.

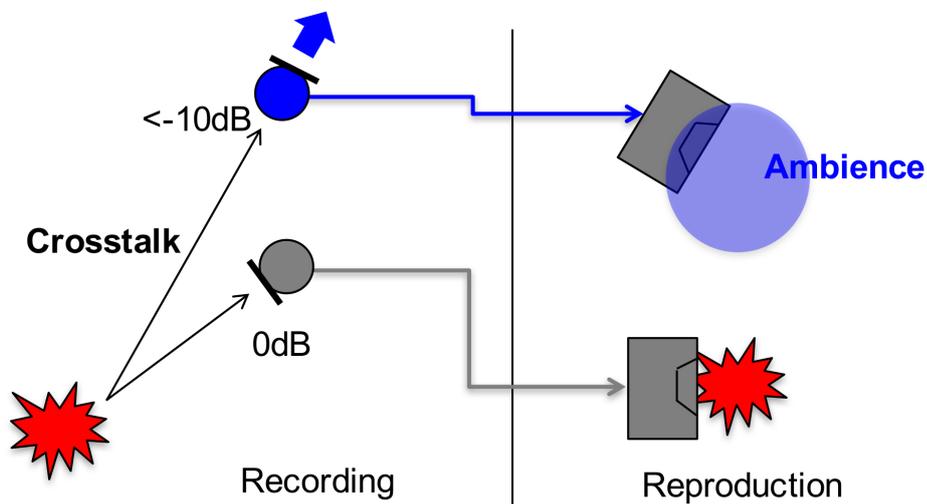


Abbildung 21 - Reduktion des vertikalen Signalübersprechens<sup>222</sup>

<sup>218</sup> vgl. Lee (2018)

<sup>219</sup> Bildquelle: Ebenda

<sup>220</sup> vgl. Wallis; Lee (2015), S. 775-776

<sup>221</sup> vgl. Lee (2018)

<sup>222</sup> Bildquelle: Ebenda

Lees Untersuchungen wurden mit einem vertikalen Lautsprecheraufstellwinkel von 30° vorgenommen, welcher auch in der horizontalen Ebene einer Standard-Lautsprecheraufstellung nach EBU R-3276 zu finden ist<sup>223</sup>. Für die Produktion von 360°-Video kommen hauptsächlich Ambisonics-Formate zum Einsatz, weshalb die Mikrofonsignale des 3D-Arrays nicht direkt auf Lautsprechern ausgegeben werden, sondern als einzelne Punkt-schallquelle in ein Ambisonics-Schallfeld encodiert werden. Die Ergebnisse Lees sind somit nicht direkt auf diesen Anwendungsfall zu übertragen, da die Präzision der räumlichen Abbildung in Ambisonics bei niedriger Ordnung geringer ist<sup>224</sup> und nicht direkt mit einem Lautsprecher-Setup vergleichbar. Dieser Umstand wird dadurch ein Stück weit kompensiert, dass die Signale des Mikrofonarrays äquidistant um die zentrale Hörposition in Ambisonics angeordnet werden, wodurch die Abbildungsbreite größer wird. Die Wiedergabewinkel betragen hier 90°. Eine solche Anordnung der Schallquellen ist in Abbildung 22 dargestellt.

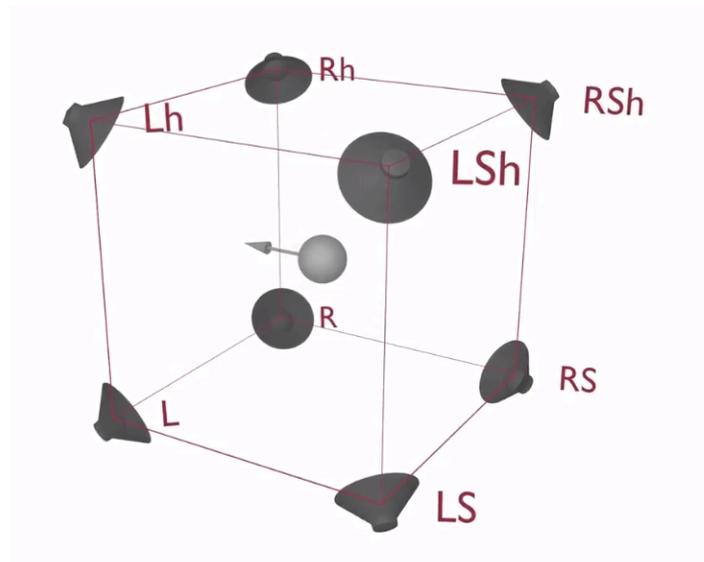


Abbildung 22 - Virtuelle Lautsprecher Anordnung<sup>225</sup>

Aus den oben genannten Gründen soll die Signaltrennung zwischen oberer und unterer Ebene des Mikrofonarrays mindestens 6 dB für möglichst viele Schalleinfallrichtungen betragen. Abbildung 23 zeigt, dass beim Einsatz von Mikrofonen mit Nierencharakteristik und einem Versatzwinkel von ca. 53° die gewünschte Signaltrennung von 6 dB für Schalleinfallswinkel ab ca. 40° erreicht wird. Beim Einsatz von Mikrofonen mit Supernierencharakteristik können vergleichbare Ergebnisse mit einem Versatzwinkel von 45° erreicht werden.

<sup>223</sup> vgl. EBU Technical Recommendation (1998), S. 9-10

<sup>224</sup> vgl. Audiokinetic Inc. (2018)

<sup>225</sup> Bildquelle: Dieterle u.a. (2018)

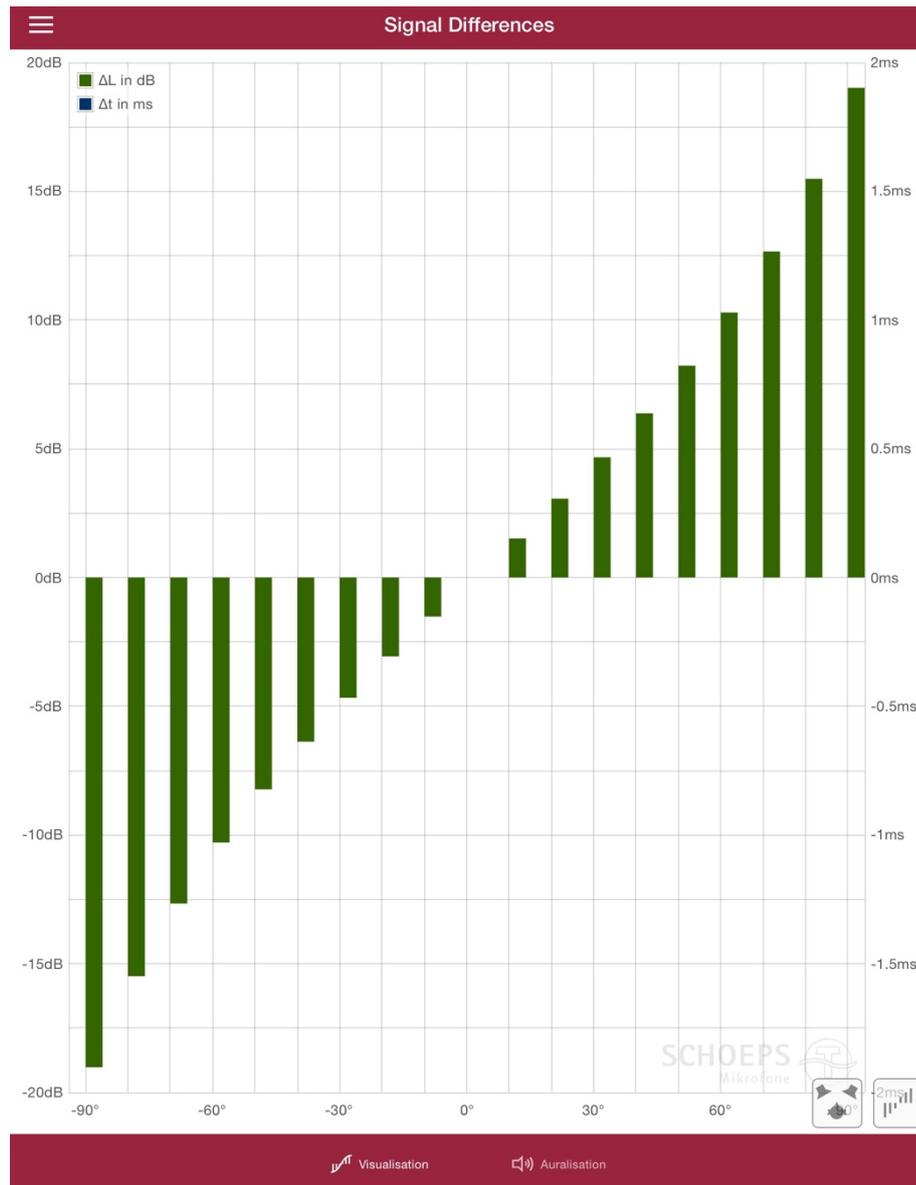


Abbildung 23 - Signaldekorrelation bei Nierenmikrofonen mit einem Versatzwinkel von 53°<sup>226</sup>

### 10.1.2 Praktische Umsetzung

Für den Aufbau des Mikrofonarrays kommen drei kurze Mikrofonbahnen zum Einsatz. Diese gehören zur Grundausstattung eines Tonstudios und sind in der Regel in ausreichender Anzahl im Rundfunk vorhanden. Sie besitzen eine Länge von knapp unter zwanzig Zentimetern und lassen sich direkt auf Stativgewinden befestigen. Eine der Bahnen dient als Träger für die anderen zwei und sie werden, wie in Abbildung 24 dargestellt, um 90° versetzt angebracht, wodurch sich optisch die Struktur eines „H“ ergibt. Optimalerweise wird der Abstand der oberen Bahnen so angepasst, dass sich gleiche Abstände zu allen Eckpunkten ergeben. Nun müssen die Bahnen an jeder Seite um eine weitere 3/8 Zoll Aufnahmeschraube ergänzt werden, um insgesamt acht Mikrofone befestigen zu können. Es ist darauf

<sup>226</sup> Bildquelle: Weiss; Wittek (2018)

zu achten, pro Seite jeweils eine Aufnahmeschraube mit Ausrichtung nach oben und eine nach unten anzubringen, um eine gleichmäßige Ausrichtung der Mikrofone zu gewährleisten.



Abbildung 24 - Grundaufbau der Mikrofonschienen<sup>227</sup>

Beim Ausrichten der Mikrofone ist zunächst auf einen symmetrischen Aufbau zu achten. Die Mikrofone, die für die untere horizontale Ebene zuständig sind, werden an den oberen äußeren Klammern befestigt und die Mikrofone für die Höhensignale an den unteren inneren Klammern. Die Abstände der Mikrofone zueinander sollten beispielsweise mithilfe eines Zollstocks kontrolliert werden. Das Anbringen von Markierungen an den Mikrofonen hilft bei der genauen Ausrichtung und verringert die Aufbauzeit. Darüber hinaus sollten sowohl die Versatzwinkel der koinzidenten Mikrofonpaare für die vertikale Ebene übereinstimmen, als auch die jeweilige Neigung der Paare zueinander, um eine gleichmäßige Abbildung zu allen Seiten zu ermöglichen. Hierbei kann es hilfreich sein, die Mikrofone für die Vertikalebene maximal nach oben auszurichten, um einen festen Bezugspunkt zum Einstellen des Versatzwinkels zu schaffen. Kommen Mikrofone mit Supernierencharakteristik zum Einsatz beträgt der Öffnungswinkel  $90^\circ$ , was sich noch recht gut mit dem bloßen Auge abschätzen lässt. Beim Einsatz von Mikrofonen mit Nierencharakteristik beträgt der Öffnungswinkel ca.  $106^\circ$ . Für die gleichmäßige

<sup>227</sup> Bildquelle: Eigene Abbildung

Ausrichtung der vier Mikrofonpaare in diesem Winkel ist der Einsatz eines einfachen Winkelmessers, z.B. in Form eines Zollstocks<sup>228</sup>, empfehlenswert. Dieses Vorgehen ist in Abbildung 25 dargestellt.

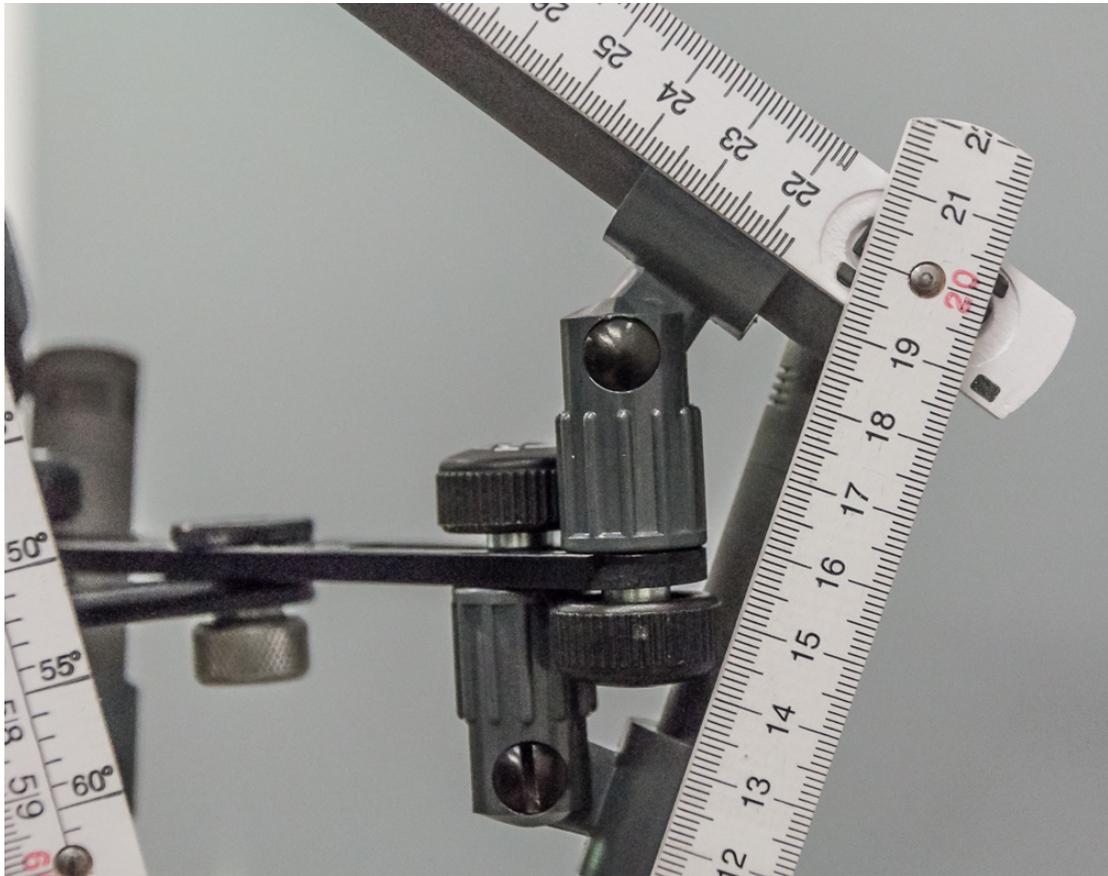


Abbildung 25 - Kontrolle des Öffnungswinkels mit einem Zollstock<sup>229</sup>

Der Grundaufbau des Mikrofonarrays aus kurzen Stereoschienen gewährleistet zwar einen einfachen und kostengünstigen Aufbau, ist aber äußerst beweglich, wodurch es schnell zu Fehlanpassungen kommen kann. Daher ist bei erstmaligem Aufbau mit der nötigen Akribie vorzugehen, um die oberen Anforderungen erfüllen zu können. In stressigen Live-Bedingungen, bei eventuell schlechten Lichtverhältnissen kann es sinnvoll sein, den Aufbau vorzubereiten. Mit Hilfe eines Schnelladapters oder einer kurzen Distanzstange kann das aufgebaute Mikrofonarray als Ganzes auf einem Bühnenstativ befestigt werden. Abbildung 26 zeigt den kompletten Aufbau des Mikrofonarrays.

<sup>228</sup> Besitzt der Zollstock keine Winkeleinteilung, so kann die Einteilung für den gewünschten Öffnungswinkel über Winkelfunktionen berechnet und angezeichnet werden

<sup>229</sup> Bildquelle: Eigene Abbildung



Abbildung 26 - Der komplette Aufbau des Mikrofonarrays<sup>230</sup>

## 11. Audiopostproduktion für 360°-Video

Dieses Kapitel behandelt die Grundlagen der Audiopostproduktion für 360°-Video. Neben dem grundsätzlichen Konzept wird eine Auswahl an Software-Tools vorgestellt, die im Rahmen der Testproduktionen zu dieser Arbeit eingesetzt wurden.

### 11.1 Produktionsformat

Für die Audioproduktion bei 360°-Video stellt Ambisonics derzeit das am besten geeignete Format dar. Ambisonics wird grundsätzlich von den meisten Plattformen für 360°-Video unterstützt und arbeitet unabhängig von einem festen Wiedergabe-Setup. Zusätzlich ist eine ständig wachsende Anzahl an Postproduktions-Tools für Ambisonics am Markt erhältlich. Diese Tools lassen sich in den Produktionsworkflow mit gängigen DAWs einbinden und ermöglichen so einen schnellen Einstieg in die Produktion von 360°-Medien. Das Ambisonics-Ambix-Format hat sich als Quasi-Standard in der Audiopostproduktion etabliert<sup>231</sup> und die gängigen DAWs und Plugins arbeiten nach diesen

<sup>230</sup> Bildquelle: Eigene Abbildung

<sup>231</sup> vgl. Audio Ease (2018)

Spezifikationen oder lassen sich darauf einstellen, weshalb es für die Postproduktion zu empfehlen ist. Grundsätzlich sollte dabei in Ambisonics höherer Ordnung produziert werden. Eine Mischung in HOA erfordert zwar mehr Rechenleistung, unterscheidet sich aber im Workflow nicht von einer Mischung in FOA. Durch HOA kann eine höhere Qualität erreicht werden und HOA eignet sich auch als Speicherformat, da eine HOA-Mischung für vielfältige Wiedergabe-Setups dekodiert werden kann und keine neue Mischung erfordert. Abwärtskompatibilität von Ambisonics ist gegeben, da nicht unterstützte Komponenten einfach weggelassen werden können<sup>232</sup>.

## 11.2 DAW-Anforderungen

Soll Audio für 360°-Video im Ambisonics-Format produziert werden, so muss die DAW in der Lage sein, Audiobusse in der benötigten Größe zu Verwalten. Pro Ambisonics-Komponente wird ein Übertragungskanal benötigt. Für die Bearbeitung von Audiomaterial in Ambisonics dritter Ordnung werden somit nach Formel (3) Audiobusse mit 16 Kanälen benötigt. Diese Möglichkeit bieten aktuell die DAWs Avid Pro Tools Ultimate, Steinberg Nuendo und Cockos Reaper. Pro Tools Ultimate<sup>233</sup> und Nuendo haben dazu spezielle Ambisonics-Tracks bis zur dritten Ordnung integriert<sup>234</sup> und Reaper ermöglicht die freie Verwaltung von Audiobussen mit bis zu 64 Spuren<sup>235</sup>.

In diesem Zusammenhang ist die interne Verwaltung von Audiostreams in Pro Tools zu beachten. Soll ein Monosignal in eine Ambisonics-Mischung dritter Ordnung integriert werden, so wird hierzu ein spezieller Ambisonics-Panner benötigt. Dieser erweitert den Audiobus des Monosignals zu einem Ambisonics-Bus mit entsprechender Spuranzahl. Wird in Ambisonics dritter Ordnung gemischt, wird somit für jedes Signal ein 16-Spur Audiobus erstellt. Im Falle einer umfangreichen Mischung mit vielen Mono-Signalen ist die Anzahl maximal verfügbarer Busse in Pro Tools schnell erreicht. Daher muss eine größere Mischung ggf. in mehrere Sub-Mischungen aufgeteilt werden. Audiobusse in Pro Tools werden Voices genannt. Ein natives Pro Tools Ultimate System bietet 256 Voices und ein HDX-System 256 Voices pro DSP-Karte<sup>236</sup>.

## 11.3 Mischungs-Konzept für 360°-Video

360°-Video ist ein lineares Medium. Das grundsätzliche Konzept der Mischung ist daher dem einer Film- oder Videovertonung ähnlich. Bei der Mischung kann eine Rektangular-Projektion des sphärischen 360°-Videos genutzt werden, um die gewohnten Videospuren und Vorschau-monitore zu verwenden. Dadurch kann wie üblich zum Bild gemischt werden.

---

<sup>232</sup> vgl. Nicol (2018), S. 185-189

<sup>233</sup> Pro Tools HD wurde mit dem Update auf Version 2018.4 in Pro Tools Ultimate umbenannt

<sup>234</sup> vgl. Steinberg Media Technologies GmbH (2018), S. 728-730; Avid Technology (2018a), S. 1257 ff.

<sup>235</sup> vgl. Cockos (2018)

<sup>236</sup> Sound on Sound Magazine (2018)

Gemischt wird jedoch nicht in einen kanalbasierten Bus, sondern in einen Ambisonics-Bus, der je nach Ordnung aus unterschiedlich vielen Spuren besteht. Dieser Ambisonics-Bus enthält die Mischung und stellt somit das Zielformat dar. Dieser Bus wird als Multichannel-Audiodatei exportiert und in die Videodatei integriert. Zusätzlich müssen Metadaten in die Datei eingefügt werden, sodass die Video Plattform das Bild- und Tonmaterial richtig interpretieren kann. Die fertig konfektionierte Videodatei kann über einen lokalen 360°-Videoplayer kontrolliert und anschließend auf der gewünschten Plattform veröffentlicht werden.

Um das Audiomaterial während der Mischung über die DAW abhören zu können, muss ein Spatializer als letztes Glied in den Ambisonics-Bus eingeschliffen werden. Der Spatializer ist für die Binauralsynthese des Ambisonics-Signals verantwortlich. Abbildung 27 zeigt schematisch das Monitoring mittels Spatializer. Anhand von Orientierungsdaten eines Head-Trackers oder sonstiger Eingabegeräte, wird das Ambisonics-Schallfeld rotiert. Es folgt die Dekodierung für das virtuelle Lautsprechersetup. Die Signale der virtuellen Lautsprecher werden mit den entsprechenden HRTFs für Richtung und Ohr gefaltet und summiert. Am Ausgang steht anschließend ein binaurales Stereosignal für die Kopfhörerwiedergabe zur Verfügung. Wird innerhalb der DAW nicht mit einem separaten Abhörbus gearbeitet, so muss der Spatializer vor der Ausspielung entfernt werden. Steht in der Mischregie ein 3D-Lautsprechersystem zur Verfügung, kann statt des Spatializers auch ein Ambisonics-Decoder eingesetzt werden, der die benötigten Lautsprechersignale aus dem Ambisonics-Schallfeld berechnet.

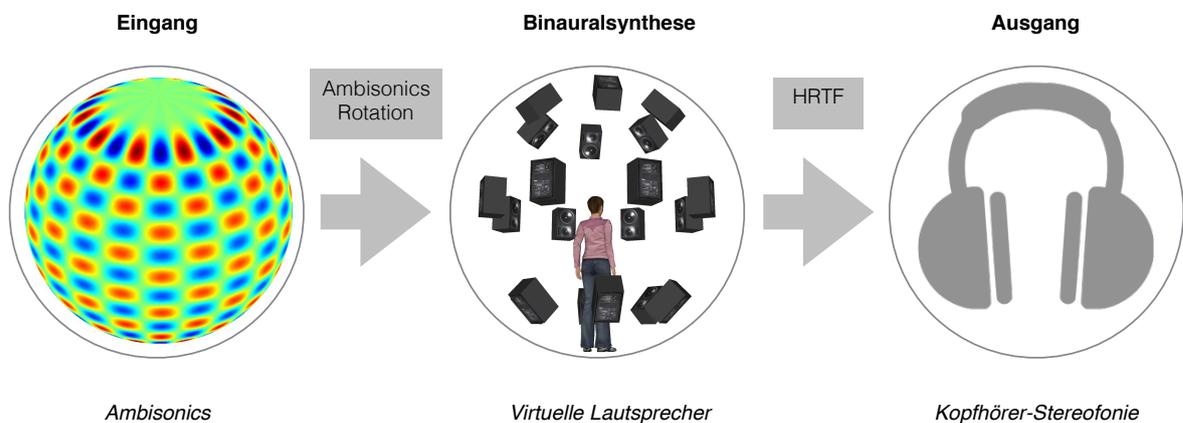


Abbildung 27 – Ambisonics-Monitoring via Spatializer<sup>237</sup>

Für die Klangbearbeitung müssen keine besonderen technischen Umstände beachtet werden und Einzelsignale können wie gewohnt mit den Mitteln der DAW bearbeitet werden. Eine wichtige Rolle in der Audiopostproduktion für 360°-Video spielt die Verteilung der Signale im dreidimensionalen Schallfeld. Mit Hilfe spezieller Ambisonics-Panner werden die Signale in Abstimmung mit dem Videobild im Ambisonics-Schallfeld platziert und encodiert. Mono-Signale, wie beispielsweise von

<sup>237</sup> Bildquelle: Eigene Abbildung

einem Stützmikrofon, stellen Punktschallquellen im Ambisonics-Schallfeld dar und werden über die Einstellung von Elevations- und Azimuthwinkel im Ambisonics-Panner platziert. Einige Ambisonics-Panner bieten die Möglichkeit, Schallquellen anhand einer Rektangular-Projektion, die sich über das Videobild legt, auszurichten. Mehrkanalige Signale, beispielweise von einem Mikrofonarray, besitzen eine räumliche Ausdehnung und sind prinzipiell zu Wiedergabe über eine Lautsprecheranordnung gedacht. Daher wird mit den Signalen eines stereofonen Arrays eine Lautsprecherwiedergabe in Ambisonics simuliert. Jedes Mikrofonsignal repräsentiert damit eine Punktschallquelle aus einer bestimmten Richtung. Für ein achtkanaliges stereofones 3D-Array, wie in Kapitel 10 vorgestellt, bietet sich eine würfelförmige Anordnung mit  $\pm 45^\circ$  Elevationswinkel für Azimuthwinkel von  $\pm 45^\circ$  und  $\pm 135^\circ$  an. Mit dem gleichen Verfahren können beispielweise auch Surroundmischungen in 5.1 oder 7.1 in Ambisonics abgebildet werden.

Wird eine Kombination aus 3D-Hauptmikrofon und Stützmikrofonen eingesetzt, so muss beim Panning auf eine exakte Abstimmung geachtet werden, damit es nicht zur Mehrfachabbildung einer Schallquelle kommt. Es bietet sich an, die Mischung mit dem Hauptmikrofon zu beginnen. Wurden bei der Aufnahme Kamera-Achse und Hauptmikrofonachse nicht aufeinander angepasst, muss diese Anpassung als erster Schritt in der Nachbearbeitung erfolgen. Dazu werden die Signale des Hauptmikrofons nach obigem Verfahren zunächst in ein Ambisonics-Schallfeld überführt. Mit Hilfe eines Ambisonics-Rotators kann dieses Schallfeld nun so gedreht werden, dass akustische und optische Perspektive zueinander passen. Anschließend werden die Stützen zugemischt und ihre Position anhand von Hauptmikrofon und Videobild angepasst.

Wird ein Ambisonics-A-Format-Mikrofon eingesetzt, so muss dieses zunächst ins Ambisonics-B-Format konvertiert werden (vgl. Kapitel 9.1.1). Dazu wird ein Ambisonics-A/B-Konverter Plugin in die Spur eingeschliffen. Neben der Konversion kann auch die Ausrichtung des Mikrofons angepasst werden, wenn eine frontale Ausrichtung bei der Aufnahme nicht möglich war.

## 11.4 Softwaretools

Im folgenden Abschnitt soll ein kurzer Überblick von Audiopostproduktions-Tools gegeben und ihr Funktionsumfang vorgestellt werden. Aufgrund der Fülle von am Markt erhältlichen Tools ist diese Auflistung beispielhaft und umfasst nur Software, die im Rahmen dieser Arbeit getestet wurde. Weitere interessante Tools für zukünftige Tests werden am Ende aufgeführt.

### 11.4.1 Sennheiser AMBEO A-B Format Converter

Der *Sennheiser AMBEO A-B Format Converter* ist ein frei erhältliches Software-Plugin um Aufnahmen eines A-Format-Mikrofons in das B-Format zu überführen. Das Tool ist für die Verwendung mit dem *Sennheiser AMBEO VR Mikrofon* entwickelt worden, kann aber auch mit anderen A-Format-Mikrofonen genutzt werden, wenn Anzahl und Ausrichtung der Kapseln übereinstimmen.

Wie Abbildung 28 zu entnehmen ist, ist zusätzlich ein Korrektur- und ein Hochpassfilter integriert. Ferner lässt sich das Ambisonics-Schallfeld rotieren und über Schaltflächen kann die verwendete Mikrofonposition eingestellt werden. Das Ausgabeformat lässt sich zwischen Fuma und Ambix umschalten und es ist eine einfache Pegelanzeige integriert<sup>238</sup>. Aktuell lässt sich das Plugin in Pro Tools nur in Quad-Spuren einschleifen, jedoch nicht in 1<sup>st</sup> Order Ambisonics-Spuren.

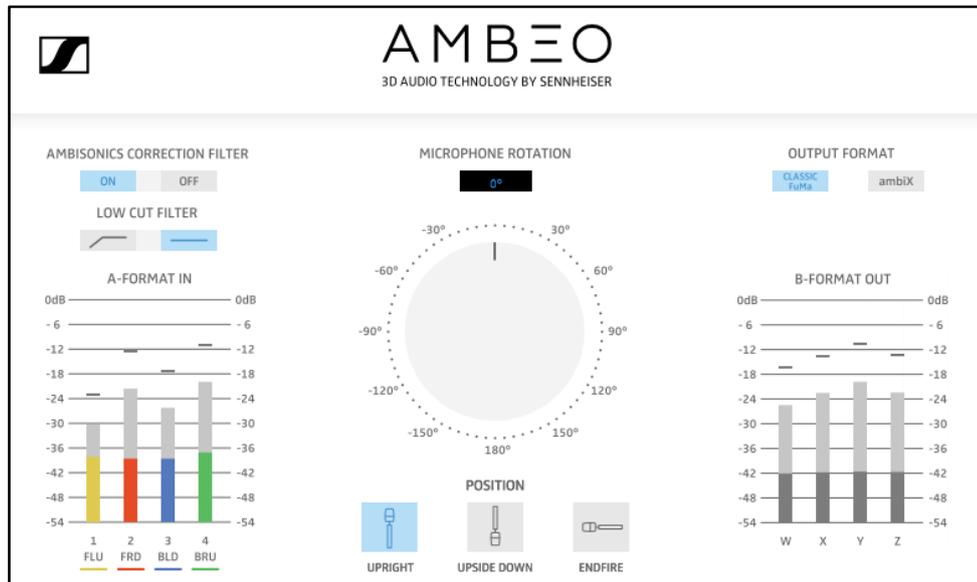


Abbildung 28 - Sennheiser AMBEO A - B Format Converter<sup>239</sup>

#### 11.4.2 Audio Ease 360 Pan Suite

Die *Audio Ease Pan Suite* besteht aus mehreren Plugins und unterstützt Ambisonics bis zur dritten Ordnung. Folgende Tools sind enthalten: *360 Pan*, *360 Monitor*, *360 Reverb*, *360 Radar*, *360 Turner* und *360 Converter*<sup>240</sup>. Abbildungen sämtlicher Plugins sind im Anhang verfügbar.

*360 Pan* ist ein Ambisonics-Panner mit dem Audioquellen in einem Ambisonics-Schallfeld platziert werden. Dazu werden die Schallquellen als farbiger Kreis auf einer Rektangular-Projektion dargestellt, wo sie grafisch platziert werden können. Die farbigen Kreise der Schallquellen können zusätzlich als Überlagerung auf dem Videobild innerhalb der DAW dargestellt und ausgerichtet werden, was vor allem bei Einzel- und Stützmikrofonen ein korrektes Panning erleichtert. Diese Funktionalität zeigt Abbildung 29.

<sup>238</sup> vgl. Sennheiser Electronic GmbH und Co. KG (2018b)

<sup>239</sup> Bildquelle: Ebenda

<sup>240</sup> vgl. Audio Ease (2017)

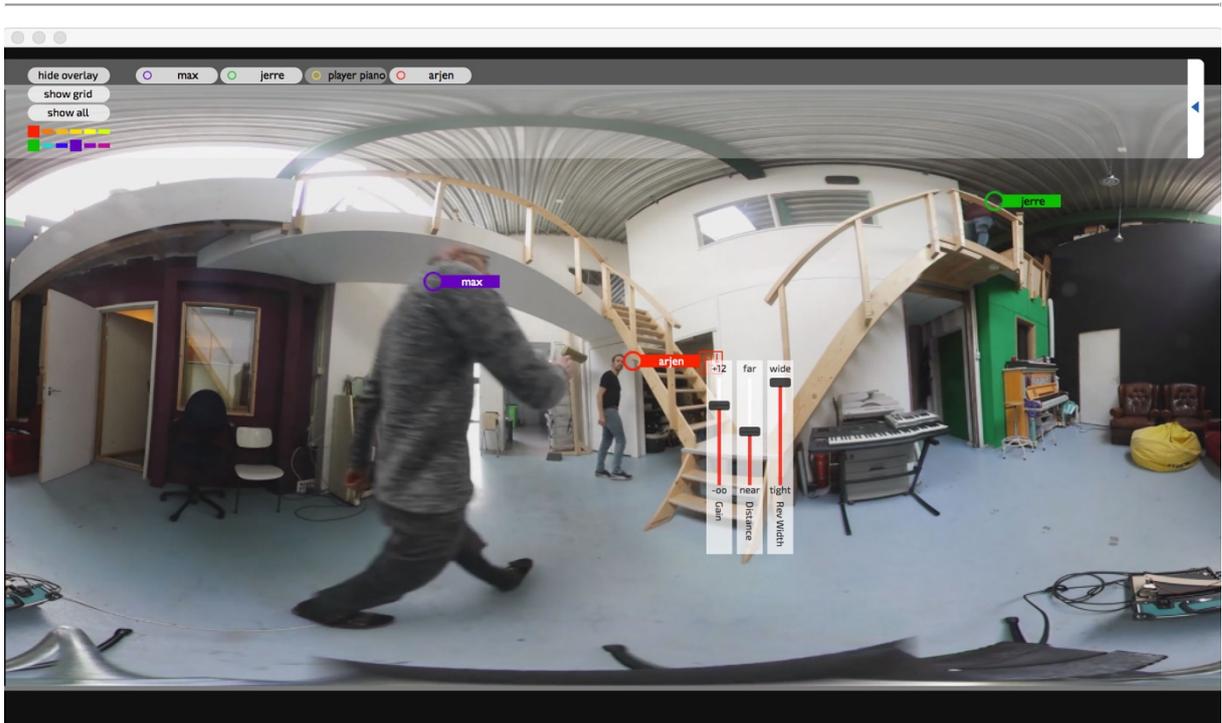


Abbildung 29 - Darstellung von Schallquellen als farbige Kreise<sup>241</sup>

*360 Monitor* ist ein Spatializer und wird zum Abhören der Mischung in den Master-Bus eingeschliffen. Er verarbeitet Orientierungsdaten, die wahlweise von einem Bluetooth Head-Tracker oder von Mausbewegungen generiert werden.

Bei *360 Reverb* handelt es sich um einen FOA-Faltungshall, dessen Raumimpulsantworten mit einem FOA-Mikrofon aufgenommen wurden. Eigene Impulsantworten lassen sich nicht einbinden, doch eine große Auswahl verschiedenster Impulsantworten ist bereits enthalten. Neben einer einfachen Klangregelung lassen sich Nachhallzeit, Streuung und Mischungsverhältnis anpassen. Eine Besonderheit ist, dass der Hall direkt aus jeder Instanz des *360 Pan* angesprochen werden kann, ohne dass ein Send-Return-Weg eingerichtet werden muss.

Mit *360 Radar* werden Schallquellen anhand ihrer Intensität farblich umrandet und im Videofenster dargestellt. Darüber lässt sich die Position einer Schallquelle visualisieren und erleichtert beispielsweise die nachträgliche Ausrichtung eines FOA-Mikrofons. Die Ausrichtung kann mit Hilfe des *360 Turner* geschehen. Der *360 Converter* wandelt zwischen den Ambisonics-Formaten Ambix und Fuma.

<sup>241</sup> Bildquelle: Audio Ease (2018)

### 11.4.3 Noisemakers Ambi HD Bundle

Das *Noisemakers Ambi HD Bundle* ist ein Paket von Audiopostproduktionstools für Ambisonics und enthält die Plugins *Pan*, *Head*, *Verb*, *Limiter* und *Converter*<sup>242</sup>. Es unterstützt Ambisonics bis zur dritten Ordnung. Abbildungen sämtlicher Plugins sind im Anhang verfügbar.

*Ambi Pan* ist ein Ambisonics-Panner, mit dem Audiosignale in einem Ambisonics-Schallfeld platziert werden können. Dies geschieht über Einstellräder für Distanz, Elevations- und Azimuthwinkel. Verschiedene Presets zur Platzierung kanalbasierter Mischungen sind vorhanden. Das Panning der Schallquelle kann wahlweise auf einer Rektangular-Projektion oder in Form einer sphärischen Aufsicht visualisiert werden; dies zeigt Abbildung 30. Zusätzlich kann ein transparentes Fenster über das Videobild der DAW gelegt werden, auf dem die Schallquellen als farbige Kreise mit Beschriftung darstellbar sind.



Abbildung 30 - Sphärische Aufsicht im Ambisonics-Panner "Ambi Pan" von Noisemakers<sup>243</sup>

*Ambi Head* ist ein Spatializer Plugin zum Abhören einer Ambisonics-Mischung über Kopfhörer. Die Orientierung kann über Mauseingaben gesteuert werden. Die HRFTs, die *AMBI Head* bei der Binauralsynthese einsetzt, lassen sich austauschen. So können beispielsweise die HRTFs der Zielplattform eingebunden werden, was eine verlässlichere Endkontrolle der Mischung ermöglicht.

<sup>242</sup> vgl. Noisemakers (2018b)

<sup>243</sup> Bildquelle: Noisemakers (2018a)

Ebenfalls besteht die Möglichkeit für jeden Nutzer individuelle HRTFs oder für sich gut funktionierende HRTFs einzubinden.

Mit *Ambi Verb* steht ein Ambisonics-Faltungshall zur Verfügung. Es werden dreißig Ambisonics-Impulsantworten mitgeliefert. Eigene Impulsantworten in Mono oder FOA-Ambisonics lassen sich nach der Konversion durch den Hersteller einbinden.

*Ambi Limiter* ist ein Ambisonics-Peak-Limiter mit einer Release Zeit von bis zu 1000 ms und Look-ahead Funktionalität mit bis zu 40 ms. Mit dem *Ambi Converter* kann zwischen den Ambisonics-Formaten Ambix und Fuma gewandelt werden.

#### 11.4.4 Facebook 360 Workstation

Die *Facebook 360 Workstation* ist frei verfügbar und die Plugins unterstützen das AAX und VST Plugin-Format. Sie ist in *Pro Tools HD* ab Version 12.8.2 bereits integriert<sup>244</sup>. Die *Facebook 360 Workstation* besteht aus vier Plugins – *FB360 Spatializer*, *FB360 Control*, *FB360 Converter* und *FB 360 Meter*. Zusätzlich sind die beiden standalone Apps *VR Video Player* und *360 Encoder* enthalten. Es wird Ambisonics bis zur zweiten Ordnung unterstützt<sup>245</sup>. Im Anhang befinden sich Abbildungen der Tools.

Beim *FB360 Spatializer* handelt es sich um einen Ambisonics-Panner. Die Platzierung der Schallquellen kann über Winkeleinstellung oder die Ausrichtung auf einer Rektangular-Projektion des Videos erfolgen. Ferner ist es möglich, einen Doppler-Effekt für sich bewegende Schallquellen zu aktivieren, mit einer Raumsimulation künstliche Reflexionen zu erzeugen und mit einem Distance-Effekt Tiefpassfilter für Schallquellen zu generieren. Für Schallquellen außerhalb des Sichtfeldes kann eine zusätzliche Dämpfung eingestellt werden, die jedoch nicht Bestandteil der Ambisonics-Mischung ist, sondern lediglich Metadaten geniert, die bei der Verbreitung über *Facebook* ausgewertet werden können.

Das *FB360 Control* Plugin ist für die Binauralsynthese zuständig und wird in den Ambisonics-Master-Bus eingeschliffen. Die Orientierung kann wahlweise über Winkeleingaben erfolgen oder vom gekoppelten VR Videoplayer empfangen werden. Weiterhin sind die globalen Einstellungen der Raumsimulation im *FB360 Control* Plugin integriert.

Des Weiteren kann mit Hilfe des *FB360 Converters* ein Ambisonics-Schallfeld rotiert werden, um so die Perspektive von Audio und Video anzugleichen.

---

<sup>244</sup> vgl. Facebook (2017)

<sup>245</sup> vgl. Facebook (2018)

Das *FB360 Meter* Plugin dient zur Lautheits-Analyse der binauralisierten Mischung und wird im Master-Bus hinter dem *FB360 Control* Plugin eingeschliffen. Die Lautheitsanalyse geschieht nach den Empfehlungen der EBU-R128. Die Messung erfolgt unabhängig von der Orientierung immer für die lauteste Richtung.

Der *VR Videoplayer* kann als standalone Wiedergabesoftware für 360°-Videos mit Ambisonics-Audio genutzt werden. Es ist ebenfalls möglich den *VR Videoplayer* mit dem *FB360 Control* Plugin innerhalb einer DAW zu koppeln, um so eine 360°-Videovorschau zu ermöglichen und Orientierungsdaten über Maussteuerung zu genießen.

Der *FB360 Encoder* dient zur Integration einer Ambisonics-Mischung in ein 360°-Video. Eine Ambisonics-Mischung kann als Mehrspur-Audiodatei aus der DAW exportiert werden und mit Hilfe des *FB360 Encoder* mit einer Videodatei verknüpft werden. Es werden ebenfalls die zur Wiedergabe notwendigen Metadaten erzeugt. Es sind Ausgabeformate für verschiedene Wiedergabeplattformen verfügbar.

#### 11.4.5 Weitere Tools

Die frei verfügbare Plugin-Library *O3A* von *Blue Ripple Sound* unterstützt Ambisonics bis zur dritten Ordnung und ist im VST- und AAX-Plugin-Format erhältlich. Es sind verschiedene Ambisonics-Panner und -Decoder sowie Werkzeuge für Schallfeldrotation, Metering und Visualisierung enthalten. Ein Werkzeug zur Binauralsynthese fehlt<sup>246</sup>.

Die *IEM Plugin Suite* wurde vom Institut für Elektronische Musik und Akustik der Kunstuniversität Graz entwickelt. Sie ist frei verfügbar und unterstützt Ambisonics bis zur siebten Ordnung. Die Plugins sind nur im VST-Plugin-Format erhältlich. Neben einem Plugin zur Binauralsynthese, sind vielfältige Werkzeuge zur Ambisonics-Schallfeldmanipulation enthalten. Lediglich ein Plugin für videobasiertes Panning ist nicht enthalten.<sup>247</sup>

Die *Ambix Plugin-Suite* wurde von Matthias Kronlachner entwickelt. Sie ist nur im VST-Plugin-Format erhältlich. Es wird Ambisonics bis zur siebten Ordnung unterstützt. Verschiedene Manipulationen des Ambisonics-Schallfeldes sind möglich und Werkzeuge für die Binauralsynthese und das Panning sind enthalten. Videobasiertes Panning wird jedoch nicht unterstützt. Die *Ambix Plugin-Suite* ist ebenfalls frei erhältlich.<sup>248</sup>

---

<sup>246</sup> vgl. Blue Ripple Sound (2018)

<sup>247</sup> vgl. Rudrich (2018)

<sup>248</sup> vgl. Kronlachner (2014)

## 12. Testproduktionen

In diesem Kapitel werden Testproduktionen beschrieben, die im Rahmen dieser Arbeit durchgeführt wurden, um in verschiedenen Szenarien Mikrofonierungen und Methoden der Audiopostproduktion für 360°-Video zu erproben. Audio- und Videobeispiele zu den Produktionen sind auf dem Datenträger im Anhang zu finden.

### 12.1 Teil 1 – Studio-Session Akustik Duo

Dieser Test wurde im Rahmen einer Musikproduktion im Tonstudio U54 an der Hochschule der Medien Stuttgart durchgeführt. Mit zwei Musikerinnen wurde ein Song im Live-Recording-Verfahren produziert und die Aufnahme wurde mit einer 360°-Kamera begleitet.

#### 12.1.1 Aufnahme

Die Besetzung bestand aus einer Sängerin und einer Gitarristin, die auch die zweite Gesangsstimme beisteuerte. Es wurde eine Cover-Version des Songs „Home“ von Philip Phillips aufgenommen. Da bei dieser Produktion die Energie des Zusammenspiels im Vordergrund stand, wurde der Song live eingespielt. Weil die Performance der Musikerinnen und die Kommunikation oberste Priorität hatten, wurden beide in einem Raum aufgenommen und das dadurch entstandene Signalübersprechen in Kauf genommen. Zielsetzung war ein möglichst authentischer, natürlicher Klang ohne aufwendige Nachbearbeitung.

Die Musikerinnen wurden so aufgestellt, dass sie sich gegenüberstanden und mit der 360°-Kamera ein gleichseitiges Dreieck bildeten. Die Kamera wurde in Augenhöhe zwischen den Musikerinnen und der Regie-Scheibe aufgebaut, um diese mit in die Videoaufnahme zu integrieren. Den schematischen Aufbau zeigt Abbildung 31.

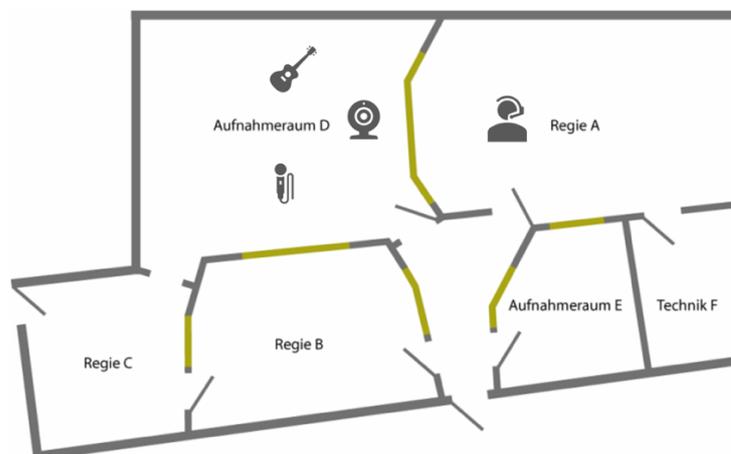


Abbildung 31 - Produktionsaufbau "Akustik Duo" im Studio U54 der HdM<sup>249</sup>

<sup>249</sup> Bildquelle: Mayr (2008)

Zur Aufnahme der ersten Gesangstimme wurde ein *Schoeps V4* eingesetzt, für die zweite Stimme ein *AKG C414* mit Nierencharakteristik. Die Gitarre wurde mit zwei *Schoeps CMC6/MK2* mikrofoniert. Diese Mikrofonsignale dienten zunächst zur Mischung der Stereo-Version. Zur Produktion des 360°-Videos wurde zusätzlich ein FOA-Mikrofon und ein 3D-Mikrofonarray nach dem Verfahren aus Kapitel 10 aufgebaut. Für das 3D-Array kamen *Schoeps CMC6/MK4* mit Nierencharakteristik zum Einsatz. Sowohl das FOA-Mikrofon, als auch das 3D-Array wurden direkt unterhalb der Kamera angebracht. Die Höhe betrug ca. 1,8 m und der Abstand zu den Musikerinnen etwa 2 m.

Es wurden mehrere komplette Takes des Stückes aufgenommen und für den finalen Take wurden Teile mehrerer Takes zusammengeschnitten. Da das Stück mit Metronom-Klick eingespielt wurde, konnten Schnitte im Audiomaterial durchgeführt werden ohne die Synchronität zum Videobild zu verlieren.

### 12.1.2 Nachbearbeitung

In der Nachbearbeitung wurde zunächst die Stereo-Mischung angefertigt, wobei das 3D-Array und das FOA-Mikrofon nicht verwendet wurde. Neben Klang- und Dynamik-Bearbeitung wurde auch die Intonation der zweiten Stimme leicht korrigiert, was aufgrund des Übersprechens jedoch nur in geringem Maße erfolgen konnten. Des Weiteren wurde künstlicher Nachhall hinzugefügt.

Die Ambisonics-Mischung für das 360°-Video wurde in dritter Ordnung angelegt und es kam das *Noisemakers Ambi HD Bundle* zum Einsatz. Begonnen wurde mit dem Vergleich des FOA-Mikrofons und des 3D-Arrays. Die Mikrofone sollten die Aufgabe eines Hauptmikrofons übernehmen und sowohl die musikalische Darbietung als auch einen realistischen Raumeindruck abbilden. Die Signale des A-Format-Mikrofons wurden mit dem *Sennheiser AMBEO A-B Converter* in das benötigte B-Format konvertiert. Die Mikrofonsignale des 3D-Arrays wurden mit dem *Ambi Pan* Plugin in Würfelform platziert und nach Ambisonics codiert. Beim Vergleich zeigte sich, dass die Abbildung des FOA-Mikrofons etwas direkter wirkte, aber die räumliche Abbildung, vor allem die der Höhen-Ebene, kein überzeugendes Raumgefühl lieferte. Von der Klangfarbe her wirkte das FOA-Mikrofon etwas dünner und es wies darüber hinaus einen stärkeren Rauschanteil auf. Das 3D-Array überzeugte mit einer guten Abbildung der Räumlichkeit und im Direktvergleich öffnete sich das Klangbild merklich und die Höhen-Ebene war deutlich wahrnehmbar. Die Abbildung des 3D-Arrays war insgesamt jedoch diffuser.

Nach diesem Vergleich wurde für die weitere Mischung das 3D-Array als Grundlage ausgewählt. Da die Ausrichtung der Kamera nicht exakt mit der Ausrichtung des 3D-Arrays abgestimmt wurde, passte die akustische Perspektive des 3D-Arrays nicht zur optischen Perspektive, weshalb das Schallfeld mit den Signalen des 3D-Arrays gedreht werden musste. Dabei kam der *Audio Ease 360 Turner* zum Einsatz. Um eine größere Nähe zu den Schallquellen und eine präzisere Richtungsabbildung zu erzielen,

wurden die Einzelsignale hinzugemischt. Dazu wurde die Signalbearbeitung aus der Stereo-Mischung übernommen. Mit Hilfe des Ambisonics-Panner wurden die Einzelsignale so platziert, dass sie mit der Abbildung im Hauptmikrofon übereinstimmten. Zusätzlich wurden die Einzelsignale auf das Hauptmikrofon verzögert, was der räumlichen Abbildung dienlich war. Ferner konnten durch die Einzelsignale Lautstärkeunterschiede zwischen den Stimmen und der Gitarre ausgeglichen werden und ermöglichten so erst eine ausgewogene Balance der Mischung. Die Klangbearbeitung der Einzelsignale durfte nicht zu intensiv ausfallen, um in Kombination mit dem Hauptmikrofon zu funktionieren. Vor allem bei der Korrektur der Intonation galt es einen Kompromiss zu finden.

Aufgrund der trockenen Akustik im Aufnahmerraum, wurde zusätzlich Ambisonics-Faltungshall zugemischt, um den räumlichen Eindruck zu verstärken. Es zeigte sich schnell, dass die Mischung für das 360°-Video wesentlich trockener ausfallen musste, als die Stereo-Mischung, um einen zum Videobild glaubwürdigen Ton zu gestalten. Aus diesem Grund kam die Impulsantwort eines Studio-Raumes mit recht kurzer Nachhallzeit zum Einsatz.

### **12.1.3 Beobachtungen**

In dieser ersten Testproduktion konnte das 3D-Mikrofonarray mit einer offenen, dreidimensionalen Abbildung des Raumes überzeugen. Aufgrund hochwertiger Schallwandler war auch die Klangfarbe ausgewogen. In beiden Punkten lieferte das FOA-Mikrofon subjektiv schlechtere Ergebnisse.

Trotz des geringen Mikrofonabstandes des Hauptmikrofons von etwa 2 Metern wirkten die Schallquellen im Zusammenhang mit dem 360°-Video zu entfernt. Dies galt auch für das FOA-Mikrofon, wenn auch weniger stark ausgeprägt. Durch das Zumischen der Einzelsignale konnte die Präsenz der Schallquellen gesteigert werden. Dies verstärkt auch die Änderung des Klangbildes bei Perspektivwechseln im 360°-Video, wodurch die Immersion gesteigert werden kann. Einzelsignalen kommt somit eine große Bedeutung innerhalb der Produktion zu. Allerdings dürfen diese nicht zu stark bearbeitet werden, sodass sie gemeinsam mit einem Hauptmikrofon verwendet werden können.

Weiterhin fiel auf, dass das nachträgliche Ausrichten des Schallfeldes in Abstimmung zum Bildmaterial ein kritischer und aufwendiger Prozess sein kann. Daher empfiehlt es sich, bereits bei der Aufnahme auf die Abstimmung der Ausrichtung zu achten oder zumindest die Versatzwinkel von Mikrofon und Kamera-Achse zu erfassen, um dadurch einen Ausgangspunkt in der Audiopostproduktion vorliegen zu haben.

## **12.2 Teil 2 – Livekonzert hr Sinfonieorchester**

Eine zweite Testproduktion wurde mit dem hr Sinfonieorchester durchgeführt. Es wurde das Eröffnungskonzert der Darmstädter Ferienkurse für Neue Musik am 14.07.2018 aufgezeichnet. Das Konzert fand in der großen Sporthalle der Lichtenbergschule Darmstadt statt. Der Hallenboden wurde zum Schutz großflächig mit Gummimatten ausgelegt und die Halle verfügte über eine für musikalische

Aufführungen relativ kurze Nachhallzeit. Es war eine P.A.-Anlage im Einsatz die unter anderem der Zuspiegelung von Sampler-Klängen diente. Abbildung 32 zeigt den Aufbau der Halle. Das Orchester spielte in amerikanischer Sitzordnung. Es wurden die Stücke „VAPE“ von Bára Gísladóttir, „dropped.drowned“ von Sarah Nemtsov und „Piano Concerto“ von Simon Stehen-Anderson gespielt. Die musikalische Leitung hatte Baldur Brönnimann. Im Rahmen dieser Arbeit wurde ein 360°-Video von der Aufführung des Stückes „VAPE“ produziert. Bára Gísladóttir beschreibt ihr Stück folgendermaßen:

*„Anregung für VAPE war der Sarin-Anschlag, den die Sekte Ömu Shinrikyö 1995 in der Tokioter U-Bahn verübte. Fünf Terroristen betraten jeweils einen Zug, ausgerüstet mit in Kunststoffbeuteln abgefülltem Sarin. Sarin ist eine durchsichtige, schnell verdampfende Flüssigkeit. Die Männer ließen die Beutel fallen und stachen mit den Spitzen ihrer Regenschirme auf sie ein, wodurch die Substanz in die Umgebung austrat und zu schrecklichen Folgen führte. In diesem Stück sind die Instrumente in fünf Gruppen aufgeteilt, die die einzelnen Angriffe repräsentieren, die Umstände und die Stimmung sowohl buchstäblich als auch symbolisch beschreiben, und die tatsächliche Verdampfung konkretisieren“<sup>250</sup>.*

---

<sup>250</sup> vgl. Internationales Musikinstitut Darmstadt; Gísladóttir (2018), S. 2

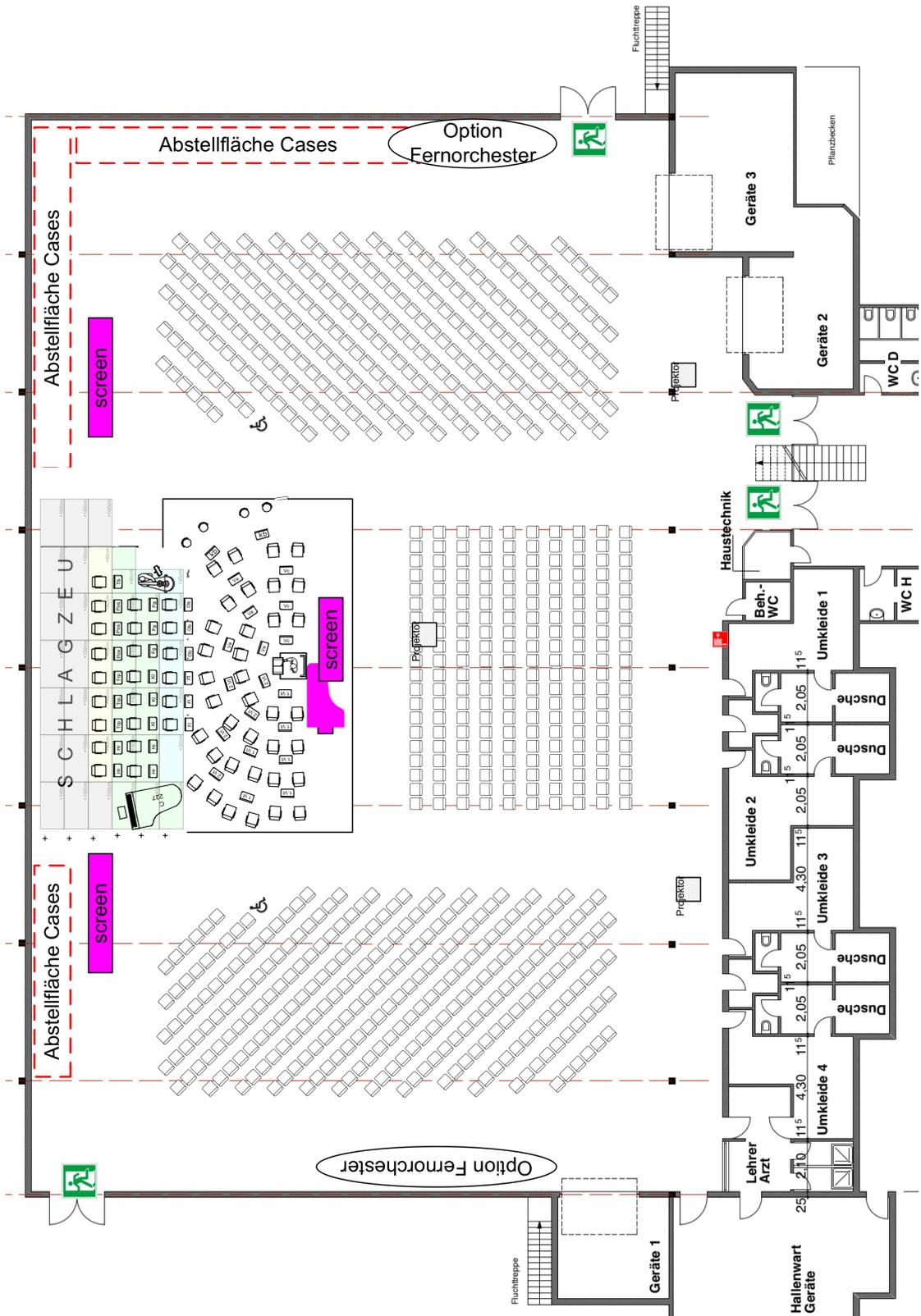


Abbildung 32 - Hallenplan des Eröffnungskonzertes<sup>251</sup>

<sup>251</sup> Bildquelle: Disposition des Hessischen Rundfunks

### 12.2.1 Aufnahme

Die Aufzeichnung des Konzertes für eine spätere Radioausstrahlung erfolgte durch die Hörfunkproduktion des Hessischen Rundfunks mit einem Ü-Wagen. Neben der Mikrofonierung für die Stereoproduktion wurde ein FOA-Mikrofon aufgebaut und ebenfalls das in Kapitel 10 vorgestellte stereofone 3D-Mikrofonarray. Für das 3D-Array kamen Mikrofone mit Supernierencharakteristik zum Einsatz. Der Mikrofonierungsplan ist Tabelle 1 zu entnehmen. Auf dem Ü-Wagen erfolgte eine Vormischung und Mehrspuraufnahme der Signale. Um die Signale des FOA-Mikrofons und des 3D-Arrays während der Aufzeichnung über Kopfhörer abhören zu können, wurden diese via MADI-Leitung an ein Laptop mit Audio-Interface weitergeleitet. Dort konnte innerhalb von *Pro Tools* die Encodierung nach Ambisonics und die Binauralsynthese durchgeführt werden.

Abbildung 33 zeigt den Aufbau des Orchesters und die Position der 360°-Kamera. Diese wurde auf Höhe des zweiten Pultes zwischen erster und zweiter Violine platziert, mit Blickrichtung zum Dirigenten. Durch diese Position kann zum einen der Dirigent von vorn gesehen werden und zum anderen befindet man sich unmittelbar zwischen den Musikern und kann sie beim Spielen ihrer Instrumente beobachten. So wird ein deutlich anderes Konzerterlebnis möglich, als auf einem der Zuschauerplätze.

FOA-Mikrofon und 3D-Array wurden nicht an der Kameraposition aufgestellt, sondern auf ca. 3,30 m Höhe hinter dem Dirigenten, um die Funktion eines Hauptmikrofons zu übernehmen und das Orchester möglichst ausgewogen in seiner Gesamtheit zu erfassen. Überdies hätte der Mikrofonaufbau an der Kameraposition die Sichtverhältnisse der Musiker zu stark eingeschränkt.

*Tabelle 1 - Mikrofonierungsplan für die Aufzeichnung des hr Sinfonieorchesters*

| Position       | Mikrofon     | Position           | Mikrofon     |
|----------------|--------------|--------------------|--------------|
| Haupt A        | Schoeps MK2s | Xylophon           | Schoeps MK4  |
| Haupt B        | Schoeps MK2s | Diverse Perkussion | Schoeps MK4  |
| LL             | Schoeps MK2s | Vibraphon          | Schoeps MK4  |
| RR             | Schoeps MK2s | Piano_L            | Neumann KM84 |
| 1. Violine     | Schoeps MK4  | Piano_R            | Neumann KM84 |
| 2. Violine     | Schoeps MK4  | Fernorchester_L    | Schoeps MK21 |
| Bratsche       | Schoeps MK4  | Fernorchester_R    | Schoeps MK21 |
| Violoncello    | Schoeps MK4  | Sampler1_L         | -            |
| Kontrabass     | Schoeps MK21 | Sampler1_R         | -            |
| Flöte_3        | Schoeps MK4  | Sampler2_L         | -            |
| Flöte_2_1      | Schoeps MK4  | Sampler2_R         | -            |
| Oboe_1_2       | Schoeps MK4  | Zuspielung_L       | -            |
| Oboe_3         | Schoeps MK4  | Zuspielung_R       | -            |
| Klarinette_3   | Schoeps MK4  | Solo_Piano_A_L     | Schoeps MK4  |
| Klarinette_2_1 | Schoeps MK4  | Solo_Piano_A_R     | Schoeps MK4  |
| Fagott_1_2     | Schoeps MK4  | Solo_Piano_B_L     | dpb 4099     |
| Fagott_3       | Schoeps MK4  | Solo_Piano_B_R     | dpa 4099     |
| Hörner_2_1     | Schoeps MK21 | 3D-Array_L         | Schoeps MK41 |

|                 |              |                |                     |
|-----------------|--------------|----------------|---------------------|
| Hörner 4 3      | Schoeps MK21 | 3D-Array R     | Schoeps MK41        |
| Trompete 3 2    | Neumann KM84 | 3D-Array Ls    | Schoeps MK41        |
| Trompete 1      | Neumann KM84 | 3D-Array Rs    | Schoeps MK41        |
| Posaune 1       | Neumann KM84 | 3D-Array Lt    | Schoeps MK41        |
| Posaune 2 3     | Neumann KM84 | 3D-Array Rt    | Schoeps MK41        |
| Tuba            | Schoeps MK4  | 3D-Array Lts   | Schoeps MK41        |
| Harfe_nah       | dpa 4061     | 3D-Array Rts   | Schoeps MK41        |
| Harfe_normal    | Schoeps MK4  | FOA-Mikrofon 1 | Sennheiser AMBEO VR |
| Gran Casa       | Schoeps MK4  | FOA-Mikrofon 2 | Sennheiser AMBEO VR |
| Cowbells        | Schoeps MK4  | FOA-Mikrofon 3 | Sennheiser AMBEO VR |
| Glockenspiel    | Schoeps MK4  | FOA-Mikrofon 4 | Sennheiser AMBEO VR |
| Pauke           | Schoeps MK4  | Spion          | Neumann KM84        |
| Röhren-Xylophon | Schoeps MK4  |                |                     |

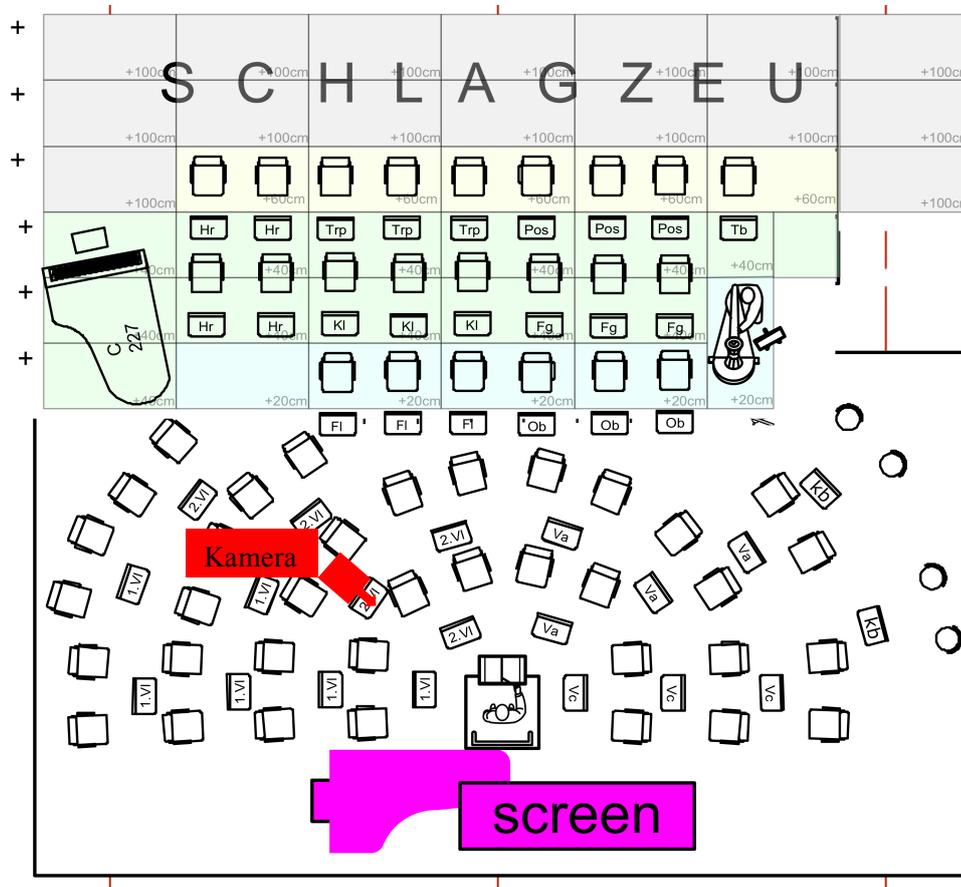


Abbildung 33 - Orchesteraufbau für das Eröffnungskonzert der Darmstädter Ferienkurse für neue Musik<sup>252</sup>

### 12.2.2 Nachbearbeitung

Für die Nachbearbeitung standen die Einzelspuren der Vormischung vom Ü-Wagen zur Verfügung, sowie die Signale des FOA-Mikrofons und des 3D-Mikrofonarrays. Für die Bearbeitung wurde ein *Pro Tools HDX-System* mit einer DSP-Karte eingesetzt, weshalb 256 Voices genutzt werden konnten. Die Ambisonics-Mischung erfolgte in dritter Ordnung. Bei dieser Testproduktion wurde die *Audio Ease 360*

<sup>252</sup> Bildquelle: Disposition des Hessischen Rundfunks

*Pan Suite* eingesetzt. Zusätzlich wurde ein Head-Tracker genutzt, der mit dem Audio Ease Spatializer gekoppelt werden konnte.

Die Ambisonics-Mischung wurde mit den 3D-Mikrofonen begonnen. Das FOA-Mikrofon musste zunächst mit dem *Sennheiser AMBEO A-B Converter* ins Ambisonics-B-Format konvertiert werden. Innerhalb des Plugins wurde auch das Schallfeld um etwa  $140^\circ$  nach rechts gedreht, um sich mit der Kameraperspektive zu decken. Die genaue Einstellung musste häufig kontrolliert und angepasst werden, da es mitunter schwierig war, einzelne Instrumente klar zu lokalisieren. In diesem Zusammenhang erwies es sich als hilfreich, die Bereiche mit hoher Schallenergie durch das *360 Radar* Plugin visuell darstellen zu lassen. Die Signale des 3D-Mikrofonarrays wurden mit Hilfe des *360 Pan* Plugins zunächst, wie in Abbildung 34 dargestellt, würfelförmig angeordnet und anschließend mit Hilfe des *360 Turner* an die Kameraperspektive angepasst. Beim Vergleich der beiden Mikrofone konnte festgestellt werden, dass das 3D-Array eine räumlichere Abbildung erzeugte und der Raum offener und größer wirkte. Das FOA-Mikrofon lieferte ein engeres, aber höhenreicheres Klangbild. Die gewünschte Präsenz und Nähe, vor allem zu den Streichern in unmittelbarer Nähe der Kamera, wurde von keinem der Mikrofone ausreichend abgebildet.



Abbildung 34 - Verteilung der Signale des 3D-Mikrofonarrays<sup>253</sup>

<sup>253</sup> Bildquelle: Audio Ease (2018)

In einem weiteren Schritt wurden die Stützmicrofone zugemischt, um die gewünschte Präsenz der Instrumente zu erreichen. Aufgrund der größeren Räumlichkeit wurde das 3D-Mikrofonarray als Basis genutzt. Die zur Verfügung stehenden 256 Voices in Pro Tools reichten nicht aus, um alle Stützmicrofone gleichzeitig in Ambisonics zu mischen. Neben den 16 Voices die pro Audiotrack in Ambisonics dritter Ordnung benötigt werden, fallen zusätzliche Voices weg, da die Plugins der *Audio Ease 360° Pan Suite* nicht auf den *Pro Tools HDX* DSP-Karten berechnet werden können<sup>254</sup>. Daher mussten die Stützmicrofone in drei Gruppen zusammengefasst und gemischt werden, was die Flexibilität der Mischung insgesamt einschränkte, da bei einer nachträglichen Anpassung einer Gruppe deren Sub-Mischung neu ausgespielt werden musste. In einer ersten Gruppe wurden Percussions, Sampler, Zuspierer und Fernorchester zusammengefasst. Streicher, Harfe und Pianos bildeten eine zweite Gruppe. Die dritte Gruppe bestand aus den Holz- und Blechbläsern. Das Panning der Stützmicrofone erfolgte mit Hilfe des *360 Pan Ambisoncis-Panners* anhand der Rektangular-Projektion des Videobildes. Dabei wurden, wie in Abbildung 35 aufgezeigt, die Mikrofonensignale als farbige Kreise auf dem Videobild dargestellt und ausgerichtet. Aufgrund des Signalübersprechens und der starken Verzerrungen im Randbereich des Videobildes mit vielen Schallquellen war eine eindeutige Platzierung der Signale nicht immer möglich. Ferner wurden die Signale nicht exakt auf die Position des Stützmicrofons angelegt, auch wenn dieses im Bild klar zu erkennen war. Stattdessen wurden die Signale direkt einem der Musiker zugeordnet. Hierdurch konnte vor allem bei den Instrumenten in unmittelbarer Nähe zur Kamera eine gute Deckung von erwarteter und tatsächlicher Schallereignisrichtung erreicht werden.

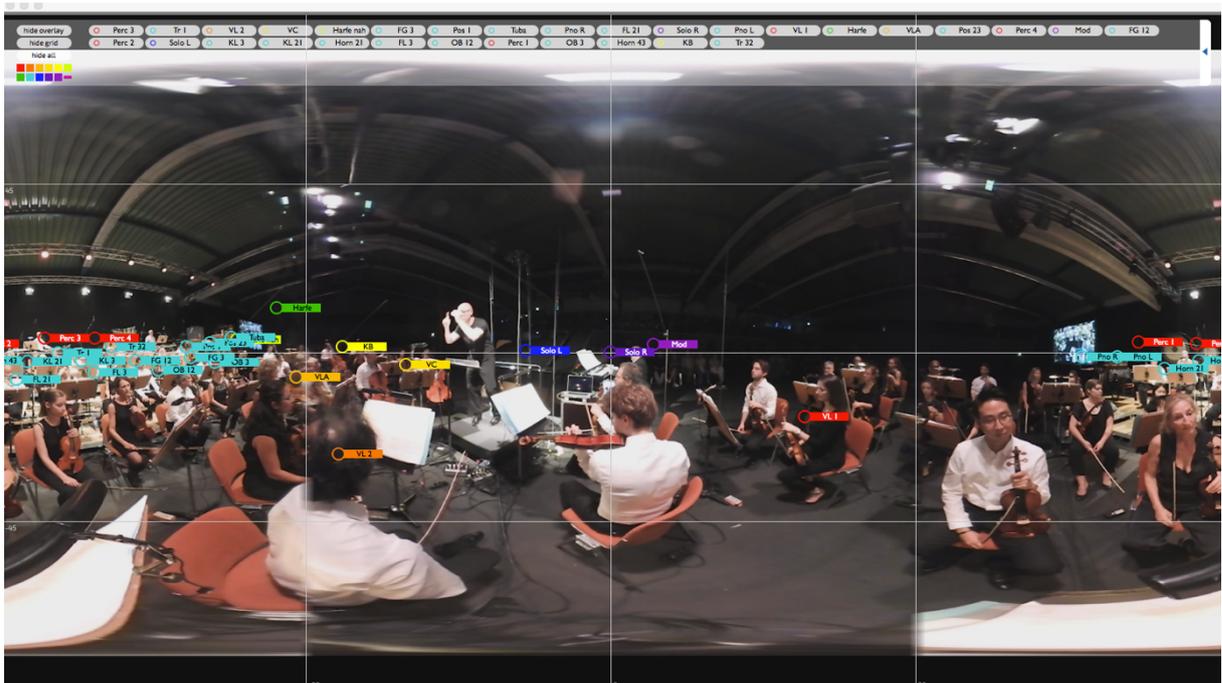


Abbildung 35 - Videobasiertes Panning der Stützmicrofone

<sup>254</sup> Sound on Sound Magazine (2018)

Die Akustik der Turnhalle war insgesamt recht trocken für ein Orchesterkonzert. Daher wurde in der Nachbearbeitung künstlicher Hall zugemischt. Dabei kam der Ambisonics-Faltungshall *360° Reverb* von *Audio Ease* zum Einsatz. Der größte Hallanteil wurde dem 3D-Mikrofonarray zugemischt, um die Räumlichkeit künstlich zu verbreitern. Es wurde die Impulsantwort eines mittelgroßen Konzertsaals genutzt, um den Abmessungen der Turnhalle einigermaßen zu entsprechen. Den Stützmikrofonen, die sich weiter entfernt von der Kamera befanden, wurde in geringem Maße ebenfalls künstlicher Nachhall zugemischt, wodurch eine Tiefenstaffelung der Signale erreicht werden sollte. Stützmikrofone in Kameranähe wurden trocken zugemischt.

### 12.2.3 Beobachtungen

Bei dieser Testproduktion fiel auf, dass das FOA-Mikrofon nicht den gewünschten offenen Raumklang erzielen konnte und das 3D-Mikrofonarray diese Anforderungen besser erfüllte. Die Abbildung des gesamten Orchesterklangs gelang mit dem 3D-Mikrofonarray eher und bei Orientierungswechseln waren die Klangänderungen ausgewogener und weniger phasig.

Zudem war zu beobachten, dass den Stützmikrofonen bei dieser Produktion eine große Bedeutung zukam. Ohne die Signale der Stützmikrofone wäre es nicht möglich gewesen, die benötigte Präsenz und Nähe der Instrumente darzustellen, die sich unmittelbar an der Kamera befanden. Überdies ließ sich feststellen, dass die Stützmikrofone für eine exakte Lokalisierung äußerst relevant sind. Erst mit den Direktsignalen der Stützmikrofone gelingt bei Kopfdrehungen die genaue Zuordnung von akustischer und optischer Schallquelle.

Es wurde ferner versucht eine Mischung zu erstellen und lediglich die Signale der Stereo-Mischung zu verwenden. Anstelle des 3D-Arrays oder des FOA-Mikrofons wurde das Stereo-Hauptmikrofon mit zwei Kugeln genutzt. Bei dieser Mischung fiel die Abbildung der Räumlichkeit weniger plastisch und flacher aus, da keine Signale zur Abbildung der hinteren und oberen Ebene vorhanden waren.

Der Einsatz des künstlichen Nachhalls zeigte, dass die Signale der Stützmikrofone sehr sensibel auf den Hallanteil reagierten und die durch Zumischung der Stützmikrofone erzeugte Präsenz schnell darunter litt. Daher sollte eine Tiefenstaffelung hauptsächlich über Pegeldifferenzen und ggf. über Klangbearbeitungen erfolgen und künstlichen Nachhall nur sparsam verwendet werden. Die Verbreiterung des Raumeindrucks gelingt gut über das Zumischen von künstlichem Nachhall zum Hauptmikrofon.

---

Zusammenfassend scheint die Kombination aus stereofonem 3D-Mikrofonarray, Stützmikrofonen und gezielt eingesetztem künstlichen Nachhall ein praktikabler Ansatz für die Produktion von 360°-Videos im Kontext orchesterlicher Musikaufführungen zu sein.

## 13. Workflowempfehlung für den Hessischen Rundfunk

Eine wichtige Frage bei der Produktion von 360°-Inhalten ist die nach der Länge des finalen Werkes. Basierend auf den Erfahrungen, die während der Testproduktionen gesammelt wurden und in Anbetracht der Wiedergabemöglichkeiten von 360°-Videos, ist eine Länge von maximal 15 Minuten empfehlenswert. In diesem Zusammenhang weist auch Bailson darauf hin, dass verfügbare HMDs unkomfortabel sind und daher nach dem Ansatz „less is more“ gearbeitet werden sollte<sup>255</sup>. Aufgrund der Größe und des Gewichts, sind HMDs ungeeignet für eine stundenlange Nutzung. Auch das Halten und Bewegen eines Smartphones oder das Nachführen des Videobildes per Maus ermüdet bei langen Videos. Daher sollten bei Konzertaufzeichnungen besser mehrere 360°-Videos einzelner Musikstücke oder Passagen produziert werden, anstatt des ganzen Konzerts am Stück.

Im Rahmen der Testproduktionen zu dieser Arbeit konnte gezeigt werden, dass ein stereofones 3D-Mikrofonarray einen echten Mehrwert für die Audioproduktion von 360°-Video darstellt. Der Einsatz eines solchen Mikrofonarrays wird daher empfohlen, wann immer es die Produktionsbedingungen ermöglichen. Ein praktikabler Ansatz für ein stereofones 3D-Mikrofonarray wurde in dieser Arbeit vorgestellt und unter realen, praktischen Bedingungen getestet. Der Aufbau des Mikrofonarrays ist für den Hessischen Rundfunk mit vorhandenem Material umsetzbar und bringt keine zusätzlichen Investitionskosten mit sich. Der zeitliche Mehraufwand für Aufbau, Ausrichtung und Verkabelung des 3D-Mikrofonarrays beträgt etwa 1,5 Stunden und wird durch eine dreidimensionale, ausgewogene Abbildung gerechtfertigt. Für den Einsatz des 3D-Mikrofonarrays werden acht zusätzliche Mikrofonwege benötigt. Ein FOA-Mikrofon lässt sich in der Praxis noch einfacher anwenden, bietet aber – gemessen an den durchgeführten Testproduktionen – keinen klanglichen Vorteil. Daher wird der Einsatz von FOA-Mikrofonen im Kontext musikalischer Aufführungspraxis an dieser Stelle nicht empfohlen. Ebenso soll ausdrücklich dazu ermutigt werden, mit weiteren stereofonen 3D-Mikrofonarrays zu experimentieren und praktische Erfahrungen zu sammeln, da sich hierdurch vielfältige Möglichkeiten der Klanggestaltung bieten.

Weiterhin ist die Möglichkeit, die 3D-Mikrofone bereits während der Aufnahme von einem separaten Arbeitsplatz aus, mittels Binauralsynthese, über Kopfhörer abhören zu können, empfehlenswert. Dies

---

<sup>255</sup> Bailenson (2018), S. 247 ff.

---

ermöglicht die Beurteilung und eventuelle Anpassung des Mikrofonaufbaus im Hinblick auf den Einsatz bei 360°-Videos. Nichtsdestotrotz ist diese Abhörmöglichkeit optional, solange sichergestellt ist, dass Routing und Aussteuerung der Mikrofonsignale des 3D-Arrays korrekt sind.

Für erste Produktionen und die ausschließliche Verbreitung über *Youtube* und *Facebook* ist die kostenlos erhältliche *Facebook 360 Workstation* ausreichend. Sie ist in *Pro Tools HD* ab Version 12.8.2 bereits integriert<sup>256</sup>. Mit ihr kann aktuell jedoch nur bis Ambisonics zweiter Ordnung produziert werden<sup>257</sup>, weshalb diese Lösung längerfristig nicht optimal ist. Der in *FB360 Encoder* ist in jedem Fall empfehlenswert, da er eine einfache Möglichkeit bietet eine Videodatei mit einer Ambisonics-Mischung zu verknüpfen und die benötigten Metadaten zu generieren.

Nach erfolgreicher Einarbeitung und etabliertem Workflow ist für die komplette Nachbearbeitung eines ca. 15-minütigen 360°-Videos, inklusive Audio- und Videobearbeitung, mit einem Zeitaufwand von etwa 1,5 bis 2 Arbeitstagen zu rechnen. Der Zeitaufwand ist unabhängig von den verwendeten Tools und der Ambisonics-Ordnung in der produziert wird.

In Anbetracht der vielfältigen, verfügbaren Softwaretools für die Nachbearbeitung von 3D-Audio in Ambisonics, werden an dieser Stelle zwei empfehlenswerte Möglichkeiten mit unterschiedlichen Schwerpunkten vorgestellt:

### *Postproduktions-Empfehlung 1*

Diese Empfehlung zielt auf einen möglichst einfachen und schnellen Einstieg in die Audiopostproduktion von 360°-Video ab. Demnach wird die DAW *Pro Tools* eingesetzt, da sie im Hessischen Rundfunk weit verbreitet ist und die Mitarbeiter im Umgang mit dieser DAW geschult sind. Ambisonics-Mischungen bis zur dritten Ordnung sind in *Pro Tools HD* ab Version 12.8.2 möglich<sup>258</sup>.

Die Ambisonics-Plugin-Bundles *360 Pan* von *Audio Ease* und *Ambi HD* von *Noisemakers* sind beide empfehlenswert und überzeugen mit einer intuitiven Bedienung und einem übersichtlichem User-Interface. Beide sind im AAX-Plugin-Format für *Pro Tools* erhältlich. Ihr Funktionsumfang unterscheidet sich geringfügig voneinander.

Ein Vorteil des *Noisemakers* Bundle ist, dass es einen Ambisonics-Limiter enthält, sowie die Möglichkeit den HRTF-Datensatz im Spatializer austauschen zu können. Darüber hinaus besteht die

---

<sup>256</sup> Facebook (2017)

<sup>257</sup> Facebook (2018)

<sup>258</sup> Avid Technology (2018b)

---

Option eigene Impulsantworten in den Faltungshall zu integrieren. Ein Tool um das Ambisonics-Schallfeld zu drehen ist im *Noisemakers* Bundle nicht enthalten.

Die *Audio Ease 360 Pan Suite* zeichnet sich dadurch aus, dass bereits während der Mischung mit einem Head-Tracker gearbeitet werden kann. Die Option, Bereiche hoher Schallintensität zu visualisieren, kann bei der nachträglichen Ausrichtung eines 3D-Mikrofons helfen und stellt damit einen weiteren Vorzug der *Audio Ease 360 Pan Suite* dar.

Die kostenfrei erhältliche Plugin-Library *O3A Core* von *Blue Ripple Sound* unterstützt ebenfalls Ambisonics dritter Ordnung und ist im AAX-Plugin-Format erhältlich. Sie stellt eine nützliche Ergänzung zu den obigen Tools dar, da sie weitere Optionen für das Ambisonics-Metering beinhaltet. Zusätzlich ist ein Ambisonics-Rotator enthalten, der das *Noisemakers* Bundle um diese Funktionalität ergänzen kann. Der Funktionsumfang der Plugin-Library *O3A Core* von *Blue Ripple Sound* ist aber allein nicht ausreichend für eine Produktion, da kein Spatializer enthalten ist. Alle Plugins lassen sich natürlich auch miteinander kombinieren.

### *Postproduktions-Empfehlung 2*

Die zweite Postproduktionsempfehlung zielt auf eine große Flexibilität, und die Möglichkeit, Ambisonics-Mischungen in sehr hoher Ordnung realisieren zu können, ab. Betrachtet man allein die Materialkosten, so ist diese Empfehlung zudem vergleichsweise preiswert.

Bei diesem Ansatz wird die DAW *Reaper* von *Cockos* eingesetzt. Diese ist in der Lage Audiobusse von bis zu 64 Kanälen zu erstellen, wodurch es möglich ist, Ambisonics-Mischungen bis zur 7. Ordnung zu erstellen. Auch wenn derzeit keine Wiedergabeplattform dieses Format unterstützt, so ist dies für Archivierungszwecke und Downmixing sehr interessant. Ferner gibt es keine Spurbegrenzungen oder limitierte Voices in *Reaper*. Hierdurch ist es möglich Mischungen mit vielen Spuren und hoher Ambisonics-Ordnung zu erstellen ohne Submischungen anfertigen zu müssen. Dies kann bei großen Mischungen Zeit sparen und bietet eine größere Flexibilität für nachträgliche Änderungen. Die maximale Spuranzahl wird lediglich durch die Leistungsfähigkeit des genutzten Computers begrenzt. Neben einem Audiointerface wird keine weitere DSP-Hardware benötigt. Für die professionelle Nutzung der DAW *Reaper* ist jedoch mit erhöhtem Schulungsbedarf und/oder längerer Einarbeitungszeit zu rechnen, da diese DAW vom Hessischen Rundfunk bisher nicht verwendet wird.

Da mit der DAW *Reaper* Plugins im VST-Format genutzt werden können, bietet es sich an, die kostenlos erhältlichen Plugins der *IEM Plugin Suite* und die *ambix* Plugins von *Matthias Kronlachner* zu nutzen, da sie ebenfalls Ambisonics bis zur siebten Ordnung unterstützen. Ein Videobasiertes Panning ist allerdings nicht inbegriffen und aktuell nur für Ambisonics bis zur dritten Ordnung erhältlich.

## *Kosten*

Für die *DAW Pro Tools Ultimate* müssen Kosten von ca. 2300€ angerechnet werden<sup>259</sup>. Sofern im Hessischen Rundfunk auf vorhandene Systeme zurückgegriffen werden kann, fallen diese Kosten jedoch nicht ins Gewicht. Für die *DAW Cockos Reaper* betragen die Kosten für eine kommerzielle Lizenz etwa 240€<sup>260</sup>.

Die *Audio Ease 360 Pan Suite* ist für etwa 300€<sup>261</sup> erhältlich, das *Noisemakers Ambi HD Bundle* für ca. 480 €<sup>262</sup>. Die Plugins *O3A Core* von *Blue Ripple Sound*<sup>263</sup>, die *IEM Plugin Suite*<sup>264</sup> und die *ambix* Plugins von *Matthias Kronlachner* sind kostenfrei erhältlich<sup>265</sup>.

## **14. Fazit und Ausblick**

Die Zielsetzung der vorliegenden Arbeit bestand darin, die qualitätsbestimmenden Faktoren der Audioproduktion für 360°-Videos im Kontext musikalischer Aufführungspraxis aufzuzeigen und einen geeigneten Produktionsworkflow für den Hessischen Rundfunk zu erarbeiten und zu testen.

Zunächst wurde 360°-Video als Format definiert und die Grundlagen des räumlichen Hörens und der Binauraltechnik vorgestellt. In einem nächsten Schritt wurden die wichtigsten Gestaltungsmerkmale von 360°-Video erarbeitet, aus denen Anforderungen an die Tongestaltung abgeleitet werden konnten. Anschließend wurde das Ambisonics-Tonformat erörtert, welches für 360°-Videoproduktionen eingesetzt wird. Es folgte eine Darstellung der verschiedenen Distributionswege für 360°-Videoinhalte. Im weiteren Verlauf wurden verschiedene 3D-Mikrofonaufnahmetechniken für 360°-Videoproduktionen aufgezeigt und die Entwicklung eines stereophonen 3D-Mikrofonaufnahmeverfahrens erläutert. Nach der Vorstellung ausgewählter Postproduktionswerkzeuge wurden zwei Testproduktionen durchgeführt, beschrieben und ausgewertet. Abschließend konnte eine Workflow-Empfehlung zur Produktion von 360°-Videos im Kontext musikalischer Aufführungspraxis für den Hessischen Rundfunk gegeben werden.

Es kann festgehalten werden, dass das Erzeugen von Immersion eine wichtige Aufgabe der Tongestaltung für die Produktion von 360°-Videos ist. Dabei kommt es auf eine hochwertige Abbildung der Räumlichkeit, Realismus und die Darstellung von Nähe und Präsenz einzelner Schallquellen an.

---

<sup>259</sup> Avid Technology (2018)

<sup>260</sup> Cockos (2018)

<sup>261</sup> Audio Ease (2018)

<sup>262</sup> Noisemakers (2018b)

<sup>263</sup> Blue Ripple Sound (2018)

<sup>264</sup> Rudrich (2018)

<sup>265</sup> Kronlachner (2014)

Hinsichtlich der Aufnahmetechnik lässt sich aus der vorliegenden Arbeit schlussfolgern, dass die Kombination aus einem 3D-Hauptmikrofonverfahren und Stützmikrofonen die geeignetste Methode ist, um Audio für 360°-Videoproduktionen im Kontext musikalischer Aufführungen aufzunehmen. Darüber hinaus ermöglicht dies weitreichende Gestaltungsmöglichkeiten in der Nachbearbeitung. Im Rahmen der Postproduktion kommt es auf die Abstimmung von optischer und akustischer Perspektive an. Die Verteilung im Panorama und die Gestaltung der Tiefenstaffelung haben dabei den größten Einfluss auf die Wirkung des Endproduktes.

Zusammenfassend kann festgehalten werden, dass es möglich ist, qualitativ hochwertiges Audio für 360°-Video im Kontext musikalischer Aufführungspraxis zu erstellen. Dabei müssen die Besonderheiten des Formats 360°-Video beachtet werden, weshalb sich die Produktionsweise für 360°-Video von der einer konventionellen Stereo-Produktion unterscheidet. Funktionelle Mikrofonaufnahmetechniken, Postproduktionswerkzeuge und Distributionswege für 360°-Videos sind heute in ausreichender Anzahl und in hoher Qualität verfügbar. Im Rahmen dieser Arbeit konnte ein praktikabler Workflow erarbeitet werden, sodass der Hessische Rundfunk mit wenig Mehraufwand und geringen Investitionskosten in die Produktion von hochwertigen 360°-Videos der Konzerte seiner Klangkörper einsteigen kann.

Für zukünftige Produktionen ist es sinnvoll, das vorgeschlagene stereofone 3D-Aufnahmeverfahren in weiteren Szenarien zu evaluieren und mit anderen Mikrofonaufnahmeverfahren zu vergleichen. Auch der Einsatz mehrerer 360°-Kameras kann dazu beitragen, 360°-Videos abwechslungsreicher zu gestalten und bringt weitere Herausforderungen für die Tongestaltung mit sich. Voraussichtlich werden sich Nachbearbeitungswerkzeuge und Distributionsplattformen in rasantem Tempo weiterentwickeln und immer eindrucksvollere Produktionen ermöglichen, weshalb es sich lohnt, diese Entwicklungen weiter zu verfolgen. Es bleibt noch abzuwarten ob sich 360°-Videos und VR-Brillen zum Massenmedium entwickeln, doch um es mit Bailsons Worten zu sagen: „However it shakes out, it’s a very special time to be part of this technical revolution. The coming years are going to be a wild ride.“<sup>266</sup>

---

<sup>266</sup> vgl. Bailenson (2018), S. 260

---

## Literaturverzeichnis

**3Dio (2018):** Omni Pro Binaural Microphone. 22.8.2018, URL: <https://3diosound.com/products/omni-pro-binaural-microphone> [Zugriff: 22.8.2018]

**Algazi, V. Ralph; Duda, Richard O.; Thompson, Dennis M. (2004):** Motion-Tracked Binaural Sound. In: *J. Audio Eng. Soc.*, Band 52, Ausgabe 11, 2004, S. 1142–1156

**Alistair (2018):** 24h in VR. 19.7.2018, URL: <http://solveforinteresting.com/24h-in-vr-000-023-electric-monk/> [Zugriff: 19.7.2018]

**Audio Ease (2017):** 360 Pan Suite - User Manual.

**Audio Ease (2018):** 360pan suite, full control over your 360 (spatial) audio with this professional audio plug-in suite. 27.8.2018, URL: <https://www.audioease.com/360> [Zugriff: 27.8.2018]

**Audiokinetic Inc. (2018):** Ambisonics Limitations. 2018, URL: [https://www.audiokinetic.com/library/2017.1.9\\_6501/?source=Help&id=ambisonics\\_limitations](https://www.audiokinetic.com/library/2017.1.9_6501/?source=Help&id=ambisonics_limitations) [Zugriff: 2.8.2018]

**Avid Technology (2018a):** Pro Tools® Reference Guide Version 2018.7.

**Avid Technology (2018b):** Release Dates and Versions for Pro Tools (v9 and later, includes Pro Tools HD). 27.7.2018, URL: [http://avid.force.com/pkb/articles/en\\_US/faq/en393795](http://avid.force.com/pkb/articles/en_US/faq/en393795) [Zugriff: 25.8.2018]

**Avid Technology (2018):** Pro Tools | Ultimate Subscriptions and Upgrades - Music Software | Avid. 25.8.2018, URL: [https://shop.avid.com/ccrz\\_\\_ProductDetails?viewState=DetailView&cartID=&sku=DYNA20001](https://shop.avid.com/ccrz__ProductDetails?viewState=DetailView&cartID=&sku=DYNA20001) [Zugriff: 25.8.2018]

**Bailenson, Jeremy (2018):** Experience on demand what virtual reality is, how it works, and what it can do.

**Bates, Enda; Dooney, Sean; Gorzel, Marcin; O'Dwyer, Hugh; Ferguson, Luke; Boland, Francis M. (2017):** Comparing Ambisonic Microphones—Part 2. Audio Engineering Society Convention 142, URL: <http://www.aes.org/e-lib/browse.cfm?elib=18607>

**Bates, Enda; Gorzel, Marcin; Ferguson, Luke; O'Dwyer, Hugh; Boland, Francis M. (2016):** Comparing Ambisonic Microphones – Part 1. Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control, URL: <http://www.aes.org/e-lib/browse.cfm?elib=18317>

**Berg, Jenny (2017):** Trendwende - Die Jungen haben wieder Bock auf klassische Konzerte. 6.3.2017, URL: <https://www.srf.ch/kultur/musik/die-jungen-haben-wieder-bock-auf-klassische-konzerte> [Zugriff: 26.8.2018]

**Berger; Kiani Anaraki; Rempe; omnia360 (2017):** Was ist Virtual Reality? Definitionsansätze der virtuellen Realität. , URL: <https://omnia360.de/blog/was-ist-virtual-reality/> [Zugriff: 18.7.2018]

**Blauert, Jens. (1974):** Räumliches Hören. Stuttgart: Hirzel

**Blauert, Jens (2013):** Räumliches Hören Komplettausgabe: Grundwerk, 1. Aufl. 1974 mit 1. Nachschrift 1985 (Ergebnisse und Trends seit 1972) und 2. Nachschrift 1997 (Ergebnisse und Trends seit 1982). Stuttgart: S. Hirzel Verlag

- Blauert, Jens; Braasch, Jonas (2008):** Räumliches Hören. In: Weinzierl, Stefan (Hrsg.): *Handbuch der Audiotechnik*. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 87–121
- Blue Ripple Sound (2018):** O3A Core. 29.8.2018,  
URL: <http://www.blueripplesound.com/products/o3a-core> [Zugriff: 29.8.2018]
- Böcking, Saskia (2008):** Grenzen der Fiktion? von Suspension of Disbelief zu einer Toleranztheorie für die Filmrezeption. Köln: von Halem
- Brill, Manfred (2009):** Virtuelle Realität. Berlin: Springer
- Camerer, Florian (2018):** Die Kirche, das Dorf und 3D-Audio - was gehört wohin?, Schoeps Mikroforum 2018, Karlsruhe, 06.04.2018, URL: <https://www.youtube.com/watch?v=osZ842Zaj5Q>
- Carpentier, Thibaut (2017):** Normalization Schemes in Ambisonic: Does it Matter?. Audio Engineering Society Convention 142, URL: <http://www.aes.org/e-lib/browse.cfm?elib=18645>
- Chapman, Michael; Ritsch, Winfried; Musil, Thomas; Zmölnig, Johannes; Pomberger, Hannes; Zotter, Franz; Sontacchi, Alois (2009):** A Standard for Interchange of Ambisonic Signal Sets. Including a file standard with metadata. Proc. of the Ambisonics Symposium, Graz, Austria
- Cockos (2018):** Up and Running: A REAPER User Guide v 5.94.
- Cockos (2018):** REAPER | Purchase. 25.8.2018, URL: <https://www.reaper.fm/purchase.php> [Zugriff: 25.8.2018]
- Core Sound, LLC (2017):** TetraMic. 2017, URL: <http://www.core-sound.com/TetraMic/1.php> [Zugriff: 5.8.2018]
- Craven, Peter Graham; Gerzon, Michael Anthony (1977):** Coincident microphone simulation covering three dimensional space and yielding various directional outputs. , URL: <https://patents.google.com/patent/US4042779A/en>
- Daniel, Jérôme (2003):** Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction
- Deboy, Daniel (2018):** Musik in VR, Schoeps Mikroforum 2018, Karlsruhe, 06.04.2018, URL: <https://www.youtube.com/watch?v=vDa2uL2ViQA>
- Deutsches Musikinformationszentrum (2016):** Umsatz des Musikveranstaltungsmarkts, Besucherreichweite und Ticketpreise nach Musikrichtungen. 8.2016,  
URL: [http://www.miz.org/downloads/statistik/124/124\\_Umsatz\\_des\\_Musikveranstaltungsmarkts\\_\\_Besucherreichweite\\_und\\_Ticketpreise\\_nach\\_Musikrichtungen\\_2014.pdf](http://www.miz.org/downloads/statistik/124/124_Umsatz_des_Musikveranstaltungsmarkts__Besucherreichweite_und_Ticketpreise_nach_Musikrichtungen_2014.pdf) [Zugriff: 26.8.2018]
- Deutsches Musikinformationszentrum (2018):** Besuche von Musikveranstaltungen und Konzerten. 6.2018, URL: [http://www.miz.org/downloads/statistik/106/106\\_Besuche\\_Musikveranstaltungen.pdf](http://www.miz.org/downloads/statistik/106/106_Besuche_Musikveranstaltungen.pdf) [Zugriff: 26.8.2018]
- Dickreiter, Michael (1995):** Mikrofon-Aufnahmetechnik. Stuttgart: Hirzel
- Dickreiter, Michael; Goeres-Petri, Jürgen (2013):** Schallwahrnehmung. In: *Handbuch der Tonstudiotechnik*. Berlin, Boston: DE GRUYTER, S. 115–135
- Dieterle, Hannes; Wittek, Helmut; Sagnowski, Kasper (2018):** Mikrofontechnik für VR, Schoeps Mikroforum 2018, Karlsruhe, 06.04.2018, URL: <https://www.youtube.com/watch?v=ut9rvTsxeEY>

- Dörner, Ralf; Jung, Bernhard; Grimm, Paul; Broll, Wolfgang; Göbel, Martin (2013):** Einleitung. In: Dörner, Ralf; Broll, Wolfgang; Grimm, Paul; Jung, Bernhard (Hrsg.): *Virtual und Augmented Reality (VR / AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 1–31
- Dörner, Ralf; Steinicke, Frank (2013):** Wahrnehmungsaspekte von VR. In: Dörner, Ralf; Broll, Wolfgang; Grimm, Paul; Jung, Bernhard (Hrsg.): *Virtual und Augmented Reality (VR / AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 33–63
- EBU Technical Recommendation (1998):** EBU Tech. 3276 - Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic.
- Elen, Richard (2001):** Ambisonics: The surround alternative. Proceedings of the 3rd Annual Surround Conference and Technology Showcase
- Ellermeier, Wolfgang; Hellbrück, Jürgen (2008):** Hören – Psychoakustik – Audiologie. In: Weinzierl, Stefan (Hrsg.): *Handbuch der Audiotechnik*. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 41–85
- Facebook (2017):** Spatial Workstation + Pro Tools | HD 12.8.2. , URL: <https://facebook360.fb.com/2017/10/18/spatial-workstation-pro-tools-hd-12-8-2/> [Zugriff: 25.8.2018]
- Facebook (2018):** Introduction — Facebook Audio 360 documentation. 19.8.2018, URL: <https://facebookincubator.github.io/facebook-360-spatial-workstation/Documentation/SpatialWorkstation/SpatialWorkstation.html> [Zugriff: 19.8.2018]
- Farina, Angelo (2017):** Ambisonics to TBE conversion. 2017, URL: <http://pcfarina.eng.unipr.it/TBE-conversion.htm> [Zugriff: 16.8.2018]
- Fink, Charlie (2018):** Charlie fink’s metaverse: an AR enabled guide to VR & AR. Washington, DC: Cool Blue Media
- Friesecke, Andreas (2014):** Die Audio-Enzyklopädie: ein Nachschlagewerk für Tontechniker. Berlin: De Gruyter Saur
- Geluso, Paul (2012):** Capturing Height: The Addition of Z Microphones to Stereo and Surround Microphone Arrays. Audio Engineering Society Convention 132, URL: <http://www.aes.org/e-lib/browse.cfm?elib=16233>
- Gerzon, Michael A. (1975):** The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound. Audio Engineering Society Convention 50, URL: <http://www.aes.org/e-lib/browse.cfm?elib=2466>
- Gerzon, Michael A. (1985):** Ambisonics in Multichannel Broadcasting and Video. In: *J. Audio Eng. Soc.*, Band 33, Ausgabe 11, 1985, S. 859–871
- Google; Youtube (2018a):** Use spatial audio in 360-degree and VR videos - Computer - YouTube Help. 2018, URL: <https://support.google.com/youtube/answer/6395969?co=GENIE.Platform%3DDesktop&hl=en> [Zugriff: 16.8.2018]
- Google; Youtube (2018b):** Upload 360-degree videos - YouTube Help. 2018, URL: <https://support.google.com/youtube/answer/6178631> [Zugriff: 16.8.2018]

- Google; Youtube (2018c):** spatial-media: Specifications and tools for 360° video and spatial audio. Google, URL: <https://github.com/google/spatial-media>
- Görne, Thomas (2015):** Tontechnik. München: Hanser
- Grover, Ronald (2017):** Virtual reality startup Jaunt to shake up entertainment industry. 9.6.2017, URL: <https://www.cnbc.com/2017/06/08/virtual-reality-startup-jaunt-to-shake-up-entertainment-industry.html> [Zugriff: 16.8.2018]
- Henle, Hubert (2001):** Das Tonstudio-Handbuch: praktische Einführung in die professionelle Aufnahmetechnik. München: Carstensen
- Hollerweger, Florian (2005):** An Introduction to Higher Order Ambisonic. In: *Florian Hollerweger's Website*, 2005
- Howard, David M.; Angus, Jamie (2009):** Acoustics and psychoacoustics. Amsterdam London: Focal
- Internationales Musikinstitut Darmstadt; Gísladóttir, Bára (2018):** Programmheft Eröffnungskonzert Darmstädter Ferienkurse für neue Musik.
- Jaunt VR (2018a):** Submission Guidelines. 16.8.2018, URL: <https://www.jauntxr.com/creators/guidelines/> [Zugriff: 16.8.2018]
- Jaunt VR (2018b):** Creator Benefits. 16.8.2018, URL: <https://www.jauntxr.com/creators/benefits/> [Zugriff: 16.8.2018]
- Kim, Sungyoung (2018):** Height Channels. In: Roginska, Agnieszka; Geluso, Paul (Hrsg.): *Immersive sound: the art and science of binaural and multi-channel audio*.
- Kronlachner, Matthias (2014):** ambiX v0.2.7 – Ambisonic plug-in suite. , URL: <http://www.matthiaskronlachner.com/?p=2015> [Zugriff: 29.8.2018]
- Kuhlen, Torsten (2014):** Virtuelle Realität als Gegenstand und Werkzeug der Wissenschaft. In: Jeschke, Sabina; Kobbelt, Leif; Dröge, Alicia (Hrsg.): *Exploring Virtuality: Virtualität im interdisziplinären Diskurs*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 133–147
- Lee, Hyunkook (2016):** Capturing and Rendering 360° VR Audio Using Cardioid Microphones. Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality, URL: <http://www.aes.org/e-lib/browse.cfm?elib=18511>
- Lee, Hyunkook (2018):** Psychoacoustics of 3D sound recording techniques: Research and Practice, Schoeps Mikroforum 2018, Karlsruhe, 06.04.2018, URL: <https://www.youtube.com/watch?v=5SsQdfKAhVA>
- Lindberg, Morten (2015):** 3D Recording with the “2L-cube”. In: *VdT-Magazin*, Ausgabe 3–2015, 2015, S. 2
- Maempel, Hans-joachim (2013):** Klanggestaltung. In: *Handbuch der Tonstudioteknik*. Berlin, Boston: DE GRUYTER, S. 370–421
- Malham, Dave (2003):** Higher order ambisonics. 2003, URL: [https://www.york.ac.uk/inst/mustech/3d\\_audio/higher\\_order\\_ambisonics.pdf](https://www.york.ac.uk/inst/mustech/3d_audio/higher_order_ambisonics.pdf) [Zugriff: 27.8.2018]
- Mayr, Bärbel (2008):** Technischer Leitfaden und Bedienerhandbuch für das Tonstudio der

---

Hochschule der Medien.

**Meyer, Jens; Elko, Gary (2016):** A Qualitative Analysis of Frequency Dependencies in Ambisonics Decoding Related to Spherical Microphone Array Recording. Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control, URL: <http://www.aes.org/e-lib/browse.cfm?elib=18316>

**mh Acoustics LLC (2018):** Eigenmic. 2018, URL: <https://mhacoustics.com/products> [Zugriff: 5.8.2018]

**Møller, Henrik; Sørensen, Michael Friis; Jensen, Clemen Boje; Hammershøi, Dorte (1996):** Binaural Technique: Do We Need Individual Recordings?. In: *J. Audio Eng. Soc.*, Band 44, Ausgabe 6, 1996, S. 451–469

**Mühlberger, Andreas (2014):** Virtuelle Realität in der Klinischen Emotions- und Psychotherapieforschung. In: Jeschke, Sabina; Kobbelt, Leif; Dröge, Alicia (Hrsg.): *Exploring Virtuality: Virtualität im interdisziplinären Diskurs*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 149–161

**Murray, Janet Horowitz (1998):** Hamlet on the holodeck: the future of narrative in cyberspace. Cambridge, Mass: MIT Press

**Neitzel, Britta; Nohr, Rolf F. (Hrsg.) (2006):** Das Spiel mit dem Medium : Partizipation - Immersion - Interaktion. Marburg: NoOrganization

**Nicol, Rozenn (2018):** Sound Field. In: Roginska, Agnieszka; Geluso, Paul (Hrsg.): *Immersive sound: the art and science of binaural and multi-channel audio*.

**Noisemakers (2018a):** 3D audio panner for cinematic VR. 22.8.2018, URL: <https://www.noisemakers.fr/ambi-pan/> [Zugriff: 22.8.2018]

**Noisemakers (2018b):** Interactive Audio Software. 28.8.2018, URL: <https://www.noisemakers.fr/> [Zugriff: 28.8.2018]

**Politis, Archontis (2018):** Real/Complex Spherical Harmonic Transform, Gaunt Coefficients and Rotations - File Exchange - MATLAB Central. 22.8.2018, URL: <https://www.mathworks.com/matlabcentral/fileexchange/43856> [Zugriff: 22.8.2018]

**Rieger, Martin (2017):** Spatial Audio format support for 360° Video Platforms & Players. 15.4.2017, URL: <https://www.vrtonung.de/en/spatial-audio-support-for-360-video-platform/> [Zugriff: 16.8.2018]

**Roginska, Agnieszka (2018):** Binaural Audio Through Headphones. In: Roginska, Agnieszka; Geluso, Paul (Hrsg.): *Immersive sound: the art and science of binaural and multi-channel audio*.

**Rudrich, Daniel (2018):** IEM Plugin Suite - Descriptions. 7.6.2018, URL: <https://plugins.iem.at/docs/pluginDescriptions/> [Zugriff: 29.8.2018]

**Rumsey, Francis. (2001):** Spatial audio. Focal Press

**Samsung (2018):** Samsung VR. 16.8.2018, URL: [https://samsungvr.com/portal/content/content\\_specs](https://samsungvr.com/portal/content/content_specs) [Zugriff: 16.8.2018]

**Schütze, Stephan; Irwin-Schütze, Anna (2018):** New realities in audio: a practical guide for VR, AR, MR and 360 video.

**Sengpiel, Eberhardt** Gegensaetzliche Stereo-Verfahren. , URL: <http://www.sengpielaudio.com/>

**Sengpiel, Eberhardt** Stereo für Lautsprecher und Kopfhörer 1. ,  
URL: <http://www.sengpielaudio.com/>

**Sengpiel, Eberhardt** Stereo für Lautsprecher und Kopfhörer 2. ,  
URL: <http://www.sengpielaudio.com/>

**Sennheiser Electronic GmbH und Co. KG (2018a):** AMBEO® for loudspeakers. 9.8.2018,  
URL: <http://en-us.sennheiser.com/ambeo-blueprints-loudspeaker> [Zugriff: 9.8.2018]

**Sennheiser Electronic GmbH und Co. KG (2018b):** Sennheiser AMBEO A-B FORMAT CONVERTER. 17.8.2018, URL: [http://www.sennheiser-sites.com/responsive-manuals/AMBEO\\_VR\\_MIC/EN/index.html#page/AMBEO%20VR%20MIC/VR\\_MIC\\_04\\_Software\\_EN.4.1.html](http://www.sennheiser-sites.com/responsive-manuals/AMBEO_VR_MIC/EN/index.html#page/AMBEO%20VR%20MIC/VR_MIC_04_Software_EN.4.1.html) [Zugriff: 17.8.2018]

**Shivappa, Shankar; Morrell, Martin; Sen, Deep; Peters, Nils; Salehin, S. M. Akramus (2016):** Efficient, Compelling, and Immersive VR Audio Experience Using Scene Based Audio/Higher Order Ambisonics. Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality, URL: <http://www.aes.org/e-lib/browse.cfm?elib=18493>

**Slater, Mel; Wilbur, Sylvia (1997):** A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. In: *Presence: Teleoperators and Virtual Environments*, Band 6, Ausgabe 6, 12.1997, S. 603–616

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

**Sontacchi, Alois; Höldrich, Robert** sound field reproduction by an improved holophony–ambisonic system. In:

**Sound on Sound Magazine (2018):** Voice, Track & I/O Counts In Pro Tools. 16.8.2018,  
URL: <https://www.soundonsound.com/techniques/voice-track-io-counts-pro-tools> [Zugriff: 16.8.2018]

**Specht, Philip (2018):** Die 50 wichtigsten Themen der Digitalisierung: künstliche Intelligenz, Blockchain, Bitcoin, Virtual Reality und vieles mehr verständlich erklärt. München: REDLINE Verlag

**Steinberg Media Technologies GmbH (2018):** Benutzerhandbuch Nuendo 8.

**Susal, Joel; Krauss, Kurt; Tsingos, Nicolas; Altman, Marcus (2016):** Immersive Audio for VR. Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality, URL: <http://www.aes.org/e-lib/browse.cfm?elib=18512>

**Theile, Günther; Dickreiter, Michael; Graul, Wolfram; Camerer, Florian; Spikofski, Gerhard (2013):** Tonaufnahme und Tonwiedergabe. In: *Handbuch der Tonstudiotechnik*. Berlin, Boston: De Gruyter Saur

**Villegas, Julián (2015):** Locating virtual sound sources at arbitrary distances in real-time binaural reproduction. In: *Virtual Reality*, Band 19, Ausgabe 3–4, 11.2015, S. 201–212

**Vimeo (2017):** Vimeo 360: the new home for immersive storytelling. 3.8.2017,  
URL: <https://vimeo.com/blog/post/introducing-vimeo-360> [Zugriff: 16.8.2018]

**Vimeo (2018):** 360-Grad-Videos hochladen. 27.7.2018,  
URL: <http://help.vimeo.com/hc/de/articles/115001877167-360-Grad-Videos-hochladen> [Zugriff: 27.7.2018]

**VisiSonics Corporation (2015):** VisiSonics 5/64 Audio/Visual Camera. 2015, URL: <http://visisonics.com/products-2/> [Zugriff: 5.8.2018]

**Wallis, Rory; Lee, Hyunkook (2015):** The Effect of Interchannel Time Difference on Localization in Vertical Stereophony. In: *Journal of the Audio Engineering Society*, Band 63, Ausgabe 10, 5.11.2015, S. 767–776

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

**Weiss, Benjamin; Wittek, Helmut (2018):** Schoeps Image Assist. Karlsruhe: Schoeps GmbH, URL: <http://www.ima.schoeps.de/> [Zugriff: 19.7.2018]

**Wenzel, Elizabeth M.; Begault, Durand R.; Godfroy-Cooper, Martine (2018):** Perception of Spatial Sound. In: Roginska, Agnieszka; Geluso, Paul (Hrsg.): *Immersive sound: the art and science of binaural and multi-channel audio*.

**Wikipedia (2017):** Blauertsche Bänder. , URL: [https://de.wikipedia.org/w/index.php?title=Blauertsche\\_B%C3%A4nder&oldid=172310304](https://de.wikipedia.org/w/index.php?title=Blauertsche_B%C3%A4nder&oldid=172310304), Page Version ID: 172310304

**Witmer, Bob G.; Singer, Michael J. (1998):** Measuring Presence in Virtual Environments: A Presence Questionnaire. In: *Presence: Teleoperators and Virtual Environments*, Band 7, Ausgabe 3, 6.1998, S. 225–240

**Wittek, Helmut (2000):** Lokalisationskurven online – Der „Image Assistant. Dokumentation über Prinzip, rechnerische Herkunft und Anwendungen.

**Wittek, Helmut (2015):** „ORTF-3D“: eine Mikrofontechnik für Atmoaufnahmen in 3D-Audio und VR. 20.9.2015, URL: <https://www.hauptmikrofon.de/de/3d/ortf-3d> [Zugriff: 27.8.2018]

**Wittek, Helmut (2018):** Stereo ist schon schön - immer noch., Schoeps Mikroforum 2018, Karlsruhe, 06.04.2018, URL: [https://www.youtube.com/watch?time\\_continue=44&v=GgDn\\_Ts3aw0](https://www.youtube.com/watch?time_continue=44&v=GgDn_Ts3aw0)

**Wittek, Helmut (2018):** ORTF-3D. 22.8.2018, URL: <https://www.hauptmikrofon.de/3d/ortf-3d> [Zugriff: 22.8.2018]

**Wittek, Helmut; Theile, Günther (2016):** Die Anwendung eines stereofonen Mehrkanalverfahrens für 3D-Audio und VR. Köln: 29th Tonmeistertagung

---

## Anhang

|   |   |
|---|---|
| Anhang: Abbildungen Audio-Plugins <i>Audio Ease 360 Pan Suite</i> .....   | A |
| Anhang: Abbildungen Audio-Plugins <i>Noisemakers Ambi HD Bundle</i> ..... | D |
| Anhang: Abbildungen Audio-Plugins <i>Facebook 360 Workstation</i> .....   | G |
| Anhang: Praxisbeispiele.....  | M |

## Anhang: Abbildungen Audio-Plugins *Audio Ease 360 Pan Suite*

Abb. 1 - 360 Pan

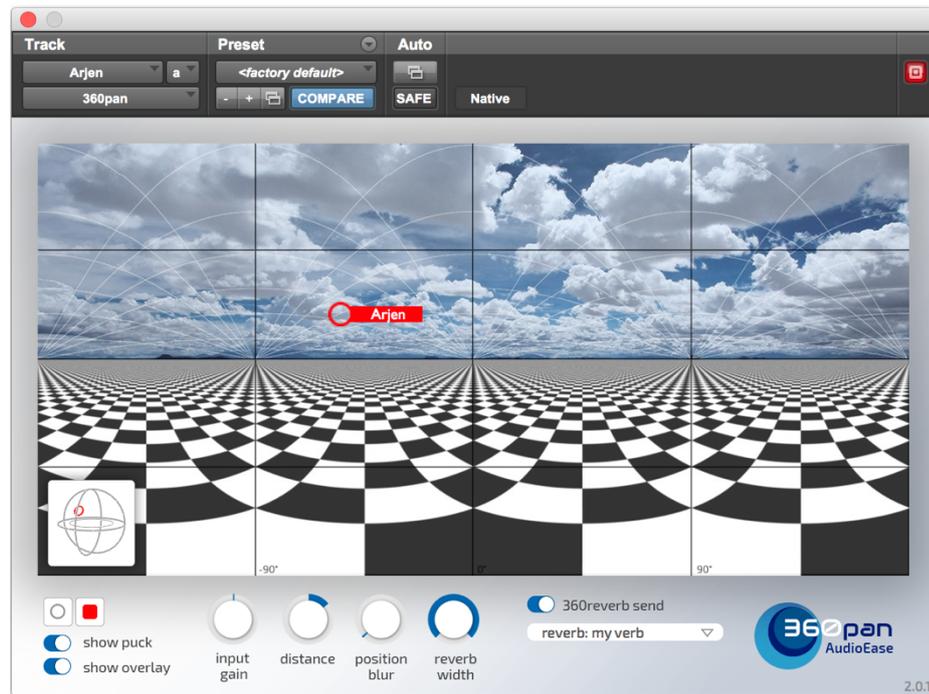


Abb. 2 - 360 Monitor

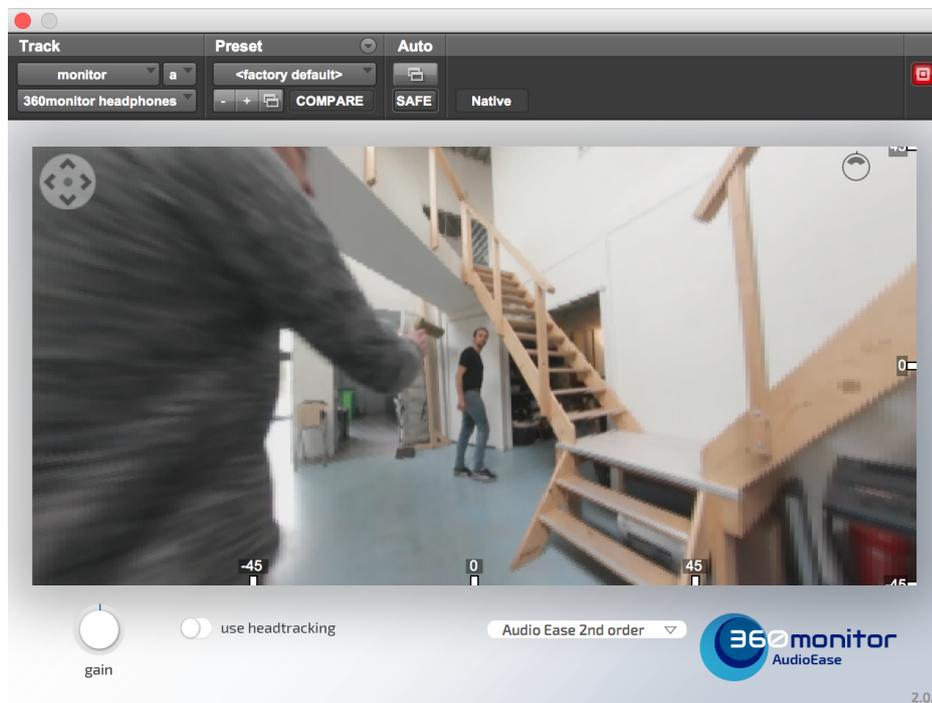


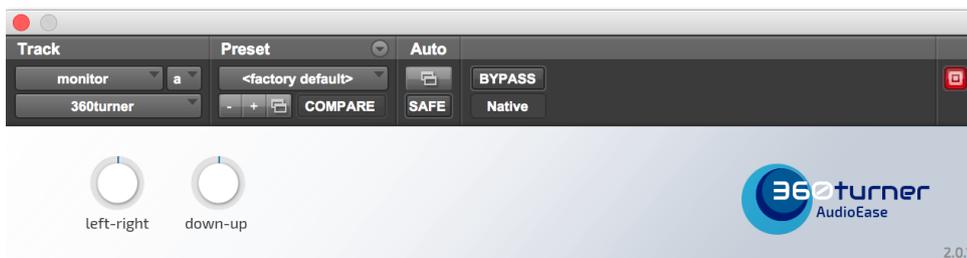
Abb. 3 - Video Overlay 360 Pan



Abb. 4 - 360 Reverb



Abb. 5 - 360 Turner



*Abb. 6 - 360 Converter**Abb. 7 - 360 Radar*

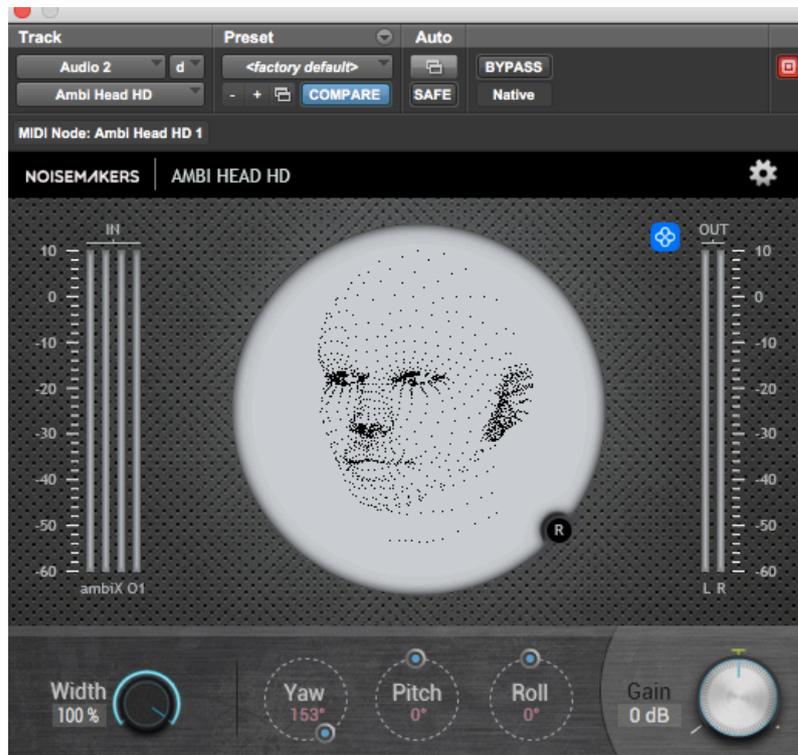
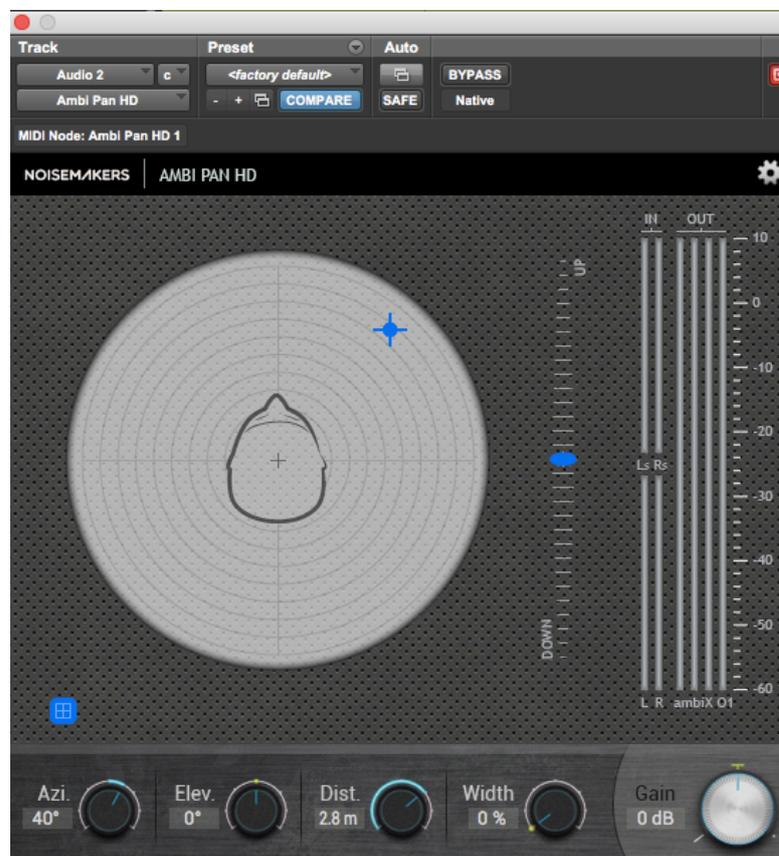
**Anhang: Abbildungen Audio-Plugins *Noisemakers Ambi HD Bundle****Abb. 8 - Ambi Head**Abb. 9 - Ambi Pan Ansicht 1*

Abb. 10 - Ambi Pan Ansicht 2

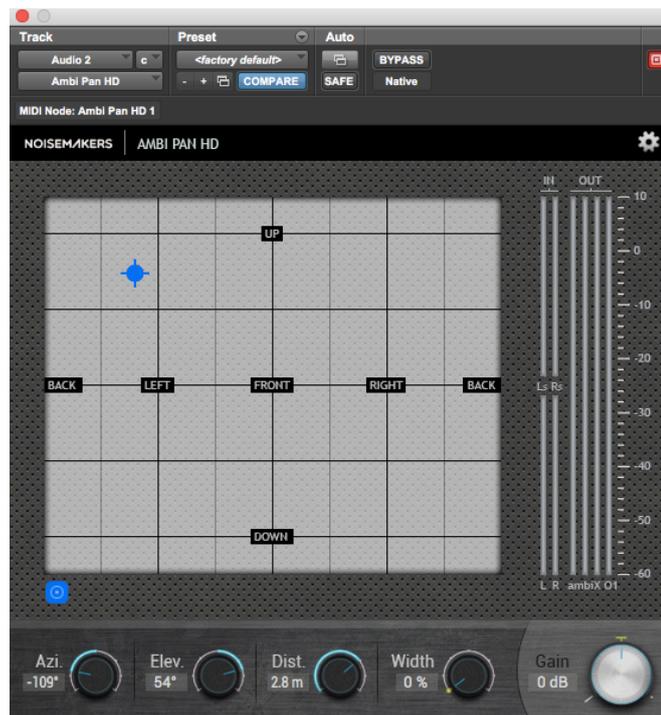


Abb. 11 - Videobasiertes Panning mit Ambi Pan



Abb. 12 - Ambi Converter

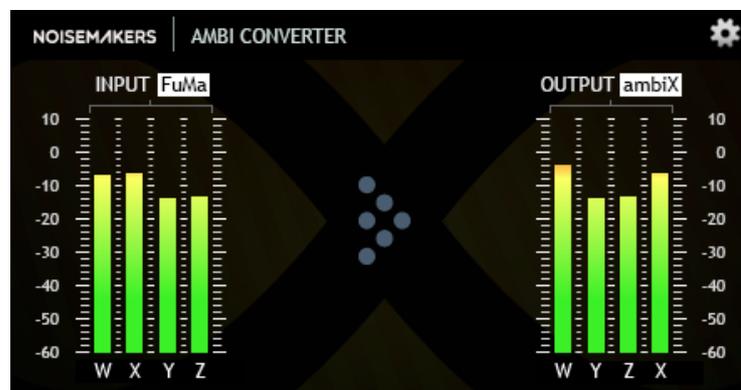


Abb. 13 - Ambi Limiter

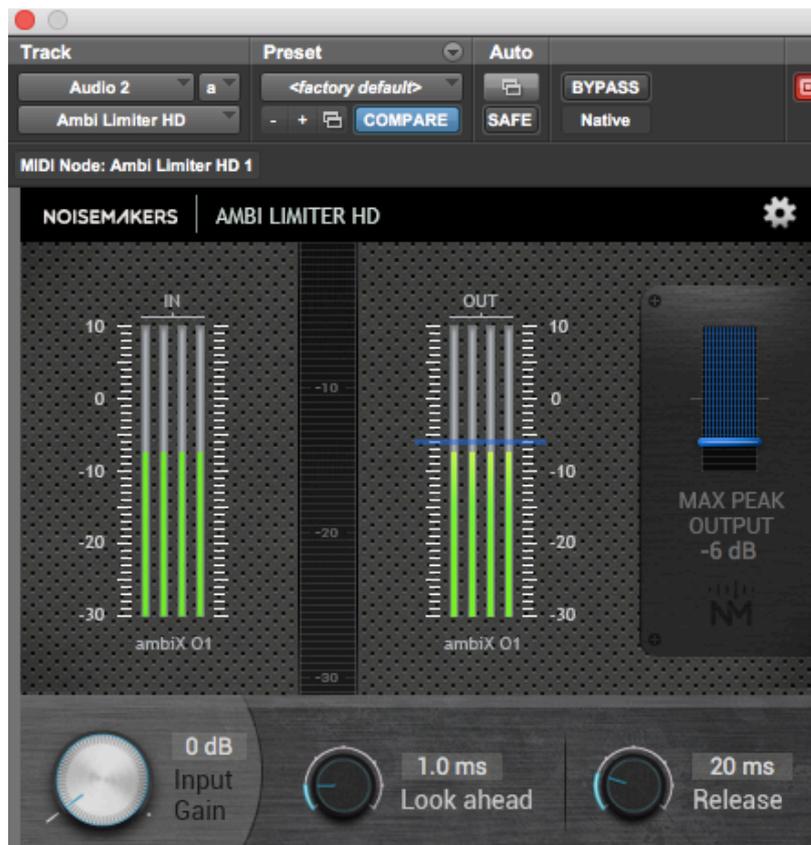
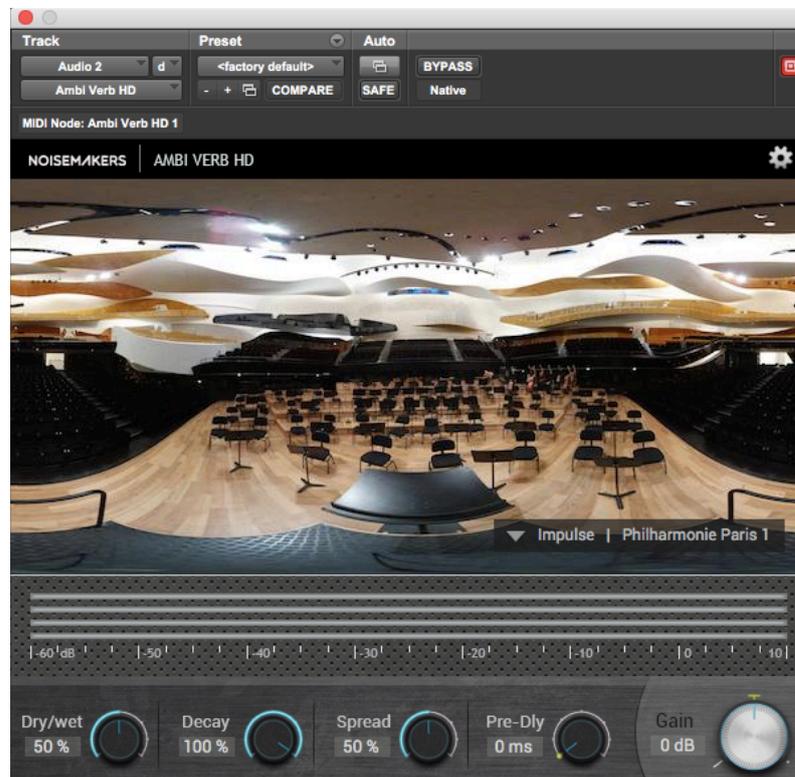


Abb. 14 - Ambi Verb



# Anhang: Abbildungen Audio-Plugins *Facebook 360 Workstation*

Abb. 15 - FB360 Spatializer

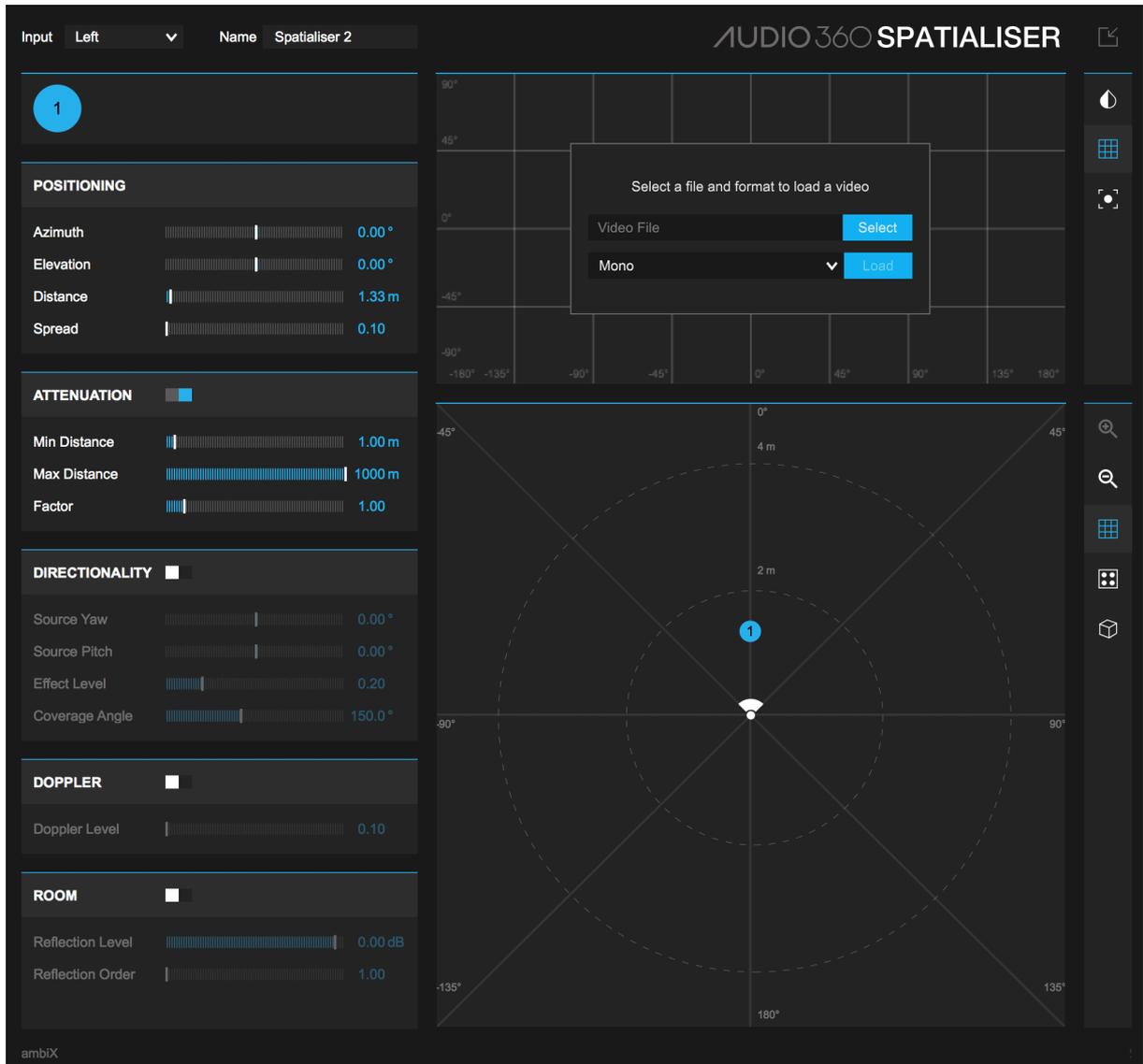


Abb. 16 - FB360 Converter

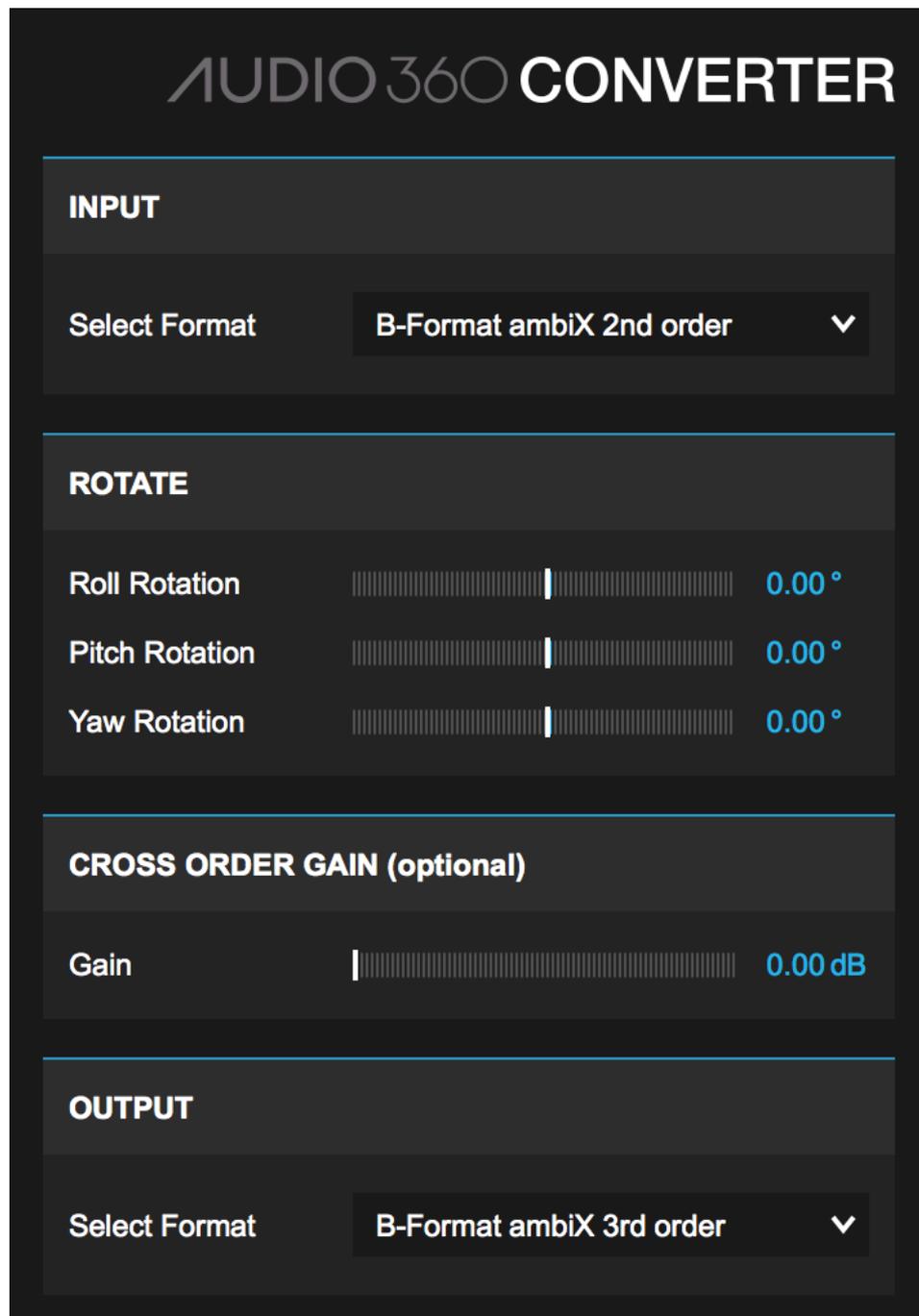


Abb. 17 - FB360 Control

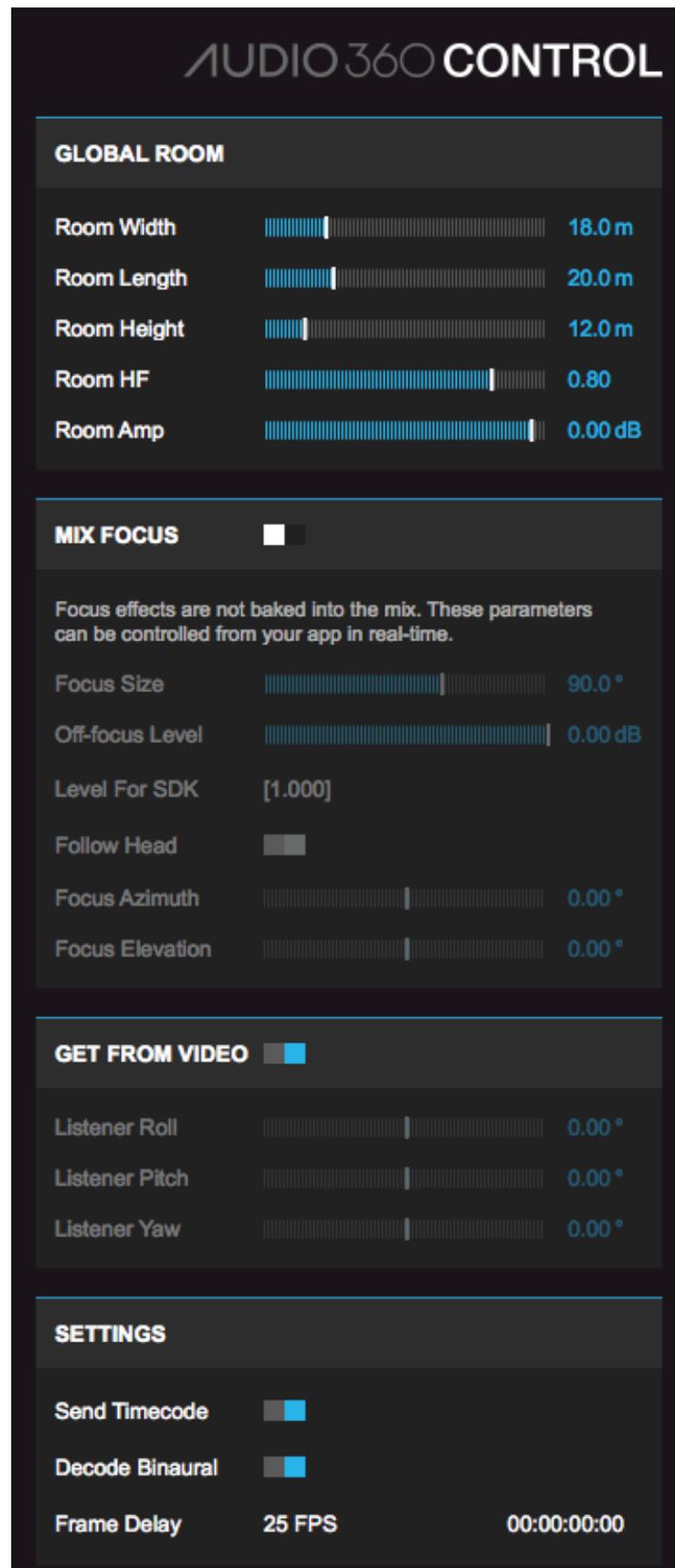


Abb. 18 - FB360 Meter

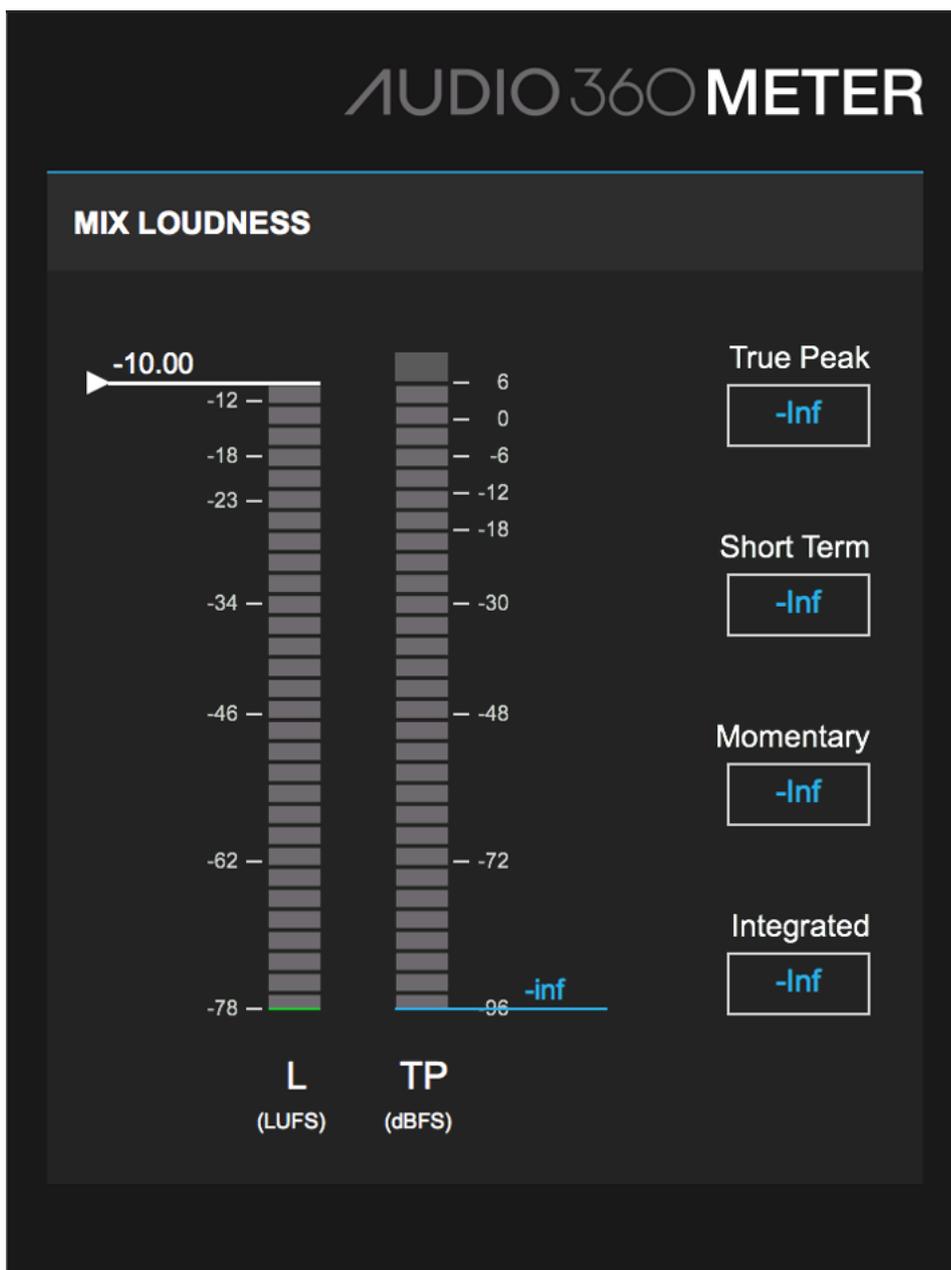


Abb. 19 - FB360 Video Player

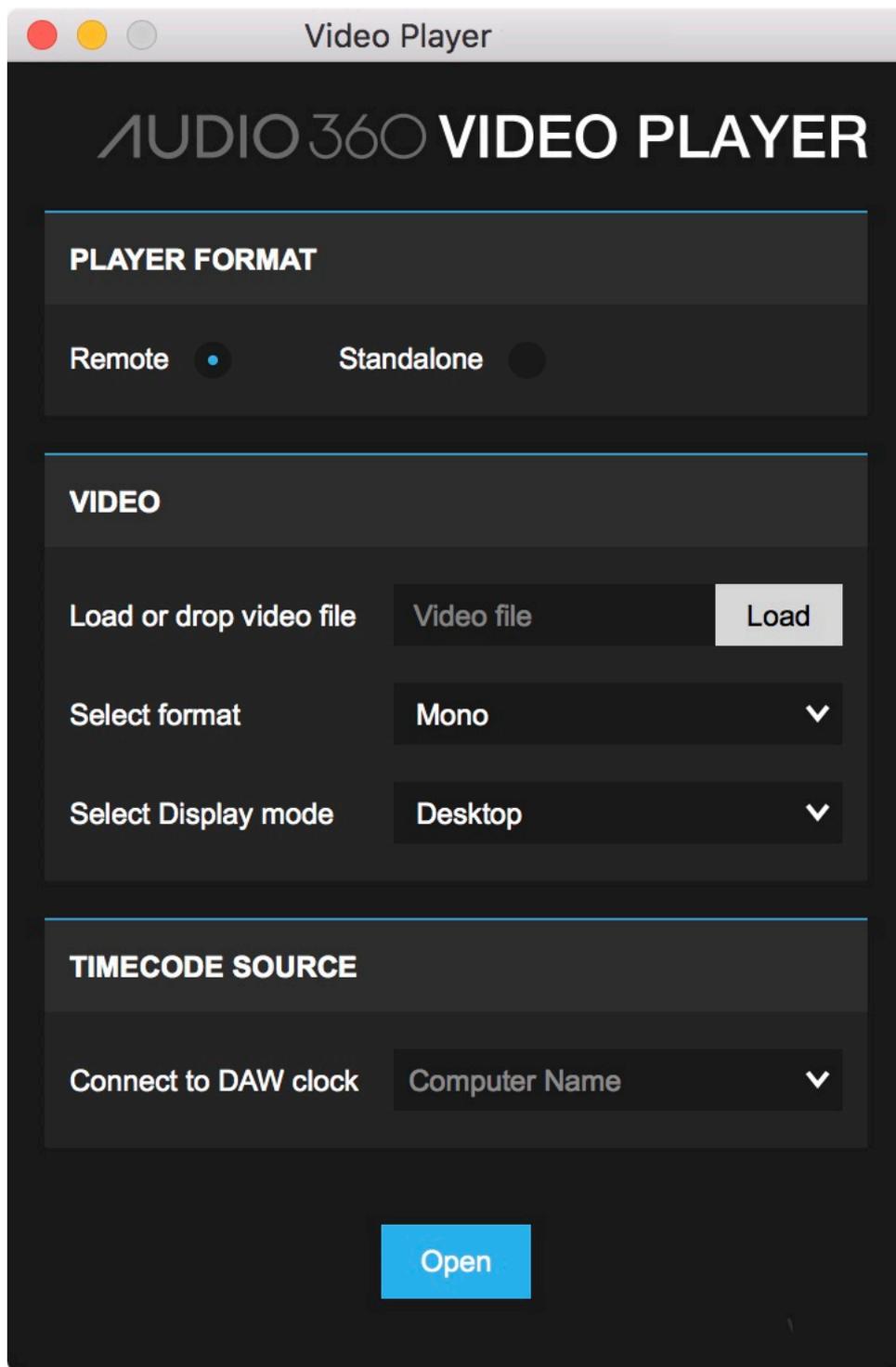
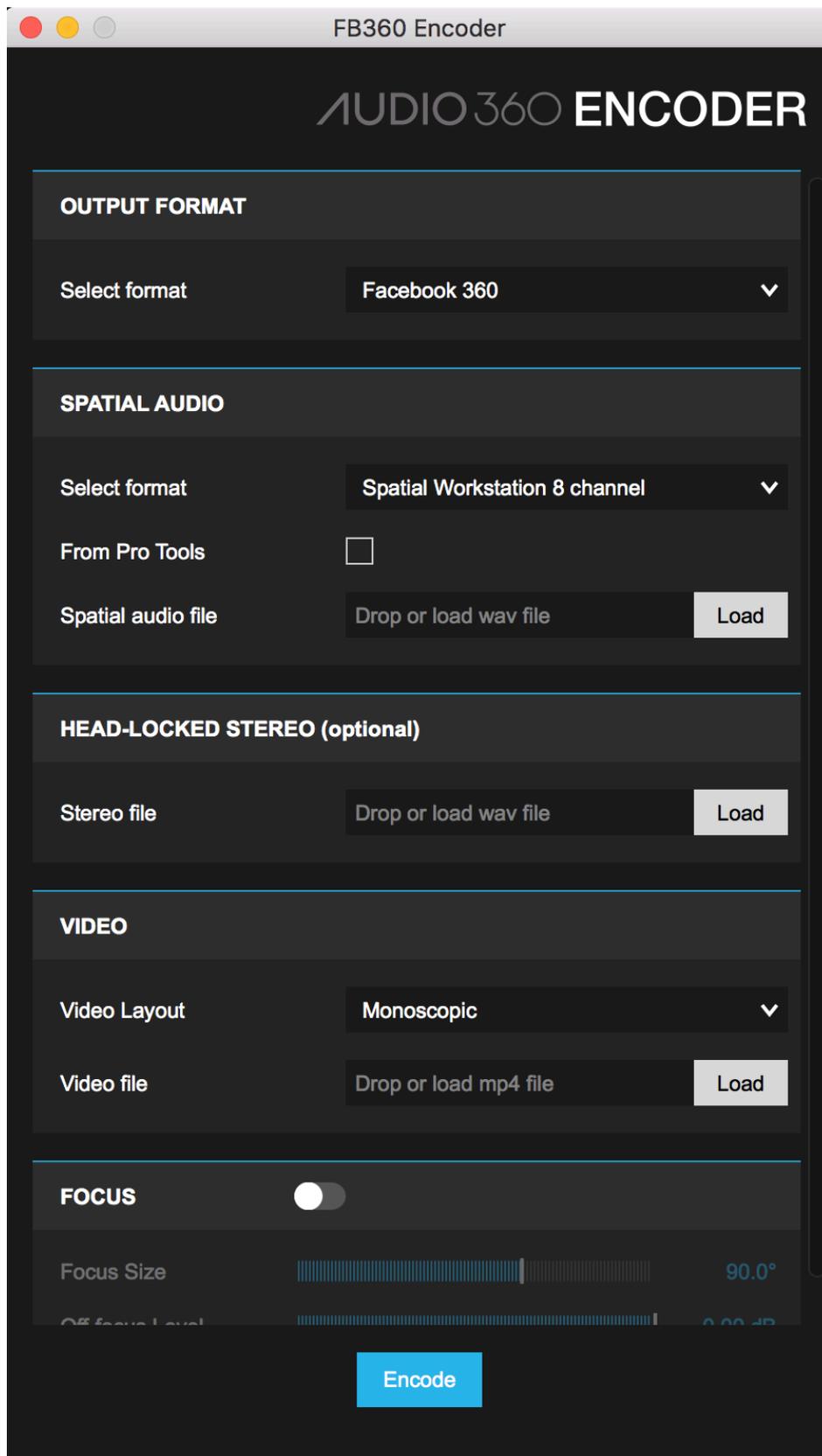


Abb. 20 - FB360 Encoder



---

## **Anhang: Praxisbeispiele**

Der beigefügte Datenträger beinhaltet eine elektronische Version dieser Arbeit sowie Audio- und Videobeispiele zu Kapitel 12. Bitte beachten sie die Hinweise der „Read\_Me“-Datei.

*Datenträger angefügt*