

Ostbayerische Technische Hochschule Amberg-Weiden

Fakultät Elektrotechnik, Medien und Informatik

Medientechnik und Medienproduktion (MP)

Technik und Anwendungen audiovisueller Medien



Masterarbeit

von

Simon L a m c h e

Adäquate Vertonung eines
360°-Promotion-Videos

Appropriate Sound Dubbing of a
360° Promotion Video

Ostbayerische Technische Hochschule Amberg-Weiden

Fakultät Elektrotechnik, Medien und Informatik

Medientechnik und Medienproduktion (MP)

Technik und Anwendungen audiovisueller Medien



Masterarbeit

von

Simon L a m c h e

Adäquate Vertonung eines
360°-Promotion-Videos

Appropriate Sound Dubbing of a
360° Promotion Video

Erstbetreuer:

Zweitbetreuer:

Bearbeitungszeitraum:

Prof. Dr. Dipl.-Ing. Maximilian Kock

M. Eng. Bernd Gerlang-Gradl

29.01.2020 bis 22.09.2020

Fakultät Elektrotechnik, Medien und Informatik

Bestätigung gemäß § 12 APO

Name und Vorname der Studentin/ des Studenten: Lamche, Simon

Studiengang: Medientechnik und Medienproduktion (MP)

Ich bestätige, dass ich die **Masterarbeit** mit dem Titel:

Adäquate Vertonung eines 360°-Promotion-Videos

Appropriate Sound Dubbing of a 360° Promotion Video

selbständig verfasst, noch nicht anderweitig für Prüfungszwecke vorgelegt,
keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt sowie
wörtliche und sinngemäße Zitate als solche gekennzeichnet habe.

Datum: 22.09.2020

Unterschrift: *S. Lamche*

Fakultät Elektrotechnik, Medien und Informatik

Masterarbeit Zusammenfassung

Studentin/ Student (Name, Vorname): Lamche, Simon
Studiengang: Medientechnik und Medienproduktion (MP)

Aufgabensteller, Professor: Prof. Dr. Dipl.-Ing. Maximilian Kock

Durchgeführt in (Firma/ Behörde/ Hochschule): evidentmedia

Betreuer in Firma/ Behörde: M. Eng. Veit Stephan

Ausgabedatum: 29.01.2020

Abgabedatum: 22.09.2020

Titel:

Adäquate Vertonung eines 360°-Promotion-Videos

Appropriate Sound Dubbing of a 360° Promotion Video

Zusammenfassung:

Inhalt der Masterarbeit ist die Konzeption und Umsetzung der 360°-Vertonung eines Promotion-Videos für eine Kommunikations- und Visualisierungsagentur. Welche Besonderheiten in Konzeption und Produktion müssen hierbei besonders im Gegensatz zu einer Stereo-Vertonung beachtet werden? Hilfreich wird dabei sein, dass die Planung und Realisierung des 360°-Videos vom Masteranden mit übernommen wird.

Schlüsselwörter:

360°-Video, 3D-Audio, Ambisonics, Binauraltechnik, Head-Tacking, immersive Audiotechniken, Sounddesign, Spatial Audio, Virtual Reality

Inhaltsverzeichnis

1	Einleitung: Potenzial von 360°-Videos für Unternehmen.....	1
2	360°-Videos	3
2.1	360°-Video-Technik	3
2.2	Bezug zur virtuellen Realität.....	5
3	Grundlagen immersiver Audiotechniken	8
3.1	Räumliches Hören	8
3.2	Binauraltechnik	12
3.3	Kopf- und raumbezügliche Verfahren.....	15
3.4	Räumliche Wiedergabetechniken im Überblick.....	18
4	3D-Audio für 360°-Videos	20
4.1	Ambisonics	21
4.2	Software und Plug-ins	26
4.3	Stand der Technik.....	28
5	Konzeption der Vertonung eines 360°-Promotion-Videos.....	30
5.1	Titel der Masterarbeit und seine Terminologien.....	30
5.2	Besonderheiten bei der Vertonung von 360°-Videos	31
5.3	Rahmenbedingungen	40
5.4	Anforderungsprofil und Planung der 360°-Vertonung.....	42
5.5	Entwicklung der Traumsequenz.....	50
6	Realisierung der Vertonung eines 360°-Promotion-Videos	54
6.1	Equipment.....	54
6.2	Video- und Tonaufnahmen.....	56
6.3	Tonpostproduktion	60
6.4	Aufbereitung für die Zielmedien	64
7	Fazit und Ausblick.....	67
7.1	Technisches Fazit und wissenschaftlicher Ausblick.....	67
7.2	Persönliches Fazit und gestalterischer Ausblick.....	69
8	Glossar.....	73
9	Verzeichnisse.....	77
9.1	Tabellenverzeichnis	77
9.2	Abbildungsverzeichnis	77
9.3	Literaturverzeichnis	78
10	Anhang.....	85
10.1	Danksagung	85
10.2	Drehbuch „Eine Nacht im Büro“	86

1 Einleitung: Potenzial von 360°-Videos für Unternehmen

Anfang 2016 präsentierte die Züricher Hochschule der Künste den ersten fiktionalen 360°-Kurzfilm der Schweiz: In „Gegen die Regeln“ begleiten die Rezipierenden ein junges Pärchen durch ein fesselndes Beziehungsdrama (Cast / Audiovisual Media, 2016). Im gleichen Jahr fiel pünktlich zur Fußball-Europameisterschaft in Frankreich bei „hagebau“ in Österreich der Startschuss zu einem besonderen Gewinnspiel: Die Aufgabe, während einer 360°-Einkaufstour möglichst viele versteckte Fußbälle zu entdecken, konnte nur gelöst werden, wenn sich die Teilnehmenden nach allen Seiten umsahen (hagebau Österreich, 2016).

Anhand der ausgewählten Beispiele lässt sich bereits ohne spezifische Vorkenntnisse die große Stärke von 360°-Videos festmachen: Die detaillierte Darstellung von Situationen und Orten, ohne sich dabei selbst an diesen zu befinden. Im Vergleich zu herkömmlichen Videos kann sich das Publikum in 360°-Videos leichter orientieren, da Positionen und Größenunterschiede in den quasi virtuellen Abbildungen der Realität akkurat wahrgenommen werden können. In dreidimensionalen Darstellungen entsteht sogar der natürliche Eindruck, diese Orte selbst schon einmal besucht zu haben (Orsolits et al., 2020, S. 268 f.). Bei der Empfindung, in einem bestimmten Maß in virtuelle Umgebungen einzutauchen, wird von Immersion gesprochen. Bei der Entwicklung von Computerspielen wird ein möglichst hoher Immersionsgrad angestrebt, indem der Spielende selbst zur Figur in der virtuellen Welt werden soll (Wulff, 2012).

Doch auch für Firmen und Unternehmen bietet der Einsatz immersiver Medien – im Rahmen dieser Arbeit insbesondere 360°-Videos – ein hohes Potenzial: Zum Beispiel zur Ausbildung oder Schulung neuer Mitarbeiter. Mit 360°-Videos lassen sich Szenarien in verschiedenen Varianten immersiv aufbereiten und können somit als Ergänzung zu Planspielen zur Vermittlung von Inhalten beitragen. Weiterhin eignen sich immersive Medien für die eindrucksvolle Präsentation auf Messen. Mit einem 360°-Video lassen sich selbst solche Einblicke in ein Unternehmen zeigen, die mit herkömmlichen Besichtigungen schwer zu realisieren sind. Vor allem, wenn dafür eine spezielle Schutzausrüstung oder Sicherheitsunterweisung nötig wäre (Orsolits et al., 2020, S. 269 ff.). Insgesamt geht der Trend klar zum immersiven Marketing. Seit dem Jahr 2016 setzen immer mehr Unternehmen auf 360°- und Virtual-Reality-Produktionen in der Werbung und erweitern somit ihre Promotion als Marketinginstrument um eine emotionale und erlebbare Facette (Göritz & Züger, 2016).

Doch trotz ihrer großen Beliebtheit schenken viele 360°-Produktionen einem entscheidenden Faktor zu wenig Gehör: Der Möglichkeit, nicht nur visuelle Eindrücke immersiv gestalten zu können, sondern auch auditive. Oft wird das Video, wie in den anfangs aufgezeigten Beispielen, lediglich mit einer Stereospur oder Musik und Off-Stimmen vertont. Dies führt dazu, dass sich der Ton nicht verändert, wenn die Blickrichtung geändert wird. Doch die auditive Komponente birgt weitaus größeres Potenzial. Nicht zuletzt, weil die Realität durch den menschlichen Hörapparat voller immersiver und dreidimensionaler Klangeindrücke ist (Rumsey, 2001, S. 1 f.). Somit stellt sich die Frage: Wie kann das Potenzial immersiver Audiotechniken effektiv zur Vertonung von 360°-Videos genutzt werden?

Diese Arbeit befasst sich mit der adäquaten Vertonung eines 360°-Promotion-Videos für eine Kommunikations- und Visualisierungsagentur. Insbesondere soll dabei untersucht werden, welche Besonderheiten in Konzeption und Produktion im Gegensatz zu einer Stereo-Vertonung beachtet werden müssen. Zu Beginn wird das Medium 360°-Video und der Bezug zur virtuellen Realität näher betrachtet. Anschließend werden die Grundlagen räumlicher Audio-techniken dargelegt und im Weiteren konkret auf 360°-Produktionen bezogen. Im praktischen Teil der Arbeit wird ein Konzept zur Vertonung eines 360°-Promotion-Videos erarbeitet, das den Aspekt des Head-Trackings berücksichtigt und die Möglichkeiten des immersiven Sound-designs diskutiert. Das Konzept wird anschließend anhand der praktischen Umsetzung des Videos in Bezug auf die Tongestaltung evaluiert.

2 360°-Videos

Die Herstellung eines 360°-Videos unterscheidet sich in vielerlei Hinsicht zu herkömmlichen Filmproduktionen: Während der Dreharbeiten fällt z. B. auf, dass die klassische Trennung zwischen vor und hinter der Kamera fehlt. Die Blickrichtung kann bei der Wiedergabe selbst bestimmt werden. Die gewonnene Freiheit des Publikums beeinflusst neben der Bildgestaltung auch Faktoren für den Ton. Deshalb ist es wichtig, zu Beginn die wesentlichen Aspekte der 360°-Videotechnik näher zu betrachten.

2.1 360°-Video-Technik

Ein 360°-Video ist ein vollsphärisches Video, das den Rezipierenden eine vollständige Sicht zu allen Seiten ermöglicht. Da 360°-Videos mit omnidirektionalen Kameras aufgenommen werden, leiten sie sich von Kugelpanoramen ab. Zur Speicherung ist eine equirektanguläre Darstellung zu wählen. Bei der Wiedergabe kann die Blickrichtung dann ausgehend von der Kameraposition frei gewählt werden. Abhängig von den verwendeten Verfahren zur Erstellung können vollsphärische Videos zudem monoskopisch oder stereoskopisch bzw. dreidimensional konsumiert werden (Mazmany, 2019, S. 34 f.). Ausgehend vom „Field of view“ (FOV), dem Sichtfeld in einer realen Situation, ergibt sich im vollsphärischen Video ein Blickfeld, das mit „Field of regard“ bezeichnet wird (Unity 3D, o. J.).

Nach der Entwicklung der omnidirektionalen Panoramafotografie entstanden in den frühen 2000er Jahren die ersten 360°-Videos. Ihren Aufschwung erlebten sie Ende 2015, als das soziale Netzwerk Facebook das Hochladen und Konsumieren von 360°-Material ermöglichte. Zusammen mit der Videoplattform YouTube, die diese Option seit 2016 anbietet, stehen 360°-Videos nun einem breiten Publikum zur Verfügung (Shukla, 2018). Doch auch auf anderen Plattformen lassen sich 360°-Videos publizieren, z. B. bei Vimeo oder Veer TV (Hoppin' Team, 2020).

Kameras zur Aufnahme von 360°-Videos bestehen aus mindestens zwei Linsen, die jeweils einen Winkel von 180° abbilden können. Werden bei der Aufnahme mehr als zwei Kameras verwendet, ergibt sich ein mosaikbasiertes Kamerasystem. Die Kameras werden mit einer entsprechenden Halterung in verschiedene Richtungen positioniert und liefern je einen kleinen Teil der 360°-Produktion (Loew-Albrecht, o. J.). Das Nadir-Bild hat einen besonderen Stellenwert, denn die nach unten positionierte Kamera filmt neben dem Boden auch das Stativ, auf dem sie sich befindet. Aufgrund dieser Besonderheit werden Nadir-Aufnahmen meist in der Postproduktion retuschiert (Loën-Wagner, o. J.).

Die folgende Abbildung zeigt ein 360°-Rig von Freedom 360 für die Aufnahme vollsphärischer Videos mit sechs baugleichen Actioncams der Firma GoPro:



Abbildung 1: 360°-Rig und Actioncams zur Aufnahme von 360°-Videos (Foto: Simon Lamche)

Stitching bezeichnet den Prozess, bei dem nach der Aufnahme aus den einzelnen Videoaufnahmen die 360°-Ansicht erzeugt wird. Eine Software sucht dabei nach übereinstimmenden Punkten zwischen den Bildern. Die Stitching-Naht, englisch als „seam“ bezeichnet, meint die Kanten, die zwangsläufig in den Randbereichen des Materials auftreten. Um ein möglichst gutes Ergebnis zu erzielen, kann das Stitching noch manuell verfeinert werden (Lyu et al., 2019, S. 55 ff.). Im Anschluss besteht die Möglichkeit, das 360°-Video in Programmen wie Adobe Premiere Pro und Adobe After Effects zu schneiden und intensiver nachzubearbeiten.

Das fertige 360°-Video kann auf unterschiedliche Arten wiedergegeben werden: Offline eignet sich dafür der GoPro VR Player, während im Internet veröffentlichte 360°-Produktionen in Webbrowsern wie Mozilla Firefox oder Google Chrome auf den entsprechenden Webseiten konsumiert werden können. Mit Smartphones oder Tablets kann das 360°-Video zudem direkt in der App von Facebook und YouTube abgespielt werden (Referenz Film, o. J.).

Für die Wiedergabe stereoskopischer 360°-Aufnahmen empfiehlt sich die Verwendung eines Head-Mounted-Displays (HMD), das auch unter der Bezeichnung VR-Headset bekannt ist. Optisch ähneln viele HMDs einer Skibrille mit zwei verbauten Linsen und einem Display, durch deren Verwendung ein besonders hoher Immersionsgrad entsteht. Zwei bekannte VR-Headsets sind die Oculus Rift und HTC Vive. Bei einigen Herstellern können Smartphones in eine spezielle Halterung eingespannt und somit als HMDs verwendet werden, wie z. B. mit der Gear VR von Samsung oder dem aus Pappe hergestellten Google Cardboard (Unity 3D, o. J.).

Zusammenfassend kann festgehalten werden, dass ein mit speziellen 360°-Kamera-Systemen aufgezeichnetes vollsphärisches Video nach dem Stitching-Prozess ein „Field of regard“ zu allen Seiten ermöglicht. Liegt das fertige 360°-Video in stereoskopischer Form vor, so kann bei der Wiedergabe mit einem Head-Mounted-Display die Realität vollständig ausgeblendet werden. Auffallend oft wird in Verbindung mit immersiven Medien der Begriff Virtual Reality bzw. VR verwendet. Im folgenden Abschnitt wird aufgezeigt, dass die virtuelle Realität jedoch nicht allein 360°-Videos einschließt, sondern weitläufig verschiedene Techniken beschreibt.

2.2 Bezug zur virtuellen Realität

Ausgehend von den bereits vorgestellten Aspekten der 360°-Video-Technik folgt, dass 360°-Videos und Virtual Reality eng miteinander verbunden sind. Doch bei näherer Betrachtung der Begriffe wird klar, dass für den adäquaten Umgang eine gezielte Differenzierung der Begriffe elementar ist. Tim Bastian Frank ist Geschäftsführer von „VRtual X“ – einer Agentur für immersive Medien mit Sitz in Hamburg. Er beschreibt diese Notwendigkeit folgendermaßen:

„Vielen ist der Unterschied zwischen 360°-Videos und Virtual Reality nicht bekannt. Die beiden Begriffe werden fälschlicherweise oft synonym verwendet. Dabei haben die Techniken gravierende Unterschiede, sowohl in der Produktion und in der Anwendung als auch im Nutzen. Deshalb ist es noch vor den ersten Planungsschritten notwendig, die Vor- und Nachteile für das jeweilige Projektziel abzuwägen.“

–

Tim Bastian Frank nach (Orsolits et al., 2020, S. 264)

Wenn im Allgemeinen von VR gesprochen wird, ist meist die „Game Based Virtual Reality“ gemeint, die virtuelle Welten vollständig mittels 3D-Software über den Computer erzeugt und in Echtzeit berechnet (Jaunt Team, 2018). Selbst wenn die Optik dabei einer fotorealistischen Darstellung entspricht, so bleibt sie technisch betrachtet eine künstlich erschaffene Welt, zum Beispiel in Form eines Computerspiels. Darüber hinaus können sich Nutzer mit einem entsprechenden Systemen frei in der virtuellen Welt bewegen, weshalb bei der computergenerierten VR auch von hochimmersiven Medien gesprochen wird (Orsolits et al., 2020, S. 265 f.).

Im Gegensatz dazu liegen einem 360°-Video filmische Aufnahmen der echten Welt zugrunde. Sie zeigen Orte, die tatsächlich existieren, auch wenn sich das Publikum zum Zeitpunkt der Wiedergabe nicht dort aufhält. 360°-Videos werden deshalb als „Cinematic Virtual Reality“ bezeichnet (Jaunt Team, 2018). Innerhalb der filmischen virtuellen Realität kann zwar die Blickrichtung frei ausgewählt werden, jedoch findet darüber hinaus keine Bewegung statt. Deutlich wird dies, wenn 360°-Videos mit VR-Headsets konsumiert werden. Passt die wahrgenommene virtuelle Welt nicht mehr zur eigenen Körperhaltung, so wird die Immersion gebrochen. Infolgedessen tritt bei einigen Rezipierenden die sog. Motion Sickness ein, die sich am ehesten mit der Seekrankheit vergleichen lässt (Orsolits et al., 2020, S. 264 f.).

Sowohl mit der computergenerierten als auch mit der filmischen virtuellen Realität in stereoskopischer Form lässt sich die reale Umgebung für ihre Betrachter ausblenden. Wesentliche Unterschiede zwischen beiden stellen jedoch die Interaktionsmöglichkeiten dar. Die „Game Based Virtual Reality“ erlaubt seinen Anwendern durch den Einsatz von VR-Headsets und Simulationen in Echtzeit, den Ablauf der Szene spontan zu beeinflussen und selbst mit der künstlichen Umgebung zu interagieren. Diese Möglichkeiten sind in „Cinematic Virtual Reality“ deutlich reduzierter, denn in 360°-Videos wird der Verlauf maßgeblich von der abgefilmten Handlung bestimmt. Durch Blickerkennung und Einsatz von 360°-Hotspots kann zwar ein interaktiver Eindruck vermittelt werden, dabei sind Wechselbeziehungen jedoch auf vorproduzierte 360°-Inhalte beschränkt. In Abhängigkeit der ausgewählten Blickrichtung kann zum Beispiel alternativer Content abgespielt oder das Video pausiert werden, bis sich die Rezipierenden erneut umsehen (Mazmany, 2019, S. 39 ff.).

Grundsätzlich werden vier Rollen unterschieden, in denen es möglich ist, virtuelle Welten zu konsumieren: Ein „Silent Observer“ betrachtet die Szene lediglich passiv, ohne Interaktionen oder eigenen Körper. Während „Silent Participants“ zwar ebenfalls eine passive Haltung einnehmen, sehen sie beim Blick nach unten zumindest einen Körper im Bild. Ein „Acting Observer“ besitzt wiederum keinen Körper, nimmt dafür aber aktiv an der Szene teil und führt Interaktionen durch. Letztlich nehmen nur „Acting Participants“ aktiv an der Handlung teil, interagieren frei in der Szene und besitzen eigene Körper (Technikum Wien Academy, o. J.).

Zur Vollständigkeit sind zwei weitere verwandte Begriffe der virtuellen Realität zu nennen: Augmented Reality (AR) und Mixed Reality (MR). Sie umfassen die Erweiterung und Vermischung der Realität mit computergenerierten Objekten durch Motion-Tracking-Systeme, so dass 3D-Objekte als Bestandteil der echten Welt aufgefasst werden und schließlich mit ihr verschmelzen können (Aichner et al., 2019, S. 1 u. 5 f.). Einige kommerzielle Anbieter richten ihren Fokus bereits weg von Virtual Reality und hin zu Augmented und Mixed Reality aus. Zum Beispiel gehörte „Jaunt VR“ einst zu den Pionieren unter den 360°-Plattformen. Neben eigenen Kamera-Systemen stand eine App im Mittelpunkt, die nach eigenen Ansprüchen der Standard für 360°-Videos auf HMD-Geräten werden sollte. Doch im Jahr 2018 wurde das Unternehmen umstrukturiert. Zwar bleibt Jaunt bestehen, jedoch künftig mit Hauptaugenmerk vor allem auf Mixed-Reality-Streams im Geschäftsbereich (Bastian, 2018).

Im Kontext der vorliegenden Masterarbeit werden 360°-Videos stets als Teil der filmischen virtuellen Realität betrachtet. Die in Kapitel 1 ausgeführten Stärken zeigen deutlich, dass diese Form der immersiven Medien nicht unweigerlich der künstlich erzeugten virtuellen Realität untergeordnet sein müssen. Es empfiehlt sich in jedem Fall, die spezifischen Vor- und Nachteile vor der Produktion abzuwägen – gleichermaßen für die visuelle und auditive Komponente, um dem jeweiligen Medium gerecht zu werden.

3 Grundlagen immersiver Audiotechniken

Ein gesundes Gehör erfüllt fundamentale Aufgaben: Dank der Fähigkeit, Sprache und Geräusche zu verstehen, können Menschen nachhaltig verbal miteinander kommunizieren und zum Beispiel Gefahrensituationen frühzeitig durch deren auditive Anzeichen erkennen. Die beeindruckende Leistung eines physiologischen Hörorgans spiegelt sich darüber hinaus in der Lebensqualität wider: Hören findet in Richtung, Entfernung und Höhe statt. Zusammen mit den visuellen Eindrücken der Augen ergibt sich eine dreidimensionale Wahrnehmung, die besonders die Klangeindrücke erlebbar macht, die den Menschen täglich umgeben. Auch Unterhaltungsmedien wie Musik, Filme und Computerspiele nutzen die Möglichkeiten immersiven Tons, ohne die der Spaßfaktor dieser Angebote wohl deutlich geringer wäre (Breebaart & Faller, 2007, S. 1). Für erfolgreiche Regisseure wie George Lucas hat die Tongestaltung demnach einen entsprechend hohen Stellenwert:

„Sound is 50 percent of the movie going experience, and I’ve always believed audiences are moved and exited by what they hear in my movies at least as much as by what they see.”

–

George Lucas nach (AZ QUOTES, o. J.)

Mit den folgenden Ausführungen soll gezeigt werden, wie der Stereoton über Surround bis zu 3D-Audio weiterentwickelt wurde. Um der Prämisse eines adäquaten Tons für Film- und Videoproduktionen Nachdruck zu verleihen, werden die Aspekte der räumlichen Audiotechniken mit wichtigen Erkenntnissen aus der Hör- und Psychoakustik veranschaulicht.

3.1 Räumliches Hören

Der Prozess des räumlichen Hörens ist binaural. Er funktioniert mit zwei Ohren und umfasst hauptsächlich die akkurate Wahrnehmung der Richtung und Entfernung einer Schallquelle sowie die Empfindung von Tiefe in geschlossenen Räumen (Görne, 2015, S. 126 f.). Dabei werden Schallereignisse anhand von Pegel- und Laufzeitunterschieden und Änderungen im Frequenzspektrum bewertet sowie in Bezug zu beiden Ohren gesetzt (Raffaseder, 2002, S. 110 ff.).

3.1.1 Physiologie des Gehörs

Während die Umwandlung der akustischen Reize in Nervensignale innerhalb der Cochlea im Innenohr stattfindet (Weinzierl, 2008, S. 43 ff.), prägen vorher die Eigenschaften des Außen- und Mittelohrs die eintreffenden Schallereignisse maßgeblich. Sie werden durch die charakteristisch geformte Ohrmuschel gebündelt und somit in den Gehörgang geleitet. Im Mittelohr agiert das schwingende Trommelfell mit den Gehörknöchelchen wie ein Druckempfänger und Signalverstärker. Die Ohrtrompete stellt dabei den nötigen Druckausgleich sicher, indem sie die Paukenhöhle mit dem Rachenraum verbindet (Blauert, 2014, S. 41 ff.).

Beim Eintreffen eines Schallereignisses in den Gehörgang treten Überlagerungen des Direktschalls mit Reflexionen der Ohrmuschel und des Körpers auf. Diese Kammfilter werden über Head-Related Transfer Functions (HRTFs) beschrieben und können, je nach Richtung und Frequenz, Pegelunterschiede von bis zu 20 dB aufweisen (Görne, 2015, S. 111). Meist unterscheiden sich die Amplituden, Frequenzen und Ankunftszeiten in beiden Ohren, sodass aus den Differenzen die Richtung der Schallquelle ermittelt werden kann. Darüber hinaus weist jeder Mensch individuelle Außenohr-Übertragungsfunktionen auf (Møller et al., 1995, S. 300).

Weiterhin gilt, dass das Ohr durch die Eigenschaften des Gehörgangs unterschiedlich sensibel auf bestimmte Frequenzen reagiert. Während die Empfindlichkeit unter 500 Hz und über 10 kHz abnimmt, so ist sie im Bereich von 2 kHz bis 4 kHz besonders hoch. Somit können zwei Signale unterschiedlich laut empfunden werden, selbst wenn sie den gleichen Schalldruckpegel aufweisen (Raffaseder, 2002, S. 90 f.). Ferner wird die empfundene Lautstärke als Lautheit bezeichnet. Für deren Bestimmung muss zudem die Dauer des Signals berücksichtigt werden sowie die Erkenntnis, dass beide Ohren für eine akkurate Beurteilung notwendig sind. Denn die monaurale Hörschwelle liegt um ca. 3 dB höher als die binaurale (Weinzierl, 2008, S. 60 ff.). Simultanmaskierung beschreibt das psychoakustische Phänomen, dass ein lautes Schallereignis die Hörschwelle eines zweiten erhöht, je ähnlicher dabei die Frequenzanteile beider Signale sind. Eine Maskierung kann auch direkt vor oder nach einem lauten akustischen Reiz auftreten, was mit der Trägheit des Gehörs zu erklären ist. Abhängig vom Pegel kann die Erholungszeit bis zu 100 ms betragen (Raffaseder, 2002, S. 93 f.).

3.1.2 Kopfbezogenes Koordinatensystem

Für das binaurale Hören sind insbesondere die interauralen Ohrsignale von Interesse, die sich in beiden Ohren unterscheiden. Die Hörereignisbildung wird im Allgemeinen durch Peilbewegungen des Kopfes und mit visuellen Eindrücken der Augen begünstigt und meist anhand des kopfbezogenen Koordinatensystems dargestellt. Es setzt sich aus der Frontal-, Horizontal- und Medianebene zusammen. Der Ursprung des Koordinatensystems ist die Mitte auf einer Achse zwischen den Eingängen beider Gehörgänge. Über die Beziehung von Azimut, Elevation und Entfernung wird die Position der Schallquelle angegeben (Weinzierl, 2008, S. 87 f.).

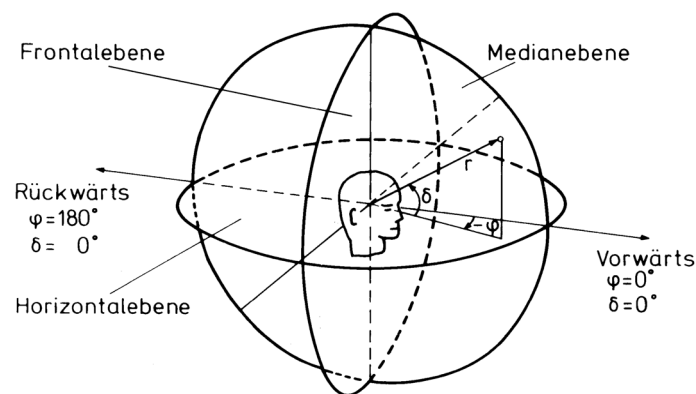


Abbildung 2: Kopfbezogenes Koordinatensystem (Weinzierl, 2008, S. 88)

Das Richtungshören in der Horizontal- und Frontalebene erfolgt sowohl durch Laufzeit- als auch durch Pegelunterschiede, die je nach Frequenz zwischen beiden Ohrsignalen auftreten. Dabei beträgt die kleinste wahrzunehmende Laufzeitänderung etwa 0,03 ms. Zur Bestimmung des Erhebungswinkels in der Medianebene sind nur Veränderungen im Frequenzspektrum zu betrachten. Bei bekannten Signalen tragen spektrale Änderungen darüber hinaus zur Differenzierung zwischen vorne und hinten bei (Dickreiter et al., 2014, S. 127 ff.).

Bei eintreffenden Schallereignissen mit einem Azimut von 0° wird das Hörereignis exakt frontal lokalisiert. In diesem Fall erhalten die Ohren das Signal gleichzeitig. Trifft ein akustischer Reiz seitlich auf den Kopf, erreicht er das eine Ohr früher als das andere. Bei ca. 0,61 ms Laufzeitdifferenz zwischen beiden Ohrsignalen tritt eine vollständige Seitwärtsortung auf. Wenn die Periodendauer einer harmonischen Schwingung jedoch kleiner ist als der Schallumweg, können Fehler in der Lokalisation auftreten. Prallt eine quasistationäre Schwingung mit einer Frequenz ab 1,63 kHz auf den Kopf und ist der Winkel geringer als 90° , kann sie fälschlicherweise vorne geortet werden (Görne, 2015, S. 126 f.).

Interaurale Pegeldifferenzen tragen ebenfalls zur Lokalisierung bei: Verschiedene Ohrschallpegel werden als Unterschied zur eigenen Position verstanden, solange die Wellenlänge kleiner gegenüber den Abmessungen des Kopfes ist. Am einen Ohr entsteht ein Schallschatten, der leiser wahrgenommen wird als der Druckstau auf der anderen Kopfseite. Ein Sinus mit 3 kHz bei einem Einfallswinkel von 90° wird z. B. an der abgewandten Seite um ca. 10 dB gedämpft. Insgesamt funktioniert die exakte örtliche Bestimmung über Pegeldifferenzen gut für Frequenzen oberhalb von 200 Hz. Da insbesondere tiefe Frequenzen nicht reflektiert, sondern um den Kopf gebeugt werden, ist deren Ortbarkeit eingeschränkt (Roginska & Geluso, 2018, S. 14 ff.). Es wird angenommen, dass reine Töne unterhalb von 100 Hz nicht mehr lokalisiert werden können (Görne, 2015, S. 127).

Aus dem sog. Elevationseffekt geht hervor, dass breitbandige Schallereignisse vor allem durch ihre dominierenden spektralen Teile lokalisiert werden. Ohrsignale mit markanten Anteilen bei 8 kHz werden z. B. oberhalb der Horizontalebene verortet (Weinzierl, 2008, S. 95).

Daraus ist zu folgern, dass Frequenzunterschiede die Differenzierung zwischen oben und unten begünstigen – zur Absicherung tragen die bereits erwähnten Peilbewegungen und der Sehsinn bei. Weiterhin ist interessant, dass exakte Angaben über die Entfernung einer Schallquelle im Freifeld schwierig sind. Das Wissen, dass weit entfernte Quellen eher leise und dumpfe Höreignisse erzeugen, lässt zumindest grobe Schätzungen zu. Innerhalb geschlossener Räume ist das Entfernungshören mittels Auswertung der Nachhallstrukturen präziser (Görne, 2004, S. 33). Im Besonderen wird dort neben den primären Schallanteilen auch der Diffusschall betrachtet, der durch Reflexionen an den Wänden entsteht (Raffaseder, 2002, S. 110 ff.).

3.1.3 Geschlossene Räume

Innerhalb des sog. Hallradius überwiegt der Anteil des Direktschalls (Weinzierl, 2008, S. 183). Da der Primärschallpegel mit steigender Entfernung sinkt, der Diffusschallpegel jedoch unabhängig von der Distanz zum Schallereignis gleichbleibt, kann die Entfernung zur Schallquelle ermittelt werden (Dickreiter et al., 2014, S. 133). Für das Richtungshören in geschlossenen Räumen ist der Präzedenzeffekt zu beachten. Er besagt, dass eine Schallquelle stets in der Richtung lokalisiert wird, aus der die erste Wellenfront die Hörenden erreicht. Der Sekundärschallpegel kann dabei laut Haas-Effekt um bis zu 10 dB höher liegen als der des Primärschalls, wenn die Verzögerungszeit höchstens 30 ms beträgt. Bei sehr großen Verzögerungszeiten wird der Sekundärschall als Echo wahrgenommen. Die Echoschwelle ist neben dem relativen Diffusschallpegel auch stark von der jeweiligen Raumbeschaffenheit abhängig (Görne, 2015, S. 128).

Deshalb sei darauf hingewiesen, dass der Klang eines Raumes durch das Zusammenspiel seiner Geometrie- und Oberflächeneigenschaften beeinflusst wird (Görne, 2004, S. 22 ff.). Die komplexen akustischen Strukturen wie Nachhall, Eigenresonanzen oder Echos können bei Bedarf durch raumakustische Werkzeuge korrigiert und angepasst werden – dazu zählen insbesondere Absorber, Reflektoren und Diffusoren (Görne, 2015, S. 90 ff.).

Anhand des Cocktailparty-Effekts kann die Stärke des binauralen Hörens innerhalb geschlossener Räume zusammengefasst werden: In einem Raum, indem sich viele Menschen angeregt unterhalten, kann eine einzelne Person eindeutig herausgehört werden, selbst wenn sie der zuhörenden Person abgewandt ist. Ist dabei jedoch ein Ohr verschlossen, so wird dies deutlich erschwert und die Sprachverständlichkeit reduziert (Blauert, 2014, S. 206). Die Erkenntnisse über die Fähigkeit, mit zwei gesunden Ohren räumlich hören zu können, macht sich die binaurale Aufnahmetechnik zu eigen. Sie wird in den folgenden Abschnitten näher beschrieben.

3.2 Binauraltechnik

Der Ansatz binauraler Systeme ist die Aufnahme beider Ohrsignale und die Berücksichtigung der spezifischen HRTFs. Bei einer unverfälschten Wiedergabe mit Kopfhörern kann ein erstaunlich natürliches Hörerlebnis entstehen, denn die räumlichen Aspekte des originalen Schallereignisses werden dabei relativ gut reproduziert. Zur Erstellung binauraler Tonaufnahmen kann beispielsweise ein Kunstkopf verwendet werden (Møller et al., 1995, S. 300).

Das nachträgliche Überführen von Audioaufnahmen in binaurale Signale ist durch die computergestützte Faltung mit Außenohr-Übertragungsfunktionen für das linke und rechte Ohr ebenfalls möglich (Görne, 2015, S. 131).

3.2.1 Kunstkopfverfahren

Während in den 1930er Jahren bereits Versuche mit einem Prototyp der Kunstkopftechnik durchgeführt wurden, feierte sie 40 Jahre später nach ihrer Weiterentwicklung in Deutschland den großen Durchbruch auf der 29. Internationalen Funkausstellung 1973 in Berlin. Die meisten Kunstköpfe sind in ihren Proportionen und akustischen Eigenschaften einem menschlichen Kopf nachempfunden, teils mit charakteristisch geformten Ohrmuscheln. Am Ende der Gehörgänge befinden sich anstelle der Trommelfelle zwei Druckempfänger (Dickreiter et al., 2014, S. 345 f.). Kunstkopfaufnahmen berücksichtigen die HRTFs direkt während der Aufnahme und durchlaufen den Bereich des Gehörgangs bei der Wiedergabe erneut.

Sie zeichnen sich deshalb im Speziellen durch Außer-Kopf-Lokalisation aus. Die interauralen Frequenz- und Intensitätsdifferenzen ermöglichen somit eine eindrucksvolle akustische Umhüllung des Kopfes. Aufgrund dieser Besonderheiten sind Kunstkopfaufnahmen eher ungeeignet für die Wiedergabe über Lautsprecher (Friedrich, 2008, S. 252).

Als Nachteil der Kunstkopfaufnahmen ist anzuführen, dass das Hörereignis eines frontal eintreffenden Schalls oft irrtümlich hinten lokalisiert wird. Die Ohren erkennen bereits kleinste Änderungen des binauralen Signals, die durch Peilbewegungen des Kopfes korrigiert werden würden. Bei der Reproduktion von Kunstkopfaufnahmen schlägt dies jedoch fehl. Zum einen wegen der individuellen Kopfform des Hörenden, zum anderen und entscheidender, da sich das Hörereignis bei der Drehung mitbewegen würde (Dickreiter et al., 2014, S. 346 f.). Mit anderen Worten ist die Hörimmersion abhängig von der Ausrichtung des Kunstkopfes.

Die folgende Abbildung zeigt den bekannten „KU 100“ der Firma Neumann:



Abbildung 3: Kunstkopf Neumann „KU 100“ (Foto: Simon Lamche)

3.2.2 Head-Tracking und Binauralsynthese

Der Sammelbegriff Head-Tracking beschreibt alle unterschiedlichen Methoden zur Erfassung der Kopf- und Halsposition. Sie dienen der Überwachung von Interaktionsmöglichkeiten in VR-Umgebungen mittels HMDs (Unity 3D, o. J.) und auditiv besonders der Steuerung der Richtung eines Hörereignisses in Abhängigkeit der Bewegungen des Kopfes sowie den resultierenden spektralen Änderungen. Die Realisierung erfolgt durch Sensoren oder Kameras, die zum Kopf hin ausgerichtet sind (Dickreiter et al., 2014, S. 346).

Weitverbreitet sind Trägheitssensoren in Form von Gyroskopen in mobilen Endgeräten aufgrund geringer Kosten und niedriger Verzögerungszeiten. Diese Technik neigt jedoch zu Ungenauigkeiten, die zur präzisen Positionsbestimmung kompensiert werden müssen. Während optische Trackingverfahren die visuellen Informationen der Kameras für Aussagen über die entsprechenden Positionen nutzen, ermitteln magnetische Sensoren die Position über Änderungen eines Magnetfelds, die sich durch verschiedene Positionen im Raum ergeben. Akustische Sensoren hingegen werten die signifikante Zeitspanne aus, in der ein akustisches Signal den Empfänger erreicht. Je mehr Empfänger dabei am Kopf befestigt sind, desto präziser erfolgt die Bestimmung der Position. Diese Informationen werden dafür genutzt, die spezifischen HRTFs anzupassen. Die dynamische Binauralsynthese erfolgt mit entsprechend hoher Rechenleistung in Echtzeit und stellt einen erheblichen Vorteil für die Gestaltung immersiver Umgebungen dar. Die kontinuierliche Anpassung der HRTFs ist auch für Schallquellen unerlässlich, die sich dynamisch um die Hörenden bewegen (Roginska & Geluso, 2018, S. 104 f.).

Binauralsynthese eignet sich, um Computermodelle von Klangräumen zu erstellen. Die notwendigen binauralen Impulsantworten des Raumes (BRIR) lassen sich entweder durch Kunstkopfaufnahmen messen oder sind durch ein spezielles Strahlenmodell simulierbar, indem die berechneten Schallereignisse mit passenden HRTFs gewichtet werden. Die Summe aller ermittelten Werte ergibt die BRIR (Weinzierl, 2008, S. 679 f.). Binaurale und transaurale Prozesse tragen ferner zur Bildung imaginärer Lautsprecher bei. Mehrkanalton kann z. B. durch kopfbezügliche Auralisation und Head-Tracking für die Kopfhörerwiedergabe aufbereitet werden. Zur Reproduktion von virtuellem Surround ist somit keine physikalische Lautsprecheranordnung notwendig (Rumsey, 2001, S. 75 ff.). Bezogen auf die in Abschnitt 3.2.1 benannten Schwächen der Kunstkopfaufnahmen müsste zur Anpassung der Kopfposition auf ein akustisches Head-Tracking zurückgegriffen werden. In diesem Fall wäre die räumliche Abbildung ohne Einschränkung des Frontalbereichs gewährleistet (Dickreiter et al., 2014, S. 347).

Um dynamisches Head-Tracking zu ermöglichen, müsste der Kunstkopf jedoch omnidirektional gestaltet sein – also rundum Ohren besitzen. Aus diesem Ansatz entstanden das binaurale Motion-Tracking-Verfahren sowie das omni-binaurale Mikrofonsystem (Roginska & Geluso, 2018, S. 98 f.). Zusammenfassend ist festzuhalten, dass die Binauraltechnik durch Einbeziehung der spezifischen HRTFs und Head-Tracking eine wegweisende Rolle innerhalb der räumlichen Verfahren einnimmt.

3.3 Kopf- und raumbezügliche Verfahren

Jedoch zählen nicht nur Kunstkopfaufnahmen zu den immersiven Audiotechniken. Bevor in den folgenden Abschnitten weitere Tontechniken zur Aufnahme und Wiedergabe im Fokus stehen, ist eine grundsätzliche Unterscheidung zwischen kopfbezüglichen und raumbezüglichen Verfahren vorzunehmen.

3.3.1 Arten der Stereophonie

Binaurale Tonaufnahmen zählen zur kopfbezüglichen Stereophonie. Da sie wie natürliche Hörsituationen gleichermaßen auf Laufzeit- und Pegeldifferenzen beruhen und die Außenohr-Übertragungsfunktionen enthalten, werden kopfbezügliche Aufnahmen meist über Kopfhörer wiedergegeben (Görne, 2015, S. 130 f.). Für die korrekte Reproduktion sollten binaurale Mikrofone und Kopfhörer für das Diffusfeld entzerrt sein. Die spektralen Färbungen einer natürlichen Hörsituation werden in diesem Fall nicht nur für frontal eintreffende Signale korrekt nachgebildet, sondern auch für seitlich einfallende Schallereignisse (Blauert, 2014, S. 1Na93 f.). Die Wiedergabe binauraler Aufnahmen über Lautsprecher ist hingegen problematisch, da sichergestellt werden muss, dass sich die interauralen Signale der binauralen Aufnahme für linkes und rechtes Ohr nicht vermischen. Sonst treten durch die gemischte Form störende Klangverfälschungen auf. Durch den Einsatz von transauralen Filtern kann dies jedoch verhindert werden (Dickreiter et al., 2014, S. 348).

Tonaufnahmen, die vorrangig zur Wiedergabe mit Lautsprechern gedacht sind, werden der raumbezüglichen Stereophonie zugeordnet. Dabei werden die Lautsprechersignale überlagert und bei der Wiedergabe mit den Eigenschaften des Raumes an den Ohren gefiltert. Durch die Überlagerung wird das Hörereignis als Phantomschallquelle an einem Punkt zwischen den Lautsprechern gebildet (Blauert, 2014, S. 163).

Ist die Verzögerungszeit eines Signals höher als 0,01 ms, so wandert die Phantomschallquelle in Richtung des ersten akustischen Reizes. Bei einer Verzögerungszeit von ca. 1 ms befindet sie sich ganz im Lautsprecher, von dem das Signal am frühesten wiedergegeben wurde. Aufgrund der relevanten Laufzeitunterschiede wird diese Methode als Laufzeitstereofonie bezeichnet. Es findet darüber hinaus eine Bewegung der Phantomschallquelle statt, wenn sich die Signalpegel verändern, aber die Summe aller Pegel gleichbleibt. Bei der Intensitätsstereofonie ist ab einem Pegelunterschied von ca. 0,6 dB eine entsprechende Bewegung festzustellen. Differenzen von mehr als 20 dB haben jedoch keinen Einfluss mehr auf die Bildung der Phantomschallquelle. Die Mischform aus dem Laufzeit- und Intensitätsverfahren wird im Allgemeinen als Äquivalenzstereofonie bezeichnet (Görne, 2015, S. 129 f.).

Die Reproduktion raumbezoglicher Aufnahmen über Kopfhörer führt zur Im-Kopf-Lokalisation (IKL), bei der das Hörereignis innerhalb des Kopfes zwischen den Ohren gebildet wird. Zur Erklärung ist anzuführen, dass bei der Wiedergabe mittels Kopfhörer die Filterwirkung des Außenohrs umgangen wird und lautsprecherstereofone Aufnahmen die HRTFs nicht wie binaurale Signale bereits berücksichtigen. Folglich sind die interauralen Reize zu ähnlich und werden deshalb im Kopf abgebildet. Eine häufige Begleiterscheinung der IKL ist ein unangenehmes Druckgefühl auf beiden Ohren (Roginska & Geluso, 2018, S. 106 f.).

3.3.2 Hauptmikrofonsysteme

Wurden separate Schallquellen im Einzelmikrofonverfahren mit gerichteten Mikrofonen in Mono aufgezeichnet, so lässt sich in einer digitalen Audioworkstation (DAW) das Stereobild mit Panorama-Potentiometern erzeugen (Dickreiter et al., 2014, S. 273). Im Folgenden wird verdeutlicht, dass Stereoaufnahmen alternativ durch spezielle Hauptmikrofon-Anordnungen gewonnen werden können:

Für die Intensitätsstereofonie sind Koinzidenzverfahren anzuführen, bei denen die Stereoinformationen nur durch Pegeldifferenzen gebildet werden. Da die Signale in ihrer Phasenlage gleich sind, können sie problemlos zu einem Monosignal summiert werden. Koinzidenzaufnahmen überzeugen durch eine sehr gute Richtungsabbildung (Görne, 2004, S. 102 ff.). Für die XY-Aufstellung werden meist Kleinmembran-Kondensatormikrofone mit Nierencharakteristik verwendet, die mit einem Versatzwinkel von 45° bis 66° dicht übereinander positioniert werden müssen. Bei falscher Anordnung kann das Stereopanorama stark leiden. Das berühmte „Blumlein Paar“ variiert die XY-Aufstellung durch die Überkreuzung zweier Mikrofone in

Achtercharakteristik im 45°-Winkel (Friedrich, 2008, S. 244 ff.). Im direkten Vergleich liefert es räumliche Aufnahmen, jedoch mit Phasendrehung der negativen Signalanteile (Ederhof, 2006, S. 188 ff.). In der MS-Anordnung wird das Panorama durch Kombinieren eines Mitten- und Seitensignals gebildet – am einfachsten mit zwei übereinander angeordneten Mikrofonen, je eins in Kugel- und Achtercharakteristik. Wichtig ist, dass die positive Seite der Acht links positioniert ist, sonst wird bei der Überführung mittels Stereomatrix in das Wiedergabeformat der linke und rechte Kanal vertauscht. Vorteilhaft ist aber, dass die Stereobreite durch das Mischungsverhältnis der Signale nachträglich angepasst werden kann (Görne, 2004, S. 106 ff.).

Durch Laufzeitstereofonie kann im Allgemeinen die räumliche Tiefe gut abgebildet werden. Aufgrund unterschiedlicher Phasenlagen ist eine Monomischung jedoch nur eingeschränkt möglich (Görne, 2004, S. 114). Bei der AB-Aufstellung werden zwei Mikrofone mit Kugelcharakteristik im Abstand von etwa 20 cm bis 1,5 m parallel positioniert. Je größer der Abstand, desto mehr Raumanteil ist in der Aufnahme enthalten – allerdings leidet darunter besonders die Richtungsdarstellung (Friedrich, 2008, S. 249). Als wichtiger Vertreter der Äquivalenzstereofonie ist das ORTF-Verfahren zu nennen, welches sowohl Pegel- als auch Laufzeitunterschiede aufzeichnet. Zwei Mikrofone mit Nierencharakteristik werden im Abstand von 17 cm und einem Öffnungswinkel von 110° positioniert (Ederhof, 2006, S. 208 f.).

Das Kugelflächenmikrofon (KFM) hingegen verfolgt einen eher kopfbezüglichen Ansatz: Zwei Grenzflächenmikrofone sind dort mit ihrer Membran bündig an den Seiten einer Kugel mit Durchmesser 18 cm oder 20 cm verbaut. Da die Kugel in Anlehnung an einen Kopf als Trennkörper fungiert, orientiert sich das KFM an den Reflexionen und Schallbeugungen einer natürlichen Hörsituation (Dickreiter et al., 2014, S. 265 ff.).

Mit Aufkommen des mehrkanalstereofonen Tons wurden die bestehenden Stereoprinzipien um zusätzliche Kanäle erweitert, damit die beobachtende Rolle der Hörenden zunehmend erlebbarer werden konnte (Friedrich, 2008, S. 253). Folgende Beispiele symbolisieren den Ausbau bekannter Hauptmikrofon-Verfahren: Zum einen ermöglicht die Kombination zweier MS-Systeme die Aufnahme des Surround-Tons, wobei das M-Signal hierfür durch je ein nach vorne und hinten ausgerichtetes Mikrofon mit Nierencharakteristik repräsentiert wird. Die hierfür notwendigen Surround-Eigenschaften können beim Doppel-MS-Verfahren aus dem Zusammenspiel zwischen M- und S-Anteilen gewonnen werden (Weinzierl, 2008, S. 593 f.). Ebenfalls gelingt es mit gemischten Verfahren, etwa durch Kombination von fünf Nierenmikrofonen in unterschiedlichen Winkeln und idealer Nierenanordnung, Aufnahmen für Mehrkanalanwendungen zu erstellen (Friedrich, 2008, S. 254 f.).

Ferner hat sich das IRT-Kreuz etabliert, bei dem vier Mikrofone mit Nierencharakteristik nach außen an den Eckpunkten eines imaginären Quadrats mit der Kantenlänge zwischen 20 cm und 25 cm positioniert sind. Diese Aufstellung verhält sich ähnlich wie die Doppelung einer ORTF-Anordnung mit Ausrichtung nach vorne und hinten. Es eignet sich besonders für Surround-Atmosphäre (Ederhof, 2006, S. 225 f.). Im kopfbezüglichen Ansatz können mehrkanalstereofone Signale zum Beispiel durch die Ergänzung des KFM's um zwei Mikrofone mit Achtercharakteristik gewonnen werden. Die benötigten Informationen für eine entsprechende Surround-Anordnung ergeben sich durch Matrizenbildung, wobei der Center-Kanal in der einfachsten Variante des Surround-Kugelflächenmikrofons außer Acht gelassen wird (Dickreiter et al., 2014, S. 300 f.).

3.4 Räumliche Wiedergabetechniken im Überblick

Dank der binauralen Technik steht im kopfbezüglichen Kontext ein qualifiziertes Mittel zur Reproduktion immersiver Tonaufnahmen mittels Kopfhörer zur Verfügung. Doch auch mit den raumbezüglichen Lautsprecheraufstellungen wurde früh experimentiert, um sie hinsichtlich ihrer Räumlichkeit weiterzuentwickeln. Der britische Ingenieur Alan Blumlein meldete im Jahr 1931 ein Patent an, das oft als Geburtsstunde dessen betrachtet wird, was im Allgemeinen die Stereophonie beschreibt. Als Stereo in den 50er Jahren kommerziell erfolgreich wurde, entsprach das Forschungsziel bereits der Entwicklung mehrkanaliger Techniken. Während die „Quadrofonie von 1968“ erfolglos blieb, öffneten Verfahren wie Cinemascope und Dolby Stereo bzw. Dolby Surround das Tor zum breiten Markt. Die vierkanalige Aufstellung mit Links, Mitte, Rechts sowie einem hinteren Kanal wurde Ende der 1970er Jahre zu einer sechskanaligen Methode weiterentwickelt (Roginska & Geluso, 2018, S. 50 ff. u. 186). Der folgende Überblick bildet insbesondere eine Brücke zur dreidimensionalen Tondarstellung:

Nachdem digitale Formate wie DTS (Digital Theater System) und Dolby Digital bereits im Kino Einzug hielten, fand in den 1990er Jahren das 5.1 Surround auch im häuslichen Umfeld Anklang: Filme konnten fortan über entsprechende Anlagen im Wohnzimmer mit Audio von Links, Mitte, Rechts sowie einem linken und rechten Surround-Kanal und dem optionalen „LFE-Kanal“ für tiefe Frequenzen konsumiert werden. (Görne, 2015, S. 307 f.). Diese kanalbasierte Methode wurde wegen des Kino-Breitbildformats um zwei Kanäle zur 7.1-Aufstellung erweitert (Roginska & Geluso, 2018, S. 190). Es sei angemerkt, dass der Begriff „Kanal“ in diesem Fall auch für eine Lautsprechergruppe stehen kann. Trotzdem funktioniert die Bildung der Phantomschallquellen bei dieser Technik nur im sog. Sweet Spot korrekt, also an einem

bestimmten Punkt wie bei der optimalen Stereohörposition an der Spitze eines gleichseitigen Dreiecks (Dickreiter et al., 2014, S. 230). Folglich sind nur die mittleren Plätze eines Kinosaals optimal nutzbar. Darüber hinaus ist das Hörerlebnis bei herkömmlichen Surround-Verfahren auf die Horizontalebene beschränkt (Roginska & Geluso, 2018, S. 222).

Auro-3D schließt als kanalbasierte Technik die Medianebene mit ein. Ausgehend von einer herkömmlichen 5.1- oder 7.1-Anordnung werden weitere Lautsprecher über der bestehenden Ebene positioniert. Optional ist ein weiterer Kanal vorgesehen, der direkt von der Decke abwärts strahlt. Er wird oft als „voice of god“ bezeichnet. Auro-3D kann durch Höhendarstellung als dreidimensionales Audiosystem bezeichnet werden (Roginska & Geluso, 2018, S. 227 f.).

Das „Sweet Spot Problem“ wird aber erst durch den objektbasierten Ansatz gelöst: Klangobjekte werden als Audio-Wellenformen mit zugeordneten räumlichen Parametern verstanden. Sie dienen der korrekten Umsetzung von Schallereignissen in passende Lautsprechersignale. Die künstlerische Herausforderung liegt in der Erzeugung komplexer Tonwelten aus einzelnen Klangobjekten (Roginska & Geluso, 2018, S. 244 f.). Eine Vielzahl von Lautsprechern kann hier flexibel angesteuert werden, ohne dass sie diskret bestimmten Kanälen zugeordnet sind (Overschmidt & Schröder, 2013, S. 329). Das 2012 eingeführte Dolby Atmos stellt eine Mischform zwischen kanal- und objektbasierter Methode dar. Abhängig davon, für welches Objekt welcher Ansatz verwendet wird, können bis zu 128 Klangobjekte gleichzeitig wiedergegeben werden. Objektbasierte Elemente müssen jedoch in Echtzeit oder zuvor für den jeweiligen Saal berechnet werden. Außerdem arbeitet das Soundsystem vor allem mit der verlustbehafteten AC-4-Kompression, die bereits bei Dolby Digital zum Einsatz kam (Ziemer, 2019).

Eine weitere objektbasierte Technik ist Iosono, die auf der aus den 1980er Jahren bekannten Wellenfeldsynthese beruht. Unter Einsatz des sog. Huygens-Fresnel-Prinzips werden viele Lautsprechersignale miteinander überlagert. So dient jeder Punkt einer Wellenfront als Ursprung einer neuen kugelförmigen Elementarwelle. Aufgrund der schwer zu realisierenden Lautsprecheranordnung wurde die Wellenfeldsynthese in der Umsetzung nur auf die Horizontalebene beschränkt. Durch intensive Forschung und moderne technische Möglichkeiten entstand um das Jahr 2010 das Iosono-Verfahren, das durch den objektorientierten Ansatz ein vollständig dreidimensionales Hörerlebnis anbietet. Hierbei werden die immersiven Eigenschaften einer virtuellen Klangszene in das Format eingebettet und können mit einem Spatial Audio Prozessor flexibel an reale Lautsprecheraufstellungen angepasst werden (Overschmidt & Schröder, 2013, S. 340 ff. u. 353 ff.).

4 3D-Audio für 360°-Videos

Kernfrage: Nach welchem Ansatz soll der Ton bei neuen filmischen Darstellungsformen wie 360°-Videos behandelt werden? Aus den grundlegenden Ausführungen zu räumlichen Audio-techniken kann gefolgert werden, dass die auditive Komponente einer 360°-Produktion mit einem rein zweikanaligen Stereoton nicht erlebbar genug wäre. Zwar lösen Techniken wie 5.1- und 7.1-Surround in der Horizontalen besser auf, aber lassen dabei den immersiven Faktor der Höhe außer Acht. An dieser Stelle setzen die dreidimensionalen Audiotechniken an, durch die eine adäquate räumliche Reproduktion im kopfbezogenen Koordinatensystem möglich ist.

Räumliche Audiotechniken werden oft unter dem Begriff „Spatial Audio“ zusammengefasst. Je nachdem, ob Höhenunterschiede berücksichtigt werden, kann damit aber ein zwei- oder dreidimensionaler Ton gemeint sein (Overschmidt & Schröder, 2013, S. 335).

Im Weiteren wird „Immersive Audio“ unter Einbeziehen der Medianebene als „3D-Audio“ bezeichnet und schließt diejenigen Techniken ein, die Schallereignisse wie beim natürlichen Hören von allen Seiten berücksichtigen. Dazu zählt auch die Binauraltechnik, die eine Kopfhörerwiedergabe von 3D-Audioinhalten ermöglicht und im Speziellen für den Konsum von Virtual Reality und 360°-Produktionen wichtig ist (Wittek & Theile, 2016, S. 1).

In der Postproduktion kann die Platzierung einzelner Audiodateien im dreidimensionalen Raum mit dem Vector Base Amplitude Panning (VBAP) realisiert werden. Das Verfahren beschreibt die Erweiterung des klassischen Panorama-Potentiometers für eine 3D-Klangszene. Die Grundlage von VBAP ist wie die der stereofonen Panoramaregelung die Bildung der Phantomschallquelle zwischen den Lautsprechern. Da die Gewichtung mittels dreidimensionaler Vektoren ausgedrückt wird, ist VBAP nicht an eine feste Lautsprecheranzahl gebunden. Aufgrund bekannter Probleme der Summenlokalisation kann reines VBAP jedoch zur Bildung ungenauer Hörereignisorte führen (Weinzierl, 2008, S. 567 ff.).

Es wird gefolgert, dass zur detailgetreuen Abbildung der Räumlichkeit eine Methode dient, die anders als VBAP auf einem physikalischen Arrangement von Schallfeldern beruht. Im Folgenden wird das schallfeldbasierte Ambisonics-Verfahren näher diskutiert, da es diesen Ansatz berücksichtigt und oft zur Vertonung von 360°-Videos genutzt wird.

4.1 Ambisonics

Das Ambisonics-Verfahren ermöglicht die Erzeugung von 3D-Schallfeldern. Je mehr Kanäle bei der Aufnahme und Wiedergabe zur Verfügung stehen, desto genauer ist die auditive Darstellung der Räumlichkeit. Obwohl die zugrunde liegenden Annahmen schon in den 1970er Jahren durch den britischen Mathematiker Michael Gerzon theoretisch beschrieben wurden, fand die praktische Umsetzung von Ambisonics erst Ende der 1990er Jahre in größerem Maße Zuspruch. Durch flexible Möglichkeiten und Berücksichtigung von Höheninformationen eignet sich die heute frei nutzbare Technik für 3D-Vertonungen (Roginska & Geluso, 2018, S. 53).

4.1.1 Funktionsprinzip von Ambisonics

In Abhängigkeit der Übertragungskanäle ergeben sich für Ambisonics verschiedene harmonische Ordnungen, die mit zunehmendem Anstieg die Genauigkeit in den Raumachsen erhöhen. Das Schallfeld lässt sich als Überlagerung radial gewichteter Ordnungen mit kugelgeometrischen Besselfunktionen und einem Phasenfaktor beschreiben. Die folgende Darstellung von Kugelflächenfunktionen soll den Zusammenhang der Ambisonics-Ordnungen visualisieren:

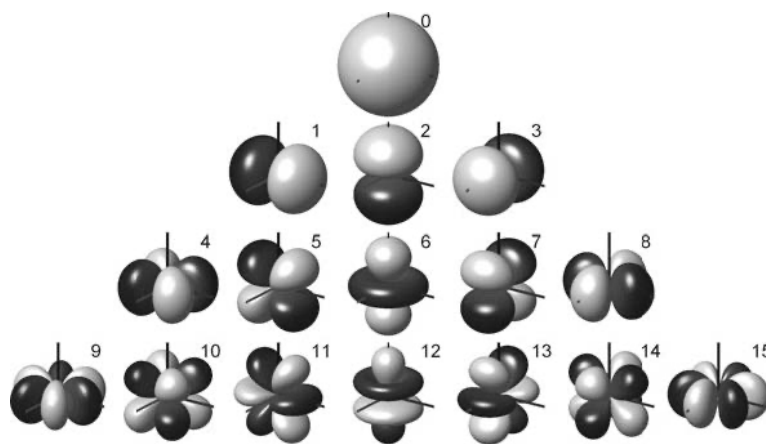


Abbildung 4: Kugelflächenfunktionen mit Kanalnummern (Zotter & Frank, 2019, S. 68)

Da sich die 0. Ordnung wie ein Schalldruckempfänger verhält, ist sie als akustische Kugel zu interpretieren. Der Schalldruck wird in diesem Kanal von allen Seiten gleichmäßig und monokompatibel eingefangen. Ambisonics 1. Ordnung, als sog. B-Format bzw. First Order Ambisonics (FOA), besteht aus vier Kanälen. FOA berücksichtigt neben dem Schalldruck in W die Druckgradienten der Raumachsen X, Y und Z. Die Richtcharakteristik kann dort mit einer akustischen Acht gekennzeichnet werden. Durch die vier Kanäle der 1. Ordnung lässt sich das

Schallfeld vollständig beschreiben. Ab Ambisonics 2. Ordnung wird von Higher Order Ambisonics (HOA) gesprochen. Die Kanalfolge wird durch V, T, R, S, U usw. fortgesetzt. Im Allgemeinen werden für Ambisonics m -ter Ordnung $(m+1)^2$ Kanäle benötigt: Durch 9 Kanäle der 2. Ordnung (SOA), 16 Kanäle der 3. Ordnung (TOA) usw. können Schallereignisse in HOA noch detaillierter im dreidimensionalen Raum abgebildet werden. (Weinzierl, 2008, S. 660 ff.).

Um Ton in Ambisonics aufzuzeichnen, wird z. B. ein sog. Soundfield-Mikrofon verwendet. Die Funktionsweise ist grundsätzlich ähnlich zur MS-Stereofonie, bei der ein Mittensignal durch Schalldruck und Seitensignale durch Druckgradienten entlang der Ohrachse aufgenommen werden. Das Soundfield-Mikrofon ergänzt diesen Ansatz um die Gradientenanteile „vorne-hinten“, „links-rechts“ und „oben-unten“. Die Umsetzung erfolgt durch vier Mikrofonskapseln mit breiter Nierencharakteristik, die entsprechend einer Tetraeder-Form ausgerichtet sind (Roginska & Geluso, 2018, S. 278 ff.). Die hierbei aufgenommenen Signale bilden zusammen das Ambisonics A-Format. Durch nachträgliches Summieren und Subtrahieren werden die Aufnahmen in das bereits erwähnte B-Format überführt und für die Postproduktion aufbereitet (Weinzierl, 2008, S. 591 ff.).



Abbildung 5: Soundfield-Mikrofone: Array, SPS200, MK4012, ST450 (Zotter & Frank, 2019, S. 11)

Zu beachten ist, dass für das B-Format verschiedene Konventionen gebräuchlich, jedoch untereinander nicht kompatibel sind. Hauptsächlich unterscheiden sie sich in ihrer Kanalreihenfolge: Während der FuMa-Standard die Kanäle in W, X, Y, Z anordnet, so gibt AmbiX die Folge mit W, Y, Z, X vor. Darüber hinaus assoziiert FuMa eine relative Dämpfung des W-Kanals um 3 dB. Bei AmbiX werden die Signale nach der SN3D-Normalization kanalübergreifend identisch behandelt (Rodgers, 2020).

Die folgende Tabelle stellt den Zusammenhang zwischen A- und B-Format vereinfacht dar:

B-Format	A-Format
	Ausrichtung der Kapseln oben: links vorne, rechts hinten
	Ausrichtung der Kapseln unten: links hinten, rechts vorne
W	links vorne + links hinten + rechts vorne + rechts hinten
X	links vorne – links hinten + rechts vorne – rechts hinten
Y	links vorne + links hinten – rechts vorne – rechts hinten
Z	links vorne – links hinten – rechts vorne + rechts hinten

Tabelle 1: Zusammenhang Ambisonics A- und B-Format (Roginska & Geluso, 2018, S. 279 f.)

Es ist anzumerken, dass weitere Kanalformate existieren, um Aufzeichnungen in Ambisonics zu beschreiben. Nachdem die Aufnahme für eine bestimmte Lautsprecheranordnung dekodiert wurde, werden die Signale als D-Format bezeichnet. Für beliebige Mehrkanalsysteme wie 5.1 Surround wurde das G-Format geschaffen, welches anstelle des D-Formats einzusetzen ist. Darüber hinaus finden sich in der Literatur das C-Format und E-Format (Hodges, 2011), auf die nicht weiter eingegangen werden soll. Jedoch wird auf diese Weise die Weitläufigkeit von Ambisonics deutlich.

Während die Signale zur Wiedergabe von FOA oft mit sog. Vienna Decodern dechiffriert werden, hat sich zur Dekodierung von HOA zum Beispiel das All-Round Ambisonics Decoding durchgesetzt. Es bildet die Signale als Integral unendlich vieler virtueller Lautsprecher ab und berechnet dann den entsprechenden Downmix für eine flexible Anzahl physikalischer Lautsprecher anhand von VBAP. Um die Klangobjekte bei der Abbildung stabil zu halten, entwickelte es sich zum sog. Multipl-Direction Amplitude Panning weiter (Zotter & Frank, 2019, S. 44 f. u. S. 72 ff.). Des Weiteren werden Triple-Balance-Panner einbezogen, die statt der Richtung eher die Position eines Klangobjekts fokussieren (Roginska & Geluso, 2018, S. 250).

Zur korrekten Reproduktion des Schallfelds müssen die Aufnahmen oder Simulationen aus ebenen Wellen bestehen und die Lautsprecher wiederum ebene Wellen abgeben. Umso mehr Lautsprechersignale überlagert werden, desto größer wird der Sweet Spot. Mit wenig Kanälen ist dieser Bereich jedoch eher klein (Weinzierl, 2008, S. 662). Die Ausführungen zu kopf- und raumbezüglichen Verfahren (vgl. Kapitel 3.3) legen aber den Schluss nahe, dass bei Kopfhörerwiedergabe der Sweet Spot vernachlässigt werden kann, da die Signale dort direkt an den Ohren anliegen. Die Dekodierung von Ambisonics für Kopfhörer erfolgt grundsätzlich auf die

gleiche Weise wie für Lautsprecher, außer dass die Signale zusätzlich mit den kopfbezüglichen Impulsantworten der entsprechenden Richtung gefaltet werden (Zotter & Frank, 2019, S. 85). Folglich wird das Ambisonics-Signal in einem Zwischenschritt auf virtuelle Lautsprecher geroutet und anschließend mit den HRTF-Eigenschaften versehen. In Abschnitt 3.2.2 wurde bereits auf die hierfür verwendete Binauralsynthese und Head-Tracking-Technik hingewiesen.

4.1.2 Ambisonics für 360°-Videos

Hochimmersive VR-Anwendungen werden normalerweise über HMDs rezipiert. Dabei wird die Position des Kopfes in Echtzeit durch Head-Tracking verfolgt. Bei stereoskopischen 360°-Videos wird über das VR-Headset zumindest die Kopfdrehung berücksichtigt und mittels binauraler Synthese für die Wiedergabe über Kopfhörer angepasst (Wittek & Theile, 2016, S. 5). Aus den Aspekten der Binauraltechnik kann wiederum gefolgert werden, dass der Ton monoskopischer 360°-Videos ohne HMD zielführend über Kopfhörer reproduziert werden sollte.

Zur Aufnahme wäre ein Kunstkopf denkbar. Dieser hätte aber den in Kapitel 3.2 formulierten Nachteil, dass im Nachhinein weder die Drehung des Kopfes noch die HRTF-Eigenschaften veränderbar wären. Ambisonics ermöglicht hingegen die flexible Speicherung von auditiven Eigenschaften. Bei der binauralen Umrechnung müssen sie entsprechend der Perspektive korrigiert werden (vgl. Abschnitt 4.1.1).

Bei den Aspekten zur 360°-Video-Technik wurde bereits angeführt, dass Onlineplattformen wie YouTube und Facebook maßgeblich zum Erfolg von 360°-Videos beitrugen (vgl. Abschnitt 2.1). Durch ihre große Popularität nehmen sie auch im Umgang mit Spatial Audio für vollsphärische Videos eine wichtige Rolle ein. Mit Verwendung von Ambisonics 1. Ordnung als Raumklangformat gibt YouTube einen inoffiziellen Standard für den Ton in 360°-Videos vor. Aktuell wird FOA in AmbiX vorausgesetzt. Zudem besteht die Möglichkeit sog. Head-Locked Stereo zu ergänzen, das sich bei der Wiedergabe nicht der Blickrichtung anpasst. Das Audio sollte sich für den Upload in einem MOV- oder MP4-Container befinden und AAC-codiert sein. Für die MOV-Datei wird ebenso unkomprimiertes Audio mit einer Abtastrate von 48 kHz und Pulse-Code-Modulation akzeptiert (YouTube, 2020).

Die Vermutung liegt nahe, dass im Fall von YouTube die Möglichkeit mit vier Kanälen einen 3D-Ton bereitstellen zu können, der weitaus besseren Auflösungsgenauigkeit höherer Ordnungen vorgezogen wird. So steht 3D-Audio auch privat Produzierenden zur Verfügung. Bei genauerer Betrachtung fällt auf, dass FOA bereits durch eine doppelte MS-Mikrofonierung

gewonnen werden könnte, wenn dabei Höheninformationen keine Rolle spielen. Aus diesem Grund wird das aus Abschnitt 3.3.2 bekannte Doppel-MS-Verfahren als zweidimensionales First Order Ambisonics aufgefasst. 3D FOA ergibt sich demnach aus einer Triple-MS-Aufstellung (Zotter & Frank, 2019, S. 6 ff.). Im einfachsten Fall ist also kein Soundfield-Mikrofon nötig, um Atmo für 360°-Videos aufzunehmen. Zum Beispiel bietet der handliche Surround-Recorder Zoom h2n mit entsprechend neuer Firmware-Version einen Modus zur Aufnahme und Speicherung von 2D FOA an (Hodges, 2018). Der seit 2018 erhältliche Zoom H3-VR ist wiederum speziell für dreidimensionales First Order Ambisonics konzipiert. Er kombiniert Soundfield-Mikrofon und passendes Aufnahmegerät (Schröder, 2018). Für den 2020 vorgestellten Multi-Track-Recorder Zoom H8 ist eine optionale Ambisonics-Mikrofonkapsel vorgesehen, um Spatial Audio produzieren zu können (Hautz, 2020).

Grundsätzlich ist Spatial Audio für 360°-Videos nicht auf Ambisonics 1. Ordnung beschränkt. Im Jahr 2016 übernahm Facebook das schottische Unternehmen „Two Big Ears“ und bietet deren Software für 3D-Audio seitdem kostenlos unter der Bezeichnung „Facebook 360 Spatial Workstation“ an (Greif, 2016), auf die zu einem späteren Zeitpunkt näher eingegangen wird. Two Big Ears entwickelte das TBE-Format, welches ein hybrides HOA darstellt. Anstelle von neun Kanälen für SOA agiert das Format mit acht. Die Einsparung des Ambisonics-R-Kanals sollte vor einigen Jahren die Kompatibilität mit DAWs erhöhen. Allerdings ergeben sich durch die Transkodierung für einige Kanäle invertierte Vorzeichen und im Speziellen in älteren Versionen Darstellungsprobleme (Farina, 2017). Facebook empfiehlt Spatial Audio im FB360 Matroska-Format hochzuladen, das 10 Kanäle nutzt: Das weiterentwickelte TBE-Format und zwei zusätzliche Kanäle für Head-Locked-Audio. Als Audiocodierung wird hierfür AAC vorgeschlagen (Facebook, 2020). Da in einem Matroska-Container im Allgemeinen etliche Codecs verschachtelt werden können, wurde das Format ähnlich der russischen Matroschka-Figur benannt (Rinne, 2010). In der aktuellen Version bietet die Facebook 360 Spatial Workstation natives Ambisonics bis zur 3. Ordnung an, auch in den Exporteinstellungen. Es ist deshalb davon auszugehen, dass Facebook TOA beim Upload berücksichtigt (vgl. Kapitel 6.4.).

Eine Initiative der Universität für Musik und darstellende Kunst in Graz entwickelte die Internetseite „HOAST Library“ zur Veröffentlichung von Audio und 360°-Videos, die mittels HOA vertont wurden. Der Webplayer akzeptiert Ambisonics bis zur 4. Ordnung, bietet Head-Tracking, binaurale Wiedergabe und ist mit HMDs wie HTC Vive und Oculus Rift kompatibel. Er bietet bislang aber keine Unterstützung für mobile Endgeräte. Um Inhalte zu veröffentlichen, muss eine schriftliche Anfrage eingereicht werden. Somit richtet sich der Service derzeit eher an fachinteressiertes Publikum (Institut für Elektronische Musik und Akustik, 2020).

4.2 Software und Plug-ins

Damit eine digitale Audioworkstation zur Bearbeitung von Ambisonics-Signalen geeignet ist, muss die jeweilige DAW mit mehreren separaten Kanälen in einer einzelnen Spur oder einem Bus umgehen können. Für Ambisonics 3. Ordnung wird z. B. eine Audiospur mit 16 Kanälen vorausgesetzt. Das weitverbreitete Avid ProTools HD unterstützt dies seit Ende 2017 (Farina, 2017). Im Vorgriff auf die unten vorgestellten Plug-ins werden neben ProTools oft DAWs wie Cockos Reaper und Steinberg Nuendo für die Arbeit mit Ambisonics empfohlen.

Es sei angemerkt, dass auch Programme zur Videonachbearbeitung wie z. B. Adobe Premiere Pro in der Version CC 2020 mit fertigen Ambisonics-Tonspuren umgehen können. Für professionelle Tongestaltung lohnt jedoch der Einsatz einer DAW mit entsprechenden Plug-ins. Da hierfür zahlreiche Tools erhältlich sind, stellen die folgenden Vertreter lediglich eine Auswahl möglicher Werkzeuge zur Erstellung einer Ambisonics-Mischung dar:

Die kostenlose O3A-Core-Library des Anbieters Blue Ripple Sound fasst viele Plug-ins für die Arbeit mit Ambisonics zusammen, zum Beispiel verschiedene Panner, Decoder, Meter und ein Konverter zwischen unterschiedlichen Ambisonics Konventionen. Die Bibliothek unterstützt HOA bis zur 3. Ordnung und ist kompatibel mit z. B. Pro Tools Ultimate und Reaper als VST- und AAX-Plug-ins. Die O3A-Core-Library wurde um einen Panner mit Videobezug erweitert. Weitere Tools sind kostenpflichtig beim Anbieter erhältlich (Blue Ripple Sound, o. J.).

Eine weitere Sammlung ist die „Ambisonic plug-in suite“. Die Open-Source Lösung steht plattformübergreifend als VST-Plug-ins zur Verfügung. Je Version und DAW lassen sich damit sogar Ambisonics bis zur 7. Ordnung verarbeiten, z. B. mit Reaper. Enthalten sind klassische Tools zum Panning und Dekodieren von Ambisonics nach dem AmbiX-Standard und zur binauralen Aufbereitung für Kopfhörer. Die Möglichkeit, eine Videovorschau zu nutzen, besteht jedoch nicht. Dennoch ist positiv anzumerken, dass die Plug-ins neben DAW-Systemen unter Windows und MacOS auch mit Ardour und dem Audio Connection Kit JACK unter Linux-Distributionen voll nutzbar sind (Kronlachner, 2014).

Das Plug-in „dearVR Pro“ des Anbieters „DearReality“ kostet derzeit 349 Dollar zuzüglich Mehrwertsteuer. Es steht dabei neben VST- und AAX- auch als AU-Plug-in zur Verfügung. Der 3D-Panner bietet eine Option, mit der die Verdeckung eines Klangobjekts durch andere Objekte im Raum simuliert werden kann. Zudem sind passende Hallstrukturen wählbar.

Das Plug-in unterstützt Ambisonics bis zur 3. Ordnung mit Auralisation in Echtzeit. Außerdem stehen umfassende Downmix-Varianten zur Verfügung. Jedoch besteht aktuell leider nicht die Möglichkeit ein Referenzvideo für das 3D-Panning zu hinterlegen (DearReality, 2020). Eine Abwandlung des Plug-ins ist das in Kapitel 4.3 erwähnte dearVR AMBI MICRO.

Mit der „360pan suite“ von „AudioEase“ für derzeit rund 300 € stehen umfassende Möglichkeiten für Ambisonics-Mischungen bis zur 3. Ordnung zur Verfügung. Da sowohl im AmbiX- als auch im FuMa-Standard gearbeitet werden kann, bieten die Plug-ins vielfältige Möglichkeiten. Die Suite besteht aus einem 3D-Panner, einem 360°-Video-Monitor mit binauraler Abhörmöglichkeit, verschiedenen Plug-ins für Konvertierung, Faltungshall, Limiter und Lautheitsmesser sowie einer Radar- und Rotationsfunktion. Die Tools können als VST- und AAX-Plug-ins z. B. mit ProTools, Reaper und Nuendo in einer entsprechend aktuellen Version genutzt werden. Für die fertigen Mischungen stehen direkte Export-Optionen für Facebook, YouTube, Jaunt etc. zur Verfügung (AudioEase, 2020).

Die in Abschnitt 4.1.2 erwähnte 360 Spatial Workstation von Facebook steht als VST- und AAX-Plug-in für ProTools, Reaper und Nuendo zur Verfügung. Sie besteht aus einem 3D-Panner mit Video, Control-Plug-in mit binauraler Abhörfunktion, Konverter, Encoder und Lautheitsmeter sowie einem Video-Player. Ambisonics kann dabei bis zur 3. Ordnung verarbeitet werden (Facebook Audio 360, 2020). Auffällig ist, dass sich der beschriebene Funktionsumfang ähnlich zu den oben aufgeführten Plug-ins von „AudioEase“ verhält. Da die Spatial Workstation jedoch kostenlos bereitgestellt wird und ebenso direkte Export-Einstellungen für Facebook und YouTube anbietet, hat sie im Rahmen der vorliegenden Masterarbeit eine besondere Stellung. Die Funktionen werden deshalb in Kapitel 6 näher beschrieben.

4.3 Stand der Technik

Der aktuelle Stand der Technik zur Aufnahme und Bearbeitung von Ambisonics zur Vertonung von 360°-Videos lässt sich folgendermaßen zusammenfassen:

Auf der NAMM Show 2019, einer Fachmesse für Musik und Audio in Kalifornien, präsentierten die Firmen Sennheiser und „DearReality“ den bis dahin ersten vollständigen Workflow für 3D-Audio für szenenbasierte 360°-Produktionen. Alle Schritte, von der Aufnahme bis zur Wiedergabe, sind mit eigenen Innovationen oder Partnerprodukten umzusetzen:

- Aufnahme in Ambisonics: Sennheiser AMBEO VR Mic
- Konvertierung: AMBEO A-B-Konverter-Plugin
- DAW-Plug-ins: AMBEO-Library, dearVR AMBI MICRO, dearVR SPATIAL Connect
- Binaurale Wiedergabe und Head-Tracking: AMBEO Smart Headset

Technisch sind hier Produktionen bis zu Ambisonics 3. Ordnung möglich (Sennheiser, 2019). Das AMBEO VR Mikrofon ist ohne Zweifel ein hochwertiges Soundfield-Mikrofon, doch es zeichnet lediglich First Order Ambisonics auf (Schröder, 2017). Folglich sind damit der direkten Aufnahme in 3D-Audio Grenzen gesetzt.

Bereits 2016 wurde in Fachkreisen kritisiert, dass Ambisonics oft vorschnell als geeignete Methode zur Vertonung virtueller Welten feststeht. Im Speziellen Ambisonics 1. Ordnung ermöglicht keine fehlerfreie auditive Immersion, da die Auflösungstiefe der vier Kanäle zu gering ist. Zudem existierte zu dieser Zeit kein Studiomikrofon, mit dem sich direkt Ambisonics höherer Ordnung aufzeichnen ließen (Wittek & Theile, 2016, S. 2). An der Weiterentwicklung der tetraederförmigen FOA Soundfield-Mikrofone wurde in den letzten Jahren intensiv geforscht. Der Ansatz, die räumlichen Komponenten mit einer Reihe von Richtmikrofonen zu bilden, deren Wirkung sich aus den sphärischen harmonischen Ordnungen ergibt, ist aufgrund der steigenden Anzahl zusammenfallender Mikrofone hoch komplex (Roginska & Geluso, 2018, S. 289). Das Ende 2019 vorgestellte „Spatial Mic“ der amerikanischen Firma Voyage Audio ermöglicht nun zum Beispiel Aufnahmen in Ambisonics SOA und nutzt dafür insgesamt acht verbaute Mikrofonkapseln. Optisch erinnert es an ein Soundfield-Mikrofon, das um weitere Komponenten ergänzt wurde (Sound & Recording, 2019). Darüber hinaus ist das „em32 Eigenmike“ des Unternehmens „mh acoustics“ erwähnenswert, welches durch zahlreiche in einer Kugel verbauten Mikrofone und integrierter digitaler Signalverarbeitungsprozesse bereits Ambisonics der 4. Ordnung ermöglicht (mh acoustics, o. J.).

Eine Alternative zur Aufnahme in Higher Order Ambisonics ist, Mikrofon-Arrays in sphärischer Anordnung zu verwenden und die Komponenten durch entsprechende Matrixbildung aus den Mikrofonsignalen abzuleiten (Roginska & Geluso, 2018, S. 289). Für dieses Vorgehen ist zum Beispiel das von Helmut Wittek und Günther Theile entwickelte ORTF-3D Verfahren zu nennen, das Ende 2017 auf dem Audio Engineering Society Kongress in New York vorgestellt wurde. Es entspricht einer Dopplung der ORTF-Surround-Aufstellung in zwei Ebenen direkt übereinander. Zusätzlich versorgen zwei Mikrofone die vertikale Ebene. Daraus resultiert eine Anordnung mit acht Kanälen die direkt auf die Lautsprecher geleitet oder virtuell in binauraler Form aufbereitet werden kann. Solche Mikrofonssysteme stellt zum Beispiel die Firma SCHOEPS her (Wittek & Theile, 2017, S. 4 ff.).

Die aufgezeigten Entwicklungen offenbaren die Chance, dass der Einsatz von HOA in 3D-Audio für 360°-Videos künftig stärker fokussiert wird. Es kann jedoch nur spekuliert werden, ob und wann dominierende Plattformen wie YouTube die Unterstützung von Ambisonics höherer Ordnung ermöglichen (Dieterle et al., 2018). Da das einst von Google gegründete „Jump-VR“-Projekt zur Entwicklung von 360°-Lösungen im Sommer 2019 eingestellt wurde (Moon, 2019) und darüber hinaus derzeit keine offizielle Ankündigung zu finden ist, scheinen die Erwartungen allerdings gedämpft.

Positiv anzumerken ist, dass der vom Fraunhofer-Institut für Integrierte Schaltungen in 2015 vorgestellte Audiokodierungsstandard MPEG-H für 3D-Audio rein technisch die Möglichkeit für Ambisonics bis zur 29. Ordnung bereithält. Das Kodierungsverfahren wird als „Next Generation Audioformat“ bezeichnet und ist nicht an feste Lautsprecheranordnungen gebunden. Ähnlich dem „Spatial Audio Object Coding“-Standard bietet MPEG-H dem Publikum die Möglichkeit, die Lautstärke einzelner Klangobjekte bei der Wiedergabe separat anzupassen. Es unterstützt zudem alle gängigen kanal- und objektbasierten Audiotechniken, auch Head-Tracking und binaurale Wiedergabe (Fleischmann, 2017 u. Roginska & Geluso, 2018, S. 196 f.)

5 Konzeption der Vertonung eines 360°-Promotion-Videos

Zu Beginn dieses Kapitels werden wichtige Begrifflichkeiten in Bezug auf den Titel der Masterarbeit erklärt. Anschließend sind die Besonderheiten bei der Vertonung von 360°-Videos zu diskutieren und die Rahmenbedingungen der praktischen Arbeit festzuhalten. Die zentralen Aspekte werden in ein Anforderungsprofil für die zu realisierende 360°-Vertonung überführt und die Planung des 360°-Promotion-Videos in Bezug auf die Tongestaltung erläutert.

5.1 Titel der Masterarbeit und seine Terminologien

Die vorliegende Masterarbeit trägt den Titel „Adäquate Vertonung eines 360°-Promotion-Videos“. Zu Beginn wurde die Prämisse formuliert, das Potenzial immersiven Tons effektiv zur Vertonung von 360°-Videos zu nutzen. In den darauffolgenden Ausführungen kamen grundlegende Aspekte zu vollsphärischen Videoproduktionen (vgl. Kapitel 2), immersiven Audio-techniken (vgl. Kapitel 3) und 3D-Audio für 360°-Videos (vgl. Kapitel 4) zur Sprache. Um das erworbene Wissen zielführend für die praktische Bearbeitung des Themas zu nutzen, sind weitere spezifische Begrifflichkeiten in Bezug auf den Titel der Arbeit zu klassifizieren:

Im Allgemeinen beschreibt der Begriff Adäquanz die Angemessenheit eines Verhaltens und dessen Üblichkeit. Das Adjektiv adäquat impliziert also, einer Sache in angemessener Weise zu entsprechen (F.A. Brockhaus, 1986, S. 128). Übertragen auf die Themenstellung der Masterarbeit wird festgehalten, dass ein angemessenes Vorgehen zur Vertonung eines 360°-Videos erarbeitet werden soll, das den Anforderungen des Mediums gerecht wird und der üblichen Anwendung entspricht. Für die Vertonung ist der Begriff des Sounddesigns essenziell, der im historischen Kontext erheblich durch die Kinoproduktionen der 1970er Jahre geprägt wurde. Das klassische Tonhandwerk wurde in Filmen wie „Star Wars“ oder „Apocalypse Now“ erstmals mit künstlerischen Aufgaben ergänzt. Die Tongestaltung orientiert sich seitdem bewusst an der Wirkung, die sie beim Publikum erzielen soll und reicht heute weit über die lediglich technischen Gesichtspunkte hinaus (Görne, 2017, S. 11).

Promotion ist eine Begrifflichkeit aus dem Marketing, durch die ein bestimmtes Produkt besonders in den Fokus der Aufmerksamkeit gestellt wird. Sie soll im Allgemeinen der Verkaufsförderung dienen (Pirntke, 2006, S. 122). Somit handelt es sich um eine Werbemaßnahme. Der Begriff wird im Rahmen der vorliegenden Arbeit in Zusammenhang mit 360°-Videos genutzt und ist in diesem Fall als Werbefilm zu verstehen.

Eine besondere Methode zur Erstellung von Werbefilmen ist das Storytelling. Werbebotschaften werden bewusst in Geschichten verpackt, die dramaturgisch an den Aufbau eines Spiel- bzw. Kurzfilms erinnern: Der Protagonist verfolgt ein klares Ziel und wird zunächst durch einen Konflikt aufgehalten. Meist folgt eine unvorhergesehene Wendung. Der zentrale Ansatz beim Storytelling ist somit die Informationsvermittlung über die emotionale Komponente einer fiktionalen oder wahren Geschichte (Idl, 2016). Das folgende Beispiel soll das Prinzip verdeutlichen: Bis heute bleibt der für EDEKA produzierte Werbefilm „heimkommen“ in guter Erinnerung. Ein einsamer älterer Herr inszeniert seinen Tod, um das kommende Weihnachtsfest im Kreis seiner Familie verbringen zu können – zuvor wurde er jedes Jahr aufs Neue von seinen Kindern und Enkeln getröstet (EDEKA, 2015).

Zusammenfassend ist festzuhalten, dass im Rahmen der vorliegenden Masterarbeit die angemessene Vertonung eines 360°-Werbefilms fokussiert wird, welche auf die Erkenntnisse der Grundlagenkapitel und 3D-Audio-Technik aufbaut. Für immersives Marketing mittels eines 360°-Promotion-Videos bietet sich der Storytelling-Ansatz an, da komplexe Werbebotschaften durch Geschichten effektiv kommuniziert werden können. Die Tongestaltung bezieht dabei verschiedene Aspekte des Sounddesigns mit ein.

5.2 Besonderheiten bei der Vertonung von 360°-Videos

Die Eigenheiten für den Ton, welche sich in vollsphärischen Inhalten ergeben, sind anfangs nicht nur für die Konsumenten ungewohnt. Die Schaffenden sind in gleicher Weise mit der neuen Situation konfrontiert, dass die Handlungsfläche nicht auf einen Kameraausschnitt begrenzt ist. Der Gestaltungshorizont erstreckt sich in 360°-Videos über den gesamten Raum. Deshalb muss das Zusammenspiel von Bild und Ton präzise durchdacht sein (Overschmidt & Schröder, 2013, S. 343).

5.2.1 Finesse des Sounddesigns in Spiel- und Werbefilmen

In Kapitel 5.1 wurde festgehalten, dass der Begriff Promotion-Video im Rahmen der vorliegenden Arbeit der Gattung Werbefilm angehört und das Storytelling nutzt. Nun soll deshalb eine gemeinsame Basis zum Film-Sounddesign geschaffen, gleichzeitig die Finesse der Tongestaltung in Werbefilmen abgrenzt werden. Des Weiteren wird das Verhältnis zwischen der auditiven und visuellen Komponente eines Films beleuchtet.

Menschen wurden in Bezug auf ihre Hörgewohnheiten über Jahrzehnte hinweg an Ton-Konventionen aus Kinoproduktionen gewöhnt, z. B. hinsichtlich des kommunikativen Codes der Synchronisation im kulturellen Umfeld. Zwar wird grundsätzlich erwartet, dass Ton und Bild zueinander synchron sind, doch internationale Produktionen werden im europäischen Raum in der jeweiligen Landessprache nachsynchronisiert. Dass die Lippenbewegungen dabei nicht exakt getroffen werden, wird nur selten als störend empfunden, da wiederum erwartet wird, den Inhalt des Films zu verstehen. Der sog. kommunikative Kontrakt zwischen Schaffenden und Rezipierenden sollte nur dann verletzt werden, wenn es dem Gesamtergebnis dient oder um die Regeln des Sounddesigns zu hinterfragen. Insgesamt ist auf eine schlüssige fiktionale Klangwelt zu achten, die das Bild begleitet und das Publikum unterhält (Görne, 2017, S. 21 f.). Demzufolge legt der kommunikative Kontrakt die Grundlage für die Beziehung zwischen Bild und Ton auch in filmischen Medien, die im Fernsehen und Internet ausgestrahlt werden.

Im Allgemeinen ist die akustische und visuelle Wahrnehmung durch die Gesamtheit zusammengesetzter Strukturen bedingt. Daraus ergibt sich die sog. Gestaltqualität, die den Gestsätzen der Nähe, Ähnlichkeit, Kontinuität, Geschlossenheit und Zusammengehörigkeit unterliegt. Diese Kohärenzfaktoren werden im Sounddesign genutzt, um durch Sprache, Geräusche und Musik die Gestaltqualität der Tonspur zu bilden (Raffaseder, 2002, S. 252 ff.). Zu beachten sind jedoch Maskierungseffekte, aus denen eine Abhängigkeit der auditiven Wahrnehmung in Bezug auf Zeit, Lautstärke und Frequenz entsteht. Es muss sichergestellt werden, dass bedeutsame Klangobjekte nicht in der Gesamtmischung untergehen (vgl. Abschnitt 3.1.1). Wenn Sound-Effekte eine tiefere Bedeutung transportieren, z. B. eine aus dem Alltag bekannte Signalwirkung oder symbolische Aussage, entsteht eine Semantik höherer Ordnung. Bedeutungsverändernde Motive sind stets kultureller Bestandteil der Gesellschaft und spiegeln sich auch in der Tongestaltung wider (Flückiger, 2017, S. 158). Während Paraphrase die schlichte Übertragung der visuellen Informationen auf den Ton beschreibt, wird einem unbestimmten Bild durch akustische Polarisierung ein konkreter Wert verliehen. Gegensätzliche Beziehungen sind als dissonant zu bezeichnen. Die zentrale Bedeutung der Klangobjekte reicht von der auditiven Beschreibung der visuellen Darstellung, über die Verbindungen zwischen Figuren bis hin zur Vermittlung von Emotionen (Raffaseder, 2002, S. 267 u. S. 274).

Die Tongestaltung für Werbefilme folgt insbesondere weiteren Prinzipien. Die Strategie hängt stark vom zu bewerbenden Produkt, den Unternehmen und deren Kommunikationszielen ab. Oft wird der kreative und zeitliche Spielraum durch spezielle Vorgaben der Auftraggebenden eingeschränkt. Das Zusammenspiel von Off-Text, Musik, Dialogen, Geräuschen und Effekten

verfolgt neben der Erzeugung passender Emotionen auch das Ziel, den Werbefilm von der Konkurrenz abzuheben. Da Werbung, anders als ein bewusst gewählter Spielfilm, oft nicht freiwillig konsumiert wird, muss zudem das Publikum aktiv zum Gebrauch bewegt werden (Henze, 2005, S. 112 u. S. 118 ff.). Lange wurde dies mit einer erhöhten Lautstärke versucht. Der in der Musikbranche vorherrschende „Loudness-War“, bei dem Produktionen durch übermäßige Kompression stetig lauter und undynamischer werden, zog seine Kreise bis hin zum Film und Rundfunk. Besonders im Fernsehen war die Werbung oft unangenehm laut im Vergleich zu Spielfilmen und konkurrierenden Anbietern. Erst seit 2012 steuern öffentlich-rechtliche und private Sender ihren Programmtönen einheitlich nach der European Broadcasting Union Richtlinie R-128 aus. Die sog. „Loudness Units relative to Full Scale“ beziehen sich auf die digitale Vollaussteuerung und betragen für das TV -23 LUFS. Im Internet existieren ähnliche Bestrebungen zur Vereinheitlichung der Lautheit von audiovisuellen Inhalten, die jedoch noch vage bei etwa -16 LUFS liegen (NDR, 2020).

In der Werbebranche findet sich darüber hinaus das „Prinzip der Abweichung“, das Erwartungen durchbricht oder bewusst gegen Regeln verstößt, um bekannten Normen nicht ohne Weiteres zu entsprechen (Gaede, 2002, S. 18). Es wird abgeleitet, dass für den Werbefilm ein Klangbild geschaffen werden muss, welches den Kommunikationszielen des Unternehmens oder dem Unique Selling Proposition eines Produktes dient, aber gleichzeitig zum Bleiben verleitet. Der Ton sollte deshalb nicht zu überladen sein, aber dennoch eine gewisse Raffinesse besitzen – sei es durch den Einsatz von Stille oder ungewöhnlichen Klangeindrücken, die von einer Norm abweichen.

Es wird festgehalten, dass für die Vertonung von Werbefilmen sowohl die grundlegenden Aspekte des Film-Sounddesigns Relevanz besitzen als auch die werbetechnischen Anforderungen einzubeziehen sind. Die folgenden Ausführungen zeigen auf, dass die Tongestaltung einer 360°-Produktion außerdem vielen bewährten Strukturen des herkömmlichen Sounddesigns für stereofonen oder mehrkanalstereofonen Ton widerspricht und deshalb eine besondere Herausforderung darstellt.

5.2.2 Tonaufnahmen am Filmset

Dass ein vollsphärisches Video ein Field of regard zu allen Seiten ermöglicht, ist bereits aus Kapitel 2 bekannt. Besonders eindrucksvoll macht sich diese Eigenschaft am Filmset bemerkbar, wenn es um die Positionierung der Mikrofone geht. Zum Beispiel ist der Einsatz einer Tonangel im herkömmlichen Sinn nicht möglich, da sich der Boom Operator zusammen mit dem Mikrofon zu jeder Zeit im Bild befinden würde. Diese Erkenntnis gilt für sämtliches Equipment am Set und das gesamte Filmteam – die Bereiche vor und hinter der Kamera verschmelzen. Für den Ton bedeutet dies, dass die Mikrofone so in die Filmkulisse integriert werden müssen, dass sie im Idealfall nicht sichtbar sind. Innerhalb einer filmischen Realität existieren sie normalerweise nicht. Ausgenommen davon sind Dokumentationen (Görne, 2017, S. 21). Auf der Suche nach geeigneten Verstecken fällt der Bereich direkt unter dem Kamerasystem auf. Da die Nadir-Aufnahme in der Postproduktion meist bearbeitet wird, fällt dort platziertes Equipment nicht zusätzlich ins Gewicht (vgl. Abschnitt 2.1). Dieser Platz ist zum Beispiel für ein Soundfield-Mikrofon geeignet, da hier das 3D-Schallfeld mit dem Kamerarastandpunkt übereinstimmt. Die Aufnahmen müssen lediglich vom A- in das B-Format überführt und können sodann als FOA-Tonspur genutzt werden (vgl. Abschnitt 4.1.1).

Die Atmosphäre der Situation wird zwar originalgetreu festgehalten, jedoch sind die Gestaltungsmöglichkeiten stark eingeschränkt. Details sind, bis auf Panorama-Einstellungen im Wechsel zum B-Format und klangliche Anpassungen der gesamten Spur, nachträglich nicht mehr veränderbar. Dieses Vorgehen eignet sich folglich am besten für Dokumentationen und spontane Aufzeichnungen unterwegs oder im Speziellen für Atmo-Aufnahmen, die durch Mikrofone näher am Geschehen ergänzt werden.

Für Dialoge bieten sich Lavalier-Mikrofone an, die direkt an der Kleidung der Personen befestigt und versteckt werden können. In diesem Zusammenhang sind Funkstrecken zu empfehlen, da die Optik und Bewegungsfreiheit sonst durch Kabel gestört wird. Gleiches gilt für Richtmikrofone, die zur Aufzeichnung von Geräuschen in Mono oder Stereo in die Location integriert werden und dem Hauptmikrofon als Stütze dienen sollen. Weiterhin eröffnet Funkequipment die Chance, dass sich das Produktionsteam während der Aufnahme an einem Ort befinden kann, der für die Kameras nicht einsehbar ist. Entfällt diese Möglichkeit, so müssen kreative Verstecke innerhalb des FOV gesucht werden, die genug Platz für z. B. den Tonmeister und die nötigen Recorder bieten, damit das Signal während des Drehs entsprechend überwacht werden kann. Es sei darauf hingewiesen, dass zum Abhören von Ambisonics mit Kopfhörern eine binaurale Umrechnung in Echtzeit erfolgen muss (vgl. Abschnitt 4.1.1).

Das Monitoring der Ansteck- und Richtmikrofone lässt sich wie gewohnt über den Kopfhörerausgang des Recorders realisieren. Bietet der Drehort keine für Menschen geeignete Verstecke, könnte das Team entweder als Statisten in die Szene integriert oder die Aufnahme als Notlösung ohne Überwachungsmöglichkeit durchgeführt werden. Alternativ sollte je nach Verwendungszweck des Endprodukts eine Nachvertonung der Szene in Erwägung gezogen werden. Aufnahmen am Set beschränken sich dann auf unbedingt notwendige O-Töne. Geräusche können nachträglich aus Archiven oder durch eine individuelle Foley-Produktion gewonnen werden, bei der die Sounds passend zum vorhandenen Bild erzeugt werden (Steppat, 2014, S. 207 f.). Für aufwendige Projekte kann zudem die Nachsynchronisation der Dialoge im Tonstudio erfolgen.

5.2.3 Diegetischer Ansatz

Unter Diegese ist die Zugehörigkeit eines akustischen Reizes zu visuellen Objekten der filmischen Realität zu verstehen. Diegetische Klangeindrücke wie Dialoge und Geräusche können von einer agierenden Person im Film gehört und zugeordnet werden. Sie bilden das natürliche Soundscape der filmischen Szene. Nichtdiegetischer Ton umfasst alles, das nicht Bestandteil der filmischen Realität ist, z. B. Musik und Effekte, die nur für das Publikum zu hören sind. Innere Monologe oder traumartige Zustände werden als metadiegetisch bezeichnet, da diese Klangeindrücke in der Wahrnehmung einer Filmfigur stattfinden (Görne, 2017, S. 20).

Nach Einführung des Tonfilms entstand die Befürchtung, dass die Tonspur der Filmmontage schaden könne, da exakt zur Handlung passende Klangobjekte redundant wirken. Die zu geringe Auflösung des Lichttons führte jedoch eher zum Gegenteil: Viele Geräusche wurden vom Publikum missverstanden. Der Ausdruck „See a dog – hear a dog“ beschreibt die damit verbundene Notwendigkeit, dass jedes Lebewesen oder bewegtes Objekt entsprechend vertont sein muss (Flückiger, 2017, S. 135 ff.). Ein diegetisch geprägtes Sounddesign wird als dokumentarisch empfunden. Liegt der Fokus hingegen auf einer nicht- bzw. metadiegetischen Gestaltung wird die Bedeutung der künstlerischen Komponente eines surrealen Sounddesigns impliziert. Zu beachten ist jedoch, dass fehlende diegetische Klangobjekte möglicherweise die Erwartungshaltung des Publikums enttäuschen, wenn die Motivation für das Ausbleiben un schlüssig erscheint (Görne, 2017, S. 231).

Es wird gefolgert, dass das bewusste Spiel mit Ton außerhalb des Bildausschnitts den Reiz eines adäquaten Sounddesigns ausmacht. Immersive Techniken eröffnen hierfür zahlreiche Möglichkeiten, die sich aus den Erkenntnissen des räumlichen Hörens und der Binauraltechnik ergeben und in Form von Surround-Ton und 3D-Audio bereits ihren Platz in der Filmproduktion einnehmen (vgl. Kapitel 3). Jedoch befindet sich handlungsrelevanter Ton mittig im Panorama, unabhängig davon, wo sich eine Person im Bild aufhält (Overschmidt & Schröder, 2013, S. 345 f.). Historisch ist dies wohl durch die Sweet Spot Problematik (vgl. Kapitel 3.4) bedingt und hat sich bis heute gehalten, auch im Stereoton des Fernseh- und Werbeumfelds.

Die Erwartung, dass sich Bild- und Tondramaturgie im Gleichgewicht befinden, führt im Kontext eines 360°-Videos jedoch zum Bruch mit solchen Tontraditionen. Inmitten einer vollsphärischen Szene wäre es höchst irritierend, wenn der dramaturgisch ausschlaggebende Ton an festen Punkten im Blickfeld fixiert wäre, die potenziell nicht mit der visuellen Ortung übereinstimmen. Insbesondere da in einer 360°-Produktion die Begrenzung des Kameraausschnitts außer Kraft gesetzt wird, rückt der diegetische Ansatz der filmischen Realität zwangsläufig in den Vordergrund des Sounddesigns: „Hier muss es darum gehen, möglichst detailgenau eine auditive Welt entstehen zu lassen, in der das Publikum selbst die Objekte findet, die Umgebung mit dem eigenen Aufmerksamkeitsfokus erforschen und erfahren kann, statt wie im Film-Sounddesign durch die Szene geführt zu werden. Die immersive Technik ist dann eine gestalterische Notwendigkeit.“ (Görne, 2017, S. 206) Demnach ist in einem Umfeld, in dem das Publikum selbst die Blickrichtung bestimmen kann, der Erwartung eines diegetisch konsistenten Tons besonders im 360°-Raum Folge zu leisten. So kann den Rezipierenden angemessen das Eintauchen in virtuelle Klangwelten ermöglicht werden.

Geräusche, die im Gegensatz zu Hard-Effects nicht oder nur ungenau im Bild zu lokalisieren sind, werden als Soft-Effects bezeichnet (Raffaseder, 2002, S. 253). Da vollsphärische Videos keinen festen Bildausschnitt vorgeben, wird abgeleitet, dass sich der Einsatz von Soft-Effects im Vergleich zur herkömmlichen Herangehensweise wesentlich komplexer darstellt. Die flexible Anpassung des temporären Kameraausschnitts im FOV kann dazu führen, dass z. B. Geräusche vom On- zum Off-Ton werden und umgekehrt. Neben der grundlegenden Frage, wie ein Geräusch klingt, ist somit im Speziellen der Ort der Schallquelle elementar. Die Differenzierung zwischen Vorder- und Hintergrund ist einerseits durch die Lautstärke und andererseits von der Klangfarbe geprägt. Speziell in geschlossenen Räumen kommen raumakustische Phänomene hinzu (Flückiger, 2017, S. 108).

Klänge außerhalb des Field of regard scheinen in 360°-Videos deshalb weniger relevant zu sein. Es wird angeregt Klangobjekte, die sich nicht im FOV befinden, entsprechend am visuellen Horizont im 360°-Raum zu positionieren und in geschlossenen Räumen Fenster und Türen einzubeziehen, falls dies dramaturgisch sinnvoll ist.

Um grundsätzlich die Gefahr zu reduzieren, dass handlungsrelevante Details im vollsphärischen Video untergehen (vgl. Kapitel 2.2), sollte in Betracht gezogen werden, die Aufmerksamkeit der Rezipierenden mit dem Ton zu lenken – jedoch ohne die gewonnene Freiheit zu stark einzuschränken. Zusammenfassend lässt sich festhalten, dass das Medium 360°-Video eher dokumentarisch erscheint und sich adäquate Ansätze zur Vertonung vollsphärischer Filme an den natürlichen Hörgewohnheiten des Menschen orientieren sollten.

5.2.4 Einsatz von Off-Stimmen und Musik

Nun ist das Vorgehen zur Integration der nichtdiegetischen Komponenten eines Sounddesigns in Bezug auf die Vertonung eines 360°-Videos zu hinterfragen – insbesondere der Einsatz von Off-Stimmen und Musik. Die Ausführungen aus dem vorherigen Abschnitt suggerieren nicht-diegetische Elemente losgelöst vom 360°-Raum in Form einer herkömmlichen Mono- oder Stereospur zu betrachten. Als Konsequenz wird das Potenzial der binauralen Wiedergabe und des Head-Trackings verspielt. Im Sinne der Zielsetzung der vorliegenden Arbeit sollte dies jedoch nicht von vornherein feststehen. Es gilt stets zu überprüfen, in welchem Kontext nicht-diegetische Klangobjekte eingesetzt werden und wie für sie ein sinnvoller Platz im Sounddesign geschaffen werden kann.

Off-Stimmen, die nur für das Publikum zu hören sind und sich demnach nicht im Bild befinden, kommentieren das Geschehen meist in Dokumentationen. Ein erzählender Text wäre ebenfalls denkbar (Bruns, 2012). Darüber hinaus sind Off-Sprecher fester Bestandteil vieler Werbefilme (Pionierfilm, 2020). Da sie nicht im Bild zu sehen sind, leuchtet es ein, Off-Text unabhängig von der Szene zu betrachten. Doch nehmen zum Beispiel innere Monologe als metadiegetische Elemente einen besonderen Platz in der filmischen Realität ein. Sie werden zwar im Off vorgetragen, finden aber im Kopf einer im Film handelnden Person statt (vgl. Abschnitt 5.2.3). Eine mögliche Lösung wäre, die Stimme trotz fehlender Lippenbewegungen am Platz der betreffenden Person zu positionieren und durch den Einsatz gewisser Effekte wie z. B. Hall zu symbolisieren, dass es sich um Gedanken handelt. In diesem Fall wäre die Zuordnung einer konkreten Position im Raum möglich. Alternativ könnte die Aufnahme direkt im

Ursprung des kopfbezüglichen Koordinatensystems platziert werden, was zur Im-Kopf-Lokalisation führen kann (vgl. Kapitel 3.3). In Folge der über die Zeit gefestigten Hörerfahrung in Bezug auf Kopfhörerwiedergabe kann IKL den Effekt einer inneren Stimme als auditives Stilmittel unterstreichen. Da sich das Gehörte aber an der Position des Rezipierenden befinden würde, liegt es im Auge des Betrachters, welche Variante zu bevorzugen ist.

Der Einsatz von Musik in vollsphärischen Inhalten, die zeitlich linear ablaufen, ist ebenso abhängig von der Situation wie von den technischen Möglichkeiten. Zeigt die Szenerie musizierende Menschen wie beispielsweise in einem Musikvideo, so bietet es sich an, die passende Musik entsprechend in den Raum zu legen. Je nach Position befinden sich die Rezipierenden dann vor oder auf einer akustischen Bühne. Die Voraussetzung für diesen Effekt ist jedoch, dass das entsprechende Stück in einzelnen Komponenten vorliegt, also Instrumente wie Schlagzeug, Gitarre und Bass separat platziert werden können (Rieger, 2019). Technisch ist es zwar möglich, einen stereofonen Gesamtmix räumlich zuzuordnen (vgl. Abschnitt 5.2.5), doch dann sollte hinterfragt werden, ob dies ein ästhetisches Wahrnehmungsproblem bedingt und den kommunikativen Kontrakt verletzt. Das bewusste Brechen bedarf des Versuchs und sollte vermieden werden, wenn die Irritation schwerer wiegt als der daraus entstehende Nutzen. Gleiches gilt für die Entscheidung, ob sich der Klang der Musik mit den Kopfbewegungen verändern sollte oder eine Head-Locked-Stereomischung dem immersiven Ansatz vorzuziehen ist. Eine Kopfdrehung im 360°-Raum beeinflusst in Abhängigkeit der Blickrichtung das Verhältnis zwischen dem Musikstück und den Ohren der Rezipierenden (Rieger, 2019). Es lässt sich nur erahnen, wie komplex die Komposition von Filmmusik nach diesen Maßstäben scheint – denn Musik trägt maßgeblich zur Erzeugung passender Emotionen bei.

Ein alternativer Ansatz Musik in 360°-Videos zu nutzen wäre, sie über Lautsprecher in der filmischen Realität wiederzugeben. Die Musik wird dann zum diegetischen Bestandteil der Produktion, da die Filmfigur sie hören kann. Zum Beispiel kann der linke und rechte Kanal einer Stereomischung mit dem linken und rechten Lautsprecher einer Stereoanlage verknüpft werden, die sich im Bild befindet. Damit dieser Effekt jedoch nicht missverstanden wird, sollte der Klang des filmischen Lautsprechers durch Bandbegrenzung und dynamikreduzierende Bearbeitungsschritte deutlich zu anderen Klängen der filmischen Realität sowie den realen Lautsprechern außerhalb des Films abgegrenzt werden (Görne, 2017, S. 250). Somit lässt sich stereofone Musik diegetisch sinnvoll im Raum positionieren, ohne sie in einzelne Komponenten aufzuteilen. Jedoch sollte durch klangliche Einschränkungen kenntlich gemacht werden, dass es sich um den Sonderfall „Lautsprecher im Lautsprecher“ handelt.

Dies heißt auch, dass sie als Teil der filmischen Realität Einfluss auf die Figuren im Film nehmen und nicht ausschließlich vom Publikum wahrgenommen werden. Die aufgezeigten Möglichkeiten für den Umgang mit Off-Text und Musik lassen den Schluss zu, dass es keine eindeutige Lösung gibt, die gleichermaßen auf alle Anwendungssituationen bezogen werden kann. Wenn keine angemessene Position im dreidimensionalen Raum gefunden wird, ist es legitim nichtdiegetische Elemente als Head-Locked-Stereo zu betrachten.

Es wird gefolgert, dass dieses Vorgehen auch für diffuse Layer- und Sound-Effekte zutrifft, da sie nicht scharf begrenzt und meist nur für die Rezipierenden zu hören sind. Es wird angeregt, bereits bei der Konzeption eines 360°-Videos zu berücksichtigen, in welcher Form insbesondere Musik und Off-Text verwendet werden soll und wie sie zur Ergänzung der diegetischen Komponenten des Sounddesigns beitragen können.

Im Rahmen der vorliegenden Arbeit wird davon abgeraten, das Video ausschließlich mit einer Stereospur zu vertonen. Einerseits wäre keinerlei Head-Tracking möglich und andererseits impliziert der diegetische Ansatz die Notwendigkeit immersiver Technik für die adäquate Tongestaltung in Bezug auf 360°-Videos (vgl. Abschnitt 5.2.3).

5.2.5 Räumliche Zuordnung von Tonaufnahmen

Abgesehen von Head-Locked-Audio werden die Aufnahmen, die als O-Ton oder durch eine Foley-Produktion entstanden sind und nicht von vornherein in einem 3D-Audioformat vorliegen, durch Spatializer Plug-ins im dreidimensionalen Raum positioniert. Die hierfür in Abschnitt 4.2 vorgestellten Tools basieren auf Ambisonics. Folgende Erklärung zeigt auf, weshalb dies für das dreidimensionale Panning durchaus sinnvoll ist. Es sei jedoch darauf hingewiesen, dass „spatialized Audio“ nicht zwingend durch Ambisonics gebildet werden muss.

Da Audiofiles aus Einzelmikrofonierung meist in Mono vorliegen, erscheint es logisch, standardmäßig einkanalige Tonaufnahmen im 360°-Raum zu positionieren. Die Gesamtheit aller akustischen Reize erzeugt wie beim natürlichen Hören insgesamt ein dreidimensionales Klangbild (vgl. Kapitel 3.1). Die räumliche Zuordnung von stereofonen Aufnahmen ist weitaus komplexer. Spatializer verfolgen das Ziel virtuelle Klangobjekte an einem bestimmten Punkt im Raum abzubilden (Lane, 2019). Der Ansatz stereofoner Aufnahmen ist, dass sich durch Pegel und Laufzeitdifferenzen Phantomschallquellen zwischen zwei Punkten abbilden lassen. Bei der Wiedergabe gibt die Lautsprecheranordnung einen festen Sweet Spot vor (vgl. Kapitel 3.4).

Dies lässt sich nicht ohne Weiteres in den 3D-Raum übertragen – es kann zum Einbruch des Stereobildes oder störenden Phasenverzerrungen kommen. Je mehr Bewegungsfreiheit das Publikum hat, desto deutlicher die Auswirkungen. Sinnvollerweise agiert Ambisonics ähnlich einer Skybox, d. h. wie im Innenraum eines imaginären Würfels, den die Rezipierenden weder verlassen noch verändern können. Somit sind stereofone Aufnahmen in Bezug auf Ambisonics weitaus unproblematischer gegenüber Tools, die mit anderen Ansätzen arbeiten (Lane, 2019).

Zusammenfassend ist festzuhalten, dass für Spatializer grundlegend einkanalige Aufnahmen zu empfehlen sind und im Speziellen Ambisonics den Umgang mit stereofonen Eingangssignalen deutlich erleichtert. Zudem ist anzumerken, dass sich die Bewegungsfreiheit in 360°-Videos lediglich auf die Ausrichtung des Kopfes beschränkt. Es wird gefolgert, dass bei der Wahl stereofoner Aufnahmen Vertreter der Intensitätsstereophonie förderlich sind, da sie die Mono-kompatibilität erhöhen (vgl. Abschnitt 3.3.2).

5.3 Rahmenbedingungen

Ziel der praktischen Arbeit ist die Konzeption und Realisierung der immersiven Vertonung eines 360°-Promotion-Videos. Das Endprodukt soll einer Agentur für Kommunikation und Visualisierung als Werbematerial zur Verfügung stehen. Evidentmedia bietet neben Webdesign und Printmedien nun selbst 360°-Produktionen an. Deshalb wird diese Erweiterung der Produktpalette fokussiert. Bezug nehmend auf die in Kapitel 1 aufgezeigten Vorteile vollsphärischer Inhalte, fiel die Wahl des Produktionsformats auf ein 360°-Promotion-Video, welches die Storytelling-Methode nutzt. Die spezifischen Terminologien wurden diesbezüglich bereits in Kapitel 5.1 klassifiziert. Da das Unternehmen vor allem online in sozialen Netzwerken promotet werden soll, wurde in Absprache mit der Agentur Facebook als primäres und YouTube als sekundäres Zielmedium für den Werbefilm festgelegt. Das vollsphärische Video soll die Firma eindrücklich nach außen präsentieren, 360°-Videos als Bestandteil deren Portfolio implizieren und das Marketing um eine immersive Komponente erweitern.

Die Videoaufnahmen finden in den eigenen Räumlichkeiten der Agentur und mit personellem Bezug statt. Aufgrund der geringen Größe der Bürofläche und der technischen Anforderungen des Equipments ergeben sich spezifische Vorgaben: Die zu nutzenden Spielflächen sind abhängig von den Stitching-Kanten, die aus der Anordnung der Kameras im 360°-Rig hervorgehen. Es ist darauf zu achten, dass Bewegung nicht zu nah am Rig oder innerhalb der sich überlappenden Bereiche stattfindet, da es sonst zu Problem im Stitching-Prozess kommt.

Um die Möglichkeiten der immersiven Tongestaltung effektiv zu nutzen, soll die Position des Kamerasystems dennoch so gewählt sein, dass alle Arbeitsplätze im Büro als Spielflächen zu Verfügung stehen.

Die Entscheidung, neben der eigentlichen Vertonung auch die Videoproduktion maßgeblich zu übernehmen, soll als Vorteil für den Ton gesehen werden. So können die Möglichkeiten bereits in der Konzeption berücksichtigt werden. Es sei jedoch darauf hingewiesen, dass sich die vorliegende Masterarbeit im Speziellen mit dem adäquaten Umgang des Tons für das 360°-Video befasst. Details zur Erstellung und Bearbeitung der Videoaufnahmen werden deshalb nicht umfassend beschrieben – ausgewählte Aspekte sind jedoch in Bezug auf die Tongestaltung zu erläutern.

Weitere Rahmenbedingungen ergeben sich aus den Auswirkungen der im Frühjahr 2020 ausgebrochenen Corona-Pandemie: So konnte das 360°-Video nicht wie geplant mit allen Mitarbeitern der Agentur gedreht werden. Da am Büro als Drehort für das 360°-Video festgehalten werden sollte, wurde als Alternative vereinbart, die Produktion lediglich zu zweit durchzuführen. Das nachträgliche Maskieren der Aufnahmen ermöglicht es, eine Person mehrfach im Video abzubilden. Allerdings erfordert diese Methode intensive Vorbereitungen, z. B. dürfen sich weder die Lichtverhältnisse noch die Kameraposition der zu maskierenden Aufnahmen ändern. Daraus folgt die statische Aufstellung des Kamerasystems in Silent-Observer-Perspektive und Aufnahmen mit Deckenbeleuchtung bei geschlossenen Jalousien. Um dem Lichteinfall an den Rändern der Verdunklung vorzubeugen, werden die Dreharbeiten nachts durchgeführt. Die resultierende Position, Ausrichtung und Höhe des Kamerastativs stellt den bestmöglichen Kompromiss zwischen den Video- und Tonanforderungen dar. Infolgedessen erleben Rezipierende das Geschehen im Endprodukt von einem etwas erhöhten Standpunkt. Für die auditive Komponente bedeutet dies, dass die Hörereignisse überwiegend unterhalb des Kopfbereichs gebildet werden müssen.

Auch die zeitliche und inhaltliche Planung der 360°-Produktion wurde an die neuen Bedingungen angepasst. Ursprünglich war angedacht, den Betrachtenden das Eintauchen in die verschiedenen Dimensionen der digitalen Arbeitswelt zu ermöglichen. Es sollte auf witzige Weise der Alltag einer Agentur für Kommunikation und Visualisierung mit den Vorstellungen Außenstehender und Wünschen der Mitarbeitenden verglichen werden. Da jedoch vor allem die Inszenierung einer Büroparty entgegen der corona-bedingten Beschränkungen gewesen wäre, musste mit entsprechendem zeitlichen Mehraufwand ein alternatives Konzept für die Umsetzung erarbeitet werden. Final spiegelt sich der selbstironische Grundgedanke in einer durch

Überstunden inspirierten Szenerie wider: Der Protagonist schläft vor Müdigkeit im Büro ein, obwohl er noch viel Arbeit vor sich hat. Er träumt davon, dass mehrere Doppelgänger die liegengebliebene Arbeit für ihn erledigen. Die Kapitel 5.4 und 5.5 erläutern weitere Details zur Planung des 360°-Videos in Bezug auf die Tongestaltung. Das finale Drehbuch zu „Eine Nacht im Büro“ ist im Anhang 10.2 der vorliegenden Arbeit zu finden.

5.4 Anforderungsprofil und Planung der 360°-Vertonung

Die technischen Grundlagen von 360°-Videos und immersiven Audiotechniken sind bereits aus den ersten Kapiteln der Masterarbeit bekannt. Damit sie zielführend zur Vertonung des 360°-Promotion-Videos angewendet werden können, sind die wichtigsten Prämissen und Planungsschritte in einem Anforderungsprofil festzuhalten. Es gilt neben den grundlegenden Erkenntnissen auch die spezifischen Terminologien und Rahmenbedingungen der Masterarbeit sowie die diskutierten Besonderheiten bei der Vertonung von 360°-Videos zu berücksichtigen.

5.4.1 Technische Learnings und kreative Inspiration

Die Grundsätze der immersiven Audiotechniken ergeben sich aus den Vorgängen einer natürlichen Hörsituation. Physiologisch betrachtet nehmen die Menschen ihre Umwelt stets durch ein dreidimensionales Klangbild wahr. Elementar für das räumliche Hören in den Ebenen des kopfbezogenen Koordinatensystems sind interaurale Pegel- und Laufzeitdifferenzen sowie Frequenzunterschiede zwischen den Ohrsignalen – in geschlossenen Räumen treten zudem weitere akustische Phänomene auf. Darüber hinaus tragen die HRTFs des Außenohrs durch ihre spezifische Filterwirkung maßgeblich zum binauralen Hören bei (vgl. Kapitel 3.1). Die Binauraltechnik macht sich die Stärke des Hörens mit zwei Ohren zu eigen, indem Tonaufnahmen mit spezifischen HRTF-Eigenschaften über Kopfhörerwiedergabe einen äußerst natürlichen Klangeindruck erzielen. Die einfachste Möglichkeit für binaurale Aufnahmen bietet ein Kunstkopf. Dieser hat jedoch den Nachteil, dass er die wichtigen Peilbewegungen des Kopfes außer Acht lässt. Durch Binauralsynthese und Head-Tracking-Verfahren ist es möglich, Audiosignale nachträglich mit HRTFs zu versehen und entsprechend der Kopfposition passend in Echtzeit zu berechnen (vgl. Kapitel 3.2). Für die adäquate Vertonung des 360°-Videos ist somit wichtig, die Vorzüge der Binauraltechnik zu nutzen und ein dreidimensionales Klangbild zu schaffen. Da in vollsphärischen Produktionen die Blickrichtung bei der Wiedergabe frei gewählt werden kann (vgl. Kapitel 2.1), ist es zielführend sich bei der Vertonung am natürlichen Hören zu orientieren.

Die herkömmlichen Stereo-Techniken spiegeln das immersive Empfinden jedoch nicht in angemessenem Maß wider. Aber die Strategien zur Aufnahme und Reproduktion stereofoner Tonsignale im kopf- und raumbezüglichen Ansatz legten die Basis für mehrkanalstereofone Anordnungen (vgl. Kapitel 3.3 und 3.4). Da für einen dreidimensionalen Klang neben der vertikalen Ebene im Speziellen Höhenunterschiede zu berücksichtigen sind, wird die Verwendung von 3D-Audio fokussiert. Dies umfasst alle immersiven Audioverfahren, die eine Orientierung im gesamten kopfbezogenen Koordinatensystem ermöglichen.

Das VBAP zur Überführung von Audiosignalen in den 3D-Raum fußt auf Summenlokalisierung und unterliegt den Schwierigkeiten kanalbasierter Ansätze. Zu bevorzugen sind deshalb Methoden, die auf einer physikalischen Betrachtung des gesamten Schallfelds beruhen (vgl. Kapitel 4). Das Ambisonics-Verfahren berücksichtigt zur Synthese und Resynthese die räumlichen Achsen und eröffnet je nach Ordnung flexible Möglichkeiten zur Vertonung von 360°-Videos. Der korrekte Umgang in Bezug auf unterschiedliche Formate und Konventionen ist jedoch komplex. Dennoch macht die Erkenntnis, dass der Ton in vollsphärischen Produktionen meist über Kopfhörer reproduziert wird und somit die Sorge nach dem Sweet Spot in den Hintergrund rückt, Ambisonics in Verbindung mit Head-Tracking und der binauralen Aufbereitung virtueller Lautsprecher-signale attraktiv (vgl. Abschnitt 4.1.1).

Neben den technischen Learnings werden im Folgenden zwei Beispiele aufgezeigt, die als kreative Inspiration für die Tongestaltung herangezogen wurden. Sie nehmen Bezug auf die in Kapitel 5.2 erörterten Besonderheiten – im Speziellen auf den diegetischen und nichtdiegetischen Ansatz. Es sei darauf hingewiesen, dass zur Auswahl die visuelle Produktionstechnik zugunsten der klanglichen Besonderheiten zu vernachlässigen war:

Der im Jahr 2016 von Google präsentierte vollsphärische Animationsfilm „Pearl“ überzeugt durch emotionales Storytelling und war 2017 für einen Oscar nominiert. Der Kurzfilm erzählt die Geschichte eines Mädchens und ihres Vaters, die in einem Auto durch das Land ziehen. Diegetische Klangobjekte sind dort in der Szene platziert, während nichtdiegetische Elemente als Head-Locked Audio interpretiert werden. Besonders deutlich wird dieses Prinzip in Verbindung mit dem Vater: Solange er auf seiner Gitarre spielt, ist z. B. die Musik räumlich zugeordnet, sonst befindet sie sich außerhalb des 360°-Raums (Google Spotlight Stories, 2016). Einen eher experimentellen Anspruch hinsichtlich der Sound-Effekte erhebt hingegen der 2018 veröffentlichte VR-Kurzfilm „Do Not Touch“: Die Rezipierenden befinden sich in einer Kunstgalerie und tauchen mit den Protagonisten in die Gemälde ein. Besonders eindrucksvoll

ist hier die Musik, welche ohne einen diegetischen Bezug zur virtuellen Szene maßgeblich im dreidimensionalen Raum platziert wurde. Da die Protagonisten jedoch zum Beispiel in einer seitlich ausgerichteten Kamerafahrt verfolgt werden, wird es nicht als Irritation wahrgenommen und zeigt, was in kreativer Hinsicht auditiv umsetzbar ist (SoKrispyMedia, 2018).

5.4.2 Technische Anforderungen und organisatorische Planung

In den Rahmenbedingungen wurden Facebook und YouTube als Zielmedien für das 360°-Video festgelegt (vgl. Kapitel 5.3). Deshalb sind an dieser Stelle die technischen Vorgaben der Plattform hinsichtlich des Raumklangformats zu beachten: Während YouTube für Spatial Audio auf First Order Ambisonics im AmbiX-Standard beharrt, so bietet Facebook das Bestreben die Ansätze von HOA nutzbar zu machen. Beide ermöglichen außerdem den optionalen Einsatz von Head-Locked-Audio, das sich bei der Kopfdrehung nicht verändert. Grundlegend ist anzumerken, dass stets die höchstmögliche Ambisonics-Ordnung genutzt werden sollte, um die Auflösungsgenauigkeit zu erhöhen. Da Facebook jedoch offiziell empfiehlt, den Ton im Hybrid-Format mit zehn Kanälen bereitzustellen (vgl. Kapitel 4.1.2), wird dies im Rahmen der Masterarbeit für das Endprodukt als angemessen empfunden. Auf der Suche nach geeigneter Software zur Vertonung von dramaturgisch linearen 360°-Videos sticht die von Facebook zum Download angebotene „360 Spatial Workstation“ ins Auge. Diese Plug-ins bieten einen hohen Funktionsumfang und stehen im Gegensatz zu vergleichbaren Angeboten kostenlos zur Verfügung, ohne durch eine fehlende Videovorschau eingeschränkt zu sein. Darüber hinaus besteht die Möglichkeit den Ton für die Zielmedien Facebook und YouTube passend zu exportieren, somit fällt die Wahl auf diese Plug-ins. Als Software zur Audibearbeitung wird Reaper verwendet, da es mit den Plug-ins harmoniert und vergleichbare DAWs wie ProTools verhältnismäßig teuer sind (vgl. Kapitel 4.2).

Da sich die Zielmedien in ihren technischen Anforderungen unterscheiden und die verwendete Software natives Ambisonics bis zur 3. Ordnung unterstützt, wird die Strategie verfolgt, für die Tongestaltung das Potenzial der höchstmöglichen Auflösung zu nutzen und für die Endprodukte einen Downmix entsprechend der Anforderungen zu exportieren. Für YouTube ist zu beachten ist, dass die Auflösungsgenauigkeit von FOA nicht ausreicht, um virtuelle Welten kompromisslos zu vertonen (vgl. Kapitel 4.3). Es wird jedoch zum einen für das sekundäre Zielmedium YouTube benötigt, zum anderen handelt es sich in diesem Fall um keine stereoskopische Produktion, die über HMDs reproduziert werden soll. Die Möglichkeiten von HOA sollen über das primäre Zielmedium Facebook genutzt werden.

Es sei zudem auf die in den Kapitel 5.1 und 5.3 klassifizierten Terminologien und Rahmenbedingungen hingewiesen. Es wird eine angemessene Vertonung angestrebt, in der die Stärken von Spatial Audio entsprechend der Möglichkeiten für die Tongestaltung genutzt werden sollen, um sie von einer herkömmlichen Stereovertonung abzuheben. Demnach ist es wichtig den gesamten Raum für die klangliche Gestaltung zu nutzen, jedoch steht diese Prämisse in unmittelbarem Zusammenhang mit der Position der Kamera und den technisch bedingten Vorgaben hinsichtlich zu nutzender Spielflächen.

Das zu vertonende 360°-Promotion-Video wurde im Rahmen der vorliegenden Masterarbeit der Kategorie des Werbefilms zugeordnet. Der Ansatz die Werbebotschaft durch eine Geschichte zu transportieren, die sich deutlich an der Dramaturgie eines Spielfilms orientiert (vgl. Kapitel 5.1), kommt der Tongestaltung zugute. Die Vertonung fußt hierfür auf den Grundlagen des Film-Sounddesigns und wird durch spezifische Anforderungen des Werbefilms ergänzt (vgl. Abschnitt 5.2.1). Die Entscheidung, für das Promotion-Video auf selbstironisches Storytelling zu setzen, lockert zudem die relativ dokumentarischen Ansprüche des Mediums 360°-Video auf und erzeugt Aufmerksamkeit in sozialen Netzen. Hierin liegt das Hauptkommunikationsziel des Promotion-Videos, ergänzt durch die Sensibilisierung für 360°-Videos als Portfolio der Agentur. Die Planung war jedoch unter zeitlichem und organisatorischem Mehraufwand an die corona-bedingten Beschränkungen anzupassen (vgl. Kapitel 5.3).

Der final resultierende Handlungsbogen stellt sich wie folgt dar: Der Protagonist Veit Stephan bleibt nach Feierabend aufgrund von Überstunden allein im Büro der Agentur „evidentmedia“ zurück. Am Schreibtisch nebenan stapelt sich Papierchaos, das bis zum Morgen bearbeitet sein muss. Doch nach einigen Stunden der Arbeit sinkt Veit vor Müdigkeit am Schreibtisch zusammen und schläft ein. Den Übergang zum Traum bildet eine Animation mit verwischenden Sternen. Durch sie werden die Rezipierenden in den metadiegetischen Part des 360°-Videos geleitet. Die drei weiteren Schreibtische im Büro sind nun mit Doppelgängern besetzt, die sich optisch nur durch ihre Kleidung vom Protagonisten unterscheiden. Während das Original schläft, erledigen die drei Doppelgänger seine Arbeit und unterstreichen dies mit rhythmischen Bewegungen und selbstironischen Texten. Als Veit am nächsten Morgen von einem eingehenden Anruf geweckt wird, sind die Doppelgänger verschwunden. Traum und Realität werden hier erneut durch eine Animation getrennt. Während des Telefonats stellt der Protagonist am nächsten Morgen fest, dass dort wo vorher Papierchaos war, nun alle Unterlagen sauber in einen Ordner einsortiert und bearbeitet sind. Überrascht, aber sichtlich zufrieden freut er sich über die erledigte Arbeit (vgl. Anhang 10.2).

Da die vollsphärische Produktion ein Field of regard zu allen Seiten ermöglicht, ihre Handlung jedoch linear verläuft, muss sichergestellt werden, dass die Rezipienten alle relevanten Punkte des Plots wahrnehmen können. Für den Ton ergibt sich somit die Anforderung, durch auditive Akzente die Aufmerksamkeit zu steuern, sodass die Blickrichtung an den jeweiligen Stellen im Video richtig gewählt wird. Es ist jedoch darauf zu achten, dass der Mehrwert der Rundumsicht nicht zu stark geschmälert wird und Bild und Ton gut aufeinander abgestimmt sind (vgl. Kapitel 5.2). Konkret wird vorgeschlagen diese Akzente durch Geräusche zu setzen, die durch Semantik höherer Ordnung eine tiefere Bedeutung erhalten (vgl. Abschnitt 5.2.1). Zum Beispiel hat im Video das Summen und Rattern des Druckers oder das Telefonklingeln eine gewisse Signalwirkung, die intuitiv dazu verleitet, die Blickrichtung entsprechend zu wählen.

In Abschnitt 5.2.2 kam zur Sprache, dass der herkömmliche Einsatz von Tonangeln zur Aufnahme des Tons nicht möglich ist, da sie sich zu jeder Zeit im Bild befinden würden. Für die Sprechparts bieten sich demnach Ansteckmikrofone an, die direkt an der Kleidung der Personen angebracht werden können. Funkequipment bietet den Vorteil, dass die Recorder außerhalb des Raumes platziert werden können und keine störenden Kabel verlegt werden müssen. Darüber hinaus kann verhindert werden, dass sich das Produktionsteam während der Aufnahme im Bild befindet. Für kabelgebundene Mikrofone und Recorder müssen geeignete Verstecke gefunden werden. Da es sich beim Drehort im konkreten Fall um das Büro der Agentur handelt, bietet es sich an Equipment dort zu platzieren, wo aufgrund technischer Geräte sowieso bereits Kabel verlaufen oder die Sicht durch verschiedene Regale und Rollcontainer eingeschränkt ist. Zudem befindet sich in der Mitte des Büros ein Sideboard, dessen zu öffnende Seite nicht einsehbar ist.

Der Platz direkt unter der Kamera ist ebenso möglich, da das Nadir-Bild in der Videonachbearbeitung retuschiert wird. Für gewöhnlich ist es der ideale Ort, um dort z. B. mit einem Soundfield-Mikrofon Atmo direkt in Ambisonics aufzuzeichnen. Die Geräusch-Atmosphäre wäre in diesem Fall jedoch hauptsächlich von den Lüftern der Computer, dem Brummen des Kühlschranks und Hintergrundgeräuschen aus den umliegenden Büroräumen und der nahegelegenen Hauptverkehrsstraße dominiert. Zu Gunsten der Werbewirksamkeit des 360°-Promotion-Videos wird deshalb auf die Aufnahme von Atmo als Originalton verzichtet.

In Hinblick auf die Konsequenzen der Rahmenbedingungen wurde das Vorgehen erarbeitet, die Tonaufnahmen am Filmset so gering wie möglich zu halten und der Nachvertonung der Szene den Vortritt zu gewähren. Zudem suggeriert die Bezeichnung Promotion-Video, dass der Ton dort nach professionellen Gesichtspunkten postproduziert wurde. Um den Aufwand

jedoch im Gleichgewicht zu halten wird festgelegt, dass Dialoge und Sprechparts direkt während der Dreharbeiten aufgenommen werden und im Speziellen die Geräusche in einer separaten Foley-Produktion nachzuvertonen sind. Darüber hinaus ermöglicht diese Strategie die nötigen Mikrofone mit Zubehör und Verstecken zu reduzieren sowie die Gesichtspunkte der binauralen Wiedergabe für die Abhörsituation vor Ort zu umgehen.

5.4.3 Gestalterische Anforderungen und kreative Planung

Grundsätzlich wird für die Nachbearbeitung festgehalten, dass die Tonaufnahmen, die während der Drehs, durch die Foley-Produktion oder aus Soundarchiven gewonnen wurden, mit entsprechenden Plug-ins im 360°-Raum platziert werden müssen, um einen dreidimensionalen Gesamtmix zu erhalten. Da die Aufnahmen im konkreten Fall ausschließlich in mono oder stereo vorliegen, werden sie für das 3D-Panning durch den Einsatz von Spatializer-Tools räumlich platziert. Obgleich die Zuordnung mit Ambisonics relativ problemlos verläuft, sei an dieser Stelle nochmals auf 5.2.5 hingewiesen. Dort wurde zur Vorsicht bei der Verwendung stereofoner Tonaufnahmen aufgerufen, da sich durch die Überführung in den dreidimensionalen Klangraum möglicherweise unerwünschter Phasenverzerrungen im Klangbild ergeben. Für die Produktion „Eine Nacht im Büro“ werden deshalb zum einen räumlich platzierte Stereofilms hinsichtlich ihrer Monokompatibilität ausgewählt und zum anderen liegen die meisten Aufnahmen aufgrund der verwendeten Mikrofone und natürlichen Gegebenheiten in monofoner Form vor. Hier spiegelt sich eine Folgerung des räumlichen Hörens wider: Wahrgenommen wird eine Vielzahl von einzelnen für sich monofonen Schallereignissen, die physiologisch durch das Hören mit zwei Ohren und den binauralen Prozessen zu einem dreidimensionalen Klangerlebnis in Richtung, Entfernung und Höhe wird.

Die Tongestaltung in 360°-Videos orientiert sich maßgeblich am diegetischen Ansatz, der in Abschnitt 5.2.3 vorgestellt wurde. Das vollsphärische Videobild, in dem die Rezipierenden die Blickrichtung frei wählen können, impliziert die Erwartung, dass Klangobjekte entsprechend richtig im Raum lokalisierbar sind. Demnach ist der kommunikative Kontrakt deutlich verletzt, wenn das Publikum im vollsphärischen Kontext eine Person sprechen sieht, aber deren Stimme falsch im kopfbezogenen Koordinatensystem wahrnimmt. Hinzu kommt, dass sich innerhalb der filmischen Realität akustische Reize aus dem Off durch eine Kopfdrehung plötzlich im Bild befinden und umgekehrt.

Dies bedeutet für die Gestaltung der Klangszene, dass alle Klangobjekte in der filmischen Realität sinnvoll im Raum zugeordnet sein müssen. Dieser Forderung soll auch in „Eine Nacht im Büro“ Folge geleistet werden, um den Rezipierenden auf angemessene Weise die Möglichkeit zu geben, die Umgebung selbst auditiv zu erforschen und das Gehörte im Bild wiederzufinden. Das Publikum befindet sich inmitten der Szene und ist somit ein Teil von ihr. Es ist anzumerken, dass der visuelle Immersionsanspruch in monoskopischen 360°-Videos einerseits deutlich geringer ist als im stereoskopischen Pendant, andererseits wurde zu Beginn der Masterarbeit die filmische virtuelle Realität deutlich von der computergenerierten abgegrenzt, die einen besonders hohen Immersionsgrad zur Ausblendung der Wirklichkeit angestrebt (vgl. Kapitel 2.1 und 2.2). Unabhängig davon sollte für die auditive Komponente das immersive Potenzial bestmöglich ausgenutzt werden, um im Sinne der Masterarbeit eine adäquate Vertonung des 360°-Promotion-Videos zu fokussieren.

Konkret ist auf die Aufteilung des 360°-Videos in filmische Realität und Traumsequenz hinzuweisen. Speziell die zweite Szene bietet als metadiegetische Komponente des Videos einen gewissen Raum für experimentelle Gestaltung. Hinzu kommt, dass sich die Handlung in der ersten und letzten Szene auf einen konkreten Bereich im Bild beschränkt, während im Traum alle Arbeitsplätze als Spielflächen genutzt werden.

Klar ist jedoch, dass neben den diegetischen Klängen auch die nichtdiegetischen Elemente für eine angemessene Vertonung relevant sind, da die Wirklichkeit in monoskopischen vollsphärischen Videos nicht ohne Weiteres komplett zu vernachlässigen ist. In den Abschnitten 5.2.1 und 5.2.4 wurde aufgezeigt, dass nichtdiegetische Klänge wie Off-Stimmen und Musik, die nur für das Publikum hörbar sind, einen festen Bestandteil sowohl im Film-Sounddesign als auch in Werbefilmen darstellen. Nicht zuletzt, da sie maßgeblich zur Erzeugung und Vermittlung passender Emotionen beitragen. Für die adäquate Tongestaltung eines 360°-Videos ist der Einsatz nichtdiegetischer Komponenten stets zu hinterfragen und zu prüfen, wie für sie ein sinnvoller Platz im dreidimensionalen Raum geschaffen werden kann. Ist dies nicht möglich, so können solche Klangobjekte als Head-Locked-Audio integriert werden, das nicht durch Kopfdrehungen beeinflusst wird. Um die Vorzüge des Head-Tracking nicht zu verspielen, sollte Head-Locked-Audio nicht ausschließlich verwendet werden.

Konkret wird für „Eine Nacht im Büro“ entschieden, auf den Einsatz von nichtdiegetischen Off-Stimmen zu verzichten, denn der Plot des Videos erinnert an einem verfilmten Sketch, für den keine Stimmen außerhalb der filmischen Realität notwendig sind.

Bestärkt wird die Entscheidung durch das „Prinzip der ABWandlung“, das im Werbeumfeld dazu genutzt wird, Traditionen zu brechen, anders zu sein und aufzufallen. Da die Agentur, für die das Promotion-Video als Werbematerial zur Verfügung stehen soll, zudem keinen spezifischen Claim in Werbemaßnahmen nutzt, den ein Off-Sprecher für gewöhnlich in Werbefilmen sprechen würde, wird in diesem Fall das Mittel der Abweichung vorgezogen.

Ferner wurden in Abschnitt 5.2.4 verschiedene Strategien zum Einsatz von Musik in 360°-Videos diskutiert. Der resultierende Konsens dieser Ausführungen ist, dass es von den visuellen Gegebenheiten im Video und den technischen Möglichkeiten abhängt, in welcher Weise Musik eingesetzt werden sollte. Für das zu vertonende 360°-Promotion-Video bietet es sich an, Musik in Szene 1 und 3 als Head-Locked-Audio zu nutzen, da der Einsatz dort nicht mit visuellen Komponenten zu begründen ist. Im Speziellen soll die Hintergrundmusik den selbstironischen Grundgedanken des Werbefilms aufgreifen und unterstreichen, dass es sich um eine Geschichte handelt, die wie ein Sketch bzw. ein humorvoller Kurzfilm mit einem dramaturgischen Handlungsbogen erzählt wird. Die Traumsequenz bietet als metadiegetische Komponente die Option, die Möglichkeiten der diegetischen Vertonung im 360°-Raum zu testen. Für die Vertonung werden hauptsächlich selbsterstellte Audioaufnahmen verwendet. Ergänzende Effekte und Musik stammen aus der „iLife-Toneffekte“-Sammlung von Apple.

Da besagte Traum-Szene im Rahmen der Konzeptentwicklung und Realisierung eine besondere Stellung innerhalb der praktischen Arbeit einnimmt, werden im Folgenden weitere Details zur Planung dieser Sequenz beschrieben.

5.5 Entwicklung der Traumsequenz

Bedingt durch die bereits vorgestellte Alternativplanung des 360°-Promotion-Videos wurde festgelegt, dass personelle Abhängigkeiten für die Produktion erheblich zu reduzieren sind. Ein besonderes Augenmerk liegt dabei auf dem Vorhaben, den gesamten Raum für die Tongestaltung zu nutzen. Die verschiedenen Arbeitsplätze sind ringsum an den Wänden angeordnet. Unter der Annahme, dass sich ein 360°-Kamera-System mittig im Raum befindet, bietet die Nutzung der vier Arbeitsplätze im Video die Möglichkeit, den Raum optimal auszunutzen. Auf die Abhängigkeiten der Anforderungen zwischen Bild und Ton wurde bereits hingewiesen. Zum einen resultieren die Spielflächen abhängig von den entstehenden Stitching-Kanten und somit von der Ausrichtung des 360°-Rigs, zum anderen sollen alle Arbeitsplätze für die Tongestaltung nutzbar sein (vgl. Kapitel 5.3). Da nicht mit jedem Kamerastandort gute Stitching-Ergebnisse erzielbar sind, war die Position durch Testläufe zu optimieren. Dies führte zum Kompromiss, dass das Publikum die Szene von einem erhöhten Standpunkt aus betrachtet.

Zwar können auf diese Weise punktuell alle vier Arbeitsplätze genutzt werden, doch in überlappenden Bildbereichen darf aufgrund der Stitching-Kanten keine Bewegung stattfinden. Folglich ist es nicht möglich quer durch den Raum zu laufen und die Position außerhalb der eng eingegrenzten Spielflächen zu wechseln. Da die Produktion unter den bekannten Umständen lediglich zu zweit durchzuführen ist, musste ein geeignetes Vorgehen zur Nutzung des Raumes in Bezug auf die Tongestaltung erarbeitet werden. Es entstand die Idee, den Protagonisten mehrfach im Bild auftreten zu lassen, um den Raum bestmöglich nutzen zu können. Konzeptionell lässt sich dieser Ansatz am sinnvollsten mit Doppelgängern umsetzen, die sich an den freien Arbeitsplätzen niederlassen und für den Protagonisten die liegengebliebene Arbeit erledigen. Da dies jedoch eine deutlich surreale Situation darstellt, wurde sie durch eine Traumsequenz von der ersten und letzten Szene abgegrenzt. Der metadiegetische Aspekt der filmischen Realität schafft hierfür den nötigen Gestaltungsspielraum.

Die verschiedenen Bestandteile des Traums werden separat aufgezeichnet und in der Postproduktion maskiert, sodass alle vier Rollen gleichzeitig zu sehen sind. Die Interaktion zwischen den Doppelgängern ist jedoch wie oben beschrieben stark eingeschränkt. Aufgabe des Tons ist diese Einschränkung kreativ zu überspielen. Um zu symbolisieren, dass die Doppelgänger gemeinsam an der Bewältigung des Papierchaos arbeiten, ohne sie tatsächlich miteinander interagieren zu lassen, wird ein Rhythmus mit unterschiedlichen Bürogeräuschen gestaltet.

Jeder Doppelgänger erhält hierfür einen individuellen Part, mit dem er jeweils zum Gesamtergebnis beiträgt. Dadurch wird sowohl die Arbeit an sich als auch eine gewisse Gruppendynamik symbolisiert und somit die eingeschränkten Interaktionsmöglichkeiten kaschiert.

Diese typischen Bürotätigkeiten und Geräusche stellen mögliche Komponenten dar:

- Ausdrucke anfertigen
- Bürostühle verschieben
- Computerlautsprecher einschalten
- Kaffeemaschine betätigen
- Kataloge durchblättern
- Notizen anfertigen
- Papier lochen und tackern
- Papierstapel ordnen
- Papier zerreißen
- Papier zerschneiden
- Schubladen öffnen und schließen
- Tastatur- und Mausklicks
- Telefonklingeln
- Wasser sprudeln

Das Tempo für den Rhythmus ist auf 140 Bpm festgesetzt. Dies soll unterstreichen, dass die Arbeit bis morgens erledigt sein muss und deshalb eine schnelle Arbeitsweise gefordert ist. Die folgende Abbildung zeigt den Grundrhythmus:



Abbildung 6: Pop-/Rock-Groove im 4/4-Takt (Eisenhauer, 2007, S. 20)

Damit die Doppelgänger fokussiert arbeiten und sich nicht ablenken lassen, wurde er alle zwei Takte durch einen Doppelschlag der Bass Drum um eine treibende Komponente ergänzt.

Da die Traumsequenz aus vier Videoaufnahmen besteht, war auf die Umsetzbarkeit der Geräusch-Choreografie zu achten. Es wurden die oben genannten Komponenten in Betracht gezogen, die sich teilweise jedoch nicht sinnvoll einsetzen ließen. Beispielsweise benötigt die Kaffeemaschine nach der Betätigung zu viel Zeit und ihr Geräusch passt nicht in den Takt. Durch die klar beschränkten Spielflächen ergeben sich weitere Einschränkungen hinsichtlich der zu verwendenden Komponenten: Obwohl es sich um akustisch reizvolle Geräusche handelt, wurde deshalb darauf verzichtet, Bürostühle zu verschieben, Wasser zu sprudeln oder

Schubladen zu öffnen. Die finalen Hauptbestandteile sind Tastatur, Locher und Papier. Weitere ausgewählte Komponenten setzen Akzente. Der Protagonist trägt mit Schnarch-Geräuschen ebenfalls zum Rhythmus bei. Das geplante Telefonklingen verschiebt sich zur besseren Realisierung in den Übergang zwischen Traum und Realität.

Die nachfolgende Tabelle ordnet den verschiedenen Rollen ihren Part zu. Insgesamt ergibt sich ein Rhythmus im 4/4-Takt mit insgesamt 48 Takten in der Zählmethode „1 + 2 + 3 + 4 +“. Die Ziffern 1 bis 4 stehen hier für die Viertelschläge. Die Achtelschläge werden mit „und“ verbalisiert – zur besseren optischen Unterscheidung bietet sich stattdessen das Plus-Zeichen an:

Rolle	Geräusch	Rhythmik
Doppelgänger 1	Tastatur-Tippen (Hi-Hat)	1 + 2 + 3 + 4 + und 1, 2, 3, 4 im Wechsel je einen Takt
Doppelgänger 2	Papier lochen (Snare Drum)	Grundsätzlich auf 2, 4
Doppelgänger 3	Papier stapeln (Bass Drum)	1, 3 und alle zwei Takte zusätzlich auf 4 +
Veit	Schnarchen (Akzente)	Alle acht Takte auf 1

Tabelle 2: Zuordnung der Geräusche und Rhythmik zu den Rollen

Der entstandene Beat wird durch selbstironische Texte je zweimal seitens der Doppelgänger passend zum Takt erweitert. Dafür unterbrechen die Doppelgänger ihre jeweilige Aufgabe:

Doppelgänger 1:

„Ja, natürlich können wir. Ja, sehr gerne machen wir.
Nein, es ist kein Problem!“

Doppelgänger 2:

„So kurz vor Ende ändern?
Ja immer doch, das geht doch noch.“

Doppelgänger 3:

„Ein Problem, nein kein Problem. Zahl’n Sie ruhig die Hälfte.
Es war wie immer alles nur ein Klacks.“

Alle Doppelgänger:

„Alles für die Kunden, alles für den Job!“

Akzente ergeben sich durch das Anschalten des Lautsprechers und das Zerreißen des Papiers. Die entsprechende Zählmethode für den letzten Punkt ist „1, 1, 1, 2, 3, 4“ unter Berücksichtigung der jeweiligen Pausenzeiten. Im Wechsel zwischen den Geräusch-Komponenten und Sprechparts entstehen kleine Lücken seitens der Doppelgänger, die durch eine Schlagzeugaufnahme im Hintergrund kompensiert werden sollen. Zur Realisierung wurde eine auditive Referenzaufnahme der Choreografie mit Metronom erstellt, die während der Dreharbeiten zu nutzen war.

In der Postproduktion wird die Referenzaufnahme durch nachproduzierte Geräusche ersetzt und für das Endprodukt aufbereitet. Die Verteilung der Doppelgänger auf die verschiedenen Arbeitsplätze ermöglicht, die jeweiligen Komponenten sinnvoll im 360°-Raum zu platzieren. Konkret werden die rhythmischen Elemente in diesem Fall wie Musik behandelt. In Bezug auf die in Abschnitt 5.2.4 aufgezeigten Möglichkeiten werden Locher, Tastatur und Papier als Instrumente interpretiert, die von den Doppelgängern gespielt werden. Da für das Publikum somit eine akustische Bühne entsteht, ist der Ansatz plausibel, die Klangobjekte entsprechend der visuellen Eindrücke anzuordnen. Die Schlagzeugaufnahme wird den Lautsprechern am Arbeitsplatz von Doppelgänger 1 unter Berücksichtigung von Abschnitt 5.2.5 zugeordnet und somit in den 360°-Raum überführt.

Zusammenfassend ist festzuhalten, dass die Traumsequenz einen wesentlichen Bestandteil des 360°-Promotion-Videos darstellt, die verschiedene Ansätze nutzt, um Klangobjekte im 360°-Raum zu positionieren. Um hier bewusst das Potenzial von Spatial Audio aufzuzeigen, wird während der Rhythmus-Performance auf den Einsatz von ergänzenden Head-Locked Tonspuren verzichtet. Die Entwicklung des Traums erfolgte stets im Sinne des Ziels, die spezifischen Kompromisse kreativ zu überspielen. Eine besondere Herausforderung stellte hierbei die bestmögliche Nutzung des Raumes bei gleichzeitiger Reduzierung personeller Abhängigkeiten in Verbindung mit den technischen Vorgaben dar. Durch die Gestaltung einer Rhythmus-Choreografie wurde final die Möglichkeit geschaffen, den Doppelgängern Raum zur auditiven Interaktion zu bieten.

6 Realisierung der Vertonung eines 360°-Promotion-Videos

Dieses Kapitel gibt Einblick in die Umsetzung des 360°-Videos. Im Speziellen werden die zentralen Aspekte zur Tonaufnahme und -nachbearbeitung sowie der Aufbereitung des fertigen Videos für die Zielmedien näher erläutert. Die Videonachbearbeitung wurde zwar ebenfalls übernommen, soll aber nicht im Detail beschrieben werden.

6.1 Equipment

Das Produktionsequipment war sorgfältig auf die Rahmenbedingungen abgestimmt. Da beispielsweise der herkömmliche Gebrauch einer Tonangel aufgrund der späteren Rundumsicht im Video nicht infrage kam, wurden zur Aufnahme der Sprache hochwertige Funk-Ansteckmikrofone verwendet. Die folgende Tabelle zeigt die Hauptbestandteile des Equipments:

Equipment	Beschreibung	Verwendung
6 x GoPro Hero 4 Black	Actioncams	Aufnahme des 360°-Videos
1 x Freedom 360 Explorer	360°-Rig	Aufnahme des 360°-Videos
2 x Sennheiser MKE-40	Ansteckmikrofone für ew-Serie, Nierencharakteristik	Sprachaufnahmen während der Dreharbeiten
2 x Sennheiser ew 100 G4	Funkstrecken-Set, ew-Serie	Sprachaufnahmen während der Dreharbeiten
1 x Sennheiser ME-62	Mikrofonmodul, K6-System, Kugelcharakteristik	Stützaufnahmen während der Dreharbeiten
1 x Sennheiser ME-64	Mikrofonmodul, K6-System, Nierencharakteristik	Separate Geräuschaufnahmen für Nachvertonung
1 x Zoom H6	mobiler Audiorecorder	Sprachaufnahmen während der Dreharbeiten
1 x Zoom H4n	mobiler Audiorecorder	Stützaufnahmen während der Dreharbeiten, separate Geräuschaufnahmen
1 x Beyerdynamic DT-770 Pro	Studiokopfhörer von Beyerdynamic	Abhören der Audioaufnahmen, binaurale Mischung

Tabelle 3: Produktionsequipment für Video- und Tonaufnahmen

Insbesondere der Zoom H6 überzeugte durch Flexibilität, da er nicht nur ein mobiler Audio-recorder ist, sondern auch als Interface für Multitrackaufnahmen mit einem Computer verbunden werden kann. Mit den entsprechenden Zusatzmodulen lassen sich sechs XLR-Kanäle zeitgleich aufzeichnen oder vier Spuren um ein XY- bzw. MS-Modul erweitern. Der Aufnahmepegel ist stufenlos je Kanal wählbar. Mit dem Zoom H6 wurden während des Drehs die Signale der extern angemieteten Sennheiser MKE-40 Ansteckmikrofone mit Nierencharakteristik aufgenommen und des Weiteren Rhythmusaufnahmen durchgeführt. Der Zoom H4n war zur Aufnahme des Stützmikrofons und als Hauptrecorder für die Geräuschproduktion vorgesehen. Um für den Referenzton ungerichtete Signale von allen Seiten zu erhalten, wurde das Sennheiser ME-62 mit Kugelcharakteristik und dem zugehörigem Phantomspeiseadapter herangezogen bzw. für die gerichteten Geräuschaufnahmen das ME 64-Modul mit Nierencharakteristik. Der professionelle Studiokopfhörer DT-770 Pro von Beyerdynamic eröffnete eine ideale Abhörmöglichkeit insbesondere für das Sounddesign, denn aufgrund seiner speziellen Diffusfeldentzerrung bietet er optimale Bedingungen für die binaurale Audiowiedergabe (vgl. Abschnitt 3.3.1).



Abbildung 7: Produktionsequipment für die Video- und Tonaufnahmen (Foto: Simon Lamche)

6.2 Video- und Tonaufnahmen

Die Aufnahmen für das 360°-Video fanden in drei Produktionseinheiten statt. Die erste Einheit diente zur Aufzeichnung des Videomaterials und zur Aufnahme der Sprechparts. In der zweiten und dritten Einheit wurden die Foley- und Rhythmusaufnahmen für das Sounddesign angefertigt. Um den Anforderungen des späteren Endproduktes gerecht zu werden, galt es in jedem Fall die produktionstechnischen Rahmenbedingungen (vgl. Abschnitt 5.3) zu berücksichtigen. Durch die Umsetzung in mehreren Produktionseinheiten konnten im Speziellen personelle Abhängigkeiten reduziert werden. Im Folgenden werden die zentralen Aspekte aller Produktionstage betrachtet.



Abbildung 8: Protagonist Veit Stephan während der Dreharbeiten (Foto: Simon Lamche)

Damit im fertigen 360°-Video mehrere Doppelgänger derselben Person gleichzeitig im Bild zu sehen sein können, musste für die Videoaufnahmen sichergestellt werden, dass sich beim Dreh weder die Kamerapositionen noch die Lichtverhältnisse ändern. Daraus ergab sich eine szenenweit statische Position der Kameras sowie Dreharbeiten bei Nacht mit geschlossenen Jalousien und Deckenbeleuchtung.

Um in Folge der statischen Kameraperspektive Jump-Cuts zu vermeiden, wurden die Aufnahmen jeweils als One-Shots umgesetzt und nicht, wie bei klassischen Filmproduktionen üblich, in mehreren Einstellungen mit wechselnden Kamerapositionen und Blickwinkeln. Zudem musste präzise auf die Einhaltung der im Vorfeld definierten Spielflächen geachtet werden, da die Bewegungsfreiheit durch den technischen Aufbau des 360°-Rigs und der Größe des Büros eingeschränkt war. Im fertigen Video werden Schnitte durch Überblendungen kaschiert.

Für den Ton erwies sich die Entscheidung als hilfreich, während der ersten Produktionseinheit hauptsächlich die Sprachaufnahmen mit Funk-Ansteckmikrofonen aufzunehmen und weitere Tonaufnahmen von der Videoproduktion zu trennen. Dieses Vorgehen ermöglichte die Reduzierung des gleichzeitig benötigten Equipments. Somit war es nicht notwendig, die erforderlichen Mikrofone und Kabel im Raum zu verstecken. Die Ansteckmikrofone wurden direkt an der Kleidung befestigt und die Signale mittels Funkstrecke zum H6-Recorder übermittelt, der sich außerhalb des Raumes befand. Zusätzlich wurde als Stütze für die Nachvertonung direkt unter der Kamera ein Mikrofon mit Kugelcharakteristik und dem zugehörigen H4n-Recorder platziert. Diese zeichneten die gesamten Dreharbeiten auf und wurden in der Nachbearbeitung zusammen mit dem Kamerastativ aus dem Videobild entfernt. Um die spätere Synchronisation zwischen Video und Originalton zu erleichtern, war stets eine Filmklappe einzusetzen.



Abbildung 9: 360°-Kamera und Stützmikrofon für Referenzaufnahmen (Foto: Simon Lamche)

Vor allem für Szene 2 war die Entscheidung dienlich, den auditiven Fokus während des Drehs auf die Sprachaufnahmen zu legen. Aufgrund der speziellen Rahmenbedingungen wird diese Szene aus mehreren Aufnahmen mit je einem Doppelgänger zusammengesetzt, die jedoch im Endprodukt zeitgleich im Bild zu sehen sind. Damit das Zusammenspiel aller Doppelgänger im fertigen Video gut harmoniert, wurde jeweils eine vorproduzierte Tonaufnahme der Sprech-Geräusch-Choreografie mit Metronom bei den Dreharbeiten eingespielt. Die Einspieler waren präzise auf die jeweiligen Sprechparts abgestimmt und optimiert, damit das Übersprechen auf die Ansteckmikrofone vermieden wird. Durch die spätere Nachvertonung der Geräusche kann das Übersprechen der Spielparts an dieser Stelle jedoch vernachlässigt werden.



Abbildung 10: Zoom Recorder zur Aufnahme der Funk-Ansteckmikrofone (Foto: Simon Lamche)

Mithilfe der Stützaufnahmen, die während der Dreharbeiten angefertigt wurden, konnten die für die zweite Produktionseinheit relevanten Geräusche identifiziert werden. Um im Endprodukt einen möglichst hohen auditiven Immersionsgrad zu erzielen, fand die Foley-Produktion im selben Büro statt, in dem zuvor das Video aufgenommen wurde. Der Recorder zeichnete mit einem Low-Cut-Filter bei einer Grenzfrequenz von 80 Hz auf. Störendem Trittschall, der ggf. über das Stativ auf das Mikrophon übertragen wird, konnte so vorgebeugt werden.

Die folgende Abbildung veranschaulicht den technischen Aufbau zur Nachvertonung am Beispiel der Tastaturgeräusche und bezieht sich stellvertretend auf die gesamte Soundproduktion:



Abbildung 11: Nachvertonung der Geräusche im Büro (Foto: Simon Lamche)

Zur Ergänzung der zweiten Szene durch Rhythmus wurde in der dritten Produktionseinheit eine Sammlung selbsterstellter Schlagzeugaufnahmen angefertigt. Diese umfasst neben dem Grundrhythmus einige Fill-ins und Akzente. Im Sounddesign konnten daraus passende Teile für die Integration in den Sprech-Geräusch-Rhythmus gewählt werden. Dieses Vorgehen erhöht zudem die Flexibilität in der Tongestaltung, da an dieser Stelle Material aus externen Soundbibliotheken vermieden wird.

6.3 Tonpostproduktion

In den Abschnitten 5.4.2 und 5.4.3 sowie im Kapitel 5.5 wurden die zentralen Aspekte der organisatorischen und gestalterischen Planung dargelegt. Die nachfolgenden technischen Erläuterungen ergänzen diese Ausführungen hinsichtlich des Vorgehens in der Postproduktion.

Die Nachbearbeitung des Tons fand hauptsächlich in Reaper unter Verwendung der Facebook 360 Spatial Workstation statt. Da für die Erstellung des Geräusch-Rhythmus und der Schlagzeugaufnahmen die DAW Logic Pro X von Apple genutzt wurde, war es wiederum sinnvoll, diese Aufnahmen dort aufzubereiten und sie im vorbearbeiteten Zustand in das Reaper-Projekt zu integrieren. Dabei war stets eine Abtastrate von 48 kHz zu wählen. Das gleiche Vorgehen war für die Filterung der Sprachaufnahmen aus der Traumsequenz anzuwenden, da zum Beispiel der Noise-Gate-Effekt in Logic bessere Ergebnisse lieferte als in Reaper. So konnten übersprochene Komponenten von den Dreharbeiten weitgehend entfernt werden, die sich bei der Aufzeichnung durch eingespielte Referenzaufnahmen der Traumsequenz ergaben.

Nach Abschluss des Stitching-Prozess und des Videoschnitts wurde ein Vorschauvideo mit grober Maskierung als Referenz für das Sounddesign gerendert und in der Spatial Workstation hinterlegt. Um die Performance im Panning-Prozess zu erhöhen, empfiehlt es sich, das Video im „DNxHR LB“-Format einzubinden. Zur Konvertierung wurde das Kommandozeilen-Programm „FFmpeg“ herangezogen. Hierfür war nach erfolgreicher Installation der folgende Terminal-Befehl auszuführen:

```
ffmpeg -i INPUTFILE -map 0:v -an -c:v dnxhd -pix_fmt yuv422p -trellis 0 -  
profile:v dnxhr_lb -y OUTPUTFILE.mov
```

Bei Betrachtung des Reaper-Projekts fällt die strukturelle Unterteilung der Audiospuren in 3D- und Head-Locked Audio auf. Diese wurden grundsätzlich getrennt voneinander behandelt und erst im Muxing-Prozess zusammengefügt. Sämtliche Tracks mit Spatial Audio benötigten das essenzielle Spatialiser Tool. Jeder dieser Spuren wurde wiederum an den 3D-Master gesendet. Das Control Plug-in diente der Unterstützung des binauralen Monitorings. Elementar ist, dass es nur je eine Instanz von beiden geben durfte. Weiterhin mussten die Spuren für Spatial Audio je 16 Kanäle anbieten, um mit TOA umgehen zu können sowie entsprechend korrekt geroutet werden. Effekte wie Equalizer oder Kompressoren waren in den 3D-Audio-Tracks hierarchisch vor dem Spatialiser einzubinden, da sonst klangliche Fehler auftraten. Alle Head-Locked Audiospuren wurden separat in den HL-Master geleitet, der mit der Masterspur einer Stereomischung gleichgesetzt werden kann (Facebook Audio 360, 2020).

Es sei angemerkt, dass die Wahl der Entfernung ohne Kenntnis über den realen Raum schwer gewesen wäre. Interessant ist, dass spontane Schätzungen im Vergleich zu den realen Werten in der 360°-Ansicht zu gering und in der equirektangulären Darstellung zu hoch angesetzt wurden. Darüber hinaus bietet das Plug-in die Option, die Position eines Klangobjekts dynamisch anpassen zu lassen, wenn sich das visuelle Objekt bewegt. Dieser Funktion wurde anfangs eine große Bedeutung zugemessen, während der Umsetzung stellte sich jedoch heraus, dass das Motion-Tracking Feature nicht zielführend funktionierte. Beispielsweise war die Verfolgung sehr ungenau und wich oft von der tatsächlichen Bewegung ab. Nach einigen erfolglosen Versuchen wurde für „Eine Nacht im Büro“ auf den Einsatz dieser Funktion verzichtet. Da jede Audiospur jedoch nur eine Instanz des Spatialiser nutzen kann, waren für größere Positionswechsel mehrere Spuren anzulegen.

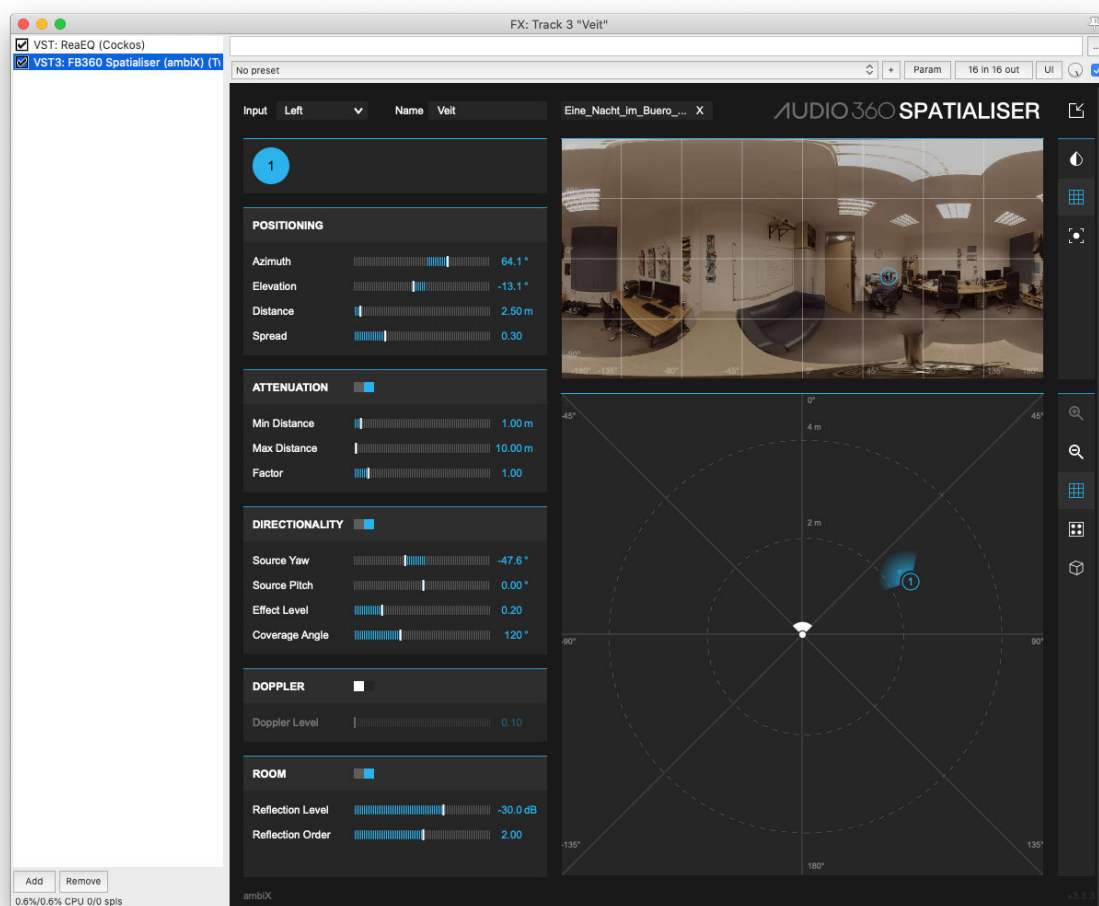


Abbildung 13: Screenshot des Spatialiser Plug-ins der Facebook 360 Spatial Workstation

Mit der Spread-Einstellung wurde festgelegt, ob sich das jeweilige Klangobjekt punktförmig oder diffus im Raum ausbreiten soll. Es bestand die Möglichkeit, eine optionale Dämpfung zu aktivieren und diese über eine minimale und maximale Entfernung zu steuern. Ferner waren gezielte Einstellungen über die Richtwirkung der Klangobjekte wichtig, da dies Schallquellen maßgeblich im Raum beeinflusst. Diesbezüglich war neben der Ausrichtung des Klangobjekts, die Tonhöhe und ein Winkel festzulegen, innerhalb dessen das Signal unverändert blieb und außerhalb des Winkels gedämpft wurde. Außerdem konnte die Effektstärke beliebig angepasst werden. Es sei angemerkt, dass im Speziellen dieser Effekt in Verbindung mit der Position im Raum großen Einfluss auf Wirkung des Head-Trackings ausübt.

Konkret galt es eine sinnvolle Kombination aus akustisch korrekter Anwendung und Erwartungen hinsichtlich des Werbefilms zu erzielen. Die Spatial Effekte sollten deutlich wahrgenommen werden, sich jedoch nicht übertrieben darstellen. Die Gefahr bei übermäßigem Einsatz wäre, dass bei binauraler Wiedergabe über Kopfhörer einer der Ohrhörer stellenweise kein Signal erhält und somit Irritationen seitens der Rezipierenden entstehen könnten. Es galt diesen Umstand abzuwägen und entsprechende Einstellungen zu wählen. Von der Nutzung des optional angebotenen Dopplereffekts wurde abgesehen, da sich konkret keine angemessene Möglichkeit ergab. Wichtig war jedoch das Einbeziehen des Räumlichkeitseffekts, der die Reflexionen innerhalb des Raumes simuliert. Hierfür verwendet die Facebook 360 Spatial Workstation eine Raummodellierung, die wiederum die ersten Reflexionsordnungen berücksichtigt. Zur korrekten Berechnung wurden die Abmessungen des Raumes im Control Tool hinterlegt.

Der Converter ermöglicht die flexible Umrechnung von einem Ambisonics-Format in ein anderes und ist zur Rotation des gesamten Signals gedacht. Es bietet sich an, dieses z. B. im 3D-Master zu platzieren, falls der gesamte Mix neu ausgerichtet oder zwischen TOA und SOA gewechselt werden soll.

Die Mischung wurde grundsätzlich wie in herkömmlichen Projekten realisiert. Die klanglich optimierten und räumlich platzierten Audiospuren waren mit den Fadern des Mixers und Automatisierungskurven aufeinander abgestimmt. Die Gesamtmischung wurde hinsichtlich ihrer Lautheit mit dem Loudness Plug-in gemessen. Durch ein separates Tool, das in den Head-Locked Master-Kanal integriert wurde, konnte die Lautheit des Stereotons in die Messung integriert werden. Für „Eine Nacht im Büro“ wurde die EBU Richtlinie R-128 herangezogen, da die Regulierung der Lautheit im Internet noch nicht einheitlich ist (vgl. Abschnitt 5.2.1).

Laut R-128 beträgt die integrierte Gesamtlautheit -23 LUFS mit einem maximalen True Peak von -1dBFS. Das Loudness Tool sollte im 3D-Master platziert werden und zeigt während der Messung neben den bereits erwähnten Informationen auch die momentane Lautheit sowie den sog. Short-Term an, der einen gemittelten Wert der letzten 3 Sekunden darstellt. Da die Messung jedoch unabhängig von Head-Tracking stattfindet, sind minimale Schwankungen im Endergebnis möglich (Facebook Audio 360, 2020).

Für den finalen Bounce der Mischung war darauf zu achten, dass das binaurale Monitoring im Control Plug-in deaktiviert wurde, da das Ambisonics-Rendering sonst nicht korrekt funktionieren würde. Das Gesamtsignal von 3D- und HL-Master war je separat im WAV-Format zu exportieren. Am einfachsten konnte dies durch die Solo-Funktion der entsprechenden Spur gewährleistet werden. Es sei darauf hingewiesen, dass für den Rendering-Prozess bezüglich Ambisonics der 3. Ordnung die Zahl der zu rendernden Kanäle manuell auf 16 eingestellt werden musste, da Reaper die Option nicht in der voreingestellten Auswahl anbot. Der Head-Locked Ton wurde wie für Stereo üblich in zwei Kanälen exportiert. Die Ausgabedateien lagen somit für 3D-Audio in TOA und Head-Locked Audio in Stereo vor. Parallel galt es, das 360°-Video durch Verfeinern der Masken, Retusche und Farbanpassungen fertigzustellen.

6.4 Aufbereitung für die Zielmedien

Anschließend konnten alle vorbereiteten Dateien zum Endprodukt zusammengefügt werden. Für das Muxing stand die Encoder Applikation zur Verfügung, die als eigenständiges Programm zur Facebook 360 Spatial Workstation gehört. Wie in Kapitel 5.3 beschrieben, sollte für das primäre Zielmedium Facebook das Matroska-Format verwendet werden. Hierfür war die entsprechende Einstellung in der Output-Sektion zu wählen. Weiterhin waren für Spatial- und Head-Locked Audio die vorbereiteten WAV-Dateien und das final aufbereitete Video in die jeweilige Sektion einzufügen. Nach dem erfolgreichen Muxing-Vorgang stand das 360°-Video im MKV-Format zur Verfügung, in dem die Streams für 3D- und Head-Locked Audio integriert sind. Es sei darauf hingewiesen, dass für Facebook auch der Export in TOA möglich ist. Darüber hinaus kann mit dem Encoder wahlweise nur Spatial- oder Head Locked Audio ausgegeben und Ambisonics-Dateien in FuMa, AmbiX und das TBE-Format überführt werden. Die für das Zielmedium benötigten Metadaten, werden im Muxing automatisch hinzugefügt. Sofern nicht Matroska als Ausgabeformat gewählt wurde, liegt das Video nach dem Muxing-Prozess im Format des Eingabevideos vor. Für gewöhnlich bieten sich MP4- oder MOV-Container an (vgl. Abschnitt 4.1.2). Darüber hinaus wird grundsätzlich eine Downmix-Option von HOA zu FOA für YouTube angeboten.

Bezüglich der Aufbereitung für YouTube sei darauf hingewiesen, dass der Encoder in der verwendeten Version der Facebook 360 Spatial Workstation v3.3.3. das Muxing von FOA, Head-Locked-Stereo und Video nicht korrekt umsetzte. Trotz richtig gewählter Einstellungen fehlten in der exportierten Datei die Spuren für Head-Locked Audio. Als Workaround wird von der Community empfohlen, alle Tonspuren manuell in eine Datei zusammenzuführen. Zur Vereinigung von Spatial und Head-Locked Audio wird das kostenlose Programm Audacity vorgeschlagen, das in den Einstellungen „erweiterte Misch-Optionen“ anbietet. Das Muxing kann mit FFmpeg über folgenden Terminal-Befehl durchgeführt werden:

```
ffmpeg -i INPUTFILE.mp4 -i "AMBIX+STEREO.wav" -map 0:v -map 1:a -  
c:v copy -c:a copy OUTPUTFILE.mov
```

Damit YouTube das 360°-Video richtig interpretiert, müssen dem Video vor dem Upload manuell Spatial Metadaten mit dem Google Metadata Injector angefügt werden (Farina, 2019).

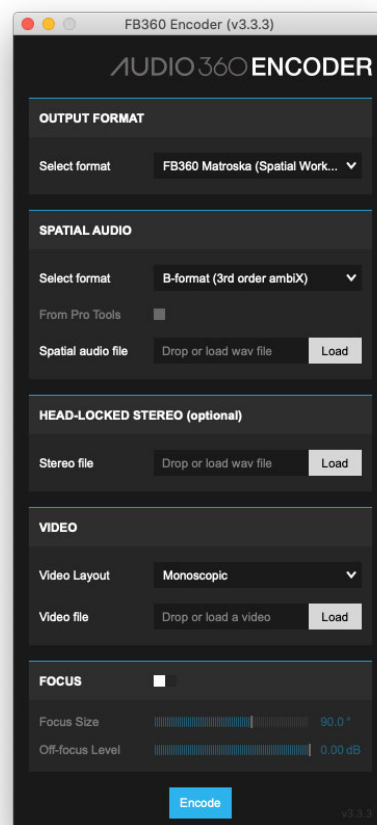


Abbildung 14: Screenshot des Encoders der Facebook 360 Spatial Workstation

Das primär für Facebook aufbereitete 360°-Video im MKV-Format kann im Standalone Video Player der Facebook 360 Spatial Workstation mit Spatial- und Head-Locked Audio sowie unter Berücksichtigung der Head-Tracking Funktion und binauraler Kopfhörerwiedergabe abgespielt werden. Optional lassen sich monoskopische und stereoskopische 360°-Videos auch im MP4- und MOV-Container starten, jedoch ohne Ton. In diesem Fall muss zusätzlich ein Audiofile im TBE-Format hinzu geladen werden. Das Abspielen anderer Tonformate ist derzeit leider nicht möglich. Ein interessantes Feature, das mit dem Video Player genutzt werden kann, ist hingegen die Fokus-Funktion. Dieser Effekt ist nicht standardmäßig in die Mischung eingeschlossen und wird bei der Wiedergabe dementsprechend in Echtzeit berechnet. Er kann jedoch optional direkt in das FB360 Matroska-Format integriert werden. Für die Fokus-Funktion wird ein Winkel in Kombination mit einem Dämpfungswert vergeben. Klangobjekte, die sich während der Wiedergabe außerhalb des Fokus befinden, werden somit um den festgelegten Wert gedämpft (Facebook Audio 360, 2020).

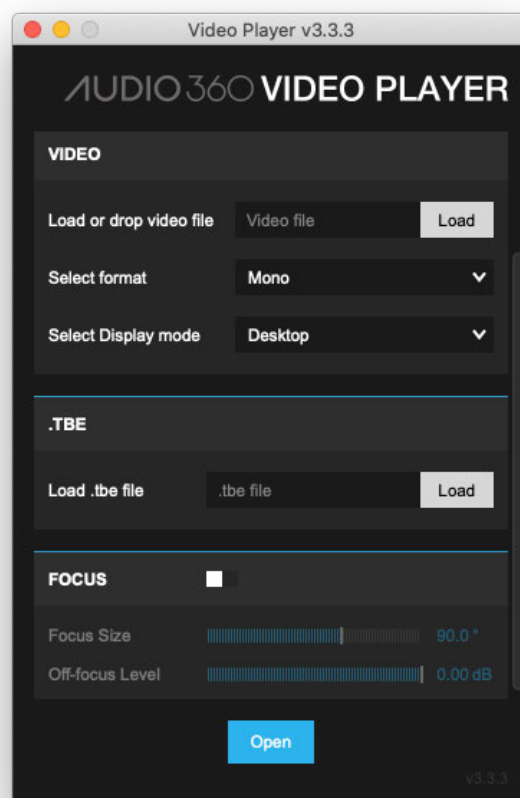


Abbildung 15: Screenshot des Video-Players der Facebook 360 Spatial Workstation

7 Fazit und Ausblick

Die vorliegenden Masterarbeit mit dem Titel „Adäquate Vertonung eines 360°-Promotion-Videos“ gibt Einblicke in die immersive Tongestaltung für ein vollsphärisches Video unter Verwendung der schallfeldbasierten Ambisonics-Technik. Den grundsätzlichen Erläuterungen zu den akustischen und technischen Hintergründen von dreidimensionalen Audiotechniken folgten detaillierte Ausführungen zu den Ambisonics-Verfahren und deren Einsatz in der Tongestaltung von 360°-Videos im Speziellen für die Zielmedien Facebook und YouTube. Nach der Diskussion spezifischer Besonderheiten wurde die Planung hinsichtlich der Vertonung des 360°-Promotion-Videos „Eine Nacht im Büro“ erarbeitet und schließlich in die praktische Umsetzung überführt.

7.1 Technisches Fazit und wissenschaftlicher Ausblick

Innerhalb der vergangenen vier Jahre wurde besonders intensiv an der Technik zur direkten Aufnahme in HOA geforscht. Inzwischen stehen Hauptmikrofone bis zur 4. Ordnung zur Verfügung. Weiterhin sind alternative Mikrofon-Arrays Gegenstand der Forschung. Dies zeigt das Bestreben der Tonbranche den Einsatz von 3D-Audio in virtuellen Welten auch in Zukunft weiter zu verbessern und den Umgang mit Higher Order Ambisonics zu fokussieren. Es steht bereits adäquate Software für das Panning und Mischen im 360°-Raum zur Verfügung, die den Umgang mit Ambisonics bis teils zur 7. Ordnung ermöglicht. Zur direkten Aufnahme werden jedoch zum aktuellen Zeitpunkt meist die klassischen Soundfield-Mikrofone herangezogen, mit denen sich Aufnahmen in First Order Ambisonics produzieren lassen (vgl. Kapitel 4.2 und 4.3). Entgegen der Kritik aus Fachkreisen, dass sich FOA aufgrund zu geringer Auflösung nicht für eine kompromisslose Hörimmersion in virtuellen Klangwelten eignet, setzt insbesondere YouTube auf dieses Format. Einerseits ist Spatial Audio somit einer breiten Masse zugänglich, andererseits wird dadurch die Veröffentlichung von Produktionen erschwert, die auf das Potenzial von Higher Order Ambisonics setzen. Es wurde ebenfalls darauf hingewiesen, dass das soziale Netzwerk Facebook TOA unterstützt (vgl. Abschnitt 4.1.2), derzeit jedoch am eigenen Matroska-Hybrid-Format als Empfehlung für Spatial Audio festhält. Auch über die Einführung von Ambisonics höherer Ordnungen bei YouTube sind derzeit nur Spekulationen möglich (vgl. Kapitel 4.3). Folglich fehlt vor allem ein international einheitlicher Standard in Bezug auf die Veröffentlichung von Ambisonics für 360°-Videos im Internet, der zu Gunsten der räumlichen Auflösung in Richtung HOA gehen sollte. Darüber hinaus scheint es derzeit nicht möglich zu sein im Speziellen die Kombination aus Spatial und Head-Locked Audio in lokalen

Playern wie GoPro VR Player oder VLC Media Player abzuspielen. Bei den eigenen Versuchen war entweder das Abspielen des Videos nicht möglich oder es fehlte die essenzielle Head-Tracking Funktionalität. Recherchen deuten lediglich auf die Unterstützung von 3D-Audio bzw. Ambisonics hin, ohne Hinweis auf das optionale Head-Locked Stereo (AudioXpress Staff, 2017) (GoPro, 2019). Dies erschwert das Testen der Endprodukte vor der Veröffentlichung und erscheint suspekt, da das Kombinieren von HOA und Head-Locked Stereo oft konzeptionell sinnvoll ist. Für die Zukunft wären lokale Videoplayer nützlich, welche beides anbieten. Es wird gefolgert, dass die adäquate Wiedergabe dieser Produktionen außerhalb des Internets eine Herausforderung darstellt und Lösungsansätze bislang nur begrenzt vorhanden sind. Mit dem Videoplayer der Facebook 360 Spatial Workstation kann z. B. die gewünschte Funktionalität hergestellt werden, aktuell jedoch nur für das TBE- und Matroska-Format (vgl. Abschnitt 4.1.2).

Für die Tonbranche ist 3D-Audio für virtuelle Welten aktueller denn je. Unter dem Gesichtspunkt, dass mit binauralem Ton bereits seit den 1930er Jahren experimentiert wurde und die Methoden zur Binauralsynthese heute maßgeblich zum Erfolg akustischer Head-Tracking-Verfahren und der binauralen Reproduktion von Spatial Audio beitragen (vgl. Kapitel 3.2), wird stetig an der Weiterentwicklung der immersiven Tontechnik geforscht. Vielversprechende Ansätze bietet z. B. das Unternehmen Gaudio Lab. Es greift bekannte Methoden auf und entwickelt Lösungen für interaktives 3D-Audio, das nicht mehr an die reine Kopfdrehung gebunden ist. Ziel ist die freiere Bewegung unter Einbeziehung der Vorteile des kanal-, objekt- und schallfeldbasierten Ansatzes mit „Spatial Upmixes“ in Echtzeit (Gaudio Lab, 2020). Der Grundgedanke, den Erfolg von dreidimensionalen Audiotechniken für virtuelle Welten nicht ausschließlich an schallfeldbasierten Methoden festzumachen, sondern die Vorteile der objektbasierten Verfahren stärker einzubeziehen, wird im Rahmen der vorliegenden Masterarbeit als wissenschaftlicher Ausblick festgehalten.

Darüber hinaus ist in Fachkreisen das zentrale Diskussionsthema weiterhin die Problemstellung der Limitierung von Ambisonics-Ordnungen. Im Sommer 2020 wurde auf dem online veranstalteten internationalen Kongress für Ton in Virtual und Augmented Reality der Audio Engineering Society ein Verfahren unter der Bezeichnung Bilateral Ambisonics vorgestellt. Dazu ergab eine kürzlich durchgeführte Studie, dass die bisherige binaurale Vorverarbeitung die Wirkung der sphärischen harmonischen Ordnungen zusätzlich reduziert. Deshalb sind die ohrbezogenen HRTFs in Bilateral Ambisonics wesentlich stärker in den binauralen Reproduktionsprozess einbezogen und mit einer neuen Ambisonics-Darstellung kombiniert.

Das Ziel ist die Auflösungsqualität in ordnungsbeschränkten binauralen Signalen deutlich zu verbessern. In Testläufen konnten bereits vielversprechende Ergebnisse erreicht werden (Ben-Hur et al., 2020, S. 1 ff.). Zwar muss sich das Verfahren erst etablieren, doch es soll im Rahmen der vorliegenden Masterarbeit ebenfalls als zukunftsweisender Ausblick herangezogen werden.

Ferner könnte sich das Medium-360°-Video im Allgemeinen verändern, zum Beispiel nach Vorbild der Photogrammetrie: Mit Hilfe vieler unterschiedlicher Fotos eines Raums kann eine spezielle Software ein nahezu fotorealistisches 3D-Modell berechnen, das anschließend in Virtual Reality weiterverarbeitet wird. Das Prinzip lässt sich grundsätzlich auf bewegte Bilder und Videokameras übertragen, ist bislang jedoch nur durch einen enormen Rechenaufwand zu realisieren. Trotzdem soll das „Volumetric Video“ als interessanter Ausblick für die Zukunft von vollsphärischen Videos festgehalten werden. (Orsolits et al., 2020, S. 272 f.).

7.2 Persönliches Fazit und gestalterischer Ausblick

Im persönlichen Fazit bezüglich der realisierten 360°-Vertonung ist im Speziellen die geringe Größe des Raumes anzumerken, in dem die Videoaufnahmen stattfanden. Die Raumgröße hatte kardinale Auswirkungen auf die visuelle und auditive Gestaltung. Dieser Umstand wurde zu Beginn der Arbeit unterschätzt. Allgemein wird empfohlen, die zu nutzende Handlungsfläche mit einem großzügigen Sicherheitsabstand zur Kamera zu wählen, um Stitching-Fehler zu reduzieren. Da in Bezug auf die Tongestaltung Wert auf die Nutzung des gesamten Raumes gelegt wurde, waren Kompromisse hinsichtlich der Kameraausrichtung nicht zu vermeiden. Der resultierende erhöhte Kamerastandpunkt hatte zur Folge, dass die Klangobjekte maßgeblich unterhalb der visuellen Position zu platzieren waren. Dass in einem größeren Raum die Kamera deutlich weiter entfernt zu den Spielflächen aufzustellen wäre, eröffnet die Chance, die Höhe an die Personen der filmischen Realität anzugleichen. Speziell der Immersionsgrad des Tons würde darin gestärkt, denn die Klangobjekte wären dann entsprechend passend auf Kopfhöhe oder über dem Kopfbereich zu positionieren.

Die Entscheidung, neben der Tongestaltung auch die Konzeption und Produktion des Videos maßgeblich zu übernehmen, sollte als Vorteil für den Ton gesehen werden (vgl. Kapitel 5.3). Tatsächlich trug dies im Speziellen positiv zur Entwicklung der Traumsequenz bei. Hier konnten die spezifischen Anforderungen für Bild und Ton des 360°-Promotion-Videos bereits in der Planung berücksichtigt werden, um die Produktion entsprechend der gestalterischen Ideen hinsichtlich des Tons auszurichten.

Durch die Umsetzung des Videos entstanden jedoch weitläufige Aufgabenfelder, die einerseits intensive Vorbereitungen bedingten und andererseits den Arbeitsfokus in gleicher Weise beanspruchten wie die Tongestaltung. Es wird in diesem Fall auf das besondere Verhältnis zwischen Bild und Ton hingewiesen: Um das selbsterstellte Video vertonen zu können, musste aus zeitkritischer Sicht zunächst der visuellen Komponente des 360°-Videos der Vortritt gewährt werden, da sie die Grundlage für die Vertonung darstellt. Darüber hinaus wurde die Vorproduktion und die Realisierung enorm von den corona-bedingten Einschränkungen beeinflusst. In Folge dessen konnte das ursprüngliche Videokonzept nicht realisiert werden. Die alternative Planung führte zu einem erheblichen organisatorischen und zeitlichen Mehraufwand.

Zwar konnten Ansätze für den Ton im Rahmen der Möglichkeiten erhalten und wegweisend zur Erstellung des finalen Videokonzepts genutzt werden, doch für die praktische Umsetzung ergaben sich, trotz enger Abstimmung mit der Agentur, große Herausforderungen: In Folge der Rahmenbedingungen war es notwendig, zeitgleich Video- und Tonaufnahmen zu koordinieren und während der Dreharbeiten neben der Aufzeichnung und Regie ebenso die Nebenrolle zu verkörpern. Die technischen Anforderungen einer 360°-Video-Produktion und das Videokonzept zur mehrfachen Abbildung einer Person im Raum führte zu speziellen Nachbearbeitungsschritten, die im Rahmen der vorliegenden Arbeit zwar angedeutet, jedoch zur Fokussierung der Themenstellung nicht detailliert ausgeführt wurden. Dennoch wird als persönliches Fazit festgehalten, dass sich der resultierende Gesamtaufwand zur Realisierung der vorliegenden Masterarbeit sehr hoch darstellte. Weiterhin ist anzumerken, dass die dreidimensionale Tongestaltung für 360°-Videos als junge Disziplin empfunden wird, die sich noch im Aufbau befindet und deshalb bislang maßgeblich durch das Hinterfragen und Anpassen herkömmlicher Vorgehensweisen bedingt ist. Spezifische Fachliteratur, die sich explizit auf Sounddesign für 360°-Videos bezieht, ist mit Blick in die Zukunft ausdrücklich erwünscht.

Insgesamt hat sich die Strategie zur Nachvertonung von 360°-Videos bewährt. Die Tonaufnahmen der Dreharbeiten und Foley-Produktion wurden als herkömmliche Mono- bzw. Stereo-Produktion durchgeführt und mit entsprechender Software räumlich zugeordnet, um einen dreidimensionalen Gesamtmix zu erzeugen. Grundsätzlich sei darauf hingewiesen, dass der zusätzliche Einsatz eines Soundfield-Mikrofons zur Erzeugung von 3D-Atmo situativ zu empfehlen ist, z. B. wenn sich die postproduktive Vertonung hauptsächlich auf nichtdiegetische Elemente fokussiert. Im konkreten Fall wurde zugunsten der Erwartungen bezüglich des Werbefilms auf den Einsatz von Atmo in Originalton verzichtet. Grundsätzlich kann der Gestaltungsspielraum durch die bewusste Kombination von Haupt- und Stützmikrofonen sowie teils

nachvertonter Komponenten wesentlich zur Steigerung des klanglich immersiven Potenzials beitragen. Positiv ist der Funktionsumfang der Spatial Workstation anzumerken: Die räumliche Platzierung der Tonsignale mittels Vorschauvideo und Einbeziehen der Raumeigenschaften in den Ambisonics-Mix wurde anhand binauraler Umrechnung und einem integrierten Tool zur Messung der Lautheit deutlich erleichtert. Im Rahmen des persönlichen Fazits wird als schade empfunden, dass im Speziellen die Motion-Tracking Funktion des 3D-Panners, um Audiosignale bewegten Objekten zuzuordnen, in den eigenen Tests keine zufriedenstellende Ergebnisse lieferte. Ein weiterer Kritikpunkt ist der derzeit fehlerhafte Export für YouTube, der jedoch durch einen Workaround zu umgehen ist (vgl. Kapitel 6.4). Es überwiegt der Vorteil, dass die vielfältigen Plug-ins sowie alle notwendigen externen Werkzeuge für die Bereitstellung des Endprodukts kostenlos nutzbar sind und Reaper verhältnismäßig günstig ist. Diese Kombination wird deshalb im Rahmen der vorliegenden Masterarbeit zur adäquaten Vertonung eines 360°-Promotion-Videos unter Berücksichtigung aller Rahmenbedingungen empfohlen.

Als gestalterischer Ausblick wird festgehalten, dass zur Überprüfung dieser Empfehlung und Optimierung des Workflows andere Plug-ins getestet werden sollen, z. B. hinsichtlich der erwähnten Motion-Tracking-Methode, zur Gestaltung des Klangraums mit Faltungshall und den Exportmöglichkeiten für verschiedene Zielmedien. Es sei ausdrücklich darauf hingewiesen, dass tiefergehende Vergleiche zwischen verschiedenen Tools aufgrund der weitläufigen Arbeitsfelder und besonderen Rahmenbedingungen der praktischen Arbeit nicht mehr während der zur Verfügung stehenden Bearbeitungszeit möglich waren. Ferner soll zur Weiterverarbeitung von „Eine Nacht im Büro“ sowie in künftigen Produktionen TOA als Audioformat für Facebook genutzt werden, um die Räumlichkeit noch effektiver zur Gestaltung virtueller Klangwelten einzusetzen. Grundsätzlich bleibt abzuwarten, wann YouTube die Unterstützung von HOA anbietet. Ferner soll aus persönlichem Interesse die weiteren Entwicklungen bezüglich der in Abschnitt 4.1.2 aufgezeigten HOAST-Library zur Veröffentlichung von Ambisonics höherer Ordnungen verfolgt werden.

Zur Steigerung des visuellen Immersionsgrads des Videos ist in Betracht zu ziehen, künftige 360°-Videos in stereoskopischer Form zu produzieren, um die immersive Reproduktion über HMDs in Virtual Reality zu gewährleisten. Darüber hinaus sind in „Eine Nacht im Büro“ die Ansteckmikrofone an der Kleidung der agierenden Personen sichtbar. Persönlich wird dies im vorliegenden Fall nicht als störend empfunden, doch zugunsten des fiktionalen Inhalts sollten Mikrofone nach Möglichkeit komplett versteckt werden. Dies trägt ebenfalls zu einer besseren visuellen Immersion bei. Ausgenommen sind Dokumentationen, in denen Mikrofone ausdrücklich erwünscht sind.

Die abschließenden Schlussfolgerungen beziehen sich auf die in Kapitel 1 formulierte Frage: „Wie kann das Potenzial immersiver Audiotechniken effektiv zur Vertonung von 360°-Videos genutzt werden?“ Zur adäquaten Tongestaltung von 360°-Videos ist der Einsatz von 3D-Audio-Techniken in Kombination mit Head-Tracking-Verfahren und binauraler Aufbereitung für Kopfhörerwiedergabe zu favorisieren. Zwar akzeptieren die meisten Zielmedien herkömmliche Stereospuren, doch die ausschließliche Nutzung von Head-Locked-Audio verspielt die Chance auf ein dreidimensionales Hörerlebnis nach Vorbild des natürlichen Hörens, welches sich in Abhängigkeit der Kopfdrehung verändert. Im Sinne einer angemessenen Vertonung vollsphärischer Videos ist die Kombination aus Spatial und Head-Locked Audio legitim, wenn dadurch dem dreidimensionalen Gesamteindruck nicht geschadet wird. Hierfür ist insbesondere auf den sinnvollen Einsatz von diegetischen und nicht diegetischen Klangobjekten zu achten und bedarf des Versuchs im Rahmen der Produktionsmöglichkeiten. Das Zusammenspiel aus technischem Fachwissen und der Wahl geeigneter Methoden für die Aufnahme und Nachbearbeitung trägt ebenso zu einer stimmigen Tongestaltung bei, wie kreative Gestaltungsansätze unter Berücksichtigung der Besonderheiten, die im Gegensatz zu einer herkömmlichen Stereovertonung entstehen. Weiterhin sind die technischen Anforderungen der Zielmedien zu beachten, aus denen sich Einschränkungen ergeben können. Hinsichtlich des kommunikativen Kontrakts ist für die Tongestaltung von 360°-Videos auf eine schlüssige Hörimmersion zu achten, die den Rezipierenden das Eintauchen in virtuelle Klangwelten ermöglicht.

Die intensive Beschäftigung mit dem Thema „Adäquate Vertonung eines 360°-Promotion-Videos“ konnte das persönliche Interesse vertiefen und wird an anderer Stelle mit Sicherheit eine fachorientierte Fortsetzung erfahren.

8 Glossar

Im Folgenden werden themenspezifische Abkürzungen und Begriffe erläutert. Es sei darauf hingewiesen, dass auf die Erklärung allgemein üblicher Verkürzungen verzichtet wird.

360°-Rig	Halterung für Kameras zur Aufnahme von 360°-Videos (Loew-Albrecht, o. J.)
360°-Video	Vollsphärisches Video in equirektangulärer Darstellung, monoskopisch (2D) oder stereoskopisch (3D), Blickrichtung bei der Wiedergabe frei wählbar (Mazmany, 2019, S. 34 f.)
3D-Audio	Audioverfahren, die Klangobjekte mit Höheninformationen im dreidimensionalen Raum abbilden (Witte & Theile, 2016, S. 1), auch als „Immersive Audio“ oder „Spatial Audio“ bezeichnet
3D-Panning	Räumliches Platzieren von Tonaufnahmen im dreidimensionalen Raum durch Spatializer Plug-ins (Lane, 2019)
A-Format	Ambisonics-Komponenten nach Aufnahme mit einem Sound-field-Mikrofon (Roginska & Geluso, 2018, S. 279 f.)
Ambisonics	Schallfeldbasierte Audiotechnik zur physikalischen Synthese eines Klangfelds unter Berücksichtigung verschiedener harmonischer Ordnungen (Roginska & Geluso, 2018, S. 53)
AmbiX	Ambisonics-Konvention mit der Kanalreihenfolge W, Y, Z, X und SN3D-Normalization (Rodgers, 2020)
Atmo	Geräusche zur Erzeugung einer Grundstimmung, Verstärken des typischen Raumeindrucks (Friedrich, 2008, S. 89 f.)
B-Format	Vierkanaliges Ambisonics 1. Ordnung mit Schalldruck W und Druckgradienten X, Y, Z (Weinzierl, 2008, S. 660 ff.)

Binaurale Wiedergabe	Für Kopfhörer optimierte Tonwiedergabe unter Berücksichtigung der spezifischen HRTFs (Møller et al., 1995, S. 300)
Binauralsynthese	Erstellung von Computermodellen virtueller Klangräume, Faltung der Tonsignale mit HRTFs (Weinzierl, 2008, S. 679 f.), Verwendung zur binauralen Aufbereitung virtueller Lautsprecher-signale (Rumsey, 2001, S. 75 ff.)
BRIR	Gesamtheit der binauralen Impulsantworten eines Raumes (Weinzierl, 2008, S. 679 f.)
DAW	Digital Audio Workstation, umfassendes Audiowerkzeug für Aufnahme, Bearbeitung u. Mischung (Friedrich, 2008, S. 295)
Diegetischer Ton	Alle Klangobjekte mit Bezug zur filmischen Realität (Görne, 2017, S. 20)
Downmix	Heruntermischen von Mehrkanalton zur Anpassung für ein Zielmedium, z. B. Reduzierung einer Mischung von Surround zu Stereo oder von Ambisonics 3. Ordnung zur 1. Ordnung
FOA	First Order Ambisonics, Ambisonics 1. Ordnung
Foley-Produktion	Nachträgliches Erstellen von Geräuschen zu einem fertigen Videobild, gezielte Nachvertonung (Steppat, 2014, S. 207 f.)
FOV	Field of view, Sichtfeld in einer realen Situation, für 360°-Videos Blickfeld bzw. Field of regard (Unity 3D, o. J.)
FuMa	Ambisonics-Konvention mit der Kanalreihenfolge W, X, Y, Z und relative Dämpfung des W-Kanals um 3 dB (Rodgers, 2020)
Head-Locked Audio	Ton außerhalb des 360°-Raums, passt sich nicht der Kopfdrehung an, meist als Stereoton zur Ergänzung von Spatial Audio

Head-Tracking	Methoden zur Erfassung der Kopf und Halsposition (Unity 3D, o. J.), Steuerung der Richtung eines Hörereignisses in Abhängigkeit der Kopfbewegungen (Dickreiter et al., 2014, S. 346)
HMD	Head-Mounted-Display, VR-Headset, Wiedergabemedium für Virtual Reality mit Head-Tracking Unity (3D, o. J.)
HOA	Higher Order Ambisonics, Ambisonics höherer Ordnung
HRTF	Head-Related Transfer Function, spezifische Außenohrübertragungsfunktion zum Ausdruck von Resonanzeigenschaften des Außenohrs (Weinzierl, 2008, S.90)
IKL	Im-Kopf-Lokalisation, Hörereignisbildung zwischen dem linken und rechten Ohr (Görne, 2015, S. 130 f.)
Immersion	Empfindung in virtuelle Welten einzutauchen (Wulff, 2012)
LUFS	Loudness Units relative to Full Scale, Einheit zur Messung der Lautheit in audiovisuellen Medien, oft in Verbindung mit der EBU Richtlinie R-128 und -23 LUFS (NDR, 2020)
Matroska	Container zur Verschachtelung etlicher Codecs (Rinne, 2010), hier spezielles HOA-Hybrid-Format von Facebook für Spatial und Head-Locked Audio (Facebook, 2020)
Muxing	Zusammenfügen von Bild und Ton, Überführen in eine Datei
Nadir-Bild	Nach unten ausgerichtetes Bild einer 360°-Produktion, zeigt das Kamerastativ und wird meist retuschiert (Loën-Wagner, o. J.)
Nichtdiegetischer Ton	Alle Klangelemente außerhalb der filmischen Realität (Görne, 2017, S. 20)

O-Ton	Originalton
Phantomschallquelle	Eingebildeter Hörort für ein Hörereignis, in der Mitte zwischen zwei Lautsprechern (Blauert, 2014, S. 163)
Räumliches Hören	Richtungs- und Entfernungshören, Empfindung räumlicher Tiefe (Görne 2015, S. 125 ff.)
SOA	Second Order Ambisonics, Ambisonics 2. Ordnung
Soundfield-Mikrofon	Mikrofon zur Aufzeichnung eines 3D-Schallfelds im Ambisonics-Format, meist in Tetraeder-Form (Roginska & Geluso, 2018, S. 278 ff.)
Stitching	Zusammenfügen der Aufnahmen zu einem 360°-Video unter Berücksichtigung überlappender Bildbereiche (Lyu et al., 2019, S. 55 ff.).
Sweet Spot	Optimale Stereohörposition an der Spitze eines gleichseitigen Dreiecks (Dickreiter et al., 2014, S. 230)
TBE	Two Big Ears, spezielles Hybrid-HOA-Format mit acht statt neun Kanälen (Greif, 2016)
TOA	Third Order Ambisonics, Ambisonics 3. Ordnung
VBAP	Vector Base Amplitude Panning, Erweiterung des klassischen Panorama-Potentiometers für die Platzierung von Klangobjekten im dreidimensionalen Raum (Weinzierl, 2008, S. 567 ff.)
Virtual Reality	virtuelle Realität, Differenzierung in computergenerierte und filmische virtuelle Realität (Jaunt Team, 2018) unter Berücksichtigung des Immersions- und Interaktionsgrades (Orsolits et al., 2020, S. 264 ff.), Augmented Reality und Mixed Reality als erweiterte Realität (Aichner et al., 2019, S. 1 u. 5 f.)

9 Verzeichnisse

9.1 Tabellenverzeichnis

Tabelle 1: Zusammenhang Ambisonics A- und B-Format (Roginska & Geluso, 2018, S. 279 f.).....	23
Tabelle 2: Zuordnung der Geräusche und Rhythmik zu den Rollen	52
Tabelle 3: Produktionsequipment für Video- und Tonaufnahmen	54

9.2 Abbildungsverzeichnis

Abbildung 1: 360°-Rig und Actioncams zur Aufnahme von 360°-Videos (Foto: Simon Lamche)	4
Abbildung 2: Kopfbezogenes Koordinatensystem (Weinzierl, 2008, S. 88).....	10
Abbildung 3: Kunstkopf Neumann „KU 100“ (Foto: Simon Lamche).....	13
Abbildung 4: Kugelflächenfunktionen mit Kanalnummern (Zotter & Frank, 2019, S. 68).....	21
Abbildung 5: Soundfield-Mikrofone: Array , SPS200, MK4012, ST450 (Zotter & Frank, 2019, S. 11) ..	22
Abbildung 6: Pop-/Rock-Groove im 4/4-Takt (Eisenhauer, 2007, S. 20).....	51
Abbildung 7: Produktionsequipment für die Video- und Tonaufnahmen (Foto: Simon Lamche)	55
Abbildung 8: Protagonist Veit Stephan während der Dreharbeiten (Foto: Simon Lamche)	56
Abbildung 9: 360°-Kamera und Stützmikrofon für Referenzaufnahmen (Foto: Simon Lamche)	57
Abbildung 10: Zoom Recorder zur Aufnahme der Funk-Ansteckmikrofone (Foto: Simon Lamche) .	58
Abbildung 11: Nachvertonung der Geräusche im Büro (Foto: Simon Lamche).....	59
Abbildung 12: Screenshot der Projektdatei mit Spatial Audio (rot) und Head-Locked Audio (blau) .	61
Abbildung 13: Screenshot des Spatialiser Plug-ins der Facebook 360 Spatial Workstation.....	62
Abbildung 14: Screenshot des Encoders der Facebook 360 Spatial Workstation	65
Abbildung 15: Screenshot des Video-Players der Facebook 360 Spatial Workstation.....	66

9.3 Literaturverzeichnis

Aichner, T., Maurer, O., Nippa, M., & Tonezzani, S. (2019). *Virtual Reality im Tourismus: Wie VR das Destinationsmarketing verändern wird*. Springer Gabler.

AudioEase. (2020). *360pan suite, full control over your 360 (spatial, ambisonics) audio with this professional audio plug-in suite*.

<https://www.audioease.com/360> (abgerufen am: 12.09.2020)

AudioXpress Staff. (2017, Dezember 22). *3D Sound Labs and VideoLAN Partner on 3D Audio for VLC Media Player*. AudioXpress.

<https://audioxpress.com/news/3d-sound-labs-and-videolan-partner-on-3d-audio-for-vlc-media-player> (abgerufen am: 18.09.2020)

AZ QUOTES. (o. J.). *George Lucas Quote*. A-Z Quotes,

<https://www.azquotes.com/quote/812678> (abgerufen am: 20.09.2020)

Bastian, M. (2018, Oktober 16). Jaunt steigt aus dem VR-Geschäft aus und entlässt viele Mitarbeiter. *MIXED / News zu VR, AR und KI*.

<https://mixed.de/jaunt-steigt-aus-dem-vr-geschaeft-aus-und-entlaesst-viele-mitarbeiter/> (abgerufen am: 14.09.2020)

Ben-Hur, Z., Alon, D. L., Mehra, R., & Rafaely, B. (2020, August 17). *Binaural Reproduction using Bilateral Ambisonics*. AES International Conference on Audio for Virtual and Augmented Reality, Online.

Blauert, J. (2014). *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) (1. Aufl.)*. S. Hirzel Verlag. <http://nbn-resolving.de/urn:nbn:de:101:1-201508062580>

Blue Ripple Sound. (o. J.). *O3A Core*. Blue Ripple Sound.

<http://www.blueripplesound.com/products/o3a-core> (abgerufen am: 12.09.2020)

Breebaart, J., & Faller, C. (2007). *Spatial audio processing: MPEG surround and other applications*. John Wiley & Sons.

Bruns, K. (2012, Oktober 13). *Off-Sprecher—Lexikon der Filmbegriffe*.

<https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=7864> (abgerufen am: 11.09.2020)

Cast / Audiovisual Media. (2016, Januar 15). *[360° KURZFILM] Gegen die Regeln—Erster fiktionaler VR-Film der Schweiz!*

<https://www.youtube.com/watch?v=69WI2-RtgqM> (abgerufen am 18.09.2020)

- DearReality. (2020). *DearVR PRO*. Dear Reality.
<https://www.dearvr.com/products/dearvr-pro> (abgerufen am 12.09.2020)
- Dickreiter, M., Dittel, V., Hoeg, W., & Wöhr, M. (2014). *Handbuch der Tonstudiotechnik* (8., überarbeitete und erweiterte Auflage). De Gruyter Saur.
- Dieterle, H., Sagnowski, K., & Wittek, H. (2018, April 15). *Hannes Dieterle, K.Sagnowski, H. Wittek (SCHOEPS): Mikrofontechnik für VR*.
<https://www.youtube.com/watch?v=ut9rvTsxeEY> (abgerufen am 12.09.2020)
- EDEKA. (2015, November 28). *EDEKA Weihnachtsclip—#heimkommen*.
<https://www.youtube.com/watch?v=V6-0kYhqoRo> (abgerufen am 11.09.2020)
- Ederhof, A. (2006). *Das Mikrofonbuch: Optimaler Einsatz im Studio und auf der Bühne* (2. Aufl.). GC Carstensen.
- Eisenhauer, G. (2007). *Bum Bum Tschack – Die neue umfassende Schlagzeugmethode für den Anfang* (1. Aufl., Bd. 1). Edition DUX.
- F.A. Brockhaus (Hrsg.). (1986). *Brockhaus Enzyklopädie: In vierundzwanzig Bänden: Bd. 1 (A-APT)* (19. völlig neu bearbeitete Aufl.). F.A. Brockhaus GmbH.
- Facebook Audio 360. (2020). *Introduction—Facebook Audio 360 documentation*.
<https://facebookincubator.github.io/facebook-360-spatial-workstation/Documentation/SpatialWorkstation/SpatialWorkstation.html#components> (abgerufen am 12.09.2020)
- Facebook. (2020). *Wie lade ich ein 360°-Video auf Facebook hoch? / Facebook-Hilfereich*
<https://de-de.facebook.com/help/828417127257368> (abgerufen am 12.09.2020)
- Farina, A. (2017, August 22). *Ambisonics to TBE conversion*.
<http://pcfarina.eng.unipr.it/TBE-conversion.htm> (abgerufen am 03.09.2020)
- Farina, A. (2019). *Ambix + HL for YouTube*.
<http://www.angelofarina.it/Ambix+HL.htm> (abgerufen am: 15.09.2020)
- Flückiger, B. (2017). *Sound Design: Die virtuelle Klangwelt des Films* (6. Auflage). Schüren.
- Friedrich, H. J. (2008). *Tontechnik für Mediengestalter: Töne hören - Technik verstehen - Medien gestalten*. Springer.
- Gaede, W. (2002). *Abweichen ... von der Norm: Enzyklopädie kreativer Werbung*. Wirtschaftsverl. Langen-Müller/Herbig.
- Gaudio Lab. (2020). AR VR and Immersive • Gaudio Lab. *Gaudio Lab*.
<https://gaudiolab.com/solution-ar-vr-and-immersive/> (abgerufen am 12.09.2020)

- Google Spotlight Stories. (2016, Mai 20). *360 Google Spotlight Stories: Pearl*. <https://www.youtube.com/watch?v=WqCH4DNQBUA> (abgerufen am: 18.09.2020)
- GoPro. (2019, Oktober 1). *FAQ zum GoPro VR-Player für den Desktop*. <https://community.gopro.com/t5/de/FAQ-zum-GoPro-VR-Player-f-252-r-den-Desktop/tap/394080> (abgerufen am 18.09.2020)
- Göritz, A., & Züger, S. (2016, August 15). *Die Revolution im emotionalen Markenerlebnis*. kress. <https://kress.de/news/detail/beitrag/135692-immersives-marketing-die-revolution-im-emotionalen-markenerlebnis.html> (abgerufen am 05.08.2020)
- Görne, T. (2004). *Mikrofone in Theorie und Praxis* (7. Aufl.). Elektor-Verl.
- Görne, T. (2015). *Tontechnik: Hören, Schallwandler, Impulsantwort und Faltung, digitale Signale, Mehrkanaltechnik, tontechnische Praxis* (4., aktualisierte Auflage). Hanser.
- Görne, T. (2017). *Sounddesign: Klang, Wahrnehmung, Emotion*. Hanser.
- Greif, B. (2016, Mai 24). *Facebook kauft Virtual-Reality-Audiospezialisten Two Big Ears*. ZDNet.de. <https://www.zdnet.de/88270175/facebook-kauft-virtual-reality-audiospezialisten-two-big-ears/> (abgerufen am: 03.09.2020)
- hagebau Österreich. (2016, Juni 10). *Hagebaumarkt 360°* [YouTube]. <https://www.youtube.com/watch?v=1bxxMqHPy4M> (abgerufen am 05.08.2020)
- Hautz, C. (2020, Juli 2). *Mobil-Recorder mit App-Unterstützung—Der Zoom H8 Handy Recorder*. bonedo.de - Das Musikerportal. <https://www.bonedo.de/artikel/einzelansicht/mobil-recorder-mit-app-unterstuetzung-der-zoom-h8-handy-recorder.html> (abgerufen am 02.09.2020)
- Henze, C. (2005). *Apropos Werbefilm: Vom kommerziellen Umgang mit der Phantasie*. UVK Verl.-Ges.
- Hodges, P. (2011, Juni 3). *Ambisonic Info / Channel Formats*. <https://ambisonic.info/ambisonics/channels.html> (abgerufen am: 10.09.2020)
- Hodges, P. (2018, Dezember 10). *Ambisonic Info / Zoom H2 and H2n*. <https://ambisonic.info/zoomh2.html> (abgerufen am 02.09.2020)
- Hoppin' Team. (2020, Januar 6). *Here Are 8 Great Places To Share VR 360 Content. Hoppin' World*. <https://hoppin.world/8-great-places-where-to-share-vr-360-content/> (abgerufen am 13.09.2020)
- Idl, I. (2016, Juni 18). *Storytelling: Das Rezept für gute Geschichten. DIE WELT*. <https://www.welt.de/icon/design/article155995347/Heute-verkauft-man-Produkte-mit-guten-Geschichten.html> (abgerufen am 11.09.2020)

Institut für Elektronische Musik und Akustik. (2020). *HOAST Library*.
<https://hoast.iem.at/> (abgerufen am 18.09.2020)

Jaunt Team. (2018). *The Cinematic VR Field Guide*. Oculus for Creators.
<https://creator.oculus.com/learn/cinematic-vr-field-guide/> (abgerufen am 28.03.2020)

Kronlachner, M. (2014, Januar 25). *AmbiX v0.2.10 – Ambisonic plug-in suite / matthiaskronlachner.com*.
<http://www.matthiaskronlachner.com/?p=2015> (abgerufen am 12.09.2020)

Lane, C. (2019, September 1). *Why Stereo Recordings And Spatial Audio Don't Usually Mix*.
<https://blog.dearvr.com/overview/why-stereo-recordings-and-spatial-audio-dont-usually-mix> (abgerufen am 05.09.2020)

Loën-Wagner, D. v. (o. J.). Nadir-Bild | Fotografie Glossar. *DVLW*,
<https://www.dvlw.de/2010/12/nadir-bild/> (abgerufen am 05.08.2020)

Loew-Albrecht, D. (o. J.). *Wie funktionieren 360 Grad Kameras?* epicphoto.de,
<https://epicphoto.de/wie-funktionieren-360-grad-kameras/> (abgerufen am 05.08.2020)

Lyu, W., Zhou, Z., Chen, L., & Zhou, Y. (2019). A survey on image and video stitching.
trlua1 Reality & Intelligent Hardware, 1(1), 55–83.

Mazmanyan, D. (2019). *Virtual Reality, Augmented Reality & 360°-Videos*. Independently published. Printed in Germany by Amazon Distribution GmbH.

mh acoustics. (o. J.). *Em32 Eigenmike® microphone*. mh acoustics – Products,
<https://mhacoustics.com/products> (abgerufen am: 12.09.2020)

Møller, H., Sørensen, M. F., Hammershøi, D., & Jensen, C. B. (1995). Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, 43(5), 300–321.

Moon, M. (2019, Mai 18). *Google is shutting down the Jump VR platform in June*. Engadget.
<https://www.engadget.com/2019-05-18-google-shut-down-jump-vr.html>
(abgerufen am: 12.09.2020)

NDR. (2020, Januar 22). *Gleich laut im Netz*.
https://www.ndr.de/der_ndr/empfang_und_technik/Gleich-laut-im-Netz-,lautheit100.html
(abgerufen am 10.09.2020)

Orsolits, H., Lackner, M., & Springer Fachmedien Wiesbaden GmbH. (2020). *Virtual Reality und Augmented Reality in der Digitalen Produktion*. <https://doi.org/10.1007/978-3-658-29009-2>

- Overschmidt, G., & Schröder, U. B. (Hrsg.). (2013). *Fullspace-Projektion: Mit dem 360°lab zum Holodeck*. Springer Vieweg.
- Pionierfilm. (2020, Januar 31). *Was ist ein Off-Sprecher? Was macht ein Sprecher?*
<https://www.filmproduktion-werbefilm.de/was-ist-ein-off-sprecher-was-macht-ein-sprecher/>
(abgerufen am: 11.09.2020)
- Pirntke, G. (2006). *Expert Praxislexikon Marketing und Controlling*. expert-Verl.
- Raffaseder, H. (2002). *Audiodesign: [Kommunikationskette, Schall, Klangsynthese, Effektbearbeitung, akustische Gestaltung]; mit ... 32 Tabellen und einer CD-ROM*. Fachbuchverl. Leipzig im Carl-Hanser-Verl.
- Referenz Film. (o. J.). 360°-Videos abspielen. *Referenz Film*.
<https://www.referenzfilm.de/filmwissen/360-videos-abspielen/> (abgerufen am 08.08.2020)
- Rieger, M. (2019, Juni 15). *Virtual Reality Musik*. VRTONUNG - Virtual Reality Sound.
<https://www.vrtonung.de/virtual-reality-musik/> (abgerufen am 12.09.2020)
- Rinne, M. (2010, Mai 20). *MKV Container Format: Der neue Standard für HD-Videos*. CHIP Online.
https://www.chip.de/artikel/MKV-Container-Format-Der-neue-Standard-fuer-HD-Videos_140101929.html (abgerufen am 17.09.2020)
- Rodgers, J. (2020, Juni 1). *Ambisonic Formats Explained—What Is The Difference Between A Format And B Format*. Production Expert.
<https://www.pro-tools-expert.com/production-expert-1/2019/12/10/ambeo-and-ambisonic-formats-a-bluffers-guide> (abgerufen am: 03.09.2020)
- Roginska, A., & Geluso, P. (Hrsg.). (2018). *Immersive sound: The art and science of binaural and multi-channel audio*. Routledge, Taylor & Francis Group.
- Rumsey, F. (2001). *Spatial audio*. Focal Press.
- Schröder, J. (2017, April 12). *Test: Geniales 3D-Mikrofon Sennheiser Ambeo VR Mic*. LowBeats.
<https://www.lowbeats.de/test-geniales-3d-mikrofon-sennheiser-ambeo-vr-mic/>
(abgerufen am: 12.09.2020)
- Schröder, J. (2018, September 17). *Ambisonics-360-Grad-Fieldrecorder Zoom H3-VR*. LowBeats.
<https://www.lowbeats.de/news-ambisonics-360-grad-fieldrecorder-zoom-h3-vr/>
(abgerufen am: 06.09.2020)

Sennheiser. (2019, Januar 24). *Weltneuheit auf der NAMM 2019: Sennheiser und Dear Reality zeigen kompletten 3D-Audio-Workflow*. Sennheiser - Kopfhörer & Headsets - Mikrofone - Business Communications.

<http://de-de.sennheiser.com/news-weltneuheit-auf-der-namm-2019-sennheiser-und-dear-reality-zeigen-kompletten-3d-audio-workflow> (abgerufen am: 12.09.2020)

Shukla, U. (2018, Februar 20). *An introduction to 360° video*. Knight Lab Studio.

<https://studio.knightlab.com/results/storytelling-layers-on-360-video/an-introduction-to-360-video/> (abgerufen am: 05.08.2020)

SoKrispyMedia. (2018, Oktober 15). *Do Not Touch (360 VR)*. <https://www.youtube.com/watch?v=ecajlIKwICg> (abgerufen am: 18.09.2020)

Sound & Recording. (2019, Dezember 16). *8-Kapsel Spatial Mic – Debüt-Mikrofon von Voyage Audio*. SOUND & RECORDING.

<https://www.soundandrecording.de/equipment/8-kapsel-spatial-mic-debuet-mikrofon-von-voyage-audio/> (abgerufen am: 12.09.2020)

Steppat, M. (2014). *Audioprogrammierung: Klangsynthese, Bearbeitung, Sounddesign ; mit 66 Listings, 33 Tabellen ... und 48 Übungsaufgaben*. Hanser.

Technikum Wien Academy. (o. J.). *360 Grad Videos & Virtual Reality: Was ist das?* Technikum Wien Academy,

<https://academy.technikum-wien.at/ratgeber/360-grad-videos-virtual-reality-was-ist-das/> (abgerufen am 28.03.2020)

Unity 3D. (o. J.). *What is AR, VR, MR, XR, 360?* Unity,

<https://unity3d.com/de/what-is-xr-glossary> (abgerufen am: 08.03.2020)

Weinzierl, S. (Hrsg.). (2008). *Handbuch der Audiotechnik* (1. Aufl.). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-34301-1>

Wittek, H., & Theile, G. (2016). *Die Anwendung eines stereofonen Mehrkanalverfahrens für 3D-Audio und VR*. 29th Tonmeistertagung – VDT International Convention.

Wittek, H., & Theile, G. (2017, Oktober 18). Development and application of a stereophonic multichannel recording technique for 3D Audio and VR. *Convention Paper 9869*. 143rd Audio Engineering Society Convention, New York, NY, USA.

Wulff, H. J. (2012, Oktober 12). *Immersion—Lexikon der Filmbegriffe*.

<https://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=942> (abgerufen am: 05.08.2020)

YouTube. (2020). *Raumklang in 360°-Videos und VR-Videos verwenden—YouTube-Hilfe*.

<https://support.google.com/youtube/answer/6395969> (abgerufen am: 02.09.2020)

Zierner, T. (2019, September 28). *Dolby Atmos: So funktioniert das Soundsystem*.
https://praxistipps.chip.de/dolby-atmos-so-funktioniert-das-soundsystem_47338
(abgerufen am: 10.09.2020)

Zotter, F., & Frank, M. (2019). *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (Bd. 19). Springer International Publishing. <https://doi.org/10.1007/978-3-030-17207-7>

10 Anhang

10.1 Danksagung

Den Betreuern der vorliegenden Masterarbeit Herrn Prof. Dr. Dipl.-Ing. Maximilian Kock, Herrn M. Eng. Bernd Gerlang-Gradl sowie Herrn M. Eng. Veit Stephan sei für die fachliche Diskussion zum Thema und der Koordination im Firmenkontext gedankt.

Darüber hinaus richtet sich ein ausdrücklicher Dank im Speziellen an:

Herrn Stefan Füg für die Erstellung der Animation zur Abtrennung der Traumsequenz und Herrn M. Eng. Andreas Keck für die Vermietung der Funk-Technik zur Tonaufnahme.

Abschließend sei allen Unterstützenden im Umfeld der Hochschule und im familiären Kreis gedankt, die beratend, logistisch und finanziell zur vorliegenden Masterarbeit und der Produktion von „Eine Nacht im Büro“ beitrugen.

10.2 Drehbuch „Eine Nacht im Büro“

EINE NACHT IM BÜRO

5. Fassung
26. Juni 2020
Autor: Simon Lamche
Co-Autor: Veit Stephan

1

AUFBLLENDE.

INT. BÜRO – ABENDS – REALITÄT | 1

Veit sitzt allein im Büro von evidentmedia (am Platz von SL). Das Deckenlicht ist an, die Jalousien an den Fenstern sind geschlossen und die restlichen Computer sind aus. Neben Veits Maus und Tastatur befindet sich eine Getränkeflasche. Veit arbeitet am Stitching eines 360°-Videos. Am Schreibtisch nebenan liegt Papierchaos und ein leerer Ordner. Simon öffnet von außen die Tür und streckt seinen Kopf herein. Er steht nur mit einem Bein im Büro.

SIMON
(humorvoll)
Also, mach nicht mehr zu lang.

VEIT
Ne, ich mach das nur noch kurz fertig.

SIMON
Okay, dann bis morgen.

VEIT
Bis dann!

Beide winken sich zu, dann schließt Simon die Tür. Veit bleibt allein zurück. Er sieht sich im Büro um. Sein Blick bleibt am Papierchaos hängen. Nach einigen Augenblicken nimmt er sein Getränk und prostet in Richtung des Papierstapels.

VEIT (FORTS.)
(ironisch)
Na dann, auf eine Nacht im Büro.

Veit nimmt einen Schluck von seinem Getränk und arbeitet zunächst motiviert weiter. Er gähnt und blickt mehrfach auf die Uhr. Die Arbeit fällt ihm zunehmend schwerer. Nach einer Weile legt Veit die Arme auf den Schreibtisch und stützt seinen Kopf auf ihnen ab, bis er vor Müdigkeit zusammensinkt und einschläft.

Das Bild wird schwarz. Sterne erscheinen und beginnen zu funkeln. Übergang zum Traum.

INT. BÜRO – NACHTS – TRAUM | 2

Als das Bild wieder hell wird, schläft Veit unverändert. Plötzlich ertönt der Drucker, der am anderen Ende des Raumes steht. Die restlichen Arbeitsplätze im Büro sind nun von drei weiteren Personen besetzt, die sich optisch nur durch ihre Kleidung und Accessoires von Veit unterscheiden. An den Computern sind Listen, Referenzmaterialien und ein Code-Editor geöffnet. Der Druckauftrag wird von Doppelgänger 2 zu Beginn der Szene gestartet. Die Szene bildet einen Loop.

Doppelgänger 1, 2 und 3 arbeiten motiviert. Die Arbeit wird durch einen Geräusch-Sprech-Rhythmus in 140 Bpm symbolisiert: Doppelgänger 3 klopft mit einem Papierstapel auf den Tisch (Bass Drum), Doppelgänger 2 locht einige Unterlagen (Snare Drum), Doppelgänger 1 tipp auf der Tastatur (Hi-Hat). Als Doppelgänger 1 die Lautsprecher an seinem Arbeitsplatz einschaltet, wird der Rhythmus durch ein Schlagzeug ergänzt. Die Doppelgänger sprechen abwechselnd Arbeitsmantren in die Kamera und pausieren dafür ihre Tätigkeit im Rhythmus.

DOPPELGÄNGER 1
(Am Platz von VS; Tastatur
abwechselnd Achtel und
Viertel; Text zweimal)
Ja, natürlich können wir. Ja, sehr gerne
machen wir. Nein, es ist kein Problem.

DOPPELGÄNGER 2
(Am Platz von HK; Locher
auf 2, 4; Text zweimal)
So kurz vor Ende ändern? Ja immer doch,
das geht doch noch.

DOPPELGÄNGER 3
(Am Platz von BGG/SF;
Papierstapel auf 1, 3 bzw.
alle 2 Takte zusätzlich
auf 4 +; Text zweimal)
Ein Problem, nein kein Problem. Zahl'n Sie
ruhig die Hälfte. Es war wie immer alles
nur ein Klacks.

Veit schnarcht als Akzent auf die 1. Doppelgänger 3 zerreit im Takt einige Blätter auf 1, 1, 1, 2, 3, 4. Die Rhythmus-Choreografie endet mit einem Schlussmantra zur Kamera.

ALLE DOPPELGÄNGER
(Text zweimal)
Alles für die Kunden, alles für den Job!

Das Bild wird schwarz. Die funkelnden Sterne verblassen. Übergang zur Realität.

INT. Büro – MORGENS – Realität | 3

Als das Bild wieder hell wird, sind die Doppelgänger verschwunden. Das Telefon neben Veit klingelt. Er wird durch das Klingeln geweckt und nimmt den Anruf entgegen.

VEIT
(erschrocken)
Evidentmedia, guten Morgen.

Während er dem Anrufer zuhört, sieht er sich im Büro um. Alles scheint wie vorher zu sein. Nur dort, wo das Papierchaos war, liegt nun der Ordner mit den einsortierten Unterlagen. Veit schüttelt unglaublich den Kopf und lächelt.

VEIT (FORTS.)
Ihre Unterlagen? Ja also, die haben wir uns schon angesehen. Wir können das gerne so umsetzen und melden uns nächste Woche nochmal bei Ihnen.
(kurze Pause)
Danke und auf Wiederhören.

Veit legt auf.

Überrascht – aber dennoch zufrieden – lehnt er sich im Stuhl zurück und verschränkt die Hände hinter dem Kopf.

VEIT (FORTS.)
Na, an solche Überstunden könnte ich mich glatt gewöhnen.

SCHWARZBLENDE.

LOGO VON EVIDENTMEDIA.