



HAW HAMBURG

ÄSTHETIK UND PHILOSOPHIE VON KOMPOSITIONS- UND PRODUKTIONSANSÄTZEN IN SPATIAL AUDIO

Fakultät DMI
Hochschule für Angewandte Wissenschaften Hamburg

Masterthesis

zur Erlangung des akademischen Grades M.A.

vorgelegt von

Leon Sudahl
(Matr.-Nr. [REDACTED])
am 17. April 2025

Erstprüfer: Prof. Thomas Görne
Zweitprüfer: Stefan Troschka

Abstract

This thesis examines compositional and production approaches in spatial audio from an aesthetic and philosophical perspective. The aim was to identify recurring patterns, listening habits and psychoacoustic effects in spatial audio work and to develop a guideline that serves as a basis for high quality spatial audio works. To answer the hypotheses, a mixed-methods approach was chosen in which qualitative interviews and secondary sources were analysed using the structured content analysis according to Kuckartz with the MAXQDA software. The formulated hypotheses were confirmed. The results clearly show that there is a frequency-dependent way of working that is orientated towards auditive perception. While many productions are still focussed on an extended front stage, spatial audio is in a transitional phase towards more flexible spatial design approaches. The statements are subject to a quantitative limitation of 25 sources. It is therefore recommended that the guidelines and the statements made be reviewed in further studies and confirmed in practice.

Zusammenfassung

Diese Arbeit untersucht Kompositions- und Produktionsansätze in Spatial Audio aus einer ästhetischen und philosophischen Perspektive. Ziel war es, wiederkehrende Muster, Hörgewohnheiten und psychoakustische Effekte in der Arbeit mit Spatial Audio zu identifizieren und einen Leitfaden zu entwickeln, der als Grundlage für qualitativ hochwertige Spatial Audio Werke dient kann. Zur Beantwortung der Thesen wurde ein Mixed-Methods-Ansatz gewählt, bei dem qualitative Interviews und Sekundärquellen mittels der strukturierten Inhaltsanalyse nach Kuckartz mit der Software MAXQDA ausgewertet wurden. Die formulierten Hypothesen wurden bestätigt. Die Ergebnisse zeigen deutlich, dass eine frequenzabhängige Arbeitsweise vorliegt, die sich an der auditiven Wahrnehmung orientiert. Während viele Produktionen noch auf eine erweiterte Frontbühne fokussiert sind, befindet sich Spatial Audio in einer Übergangsphase zu freieren räumlichen Gestaltungsmöglichkeiten. Die Aussagen unterliegen einer quantitativen Beschränkung von 25 Quellen. Es wird daher empfohlen, den Leitfaden und die getroffenen Aussagen in weiteren Untersuchungen zu überprüfen und in der Praxis zu bestätigen.

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die zum Gelingen dieser Masterarbeit beigetragen haben.

Mein besonderer Dank gilt meinem Erstprüfer Prof. Thomas Görne für das Vertrauen in mein Forschungsvorhaben und die wegweisenden Gespräche. Ebenso danke ich meinem Zweitprüfer Stefan Troschka für die engagierte Begleitung und die hilfreichen Hinweise im Verlauf der Arbeit.

Ein großer Dank geht an alle Interviewten, die durch ihre Offenheit, Zeit und Expertise einen entscheidenden Beitrag zur Datengrundlage dieser Arbeit geleistet haben.

Weiterhin danke ich meinen Mitstudierenden, Freunden und meiner Familie für die geduldige Unterstützung, den Zuspruch und das Verständnis während der Entstehung dieser Arbeit. Mein besonderer Dank gilt dem Team rund um die SALS, insbesondere Lucca Riitano, und Anna Smits, die enorm viel Zeit für die sprachliche Korrektur dieser Arbeit investiert hat.

Abschließend möchte ich mich bei der HAW Hamburg und insbesondere beim Team des Tonlabors bedanken, vor allem bei Tobias Falke, der jederzeit erreichbar war, um meine Fragen und Unsicherheiten zu klären.

Inhaltsverzeichnis

1	Einleitung	1
2	Motivation und Vorgehen	2
3	Theoretische Grundlagen	4
3.1	Definition des Begriffs Spatial Audio und weiterer relevanter Begriffe . . .	4
3.2	Geschichte und Entwicklung von Spatial Audio	6
3.3	Wahrnehmung von räumlichem Klang	9
3.4	Psychoakustische Phänomene	10
3.5	Spatial-Audio-Formate	12
3.5.1	Kanalbasierte Formate	13
3.5.2	Objektbasierte Formate	16
3.5.3	Szenenbasierte Formate	17
3.5.4	Kopfbezogene Formate	19
4	Methodik	21
4.1	Fokussierte Experteninterviews	22
4.2	Probanden	23
4.3	Fragebogen	24
4.4	Durchführung der Interviews	25
4.5	Aufbereitung der Interviews	25
4.6	Auswahl der Sekundärliteratur	27
4.7	Analyse mit MAXQDA	28
4.8	Forschungshypothesen	29
5	Ergebnisse der Interviews	30
5.1	Das Profil der Befragten und allgemeine Informationen zu Spatial Audio (Einstiegsfragen)	30
5.1.1	Erfahrung der interviewten Personen	31
5.1.2	Spatial Audio Formate	32
5.1.3	Genres	34
5.1.4	Motivation und Beweggründe zur Arbeit in Spatial Audio	35
5.2	Zusammenfassung der Kernfragen (Schlüsselfragen)	36
5.2.1	Software und Tools	36
5.2.2	Spatial-Audio-Mikrofonierungsansätze	38
5.2.3	Anzahl der Klangelemente	39
5.2.4	Bewegung und Positionierung von Klängen und Instrumenten . . .	40
5.2.5	Frequenzabhängigkeiten im Tieftonbereich	42
5.2.6	Integration der Hörenden und das Sweet-Spot-Problem	42
5.2.7	Räumlichkeit und der Umgang mit Halleffekten	43
5.2.8	Mixing in Spatial Audio	46
5.2.9	Mastering in Spatial Audio	48
5.2.10	Methoden, Ansätze und Philosophien hinter Spatial-Audio- Pro- duktionen und Kompositionen	50
5.2.11	Produktions und -Kompositionsprozesse	56
5.3	Ergänzungen durch Eventualfragen	59
5.3.1	Psychoakustik und Hörgewohnheit	59

5.3.2	Probleme und Herausforderungen	60
5.3.3	Faszination und Vorteile von Spatial Audio	63
5.4	Zusammenfassung der Abschlussfragen	64
5.4.1	Zukunft von 3D-Audio	64
5.4.2	Wünsche	65
5.4.3	Tipps	66
5.5	Meinung zu Standardisierten Spatial Audio Workflows	66
6	Analyse	68
6.1	Erkenntnisse aus den Spatial Audio Listening Sessions und der experimen- tellen Forschungsarbeit	68
6.2	Analyse der Thesen	69
6.2.1	Hypothese 1	70
6.2.2	Hypothese 2	74
6.2.3	Hypothese 3	75
7	Diskussion und Ausblick	80
8	Fazit	82
A	Digitaler Anhang	84
	Literatur	85

Abbildungsverzeichnis

1	Kopfbezogenes Koordinatensystem für die Beurteilung der Hörereignisrichtung [Dickreiter, 1997, S. 117]	9
2	Hörschwelle und Schmerzgrenze [Zwicker and Fastl, 1999, S. 17]	11
3	Kategorisierung der Experteninterviews nach Tätigkeitsfeldern, eigene Grafik.	31
4	Erfahrung in Jahren, eigene Grafik.	31
5	Kategorisierung der Experteninterviews nach Tätigkeitsfeldern, eigene Grafik.	32
6	Spatial Audio-Formate in der Musikproduktion, eigene Grafik.	32
7	Spatial Audio-Formate in Hörspiel und Sounddesign, eigene Grafik.	33
8	Spatial Audio-Formate in Experimentellen Produktionen und Installationen, eigene Grafik.	33

1 Einleitung

Als Edgar Varèse 1958 seine fast 400 Lautsprecher im Philipps-Pavillon auf der Weltausstellung in Brüssel installierte, hätte er wohl nicht gedacht, dass wir 70 Jahre später mit nur 5 Lautsprechern in der Lage sein würden, sein Werk mit anhaltender Faszination zu hören. Wir sind in der Lage, eine Hörsituation zu schaffen, die uns trotz der deutlich abweichenden Lautsprecheranzahl immer noch von seinem Werk in Bezug auf die wahrgenommene Räumlichkeit und Umhüllung begeistert¹. Dank des technischen Fortschritts stehen uns heute eine Vielzahl von Werkzeugen in Form von Software und physikalischen Renderern zur Verfügung, um räumliche Audioproduktionen in Formaten wie Dolby Atmos oder Ambisonics zu realisieren. Auch wenn Formate wie Ambisonics bereits seit 1970 existieren und es immer wieder Ansätze gab, diverse Formate marktreif zu machen, wurde Spatial Audio erst in den letzten 10 Jahren durch Dolby Atmos, Virtual Reality, Game Audio und durch die Ankündigung weiterer Formate und Codecs zunehmend für die Audiobranche und den Endkonsumenten zugänglich [Roginska and Geluso, 2018].

Es gibt inzwischen zahlreiche Methoden, um in Spatial-Audio-Formaten zu komponieren und zu produzieren. Laut dem Experten Lasse Nipkow bietet Spatial Audio im Vergleich zu Stereo, welches insbesondere unter psychoakustischen Gesichtspunkten ausgereizt ist, noch zu erforschende Aspekte, vor allem wenn es um die Ästhetik geht². Für die Erstellung eines hochwertigen Stereo-Mixes ist das Internet mit Plattformen wie YouTube eine große Hilfe. Es gibt unzählige Leitfaden, Bücher und Foren, die Tipps und Tricks sowie ganze Workflows zur Arbeit in Stereo anbieten. Für Kompositions- und Produktionsansätze in Spatial Audio gibt es vergleichsweise wenig Literatur, diese steigt in den letzten Jahren aber stetig an.

Spatial Audio bietet einen großen Gestaltungsraum für Audiophile, Forschende, Komponierende, Produzierende und Musikschaaffende. Bisher wurden lediglich Standards in Bezug auf die Lautheit von Veröffentlichungen bei Streaminganbietern wie Apple Music oder für bestimmte Lautsprecheranordnungen definiert.

Ziel dieser Arbeit ist es, wiederkehrende Muster, Hörgewohnheiten und psychoakustische Effekte in der Arbeit mit Spatial Audio zu identifizieren und einen Leitfaden zu entwickeln, der als Grundlage für qualitativ hochwertige Spatial Audio Werke dienen kann. Dieses Ziel soll anhand von Experteninterviews und einer qualitativen Inhaltsanalyse erreicht werden. Die Analyse orientiert sich an drei Thesen, die in diesem Zusammenhang erarbeitet wurden, (siehe Unterabschnitt 4.8).

¹ Dies wurde in mehreren Spatial Audio Listening Sessions im Immersive Audio Lab der HAW Hamburg beobachtet, siehe Unterabschnitt 6.1.

² Interview 14 mit L. Nipkow (Einverständnis zur namentlichen Nennung wurde schriftlich eingeholt).

2 Motivation und Vorgehen

Im Rahmen der akademischen Ausbildung gab es viele Gelegenheiten Spatial-Audio-Stücke zu hören und mit den Komponierenden und Produzierenden über die verwendete Software und deren Arbeitsprozesse zu diskutieren. Dabei ist die Variabilität der Prozesse unter den Tonschaffenden aufgefallen. Dies führte zu der Frage, welche Faktoren trotz der unterschiedlichen Formate, Software, Genres und Produktions- bzw. Kompositionsformen qualitativ hochwertige Spatial-Audio-Werke begünstigen und ob sie bewusst oder unbewusst umgesetzt werden. Aufgrund dessen wird in dieser Arbeit untersucht, welche grundlegenden Gesetze und Regeln für alle Tonschaffenden bei Spatial-Audio-Produktionen bzw. -Kompositionen gelten und ob es Ansätze gibt, die grundlegend zu einer qualitativ überzeugenden Spatial-Audio-Produktion oder -Komposition führen können.

Im Zuge zahlreicher technischer, künstlerischer und forschender Tätigkeiten im Bereich Spatial Audio, darunter auch die *Spatial Audio Listening Sessions(SALS)* im *Immersive Audio Lab(IAL)* der HAW Hamburg, konnten umfangreiche Erkenntnisse gewonnen und Herausforderungen identifiziert werden. Auf Basis der SALS ist eine experimentelle Forschungsarbeit in Bezug auf die Wahrnehmung komplexer Klänge in Spatial Audio entstanden. Auch diese Erkenntnisse werden in Unterabschnitt 6.1 erläutert. Im Rahmen dieser Masterarbeit werden die theoretischen Erkenntnisse durch einen praktischen Teil ergänzt. Dieser besteht aus drei Spatial-Audio-Kompositionen, die die in dieser Arbeit beschriebenen Herangehensweisen praktisch untersuchen sollen. Zusätzlich wird ein Leitfaden zu Spatial-Audio-Workflows abgeleitet, der künstlerische Freiheiten respektiert und zu hochwertigeren Spatial-Audio-Werken beitragen kann.

Das Hauptziel dieser Arbeit ist es, einen Überblick über verschiedene Herangehensweisen, Produktionsmethoden und die zugrunde liegende Philosophie von Spatial Audio zu schaffen. Darüber hinaus sollen aktuelle Herausforderungen erörtert und Empfehlungen für künstlerische und technische Herangehensweisen formuliert werden. Diese Empfehlungen sollen sowohl Lernenden als auch erfahrenen Schaffenden als Werkzeug dienen, um ihre Produktionen zu verbessern und grundlegende Fehler zu vermeiden. Dabei handelt es sich nicht um die Aufstellung eines festen Regelwerks, sondern um die Ermittlung von Prinzipien, die den kreativen Freiraum der Künstlerinnen bewahrt.

Zur Beantwortung der Forschungsfragen wurde eine qualitative Datenerhebung mit anschließender inhaltsanalytischer Auswertung durchgeführt. Diese basiert auf 17 qualitativen Experteninterviews sowie der Analyse von weiteren 8 Publikationen, darunter wissenschaftliche Artikel, Bücher und Live-Präsentationen. Die Interviews lieferten eine Fülle

von Informationen, die über die ursprüngliche Fragestellung hinaus weitere Aspekte in Bezug auf Spatial Audio betrachten ließen. Zu diesen gehören unter anderem die persönliche Faszination der Kreativen für Spatial Audio, die Vorteile verschiedener Formate und Software sowie Wünsche für zukünftige Entwicklungen auf Hard- und Softwareseite, um die Effizienz und Qualität ihrer Produktionen zu steigern.

Spatial-Audio-Formate werden in vielen Bereichen angewandt. Von Virtual Reality über Gaming bis zu Filmen und Musik sind viele Kategorien vertreten. Zur Eingrenzung des Themas liegt der Fokus dieser Arbeit ausschließlich auf rein auditiven Formaten.

3 Theoretische Grundlagen

Für die detaillierte Analyse der Interviews und Sekundärdaten ist eine intensive Auseinandersetzung mit der Klangwahrnehmung, dem räumlichen Hören sowie den technischen Hintergründen der verschiedenen Spatial-Audio-Formate entscheidend. Diesen Formaten liegen unterschiedliche technische und physikalische Ansätze zugrunde, die wiederum Einfluss auf die verwendete Software haben und den Rahmen für die künstlerische Arbeit vorgeben. Ästhetische Entscheidungen werden somit maßgeblich durch das gewählte Format beeinflusst [Pfanzagl-Cardone, 2023]. Im folgenden Kapitel werden diese Grundlagen erläutert und detailliert beschrieben, um die Entscheidungen und Überlegungen der Befragten nachvollziehbar zu machen.

3.1 Definition des Begriffs Spatial Audio und weiterer relevanter Begriffe

Der Begriff *Spatial Audio* wird in dieser Arbeit verwendet, um Klang im Raum der mit technischen oder ästhetischen Ansätzen ein dreidimensionales Klangerlebnis schafft. Grundsätzlich ist jeder Klang räumlich (spatial), unabhängig davon, ob er von einem oder zwanzig Lautsprechern erzeugt wird [Blauert, 1997]. Jeder Klang regt den Raum an und erzeugt ein Schallfeld, dessen Charakteristika von den akustischen Eigenschaften des jeweiligen Raumes abhängen.

Spatial Audio findet vorwiegend Anwendung in Bezug auf mehrkanalige Audioerlebnisse, die darauf abzielen, die Hörenden zu umhüllen. Dabei wird Klang aus 360° sphärisch um die Hörenden generiert, sodass der Eindruck entsteht, Klänge aus allen Richtungen wahrzunehmen. Dies wird durch die Komposition selbst oder durch die Verwendung von Höhenlautsprechern erreicht. Spatial Audio orientiert sich dabei stark an unserer natürlichen, akustischen Realität. Weitere Begriffe, die für Spatial Audio verwendet werden, sind *3D-Audio* und *immersive Audio*.

Der Begriff 3D-Audio hebt die dritte Dimension, die Höhenebene, besonders hervor. Genau genommen müsste man von 4D-Audio sprechen, da Klang sich stets durch Raum und Zeit bewegt [Pfanzagl-Cardone, 2023].

Immersives Audio beschreibt das vollständige Eintauchen in eine Klangwelt. Es ermöglicht eine an die Realität annähernde Wiedergabe realer Orte in virtuellen Umgebungen und die Schaffung neuer akustischer Welten [Roginska and Geluso, 2018]. Dieser Begriff wird häufig im Marketing verwendet, um immersive Erlebnisse zu bewerben. Genau betrachtet kann bereits ein einzelner Lautsprecher ein immersives Erlebnis schaffen. Komplexe

Lautsprecherinstallationen können den Grad der Immersion jedoch deutlich steigern und das Eintauchen in eine Klangwelt erleichtern. Spatial Audio besitzt somit das Potenzial, immersive Erlebnisse zu erzeugen, wobei dies stark von Hörer, Umgebung, Situation und anderen Faktoren abhängt, die in dieser Arbeit diskutiert werden.

Eine mögliche korrekte Einordnung der Begriffe wäre wie folgt: Spatial Audio wird mithilfe von 3D-Audio-Technologien zu immersiven Audioerlebnissen. In dieser Arbeit wird der Begriff Spatial Audio bevorzugt, da auch Arbeitsweisen in Formaten wie Quadrophonie und Oktophonie sowie in nicht als typische 3D-Audio-Formate definierten Setups untersucht werden, obwohl diese Formate in der Lage sind, einen räumlich umhüllenden Klang zu ermöglichen. Der Fokus liegt auf mehrkanaligen, szenen- und objektbasierten Formaten, die vier oder mehr Lautsprecher nutzen. Eine Ausnahme bildet das zweikanalige Format *Binaural Audio*, das zur Wiedergabe von Spatial Audio über Kopfhörer dient.

In dieser Arbeit wird von *Kompositions- und Produktionsmethoden*, sowie *Raum, Räumlichkeit* und *Nachhall* gesprochen. Nachfolgend werden diese Begriffe für das Verständnis dieser Thesis erläutert.

Unter Kompositionsmethoden werden im Rahmen dieser Arbeit gestalterische Verfahren zur Strukturierung, Organisation und künstlerischen Konzeption von Klangmaterial verstanden. Sie umfassen sowohl traditionelle kompositorische als auch algorithmische Verfahren, wie sie zum Beispiel in elektroakustischer Musik zu finden sind [Roads, 2015, Smalley, 1997]. Die Produktionsmethoden beschreiben technische Prozesse wie die Nachbearbeitung des Klangs – vor allem in Bezug auf Aufnahme, Mischung und räumliche Verteilung im Mix [Emmerson, 2007, Huber and Runstein, 2017].

Einige der Befragten haben sich gezielt nur auf die Produktionsmethoden spezialisiert und arbeiten häufig mit den *Stemmixing-Methoden* (siehe Unterunterabschnitt 5.2.11). Ein weiterer Teil identifiziert sich primär als Komponierende. Dabei ist anzumerken, dass viele Produktionsprozesse in den kompositorischen-Workflow integriert sind und diese Prozesse häufig nicht klar voneinander trennen lassen. Emmerson spricht hierbei von einem „Language Grid“, bei dem sich gestalterische und technologische Aspekte überlagern [Emmerson, 2007].

In dieser Arbeit wird oft von Raum, Räumlichkeit und Hall gesprochen. Diese Begriffe werden oft unter dem Begriff *Hall* zusammengefasst, sollten jedoch differenziert betrachtet werden, da sie unterschiedliche Aspekte beschreiben:: Der Raum wird als physikalischer Ort und als ästhetischer Gestaltungsparameter verstanden [Emmerson, 2007], während

mit der Räumlichkeit die subjektiven akustischen räumlichen Eigenschaften der Klänge beschrieben werden [Blauert, 1997]. Hall kommt in Räumen natürlich vor, kann aber auch technisch reproduziert werden und gilt als akustischer Indikator für räumliche Tiefe und Kontext [Smalley, 1997].

Der Begriff Ästhetik umfasst in dieser Arbeit die gestalterischen, wahrnehmungsbezogenen und emotionalen Dimensionen der Klanggestaltung in Spatial Audio. Er bezieht sich auf die Wirkung, Ordnung und Komposition von Klang im Raum sowie auf subjektive Hörerlebnisse, die durch technische und kreative Entscheidungen beeinflusst werden, vgl. hierzu [Pfanzagl-Cardone, 2023].

3.2 Geschichte und Entwicklung von Spatial Audio

Die ersten Räume, in denen sich frühe Formen räumlicher Klangerfahrungen entwickelten, waren Versammlungen bzw. musikalische Zusammenkünfte in den Höhlen der Steinzeitmenschen [Bagenal, 1951]. Höhlen bieten umfangreiche reflektive Eigenschaften, die oft einen langen Nachhall verursachen. Nach heutiger Definition wurden die ersten Spatial-Audio-Stücke in den Höhlen der ersten Menschen in der Steinzeit uraufgeführt. In Höhlen wurden auch die ersten Instrumente gefunden. Das sogenannte Schwirrholz oder die Knochenflöte wurden in verschiedenen Ausführungen entdeckt. Es ist anzunehmen, dass sie zu rituellen Zwecken genutzt wurden und die besondere Akustik der Höhlen, die je nach Beschaffenheit mit der von Kathedralen vergleichbar ist, das räumliche Klangerlebnis maßgeblich geprägt hat [Deutschlandfunk Kultur, 2025].

Weitere für Spatial Audio relevanten musikalischen Orte waren die minoischen Theaterstätten, auf deren Grundlagen später die griechischen Amphitheater gebaut wurden. Diese Theater wurden vor über 2200 Jahren gebaut und waren erste Versuche, den räumlichen Klang zu beeinflussen, um vor allem die Sprachverständlichkeit positiv zu verändern [Chourmouziadou and Kang, 2008]. Die nächsten bedeutenden Orte in Form von Gebäuden, in denen wir bewusst von Spatial Audio sprechen können, waren die Bauten der Menschen der letzten Jahrhunderte. Ob Paläste, Kirchen, Kathedralen, Moscheen oder Tempel – die großen Säle der Königinnen und Könige boten und bieten uns bis heute nur durch die Architektur einen komplexen räumlichen Klang. Die Akustik der Bauten und deren teilweise langen Nachhallzeiten sowie reflektierende Eigenschaften beeinflussten die Art und Weise, wie Stücke komponiert wurden. Die Werke wurden an die Raumeigenschaften angepasst, und erste Ideen, Klänge im Raum zu bewegen, entstanden im 9. Jahrhundert. Eine der Ideen wurden im *Organum* umgesetzt und erreichte eine Art erste Polyphonie durch das Singen der Paralleltöne in einem Raum mit langer Nachhallzeit [Roginska and Geluso, 2018].

Weitere Versuche, bewusst mit räumlichem Klang zu spielen, geschahen in der Zeit der Renaissance im 16. Jahrhundert. Ansätze waren zum Beispiel die Verteilung der Chöre in der venezianischen Polychoralität oder das Ruf und Antwort Prinzip [Slotki, 1936]. In den folgenden Epochen wurden weitere räumliche Effekte wie Echos und weitere Positionierungen von Chor- und Instrumentengruppen getestet. Wichtig zu erwähnen ist das *Requiem* von *Hector Berlioz* aus dem Jahr 1837, welches bewusst psychoakustische Effekte wie die Illusion einer horizontalen Umhüllung der Zuhörer durch die Reflexionen des Raumes und die Zeitunterschiede des Gesangs der Chöre ausnutzte [Broderick, 2012].

Die Kompositionstechniken wurden an die Räume angepasst. Beispielsweise wurden Stücke für Chöre an die lange Nachhallzeit der Kirchen angepasst, und Komponisten wie Joseph Haydn haben ihre Werke speziell für Konzertsäle komponiert, um den Raum in der Komposition ideal einzusetzen. Der Einbezug von Raum – der wesentliche Bestandteil der Spatial-Audio-Erfahrung – wurde ein zentraler Aspekt in der musikalischen Avantgarde, in der Dodekaphonie, im Serialismus, in der elektronischen Musik und findet sich heute vor allem in elektroakustischen Stücken wieder [Slotki, 1936].

Henry Brant (1913–2008) war ein Pionier der räumlichen Musik und ließ sich stark von Charles Ives inspirieren, wobei er nicht nur Ensembles räumlich trennte, sondern deren genaue Positionierung im Aufführungsraum detailliert vorgab. Seine Kompositionen nutzten psychoakustische Effekte wie das „Cocktail-Party-Phänomen“, das die selektive Aufmerksamkeit durch räumliche Separation erleichtert, sowie Konzepte wie die Lokalisationsunschärfe und die höhenabhängige spektrale Variation der *Head Related Transfer Function*³, die er intuitiv erkannte. Brant betrachtete den Raum jedoch nicht als primäres kompositorisches Prinzip, sondern als Mittel zur Erweiterung traditioneller tonaler Kompositionstechniken, indem er durch die räumliche Anordnung von Klangquellen klangliche Klarheit und strukturelle Flexibilität schuf [Roginska and Geluso, 2018]. Ende des 19. Jahrhunderts begann die technische Entwicklung durch Erkenntnisse in der Wissenschaft. Die Arbeit mit verschiedenen Aufnahmemethoden und Wiedergabeverfahren begann den Umgang mit räumlichem Klang zu beeinflussen [Roginska and Geluso, 2018]. Die wichtigsten Meilensteine der letzten 144 Jahre sind:

- Erstes Binaural Recording (1881): Clément Ader, ein französischer Ingenieur, präsentierte das „Théâtrophone“ auf der Elektrizitätsausstellung in Paris. Diese Technologie nutzte zwei Mikrofone, um ein binaurales Hörerlebnis über Kopfhörer zu erzeugen [Collins, 2008].
- Blumlein Stereo (1931): Alan Blumlein, ein britischer Ingenieur, entwickelte die Blumlein-Stereotechnik, die die Grundlage für moderne Stereophonie bildet. Diese

³ Kopfbezogene Übertragungsfunktion, in Unterabschnitt 3.3 erläutert.

Technik ermöglichte eine realistische Reproduktion von Richtung und Bewegung des Schalls [Blumlein, 1931].

- Erste Mehrkanal-Spatial-Audio-Installation (1958): Im Philips-Pavillon bei der Weltausstellung 1958 in Brüssel umgesetzt. Hier wurden die von Yannis Xenakis und Edgar Varèse komponierten Stücke für „Poème électronique“ über 400 Lautsprecher wiedergegeben [Ziemer, 2023, 17].
- Quadrophonie (1970er Jahre): Ein früher Versuch, Spatial Audio zu kommerzialisieren, bei dem vier Lautsprecher verwendet wurden, um Schallquellen rund um den Hörer zu platzieren. Diese Technologie fand begrenzten Erfolg, legte aber den Grundstein für spätere Surround-Sound-Systeme [Roginska and Geluso, 2018, 52].
- Ambisonics (1970er): Ein umfassenderer Ansatz für 3D-Audio, der ursprünglich in den 1970er Jahren entwickelt wurde und seitdem weiterentwickelt wurde. Ambisonics ermöglicht eine vollständige Kugelklangwiedergabe und wird oft in VR-Anwendungen verwendet [Gerzon, 1973].
- Dolby Surround (1980er Jahre): Dolby Laboratories führte Dolby Surround ein, das für Kinofilme entwickelt wurde und Mehrkanal-Audio über vier Kanäle (links, Mitte, rechts und Surround) ermöglichte [Roginska and Geluso, 2018, 52].
- Wave Field Synthesis (2000er): Systeme, die auf der Wellenfeldsynthese basieren, bilden ein Schallfeld in Form einer Kugelwelle und ermöglichen eine hohe Ortung. Diese Systeme sind aber in der Praxis nicht ohne hohen Aufwand anwendbar, daher stagniert der Fortschritt [Ziemer, 2023, 3].
- Multichannel-Formate (VBAP⁴ und DBAP⁵): Das Format mit fixer Lautsprecheranordnung (VBAP) basiert auf der Verwendung von Vektoren zwischen zwei oder drei Lautsprechern und die freie Lautsprecheranordnung bei DBAP nimmt anhand der relativen Distanz zu den Lautsprechern die Pegelverteilung der Schallquellen vor [Lossius et al., 2009].
- Dolby Atmos (2012): Als ursprünglich für das Kino entwickeltes Format ist Dolby Atmos mit seinem objektbasierten Ansatz das erste Spatial-Audio-Format mit Höhenwiedergabe, das es in den Konsumentenmarkt geschafft hat.

Der Einfluss dieser Verfahren auf die Ästhetik und Philosophie von Spatial Audio wird in den folgenden Kapiteln erörtert und in der Auswertung diskutiert.

⁴ Vector Base Amplitude Panning (1997).

⁵ Distance-Based Amplitude Panning (2009).

3.3 Wahrnehmung von räumlichem Klang

„Physikalischer Schall ist eine Druckwelle, die sowohl Schallereignisse als auch die Eigenschaften eines akustischen Raums an den Hörer überträgt und so die Außenwelt mit den Ohren des Hörers verbindet. Da die Physik des Schalls komplex ist, umfasst die Übertragung Prozesse wie Reflexion, Dispersion, Brechung, Absorption usw., die alle von den akustischen Eigenschaften des Raums abhängen. Beim Eintreffen im Innenohr werden die Schallwellen in neurologische Signale umgewandelt, die vom Gehirn verarbeitet werden; die Außenwelt wird mit dem inneren Bewusstsein verbunden“:[Blessner and Salter, 2007, S. 12].

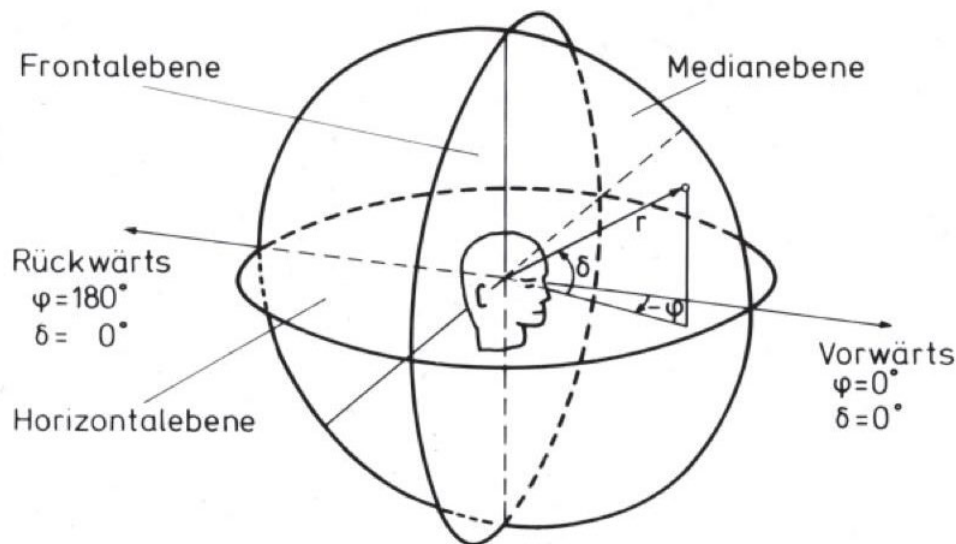


Abb. 1: Kopfbezogenes Koordinatensystem für die Beurteilung der Hörereignisrichtung [Dickreiter, 1997, S. 117]

Das kopfbezogene Koordinatensystem (siehe Abbildung 1) veranschaulicht, wie Hörereignisse lokalisiert werden. Mithilfe der Azimutebene (Horizontalebene), der Medianebene (Vertikalebene) und der Frontalebene lässt sich die Richtungswahrnehmung unterscheiden [Dickreiter, 1997]. Hauptsächlich müssen die Median- und Azimutebene differenziert werden, da im Vergleich zur Frontalebene unterschiedliche Effekte zur Lokalisierung von Schallereignissen stattfinden.

Die Frontalebene befindet sich um die Ohren und den Hals herum und beschreibt die Richtungen links, rechts, oben und unten [Frisecke, 2014]. Großen Einfluss bei der Ortung

von Schall spielt die HRTF, welche die Eigenschaften des menschlichen Ohrs berücksichtigt, denn das Ohr ist als richtungsabhängiger Schallreflektor ausgelegt [Görne, 2015]. Aufgrund der Form der Ohrmuschel und des Gehörgangs wird die Ortung von Schall ermöglicht – es entsteht ein Interferenzmuster. Mit der Bewegung unseres Kopfes erzeugen wir weitere Richtungsinformationen, da die HRTF im Moment der Bewegung verändert wird. Die räumliche Wahrnehmung ist binaural, also 'zweiohrig', und spielt eine tragende Rolle bei der Wiedergabe von Ton über Kopfhörer [Görne, 2015].

Die Azimutebene liegt in der Horizontalen und beschreibt die Signalrichtungen links, rechts, vorne und hinten [Frisecke, 2014]. Eine Schallquelle wird durch Laufzeit- und Pegeldifferenzen von Signalen in der Horizontalebene lokalisierbar. Vor allem sind die *interauralen Zeitdifferenzen* dafür verantwortlich, dass wir in der Horizontalebene Schallereignisse orten können [Dickreiter, 1997]. Ein großes Problem bei der Ortung von Schall ist die *Lokalisationsinversion*. Diese sogenannte *Vorne-Hinten-Verwechslung* tritt vor allem bei unbekannten, schmalbandigen, tiefen Signalen auf und ist auf die sich ähnelnden spektralen Eigenschaften und Laufzeitdifferenzen zwischen vorderer und hinterer Position zurückzuführen [Ziemer, 2023]. Zeit- oder Pegeldifferenzen können in der Medianebene, die in der Vertikalen liegt nicht festzustellen. Es entstehen allerdings Klangfarbenunterschiede, welche durch die Kopfform und die Ohren unbewusst wahrgenommen werden. Hierdurch werden bestimmte Frequenzbänder angehoben. Die Lokalisationsschärfe in der Medianebene ist bedeutend geringer als in der Azimutebene [Dickreiter, 1997].

Die Einschätzung der Entfernung eines Schallereignisses basiert auf der Erfahrung des Gehörs. Diese wird durch das Alter und gemachter Hörerfahrungen geprägt. Darüber hinaus spielen die Lautstärke des Schallereignisses und die Veränderung des Spektrums auf dem Weg zum Ohr eine tragende Rolle [Dickreiter, 1997].

Mithilfe unserer Hörerfahrung, die wir im Laufe unseres Lebens sammeln, und den zuvor beschriebenen Phänomenen zum Richtungs- und Entfernungshören können wir auch die Ausdehnung von Klängen, die Tiefe, Breite und Höhe, beschreiben. Die Umgebung, in der wir uns als Zuhörende befinden, ermöglicht eine Umhüllung durch einen Klang oder vieler Klangereignisse. Dies geschieht durch die Reflexionen an den Wänden, Gegenständen und Personen im Raum, die das Schallfeld beeinflussen [Kendall, 2014].

3.4 Psychoakustische Phänomene

Die Wahrnehmung von Klang, vor allem die Ortung, wird von weiteren Effekten beeinflusst – sogenannten psychoakustischen Phänomenen. Der vom Gehör aufgenommene Schall wird gefiltert, analysiert und getrennt. Im Gehirn findet ein selektiver Prozess statt,

der die wahrgenommenen physikalischen Reize reduziert [Ziemer, 2023]. Es gibt bestimmte und absolute Schwellenwerte, die den Bereich des Schalls eingrenzen, der schließlich psychologisch verarbeitet wird. Die erste Schwelle ist die Ruhe- oder Hörschwelle. Diese gibt an, ab welcher Schalldruckstärke Klang wahrgenommen wird. Sie ist – wie viele weitere psychoakustische Effekte – vom Frequenzgang abhängig. Vereinfacht gesagt werden höhere Frequenzen eher als tiefe Frequenzen bei einem niedrigeren Schalldruck wahrgenommen. Das Gehör verarbeitet Schall demnach nicht als Ganzes, sondern in Frequenzbändern [Ziemer, 2023].

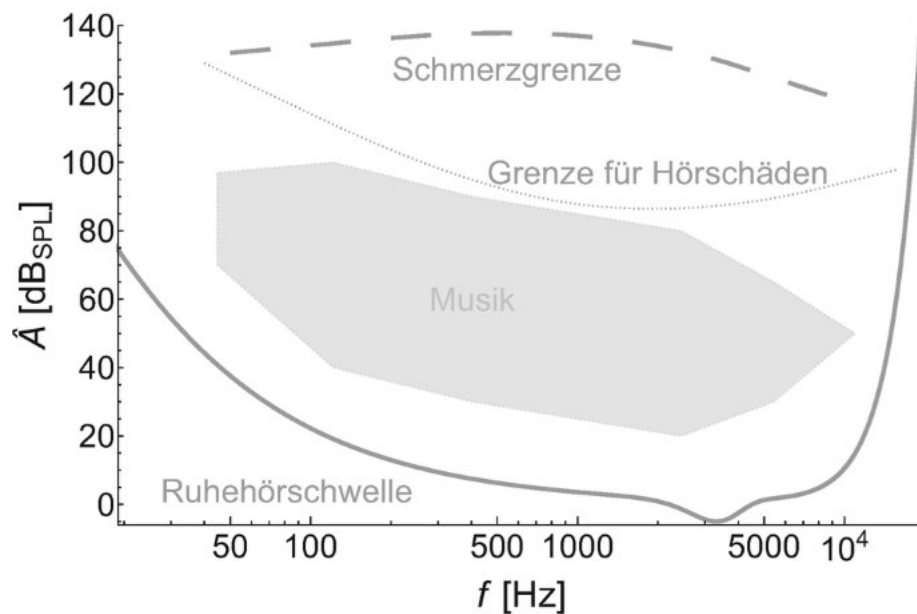


Abb. 2: Hörschwelle und Schmerzgrenze [Zwicker and Fastl, 1999, S. 17]

Ein weiteres relevantes Phänomen ist die Maskierung von Schallquellen im alltäglichen Leben, welches sich auch bei Spatial-Audio-Produktionen einsetzen lässt oder genutzt wird, um andere Schallquellen zu beeinflussen. Maskierung bedeutet die Überlagerung eines Schalls durch einen anderen. Im Gehirn werden laute Schallquellen priorisiert und solange der Pegel der anderen Quelle nicht angehoben wird, verschwindet sie in den Hintergrund. Befinden wir uns in einem mit Menschen gefüllten Raum, muss die Musik, die abgespielt wird, die sogenannte Maskierungsschwelle überschreiten, damit sie die gewünschte Aufmerksamkeit erzielt.

Es gibt verschiedene Maskierungsphänomene: die *monaurale Maskierung*, bei der der Maskierer – das Signal, das dominante Signal überlagern soll – nur von einer Quelle, zum Beispiel einem Kopfhörer, abgespielt wird. Bei der *gleichzeitigen Maskierung*, werden im Gehirn höhere Frequenzen priorisiert und verstärkt. Durch leise Klänge, die länger brauchen, um verarbeitet zu werden als laute, kann eine Vormaskierung stattfinden. Durch die Trägheit des Gehörs ist auch eine Post-Maskierung möglich. Erwähnenswert ist auch die

binaurale Maskierung. Werden unterschiedliche Signale auf beiden Kopfhörern wiedergegeben, gelangen diese gleichzeitig in den Hörstrom⁶ und werden als eins wahrgenommen. Diese Phänomene lassen sich als spektrale Maskierung (frequenzbezogen), zeitliche Maskierung und räumliche Maskierung einsetzen [Ziemer, 2023].

Ein weiteres Phänomen ist der Präzedenz-Effekt, auch als Gesetz der ersten Wellenfront bekannt. Schallereignisse werden primär durch die zuerst eintreffende Welle lokalisiert, auch wenn später eintreffende Signale lauter sind. Dieser Effekt tritt insbesondere bei transienten Signalen auf und verhindert, dass seitliche Reflexionen die wahrgenommene Richtung verfälschen – solange das Zeitintervall zwischen den Signalen unter 50 ms bleibt, da ansonsten eine Echo-Wahrnehmung einsetzt [Ziemer, 2023].

3.5 Spatial-Audio-Formate

Die für diese Arbeit untersuchten Produktions- und Kompositionsansätze werden in folgenden Formaten umgesetzt, die sich grundlegend in zwei Kategorien einordnen lassen: räumlich bezogene Formate und kopfbezogene Formate. Zu den räumlichen Formaten gehören die kanalbasierten Formate wie Quadraphon, Oktophon, 12.0, 22.2, individuelle mehrkanalbasierte Formate, szenenbasierte Formate wie Ambisonics, Higher Order Ambisonics und Wellenfeldsynthese sowie das objektbasierte Format Dolby Atmos. Das kopfbezogene Format Binaural wird über Kopfhörer abgespielt. Es gibt jedoch Ansätze, auch binaurale Inhalte über Lautsprecher wiederzugeben – siehe Unterunterabschnitt 3.5.4. Raumbezogene Formate werden über Lautsprecher wiedergegeben. Die Grundlage zur Verteilung des Klangs auf Spatial-Audio-Lautsprechersystemen erfolgt über Panningmethoden und die entsprechende Decodierung auf das benutzte System. Die Panningmethoden orientieren sich an unseren physischen und psychischen Fähigkeiten, Klang zu verarbeiten (siehe Unterabschnitt 3.3). Die in den folgenden Kapiteln beschriebenen Formate werden mit ihren Vor- und Nachteilen und somit ihrem Einfluss auf die Produktions- und Kompositionsarbeit – erläutert.

Als Werkzeuge, um diese Panningmethoden und die Verteilung des Klangs im Raum zu ermöglichen, werden viele verschiedene Programme in Form von VST-Plug-ins wie der IEM Plug-in Suite, alleinstehender Software wie MAX/MSP oder SPAT Revolution und viele weitere eingesetzt. Einige dieser Werkzeuge lassen sich mit mehreren Formaten kombinieren, während andere nur für ein bestimmtes Format entwickelt wurden. Die wichtigsten Programme werden in den jeweiligen Kapiteln erwähnt. Im folgenden Kapitel werden die relevanten Formate und deren Spatialisierungsweise erläutert.

⁶ Sequenz auditiver Ereignisse, die von unserem Gehirn wahrgenommen werden [Bregman, 1990].

3.5.1 Kanalbasierte Formate

In kanalbasierten Formaten wird jeder benutzte Kanal einem Lautsprecher zugewiesen. Die Zuweisung erfolgt über das Routing und Panoramaregler innerhalb der *Digital Audio Workstation* (DAW). Jeder Kanal enthält die Pegeleinstellungen und die Klangeinstellungen, die über Effektketten erreicht werden. Dieses Verfahren synthetisiert keine realen oder virtuellen Schallfelder, sondern erzeugt Phantomschallquellen mithilfe der Summenlokalisation. Dies wird durch das Vector Base Amplitude Panning (VBAP) ermöglicht [Weinzierl, 2008]. Da diese Arbeitsweise an feste Lautsprecheranordnungen gebunden ist, gibt es vorgegebene Standards die durch die International Telecommunication Union (ITU) definiert werden. Dies setzt voraus, dass die Abhörsituation auch nach diesen Standards gewährleistet wird. Da dies auf Konsumentenseite oft nicht der Fall ist, kann dies zu einem Nachteil werden, denn für das optimale Hörerlebnis sollten sich die Hörenden in der optimalen Hörposition, dem Sweet Spot befinden. Dieses Problem entfällt bei monophonen Wiedergabesystemen, da die Klangbalance über einen größeren Hörbereich stabil bleibt [Roginska and Geluso, 2018].

Stereo

Viele der in dieser Arbeit analysierten Herangehensweisen und Workflows lassen sich auf das nach wie vor dominierende Format Stereo zurückführen. Der Großteil der Komponierenden und Produzierenden haben in Stereo begonnen zu arbeiten, daher ist eine ausführliche Auseinandersetzung mit dem 'Ur-Format' in diesem Abschnitt angedacht.

Im Stereoformat werden zwei Lautsprecher mit den Kanälen Links und Rechts mit Audiosignalen bespielt. Nach dem ITU-Standard werden die zwei Lautsprecher im Winkel von 60° zum Hörenden aufgestellt. Durch die Verwendung mehrerer Lautsprecher tritt der Präzedenz-Effekt auf (siehe Unterabschnitt 3.4). Eine Phantomschallquelle wird erzeugt und ermöglicht es Schallquellen auf der Horizontalen zu verteilen und es entsteht das bekannte Stereobild [Görne, 2008]. Bei einem Rock-Mix kann so die Leadgitarre nach Links, die Rhythmusgitarre nach rechts und die Stimme in die Mitte gesetzt werden, ohne dass ein dritter Lautsprecher platziert werden muss. Dadurch sind die Instrumente leichter voneinander zu trennen und das Format bietet somit mehr künstlerische Freiheiten als die Arbeit in Mono. Wie bereits erwähnt, gibt es eine ideale Hörposition. Eine Abweichung des Hörenden von dieser Position führt in der Regel zum Zusammenfallen des Stereobildes. Spektrale, direktionale und räumliche Eigenschaften des Stereomixes gehen dabei verloren [Roginska and Geluso, 2018].

Stereo ermöglicht, ähnlich wie Mono, eine Tiefenstaffelung, indem Pegelunterschiede, Frequenzbalance und Hall gezielt genutzt werden. Diese ist jedoch durch das kleine gegebene

ne Panorama und fehlende Höhenlautsprecher begrenzt. Bei der Lautsprecherwiedergabe interagiert das Audiosignal mit dem Raum. Reflektierte Schallanteile können dabei eine räumliche Wahrnehmung erzeugen, die jedoch maßgeblich von den akustischen Eigenschaften des Wiedergaberaums abhängt [Roginska and Geluso, 2018]. Auch wenn Stereo durch Hall und Phantomschallquellen einen räumlichen Eindruck vermitteln kann, gilt es nicht als Spatial Audio, da die Wiedergabe auf eine horizontale Ebene beschränkt bleibt und keine explizite dreidimensionale Platzierung der Klangquellen möglich ist [Rumsey, 2001].

Quadrphonie, Surround und Oktophonie

Durch zusätzliche Lautsprecher in der Horizontalebene können virtuelle Schallquellen in 360°, um den Hörer platziert werden und der Grad der Umhüllung wird deutlich erhöht. Mit der Quadrphonie, die in den 1960er entwickelt wurde, begann der erste Versuch Musik in einem Mehrkanalformat zu kommerzialisieren. Einige herausragende Produktionen wie *The Dark Side of the Moon* von *Pink Floyd* oder *Tubular Bells* von *Mike Oldfield* profitieren laut Teilnehmenden der Spatial Audio Listening Sessions signifikant durch die Spatialisierung. Maßgebend für die Qualität war schon 1973 der Kompositions- und Produktionsansatz direkt im Spatial Audio Format zu produzieren. Die Quadrphonie konnte sich im Markt nicht durchsetzen, da der technische Aufwand zu groß war und die Systeme untereinander nicht kompatibel waren. Die Matrixsysteme der Quadrphonie wurden jedoch für die folgenden Surroundsysteme weiterentwickelt [Ziemer, 2023].

Durch die Aufstellung der Lautsprecher in großen Räumen können bei Quadrphonen Stücken Lücken in der Umhüllung des Hörenden entstehen, daher werden vor allem in der elektroakustischen Musik und vielen Klanginstallationen acht Lautsprecher in einem Kreis angeordnet. So kann ein homogenes Klangfeld um den Hörer geschaffen werden, das auch in größeren Räumen die Lücken zwischen den Lautsprechern schließen kann. Die Ausdehnung der Wiedergabe über mehrere Lautsprecher erhöht den künstlerischen Spielraum für Komponierende sowie Produzierende. Eine Besonderheit der Quadrphonie und Oktophonie ist das Spiel mit psychoakustischen Effekten [Ziemer, 2023]. Durch das gezielte Einsetzen der blauertschen Bänder⁷ kann eine Höhenwahrnehmung ohne den Einsatz von Höhenlautsprechern realisiert werden. Diese Erfahrung wurde mehrfach in den SALS und der experimentellen Forschungsarbeit bestätigt [Sudahl, 2023].

Multikanalformate und Installationen

Kanalbasierte Formate werden zum Beispiel durch die Renderer von Spat Revolution, den Renderer von Reaper64 oder durch Softwarekombinationen mit MAX MSP oder der Open-Source-Variante PureData umgesetzt. Ein Panning-Algorithmus für kanalbasierte

⁷ Blauertsche Bänder sind akustischen Frequenzbereiche, die für das menschliche Richtungshören in der Medianebene essenziell sind.

Formate ist das Vector Based Amplitude Panning, welches die Lautsprecher als Richtungsvektoren benutzt, um die Schallquelle als Vektor im Raum darzustellen. Dies wird dann durch die Änderung der Amplituden verschiedener Lautsprecher realisiert. Eine Bewegung einer Schallquelle durch den Raum wird durch eine Interpolation der Amplituden durch den VBAP-Algorithmus erreicht. Eine Alternative zu VBAP ist das Distance Based Amplitude Panning, kurz DBAP. Dieser Panning-Algorithmus unterscheidet sich dadurch, dass er die Position der Schallquelle mittels der relativen Lautstärke zwischen Lautsprecher und Schallquelle steuert. Er bietet sich deswegen für freie Lautsprecheranordnungen die oftmals bei Installationen verwendet werden an [Roginska and Geluso, 2018].

Mithilfe der zuvor erwähnten Spatialisierungsverfahren sind die Kunschtschaffenden nicht an feste Lautsprecherpositionen gebunden. Es werden meistens mindestens vier Lautsprecher in einer quadraphonieähnlichen Aufstellung oder mit dem Zusatz von Höhenlautsprechern verwendet. Diese Ansätze ermöglichen eine freie Kompositions- und Produktionsweise der Stücke und machen es möglich, sich an die Räumlichkeiten anzupassen. Es werden auch rein kanalbasierte Stücke bei bestehenden Lautsprechersystemen produziert, wie es im Immersive Audio Lab⁸ der HAW Hamburg der Fall ist. Solche Stücke sind jedoch nur für jeweilige Setups geeignet.

Loudspeaker Orchestra (Acousmonium)

Das Lautsprecherorchester wurde 1974 von François Bayle und der Groupe de recherches musicales (GRM) entwickelt und gilt als Meilenstein elektroakustischer Musikgeschichte. Es stellt einen frühen Ansatz der räumlichen Audiowiedergabe in der Musikpraxis dar und wird durch den Einsatz von mehreren Lautsprechern in unterschiedlichen Größen umgesetzt. Dieser Ansatz wurde in den heutigen Spatial-Audio-Formaten weiterentwickelt. Besonderheit des Acousmonium ist die Verteilung der Klangquellen im Raum durch eine manuelle Steuerung, welche als performativer Part gesehen werden kann. Diese Art ersten Pannings lässt sich in allen Spatial-Audio-Formaten wiederfinden [Berliner Festspiele, 2024].

Das Acousmonium etablierte die räumliche Bewegung von Klang als neue künstlerische Dimension. Der philosophische Einfluss der *musique concrète*, insbesondere durch Pierre Schaeffer und Pierre Henry, manifestierte sich im Konzept der Akusmatik, d.h. der bewussten Trennung von Klang und Quelle. Diese Idee ist auch für die binaurale Audiowiedergabe, Virtual-Reality-Audio und immersive Installationen von Bedeutung, da sie eine neue Form der Hörerfahrung schafft, bei der Klänge nicht mehr an sichtbare Quellen gebunden sind [Berliner Festspiele, 2024].

⁸ Umfasst 33 Koaxiale Lautsprecher, die in einer Halbkugel angeordnet sind und zwei Subwoofer.

3.5.2 Objektbasierte Formate

Mit der Einführung von *Dolby Atmos* im Jahr 2012 wurde gleichzeitig das erste objektbasierte Audioformat eingeführt, das in der Musikindustrie als potenzielles Spatial-Audio-Format gesehen wird, das bleibt. Mithilfe von Metadaten, die das Klangobjekt und die Vektorinformationen über die Platzierung im Raum enthalten, kann der Dolby Renderer die Signale entsprechend auf die für Dolby Atmos typischen Lautsprecheranordnungen senden [Gullö et al., 2024].

Weitere relevante objektbasierte Formate sind *MPEG-H Audio*, *DTS:X*, *AuroMax*, *Wwise Spatial Audio* oder *Sony 360 Reality Audio*. Diese Formate sind für Streaming, Virtual Reality, Gaming und auch für Musikproduktionen wichtig. Sie werden in dieser Arbeit nicht weiter beschrieben, da sie sich in der Musikkomposition und -produktion nicht etabliert haben und den Rahmen dieser Arbeit übersteigen.

Dolby Atmos

Dolby Atmos wurde als Kinoformat eingeführt und hat sich erst über die Einführung als Musikformat im Jahr 2019 auf Amazon und Tidal und schließlich mit der Integration auf Apple Music im Jahr 2021 als kommerzielles Spatial-Audio-Format durchgesetzt. Auch die technische Weiterentwicklung von Soundbars, wie die im Jahr 2019 eingeführte Atmos-fähige Ambeo-Soundbar von Sennheiser, hat dazu beigetragen, die Reichweite von Spatial-Audio-Formaten zu erhöhen. Soundbars werden immer erschwinglicher und technisch weiterentwickelt, sodass das Spatial-Audio-Erlebnis auch zu Hause zunehmend für die Konsumierende interessant wird.

Für die Arbeit in Dolby Atmos ist die Dolby-Atmos-Production-Suite oder die Dolby-Atmos-Mastering-Suite notwendig. Diese unterscheidet sich nur im Renderingprozess, der auf ein externes Hardwaregerät verlagert wird, um die Rechenleistung des Rechners mit der DAW nicht zu beeinträchtigen. Das Softwarepaket besteht bei der Production Suite aus dem Renderer und dem Panner. Diese sind mittlerweile in den meisten DAWs integriert. Dolby Atmos bietet die Chance, 128 Objekte in einem Mix zu benutzen. Um mit Dolby Atmos zu arbeiten, ist man softwaretechnisch an den Dolby-Atmos-Renderer, an den Dolby-Atmos-Panner und für das Mastering an den Dolby-Atmos-Mixer gebunden. Ein Objekt wird klassisch als eine Spur in der DAW angelegt, hier sind weitere Effekte wie EQ, dynamische Effekte oder Halleffekte anwendbar [Pfanzagl-Cardone, 2023].

Die Anzahl der Objekte ist durch die Auflösung begrenzt. Mit einer Samplerate von 48 kHz sind 128 Objekte verwendbar. Es gibt folglich Limitierungen auch in Bezug auf die benutz-

baren *Beds*⁹ und den Low Frequency Effects(LFE), der nur über das Bed Signale annimmt. Wie mit diesen Limitierungen umgegangen wird, wird in Unterunterabschnitt 5.2.10 beschrieben. Der Dolby Renderer bietet Optionen zum Heruntermischen in Echtzeit an. Ein 7.1.4-Mix kann auf 5.1 oder Stereo heruntergemischt oder durch die Dolby-HRTF binauralisiert werden. Somit entsteht die Möglichkeit eines neuen Workflows: dem Atmos-First-Workflow. Dieser schreibt vor direkt in Dolby Atmos zu produzieren und lediglich durch den Renderer in Stereo herunterzumischen [Pfanzagl-Cardone, 2023].

Dolby Atmos eröffnet neue kompositorische Möglichkeiten, da Objekte frei im Raum positioniert werden können. Die Wiedergabe von Dolby Atmos ist skalierbar. Stücke können binaural wiedergegeben werden oder für Surroundsysteme automatisch vom Renderer heruntergemischt werden. Mittlerweile sind die benötigten Werkzeuge in DAWs wie Pro Tools, Logic, Ableton und Nuendo integriert und erleichtern den Zugang zum System.

3.5.3 Szenenbasierte Formate

Ambisonics und die *Wellenfeldsynthese(WFS)*, basieren auf der Methode, ein ganzes Klangfeld zu kodieren, das unabhängig von der Anordnung der Lautsprecher, im gesamten Raum konsistent ist. Das ermöglicht einen flexiblen Einsatz, bringt aber auch Nachteile mit sich (siehe Unterunterabschnitt 5.3.2). Beide Systeme verfolgen dasselbe Ziel, basieren jedoch auf unterschiedlichen Grundlagen. Die WFS beruht auf dem physikalischen Prinzip von Huygens, während Ambisonics auf mathematischen Modellen, den sogenannten sphärisch-harmonischen Funktionen basiert [Zotter and Frank, 2019].

Ambisonics und Higher Order Ambisonics

Das auf sphärisch-harmonischen Funktionen basierende Format Ambisonics wurde 1973 von Gerzon als Alternative zu kanalbasierten und Surround-Formaten eingeführt. Anstatt Objekte im Raum zu erzeugen, wird ein komplettes Klangfeld als Sphäre um den Hörer erzeugt, das jede Richtung in Bezug auf den Schalldruck gleich behandelt. Das Ur-Format von Ambisonics ist First Order Ambisonics(FOA), bestehend aus vier Signalen:

- W – Omnidirektionaler Druck (Summensignal)
- X – Richtungskomponente auf der Horizontalachse Vorne–Hinten
- Y – Richtungskomponente auf der Horizontalachse Links–Rechts
- Z – Richtungskomponente auf der Vertikalachse (Höheninformation)

Um diese vier Signale zu erhalten, ist eine Matrizierung der aufgenommenen Signale erforderlich. Ein FOA-Recording erfolgt mit vier Mikrofonen mit Nierencharakteristik, die

⁹ Festes, kanalbasiertes Audiolayout (meist 7.1.2), in dem bestimmte Elemente des Mixes platziert werden [Pfanzagl-Cardone, 2023].

als Tetraeder angeordnet sind. Diese Aufnahme ist im sogenannten A-Format und bedarf einer Encodierung ins B-Format, welches schließlich mit einem Decoder auf das Lautsprechersystem gesendet werden kann [Pfanzagl-Cardone, 2023].

Die räumliche Auflösung von FOA ist gering, daher wurde *Higher Order Ambisonics* (HOA) – eingeführt. Je mehr Lautsprecher zur Wiedergabe verwendet werden, desto feiner wird die Ortung von Klangquellen, und desto größer wird der Sweet Spot. Die Anzahl der Aufnahme- und Wiedergabequellen spiegelt die Ordnung wider. Dabei gilt: Anzahl der benötigten Lautsprecher/Mikrofone = $(\text{Ordnung} + 1)^2$. Möchte man in Ambisonics dritter Ordnung arbeiten werden bereits 16 Lautsprecher oder 16 Mikrofone benötigt [Zotter and Frank, 2019].

Ambisonics bietet im Vergleich zu kanal- und objektbasierten Formaten einige Vor- und Nachteile. Es eignet sich, wie bei dem Workflow mit freien Lautsprecheranordnungen und Software wie SPAT, besonders für immersive Kunstformen, bei denen sich die Hörenden im Raum bewegen. Es kann auf verschiedenen Lautsprecheranordnungen oder binaural über Kopfhörer wiedergegeben werden und ermöglicht das Verwenden eigener HRTFs. Somit ist es weniger beeinflusst von Renderern, die nicht transparent arbeiten. Es bietet volle räumliche Flexibilität und Rotation durch freie Softwareprogramme wie die Plug-in Suite von IEM und hebt sich dadurch von anderen Formaten ab.

Ambisonics leidet allerdings unter begrenzter Lokalisationstreue und Klangpräzision, da Klangquellen verschwimmen können und ungenau wahrgenommen werden. Dies ist besonders bei niedrigen Ambisonics-Orders der Fall. Eine hohe Präzision ist somit nur mit hohem technischem Aufwand, ab der dritten Ordnung, gegeben. Dafür werden mindestens 16 Lautsprecher benötigt. Die Rechenleistung steigt im Gegensatz zu kanalbasierten und objektbasierten Formaten erheblich an – je nachdem, wie viele Spuren und Plug-ins im Projekt verwendet werden. Ambisonics muss für jede Wiedergabesituation neu dekodiert werden, wodurch sich Artefakte oder Qualitätsverluste ergeben können [Zotter and Frank, 2019].

Wellenfeldsynthese

Die Wellenfeldsynthese (WFS) wurde in den 1980er Jahren entwickelt und beruht auf dem Huygensschen Prinzip, welches die Ausbreitung von zwei- und dreidimensionalen Wellen und der dabei auftretenden Phänomene beschreibt. Es besagt, dass jeder Punkt einer Wellenfront als Ausgangspunkt einer neuen Kreis- bzw. Kugelwelle im dreidimensionalen Raum mit gleicher Geschwindigkeit und gleicher Frequenz – einer sogenannten Elementarwelle – betrachtet werden kann [Berkhout et al., 1993]. Ein Interferenzfeld mit kleinem Sweet Spot ist somit vermeidbar, und es entsteht eine größere Sweet Area dank

der synthetisierten Elementarwellen. Die WFS ermöglicht es, Phantomschallquellen vor den Lautsprechern zu positionieren und nicht, wie bei der kanalbasierten Variante, im dahinterliegenden Raum [Görne, 2015].

Aufgrund des hohen technischen Aufwands und der wenigen Systeme, die für Kunstschaffende zugänglich sind, wird die WFS seltener angewandt – auch wenn eine exakte räumliche Reproduktion von Klangquellen gegeben ist und Klangquellen stabil an einem festen Ort, unabhängig von der Position der Hörer:innen, erscheinen. Zusätzlich ist der Rechenaufwand enorm, und die bisher erwähnten DAWs unterstützen das Format nicht, was die Zugänglichkeit und schließlich den Workflow erheblich erschwert.

3.5.4 Kopfbezogene Formate

Das relevanteste kopfbezogene Format für diese Arbeit ist die Wiedergabe von Spatial Audio von binauralisierten Inhalten über Kopfhörer. Ein weiterer Ansatz, ist die transaurale Wiedergabe binauraler Inhalte mit zwei Lautsprechern, welche mit einem kleinem Abstand an den Kopf des Hörenden positioniert werden [Roginska and Geluso, 2018].

Binaural

Die Wiedergabe von Spatial Audio über Kopfhörer wird durch die Head Related Transfer Function, übersetzt Kopfübertragungsfunktion, ermöglicht. Die HRTF übersetzt das raumbezogene Signal in ein kopfbezogenes Signal. Dadurch werden irritierende Wahrnehmungseffekte und die Im-Kopf-Lokalisation ausgelöscht. Generell wird das Hören mit einer HRTF als angenehmer betrachtet, da es sich bei Nutzung der eigenen HRTF um das natürlich gewohnte Hören handelt [Görne, 2015].

Die HRTF simuliert interaurale Zeitdifferenzen (ITD), interaurale Pegeldifferenzen (ILD) und die spektrale Filterung, welche durch die Ohrmuschel entsteht. Während die ITD für die tieffrequente Lokalisation und die ILD für die hochfrequente Lokalisation in der Horizontalen verantwortlich sind, beeinflussen Pinna-Effekte und die Frequenzfilterung die vertikale Lokalisation [Roginska and Geluso, 2018].

Die Erstellung hochwertiger HRTFs ist ein komplexes Verfahren, das mit einem hohen technischen Aufwand einhergeht. Somit sind individuelle HRTFs schwer zugänglich, und es werden stattdessen standardisierte HRTFs verwendet. Eine oft genutzte HRTF ist die des Kunstkopfes KU100. Da dies eine Nachbildung eines generischen männlichen Kopfes ist, werden binauralisierte Inhalte nicht bei jedem mit dem gewünschten Effekt gehört, denn bei nicht passenden HRTFs kann es zu Fehlwahrnehmungen von Phantomschallquellen und Lokalisationsinversion kommen. Im schlechtesten Fall ist eine Außenkopflokalisation nicht möglich [Roginska and Geluso, 2018].

Trotz der erwähnten Komplikationen mit unstimmigen HRTFs ist die Wiedergabe von Binaural Audio die gängigste Methode, Spatial Audio zu vermarkten und verfügbar zu machen. Die meisten Menschen hören Musik über Kopfhörer – so muss kein komplexes und kostspieliges Lautsprechersystem verwendet werden. Zusätzlich wird das Sweet-Spot-Problem vermieden, und auch Crosstalk Cancellations (CTCs)¹⁰ können ausgeschlossen werden [Ziemer, 2023].

Transaural

Transaurales Audio ermöglicht die Wiedergabe von binauralem Klang über Lautsprecher, indem es mittels CTC unerwünschte Schallübersprechungen zwischen den Ohren eliminiert. Dadurch entsteht eine räumliche Klangillusion ohne Kopfhörer. Allerdings funktioniert dies nur in einem begrenzten Sweet Spot, da kleinste Kopfbewegungen die Lokalisationsgenauigkeit beeinträchtigen können. Raumakustik und Reflexionen stellen eine Herausforderung dar, da sie die CTC-Effektivität mindern und die räumliche Wahrnehmung verfälschen können. Künstlerisch bietet transaurales Audio die Möglichkeit, binaurale Klangwelten in offenen Rauminstallationen erlebbar zu machen, bleibt aber aufgrund technischer Einschränkungen auf spezialisierte Anwendungen beschränkt [Roginska and Geluso, 2018].

Headtracking

Beim Hören über HRTF-kompensierte Kopfhörer kommt es vermehrt zu einer Front-Back-Inversion: Klänge werden als Hörereignisse wahrgenommen, die von hinten kommen, obwohl sie vor dem Hörenden generiert werden und umgekehrt. Durch Headtracking kann dieser Effekt weitgehend vermieden werden, und der Höreindruck bleibt räumlich konsistent, da die Kopfbewegung die Lokalisation unterstützen [Hess, 2012].

¹⁰ Übersprecheffekte bei der Lautsprecherwiedergabe.

4 Methodik

Die Arbeit mit Spatial Audio hat in den letzten 20 Jahren deutlich an Bedeutung gewonnen. Eine zunehmende Anzahl an Audioschaffenden arbeitet mit Spatial-Audio-Formaten und testen die Grenzen und Möglichkeiten aus, um immer bessere Inhalte zu schaffen. Die Arbeit in Spatial Audio ist sehr experimentell und unterliegt einer höheren Komplexität und mit Dolby Atmos ist das erste erfolgreiche Spatial-Audio-Format im Konsumentenmarkt angekommen. Die Anzahl der Fachkundigen in Spatial Audio ist im Vergleich zu Stereo deutlich geringer, daher hat es sich angeboten, umfangreiche *Experteninterviews* zu führen. Dazu wurde ein mehrteiliger Fragebogen entworfen, der je nach Spezialgebiet der Befragten angepasst wurde. Dies ermöglichte einen umfassenden Einblick in die Arbeit und bietet die Chance auf weitere Diskussions- und Forschungsgrundlagen. Insgesamt wurden 17 Interviews geführt und 8 weitere Quellen als Sekundärdaten – bestehend aus Interviews und Publikationen zum Thema Produktions- und Kompositionsmethoden – für die Analyse ausgewählt.

Im Gegensatz zu quantitativen Datenerhebungsverfahren, die numerische Daten in Zahlenform mit standardisierten Messinstrumenten erfassen, zielt die qualitative Datenerhebung auf nicht-numerische Daten ab, die beispielsweise durch Interviews in Textform gewonnen werden. Es können auch Bilder, Grafiken oder Filme zur qualitativen Datenerhebung hinzugezogen werden [Kuckartz, 2012]. Die durch Interviews erhobenen Daten gelten als aktive Daten, da sie im Rahmen der Forschung neu generiert wurden und zuvor nicht existierten. Im Gegensatz dazu werden Sekundärdaten als natürliche Daten bezeichnet, da sie unabhängig vom konkreten Forschungskontext bereits vorhanden waren [Schreier, 2023].

Die qualitative Datenerhebung ist in diesem Fall eine ideale Methode, da sie es ermöglicht, die subjektiven Perspektiven und kreativen Prozesse der Befragten zu verstehen und in ihrem Kontext zu analysieren. Somit können individuelle Perspektiven, Erfahrungen und Bedeutungsstrukturen von Komponierenden und Produzierenden im Bereich Spatial Audio erfasst werden. Dabei liegt der Fokus darauf, wie diese Personen ihre Ansätze und Philosophien in der Praxis umsetzen und interpretieren. Es wird nach einem subjektiven *Warum* gefragt, um die Hintergründe zu erforschen, statt einem objektiven *Wie Viel*, das eher in quantitativen Studien im Vordergrund steht [Kirchmair, 2022].

Des Weiteren basiert diese Arbeit auf einem Mixed-Methods-Ansatz, um eine umfassende und aussagekräftige Analyse der Methoden zu ermöglichen. Mithilfe von Primärdaten in Form von Experteninterviews sowie ergänzenden Sekundärdaten werden mittels einer qualitativen Inhaltsanalyse Aussagen zu den aufgestellten Thesen gewonnen.

Es wurde sich bewusst für ausführliche Interviews mit einem umfangreichen Fragebogen entschieden, um fundierte Aussagen treffen zu können und darüber hinaus tiefere Einblicke über die Arbeits- und Denkprozesse der Befragten zu gewinnen. Da sich hierfür eine perfekte Chance bot, über die Thesen hinaus Daten zu generieren, werden diese im Diskussionsteil aufgegriffen.

4.1 Fokussierte Experteninterviews

Zur Generierung der Daten wurde eine Interviewkombination entwickelt, die sich der qualitativen Interviewführung zuordnen lässt. Motivation dafür war es, einen offenen Diskurs zu ermöglichen, da die Arbeit in Spatial Audio noch sehr experimentell ist und eine höhere Komplexität aufweist [Mikos, 2017]. Ein weiterer wichtiger Faktor der qualitativen Interviews ist das Prinzip der Reflexivität, welches vom Interviewer verlangt, individuelle Aussagen im Kontext übergeordneter Relevanzsysteme zu interpretieren. Deswegen wurde die Interviewführung regelmäßig reflektiert und nach Bedarf angepasst [Mikos, 2017].

Es handelt sich bei der Mischform um Komponenten aus der Experten-, fokussierten und problemzentrierten Interviewführung. Um ästhetische Entscheidungen zu erfassen, wurden Komponenten aus dem fokussierten Interview gewählt. Es wurden offene Fragen gestellt, die einem klaren thematischen Rahmen unterlagen. Die Interviewten haben dadurch die Freiheit, ihre Sichtweisen ausführlich darzustellen. Einige Aspekte der Experteninterviews wurden verwendet, um technisches Wissen über Tools, Workflows und Formate zu sammeln. Hierfür wurden die Fragen je nach der befragten Person angepasst und das Interview gezielter auf das jeweilige Fachgebiet ausgerichtet. Elemente aus den problemzentrierten Interviews wurden gewählt, um Herausforderungen zu erkennen und Lösungsansätze für bestehende Probleme zu erarbeiten. Somit bringen die Befragten eigene Perspektiven bzw. Lösungsansätze durch teils offene und strukturierte Fragen ein, die einem grundlegenden Leitfaden folgen, der allen Interviewansätzen zugrunde liegt. Diese Kombination aus den verschiedenen Interviewmethoden ermöglicht eine effektive und zielführende Gesprächsführung in komplexen Themen und den gewünschten offenen Diskurs [Mikos, 2017].

Dieser Interviewansatz wurde aus der Erfahrung aus der Arbeit „Sound in Space“, einer experimentellen Forschungsarbeit, weiterentwickelt. Hier wurde ein Ansatz des fokussierten Interviews auf eine Gruppensituation angewendet. Diese Form des fokussierten Interviews ist an ein offenes Gruppengespräch angelehnt, um eine klassische Interviewsituation und damit verbundene Biases zu vermeiden und schließlich eine angenehmere Umgebung für die Probanden zu schaffen. Obwohl die Interviewsituation offen gestaltet ist, ist sie den-

noch indirekt durch die Hypothese gelenkt und kann dadurch gezielt für die Evaluation der Wirkung von Medien auf Teilnehmende eingesetzt werden [Mikos, 2017]. Daher wurde bei der Gestaltung der Interviewform darauf geachtet, dass diese Aspekte mit einfließen.

Der Fragebogen besteht aus Schlüssel- und Eventualfragen. Schlüsselfragen sollen eine Erzählung anregen, während Eventualfragen bestimmte Aspekte ansprechen und zur Lenkung der Interviewsituation dienen. Entscheidend hierbei war, dass die Fragen den Befragten die Möglichkeit geben, eine individuelle Perspektive zu entwickeln und nicht einfach mit *ja* oder *nein* beantwortet werden können [Mikos, 2017]. Der Fragebogen ist im digitalen Anhang A zu finden.

4.2 Probanden

Wer ist eigentlich Spatial-Audio-Experte?

Aufgrund der Neuheit der Spatial-Audio-Formate und der geringen Anzahl an Langzeitforschenden sowie praktizierenden Anwendenden können Audioschaffende als Experten gelten, wenn sie sich theoretisch mit Spatial Audio auseinandergesetzt und praktische Erfahrungen in der Produktion oder Komposition in einem entsprechenden Format gesammelt haben. Dies ermöglicht Masterstudierenden, die sich mit dem Thema ausführlich befasst haben, an den Interviews teilzunehmen. Die Kontaktaufnahme erfolgte hauptsächlich per E-Mail und direkt persönlich vor Ort an der HAW Hamburg. Die E-Mails umfasste die Vorstellung des Projekts, Ziel der Interviews und Datenschutzaspekte. Es wurden insgesamt 17 Personen interviewt und in drei Kategorien eingeteilt:

Musikproduktion und Aufnahme - 6 Personen

Unter Musikproduktion und Aufnahme fallen Spatial-Audio-Produktionen, die für Formate wie Dolby Atmos produziert und komponiert wurden sowie Spatial-Audio-Aufnahmetechniken.

Experimentelle Arbeiten - 7 Personen

Unter Experimentellen Arbeiten fallen Stücke, die zu forschenden Zwecken entstehen. Darunter fällt die Arbeit mit der Wellenfeldsynthese, experimentelle Lautsprecher Installationen und die Arbeit in Genres wie elektroakustische Musik oder Music concrete.

Hörspiel und Soundscapes - 4 Personen

Diese Kategorie umfasst die Arbeit an Hörspielen und Soundscapes in Dolby Atmos und Ambisonics sowie an Installationen in diversen Mehrkanal-Formaten. In dieser Kategorie gibt es Überschneidungen zur Musikproduktion und zu Experimentellen Arbeiten. Grundsätzlich sollen diese Kategorien dazu dienen die Probanden einzuteilen und Ge-

meinsamkeiten sowie Unterschiede zwischen den verschiedenen Bereichen herauszufinden. Um den Rahmen der Arbeit einzuschränken wurden rein auditive Formate gewählt.

Zur Bewahrung des Datenschutzes wurden die Interviews nach der DSGVO anonymisiert. Es wurden über 40 Einladungen zu den Interviews versendet, von denen ein Großteil nicht beantwortet wurde oder aus zeitlichen und finanziellen Gründen die Anfrage abgelehnt wurde. Aufgrund der Nähe zur HAW und dem Tonlabor stammt der Großteil der Befragten aus dem Masterstudiengang Zeitabhängige Medien – Sound/Vision und Audioschaffenden aus dem Umkreis Hamburgs. Es ist daher anzunehmen, dass die Nähe zur HAW und dem Tonlabor einen prägenden Einfluss auf die Zusammensetzung der Stichprobe hatte. Dieser Aspekt wird in der Diskussion aufgegriffen und reflektiert (siehe Abschnitt 7).

4.3 Fragebogen

Die Erstellung des Fragebogens erfolgte in mehreren Schritten und wurde während der Interviews je nach Zielgruppe leicht abgeändert, um sich an die individuelle Situation mit den Probanden anzupassen. Durch Gespräche mit mehreren Experten in der Medien- und Kunstforschung (Kris Kuldkepp, Stefan Troschka, Robert Mores und Thomas Görne von der HAW Hamburg) wurden Fragen entfernt, hinzugefügt und optimiert. Der Fragebogen wurde an die drei unterschiedlichen Probandengruppen angepasst und eine englische Version erstellt.

Der Fragebogen wurde in drei Abschnitte eingeteilt. Zu Beginn wurden die Einstiegsfragen gestellt. Hier ging es darum, Hintergründe zur Arbeit der befragten Person zu erforschen: Seit wann arbeitet die Person in Spatial Audio, was waren ihre Beweggründe, und welche Software wird benutzt? Zu jedem Abschnitt wurde eine Übergangsfrage zur Einleitung in den nächsten Abschnitt gestellt.

Der Hauptteil beinhaltete die Schlüsselfragen des Interviews: *Wie sieht ein typischer Produktionsprozess und Kompositionsprozess eines Spatial-Audio-Stücks aus, wie gehen die Probanden mit der Technik und den Klängen in Kompositionen um, und wie ist ihre Herangehensweise in Bezug auf den Raum?* Für den Abschluss der Interviews wurden Fragen wie: *Welche Tipps würdest du einem Neueinsteiger geben?* oder *Welche Werkzeuge bzw. Plug-ins wünschst du dir, die es noch nicht gibt?* gestellt. Zusätzlich wurden Eventualfragen formuliert, die je nach Situation gestellt wurden.

4.4 Durchführung der Interviews

Die Interviews wurden teils persönlich, teils per Videocall geführt. Hier galt es zu beachten eine angenehme Interviewsituation für die Befragten zu schaffen.

Die meisten Interviews in persona fanden an der HAW Hamburg, am Ligeti-Zentrum oder im Umkreis von Hamburg statt. Die Interviews wurden mit einem Zoom Recorder H4n Pro und zwei SM58 von Shure mit Tischstativen aufgenommen, um die Konversationen für die Transkribierung trennen zu können. Als Backup wurde das XY-Stereo-Mikrofon des Zoom H4n Pro benutzt. Die Interviews per Videocall wurden mithilfe der Plattform Zoom und dem Programm Open Broadcast Studio(OBS) aufgenommen. OBS ist ein Open-Source-Programm und ist speziell zum Aufnehmen und Streamen entwickelt worden. Es wurde aufgrund des Datenschutzes auf die Aufnahme des Videos verzichtet. Dieser Interviewansatz ermöglichte es, die Interviews selbstständig durchzuführen und nicht auf eine zweite Person angewiesen zu sein, die sich um die Technik oder den Ablauf kümmert, da der Fragebogen entsprechend vorbereitet wurde.

Die Länge der Interviews richtete sich grundlegend nach dem Umfang des Fragebogens und der Zeit, die den Befragten zur Verfügung stand. So variierte die Länge der Interviews zwischen 45 und 140 Minuten. Ein weiterer Faktor, der die Länge der Interviews beeinflusst hat, war die Erfahrung der Befragten in Spatial Audio. Personen, die in mehreren Formaten gearbeitet haben und an mehreren Werken gearbeitet haben, konnten in der Regel die Fragen ausführlicher beantworten. Die Herausforderung bestand darin, das Interview so zu führen, dass in jedem Fall die Schlüsselfragen beantwortet werden und die Antworten auf Eventualfragen und weniger relevante Fragen einzuschränken, wenn die Zeit es nicht zuließ. Dieses Format ermöglichte es, bei ausführlichen Interviews über den Grundumfang hinauszugehen und somit zusätzliche Informationen zu gewinnen. Zur Orientierung der Interviews diente der Fragebogen, der durch zusätzlichen Notizen aus vorherigen Interviews erweitert wurde. Während der Interviews wurden bestimmte Fragen markiert, wenn diese in einer vorherigen Antwort beantwortet wurden oder noch einmal aufgegriffen werden sollten.

4.5 Aufbereitung der Interviews

Die Interviews wurden in mehreren Schritten für die Analyse aufbereitet. Diese bestanden aus der Aufbereitung des Audiomaterials, der Transkription und der Codierung der Interviews in *MAXQDA* einer *Computer assisted qualitative data analysis software*.

Aufbereitung des Audiomaterials

Die aufgenommenen Interviews wurden im Ersten schritt in der DAW Cubase bearbei-

tet. Diese Aufbereitung diente dazu, den Transkriptionserkennungsprozess durch die KI von MAXQDA zu optimieren¹¹. Je besser die Sprachverständlichkeit, desto besser konnte die KI die Interviews transkribieren. So wurden die Audiofiles normalisiert, die Sprecher geequalized und bei Bedarf komprimiert. Außerdem wurden Pausen und nicht relevante Passagen herausgeschnitten.

Transkription

Die Transkription der Audiofiles wurde mit der KI von MAXQDA vorgenommen. Diese wurde vor dem ersten Interview getestet und erstellt eine inhaltlich korrekte Transkription. Eine nachträgliche Bearbeitung der Transkriptionen war dennoch notwendig, da einige Fachbegriffe nicht richtig erkannt, und markante sprachliche Fehler ausgebessert werden mussten. Während der Verbesserung der Transkripte wurde gleichzeitig die Codierung der Texte vorgenommen. Dieser Workflow hat viel Zeit erspart und es ermöglicht, deutlich mehr Interviews durchzuführen. Um die Genauigkeit der KI zu optimieren, wurde ein Vokabular vorgegeben werden. Dieses Vokabular ermöglichte der KI, noch unbekannte Begriffe, die vor allem in der Audioproduktion verwendet werden, korrekt zu erkennen. Das Vokabular wurde mit der Zeit erweitert und umfasst folgende Begriffe:

Musikproduktion, Spatial Audio, HAW, Lautsprecher, Panning, Volume, Lautstärke, 3D-Audio, Software, Plug-ins, IEM Plug-ins, HRTF, Binauralisierung, Layer, High and low frequencies, Dolby Atmos, Ambisonics, Impulsantwort, Psychoakustik.

Codierung in MAXQDA

Die Codierung mit MAXQDA für Interviews ist ein systematischer Prozess, bei dem Textpassagen mit Codes versehen werden, um bestimmte Themen, Konzepte oder Muster zu identifizieren. Zunächst wurde ein Codierleitfaden erstellt, der auf den Forschungsfragen basiert und die relevanten Kategorien definiert. Während der Transkription wurden den Textstellen entsprechende Codes zugewiesen, die entweder deduktiv (basierend auf vorab definierten Kategorien) oder induktiv (aus dem Material heraus entwickelt) entstanden. MAXQDA ermöglicht es, Codes hierarchisch zu organisieren, miteinander zu verknüpfen und Zusammenhänge zwischen Kategorien zu analysieren. Dies erleichtert es, zentrale Erkenntnisse aus den Interviews systematisch zu extrahieren und mit den Forschungszielen in Verbindung zu bringen. Insgesamt wurden 1833 Codes den Textabschnitten zugeordnet. Zu einigen Fragen konnten mehrere Codes den Antworten zugewiesen werden. Zum Beispiel konnte aus der Schlüsselfrage *Welche Methoden verwenden Sie für die Erschaffung des Raumes in der Produktion?* folgende Codes abgeleitet werden:

- Umgang mit Räumlichkeit, Nachhall und Raum

¹¹ Zur sprachlichen Überarbeitung der Arbeit – hinsichtlich Rechtschreibung, Zeichensetzung und Schreibstil – wurde ChatGPT als unterstützendes Werkzeug verwendet.

Zusätzlich konnten weitere Codes aus den Antworten der Frage zugeordnet werden. Dazu gehören der Umgang mit Klängen, Herangehensweisen, Prozesse und Workflows.

4.6 Auswahl der Sekundärliteratur

Zur Erweiterung des Datensatzes und um genauere Aussagen treffen zu können, wurden die 17 Interviews mit acht weiteren Sekundärquellen ergänzt. Die Sekundärliteratur umfasst unter anderem Interviews, Artikel aus Zeitschriften wie dem VDT Magazin, Conference Paper der AES und Podcastbeiträge, in denen Experten über das Thema sprechen. Diese wurden nach den definierten Kriterien aus Kapitel 4.2 ausgewählt und anschließend in die drei verschiedenen Kategorien eingeteilt.

Sekundärquellen, die der Musikproduktion zugeordnet werden, sind:

- Podcast von HIFI.de kHz und Bitgeflüster vom 13.04.2022 in der Ausgabe #14: Interview mit David Ziegler von Dolby Atmos.
- Der VDT-Beitrag der Ausgabe 4/2024: Mastering für objektbasierte Musikproduktion, bei dem mehrere Experten involviert waren.
- Das AES Paper: An Investigation Into Immersive Contemporary Popular Music Mixing Practice.
- Interviews auf www.soundonsound.com und <https://pureaudiorecordings.com>: Morton Lindberg über 3D-Audio.
- Interviews auf www.amazona.de vom 19.12.2022: Produzent Hans Martin Buff über 3D-Audio.

Mit den sechs Primärquellen fallen insgesamt elf Quellen in die Kategorie der Musikproduktion und bilden die größte Gruppe. Zur Kategorie Hörspiel und Soundscapes kommen zu den vier Primärquellen zwei weitere Beiträge hinzu:

- Beitrag im VDT-Magazin 1/2024: 3D-Audio-Hörspiel mit dem Namen Erdsee
- Podcast von HIFI.de kHz und Bitgeflüster #58: 3D-Sound: Gibt es Alternativen zu Dolby Atmos? Mit Martin Rieger (VRTonung).

Somit sind sechs Quellen insgesamt der Hörspiel- und Soundscape-Kategorie zugeordnet. Die Kategorie der Experimentellen Arbeiten mit 7 Primärquellen wird mit einem weiteren Beitrag ergänzt und umfasst somit acht Quellen:

- Interview mit Natasha Barrett vom 6.9.2023 auf der Website Avant Music News.

Insgesamt konnten 25 Quellen als Grundlage dieser Arbeit zusammengeführt werden, die durch die Datenmenge qualitativer Antworten eindeutige Aussagen zu den Thesen ermöglichen.

4.7 Analyse mit MAXQDA

Zur Unterstützung der Analyse und für die Transkribierung der Daten wurde die bereits erwähnte Computer assisted qualitative data analysis software namens MAXQDA eingesetzt.¹² Dieses Programm unterstützt eine Vielzahl von Dateiformaten, die in verschiedenen Analyseverfahren untersucht werden können. Entscheidend für die Wahl der Software war die Unterstützung von Audiocodern wie WAV und MP3 sowie den Textformaten PDF, ODT und DOC. Außerdem ist der Analyse-Support in verschiedenen Sprachen notwendig, da auch einige Interviews auf Englisch gehalten wurden [Rädiker and Kuckartz, 2019]. Ein weiterer wichtiger Faktor für die Wahl der Software ist die Arbeit mit Codes, die für eine Strukturierung unabdingbar ist. Der Ursprungsgedanke hinter *MAXQDA* ist die Auswertung von in der empirischen Sozialforschung erhobenen Daten. Die Analysefunktionen lassen sich aber auch für Arbeiten wie diese problemlos einsetzen. Weitere Analysefunktionen sind das Verfassen von Memos und Kommentaren, paraphrasieren von Textstellen, thematische Zusammenfassungen zu erstellen, die Codesuche sowie die Unterstützung für Mixed-Methods-Ansätze [Rädiker and Kuckartz, 2019].

Analysen in Programmen wie MAXQDA durchzuführen, wird von einigen Fachkundigen kritisiert. Argumentiert wird, dass den Nutzern dieses Qualitative Data Analysis Software(QDAS)-Programm bestimmte Analysemethoden aufgedrängt und der Interpretationsraum eingegrenzt werde, da zu stark in Kategorien gearbeitet werde. Auf der anderen Seite werden QDAS-Programme als Werkzeugkasten gesehen, die den Prozess unterstützen und durch viele Funktionen zur Analyse als Erweiterung betrachtet werden können. In dieser Arbeit diene MAXQDA als Tool zur Analyse, wobei die gewählte Methodik einen flexiblen Interpretationsprozess ermöglichte.

Die strukturierende qualitative Inhaltsanalyse nach Kuckartz ist für diese Arbeit eine ideale Form der Analyse, denn sie bietet ein systematisches Arbeiten mit voller Transparenz und nachvollziehbaren Aussagen. Die Inhaltsanalyse nach Kuckartz ist in ihrer Form sehr flexibel und somit als gewählte Methode passend. So können aus den Forschungsfragen Kategorien deduktiv abgeleitet werden, und Kategorien im Prozess induktiv weiterentwickelt werden. *MAXQDA* unterstützt diesen Ansatz durch die Erstellung von Codes und Kategorien sowie durch Funktionen, diese zu visualisieren [Kuckartz, 2012].

¹² Diese Softwaretypen werden mittlerweile mit QDAS für *Qualitative Data Analysis Software* abgekürzt [Rädiker and Kuckartz, 2019].

4.8 Forschungshypothesen

Zu Beginn der Studie wurden drei zentrale Thesen formuliert, die sich aus bestehenden theoretischen Überlegungen und praktischen Erfahrungen im Bereich des Raumklangs ableiten lassen. Die qualitative Auswertung der Interviews diente nicht nur der Überprüfung, sondern auch der Konkretisierung und Weiterentwicklung dieser Thesen. Mithilfe der inhaltsstrukturierenden qualitativen Inhaltsanalyse nach Kuckartz[Kuckartz, 2012] wurden relevante Aussagen aus dem Datenmaterial extrahiert und den Thesen zugeordnet. Es wurden schließlich drei Thesen formuliert:

Hypothese 1 Unabhängig von Genre, Format oder beruflichem Hintergrund zeigen sich in der Produktion und Komposition von Spatial Audio wiederkehrende Muster und Gemeinsamkeiten in kreativen und technischen Herangehensweisen.

Hypothese 2 Die Produktion und Komposition von Spatial Audio erfordert angepasste Ansätze und Workflows, um den Übergang von Stereo zu Spatial Audio hinsichtlich unserer Hörgewohnheit zu erleichtern und in der Zukunft das umfassende Potenzial von Spatial Audio nutzen zu können.

Hypothese 3 Für die Produktion und Komposition von Spatial Audio lassen sich Guidelines formulieren, die sowohl Anfängern als auch erfahrenen Produzierenden Orientierung bieten, ohne die künstlerische Freiheit einzuschränken.

5 Ergebnisse der Interviews

In diesem Kapitel werden die Ergebnisse der Experteninterviews und der Sekundärquellen systematisch aufbereitet und in 18 alleinstehenden Kapiteln zusammengefasst. Die genauen Wortlaute der Interviews sind im Anhang einsehbar. Dieses Kapitel ist in drei Abschnitte unterteilt. Zunächst werden die Einstiegsfragen, welche sich mit allgemeinen Informationen zu den Befragten und Spatial Audio beschäftigen, zusammengefasst. Die Schlüsselfragen, der Hauptteil der Interviews, beschäftigen sich ausführlich mit den Kompositions- und Produktionsmethoden, während mit den Abschlussfragen Informationen zu zukünftigen Gedanken rund um Spatial Audio gesammelt wurden. Zusätzlich werden die Ergebnisse der Eventualfragen zusammengefasst. Diese geben weitere Einblicke in die Arbeit der interviewten Personen und erweitern die Einsicht in die Zusammenhänge innerhalb der Arbeit in Spatial Audio.

Es ist anzumerken, dass sich Themen überschneiden. Während der Umgang mit tiefen Frequenzen ein eigenes Thema ist, finden sich dort auch Überschneidungen im Kapitel Bewegung und Positionierung von Klängen und Instrumenten. Die Kapitel können somit unabhängig voneinander gelesen werden. Aufgrund der Sekundärdaten wird es vorkommen, dass zu einzelnen Fragen keine Antworten gefunden wurden, da diese Informationen aus den gegebenen Daten nicht ableitbar waren. Es wurde jedoch darauf geachtet, dass vergleichbare Daten zu den Kernaussagen der drei Thesen in allen Quellen enthalten sind.

5.1 Das Profil der Befragten und allgemeine Informationen zu Spatial Audio (Einstiegsfragen)

Die Antworten zur beruflichen Identifizierung beschreiben ein breites Spektrum an Rollen im Bereich der Tätigkeitsfelder in Spatial Audio. Der Großteil der Befragten arbeitet in Bereichen Produktion, Komposition, Sounddesign oder Klangkunst. Oft greifen die Tätigkeitsfelder ineinander. Das Sounddesign ist in vielen Fällen Teil der Produktionspraxis, und viele Komponierende sind zugleich auch performativ tätig und präsentieren ihre Werke live. Die experimentell ausgerichtete Kompositionspraxis ist vor allem im Bereich der elektroakustischen Musik angesiedelt. Einige Personen begleiten den gesamten Entstehungsprozess einer Spatial-Audio-Produktion – von der Idee bis zur Umsetzung. Ein kleiner Teil ist in Forschung und Lehre tätig und widmet sich der Weiterentwicklung technischer Formate sowie ästhetischer Fragestellungen. Insgesamt zeigt sich eine Vielfalt an Arbeitsweisen und Zuständigkeitsbereichen. Die Interviews lassen sich wie folgt den Tätigkeitsbereichen der bereits definierten Kategorien zuordnen.



Abb. 3: Kategorisierung der Experteninterviews nach Tätigkeitsfeldern, eigene Grafik.

5.1.1 Erfahrung der interviewten Personen

Viele der Befragten haben ihre Erfahrungen mit Spatial Audio in den letzten 10–15 Jahren gesammelt, einige wenige beschäftigen sich über 20 Jahre mit dem Format. Der Großteil begannen bereits im Studium oder bei Praktika, andere entwickelten ihre Expertise im Laufe ihrer Karriere durch Projekte, die immer häufiger in Spatial-Audio-Formaten durchgeführt werden. Die Befragten betonten die Bedeutung von Praxiserfahrung, Weiterbildung und der Auseinandersetzung mit den technischen Möglichkeiten und kreativen Anwendungen dieser Formate. Die Befragten sehen in Spatial Audio einen vielversprechenden Trend, insbesondere in Verbindung mit audiovisuellen Medien, während andere die Akzeptanz und Verbreitung der Technologie noch als Herausforderung wahrnehmen.

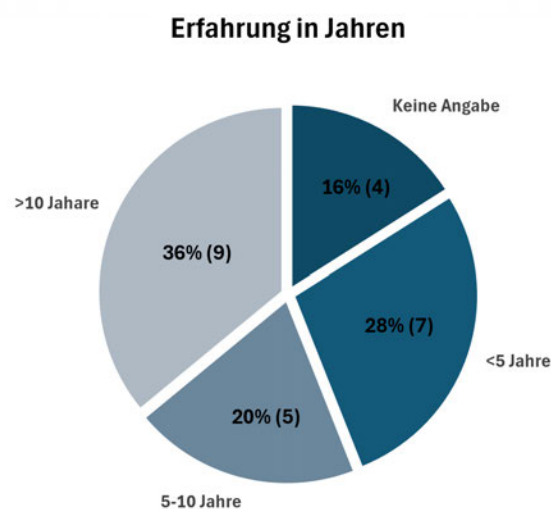


Abb. 4: Erfahrung in Jahren, eigene Grafik.

5.1.2 Spatial Audio Formate

Die folgenden Diagramme zeigen die Verteilung der benutzten Formate. Hierbei waren Mehrfachantworten möglich.

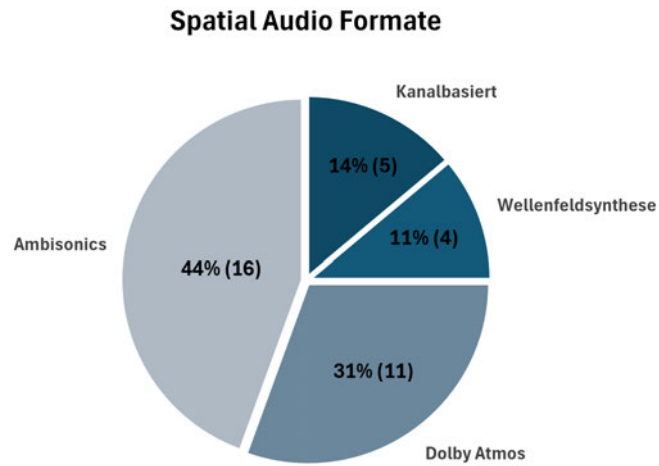


Abb. 5: Kategorisierung der Experteninterviews nach Tätigkeitsfeldern, eigene Grafik.

Die Spatial-Audio-Formate wurden in den drei Kategorien unterschiedlich oft eingesetzt:

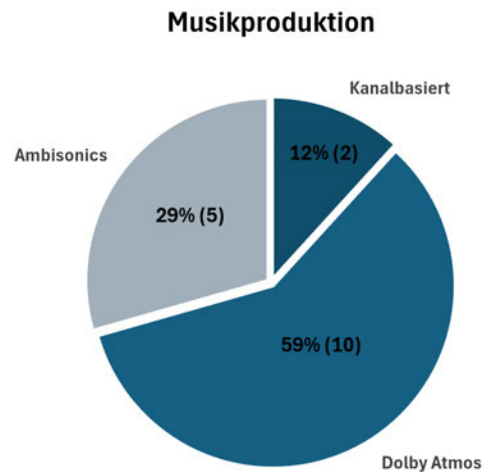


Abb. 6: Spatial Audio-Formate in der Musikproduktion, eigene Grafik.

Hörspiel und Sounddesign

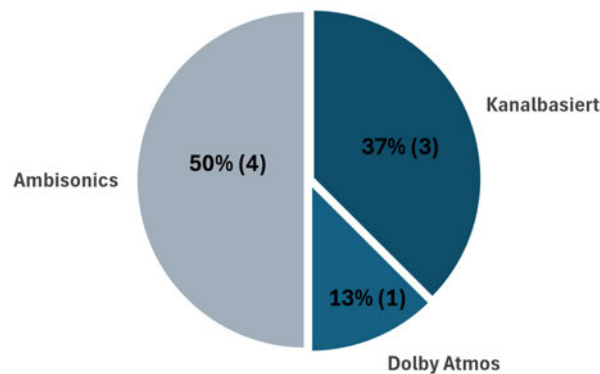


Abb. 7: Spatial Audio-Formate in Hörspiel und Sounddesign, eigene Grafik.

Experimentelle Produktionen und Installationen

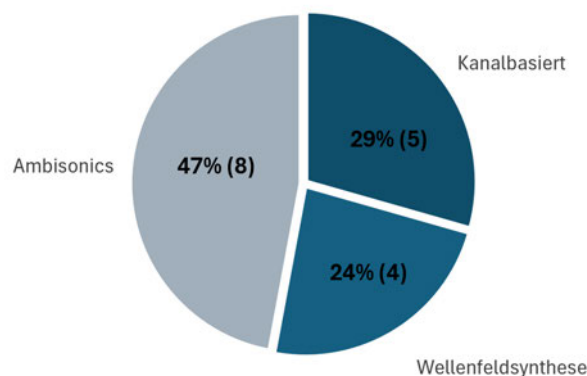


Abb. 8: Spatial Audio-Formate in Experimentellen Produktionen und Installationen, eigene Grafik.

Zu den verschiedenen Formaten wurden einige Vorteile genannt, die zur Wahl der Formate eine Rolle gespielt haben. Dolby Atmos wird Flexibilität in Bezug auf die Lautsprecherkonfigurationskompatibilität zugesprochen. Zum Beispiel kann ein Mix in 9.1.6 auch auf einem 7.1.2-Lautsprechersystem wiedergegeben werden. Außerdem kann dieser durch den Atmos Renderer auf Formate wie Surround oder Stereo heruntergemischt oder für die Wiedergabe über Kopfhörer binauralisiert werden. Erwähnt werden auch die ökonomischen Vorteile in Bezug auf den Spatial-Audio-Markt. Dolby Atmos hebt sich gegenüber anderen Formaten durch seine Lukrativität ab. Ein großer Vorteil bietet nach Meinung einiger der Befragten der standardisierte Workflow durch die Dolby Atmos Suite und das bessere Bass-Management durch den Einsatz eines LFEs. Generell wird Dolby Atmos durch das objektbasierte Arbeiten als logischer und einfacher anwendbar empfunden. In Kapitel Unterunterabschnitt 5.3.2 wird auf die Konsequenzen und Nachteile eingegangen.

Ambisonics werden ebenfalls einige Vorteile zugesprochen. Vor allem wird der Vorteil der Spatialisierungsmethode der Kugelwelle hervorgehoben, die die untere Dimension miteinbezieht, wodurch eine vollständige 360°-Umgebung kreiert wird. Der Ansatz der Kugelwelle bietet eine umfassendere Räumlichkeit als bei anderen Formaten. Dadurch, dass das Format achsenbasiert ist, bietet sich die Möglichkeit einer Szenenrotation. Ambisonics ist nicht durch ein Objektlimit eingeschränkt. In der Theorie kann eine grenzenlose Anzahl an Klängen platziert werden. Ein weiterer mehrfach erwähnter Aspekt ist der Kostenfaktor. Das Patent von Ambisonics ist abgelaufen, dadurch gibt es viele Programme und Plug-ins, als Open Source Software angeboten werden. Darüber hinaus wird die Kompatibilität mit Programmen wie Max MSP und PureData erheblich erleichtert und bietet Entwicklern die Möglichkeit, neue Wege für die Arbeit mit Spatial Audio zu beschreiten und Probleme gemeinsam mit Entwicklerteams zu lösen. Ambisonics wird als freieres Format gesehen, da die Software und Tools keine Orientierung vorgeben und der Ansatz einer Front – wie sie in Kinoformaten und dadurch bei Dolby Atmos gegeben ist – hier nicht gilt.

Kanalbasierte Formate, die auf Spatialisierungsmethoden wie VBAP basieren, sind laut einiger Befragten recheneffizienter, da sie weniger Processing der Soundquellen erfordern. Es wird vor allem betont, dass die Spatialisierung und welchen Effekt sie auf den Klang hat, einfacher nachvollziehbar ist und weniger Klangmanipulation durch die Spatialisierung erfolgt. Die Spatialisierung in kanalbasierten Formaten wird von den Befragten als transparenter und direkter wahrgenommen, da sie den ursprünglichen Klangcharakter weniger stark verändert.

Der Wellenfeldsynthese wird lediglich eine besonders gute Lokalisierbarkeit zugesprochen. Es gibt zu wenig Systeme und daher zu wenig Fachkundige, da der technische Aufwand hinter dem System zu groß ist.

In der Analyse werden die Vorteile analytisch betrachtet, bewertet und je nach Datenlage bestätigt oder als fraglich eingestuft, was durch Vorurteile oder Unkenntnis über andere Formate geschehen kann.

5.1.3 Genres

Im Bereich der Musikproduktion, insbesondere mit dem in dieser Kategorie dominierenden Dolby Atmos-Format, werden vor allem elektronische Musik, Pop und Hip-Hop produziert. Klassische Musik spielt ebenfalls eine Rolle, wenn auch eine deutlich geringere.

Im Bereich der experimentellen Produktionen und Installationen werden akusmatische, elektroakustische und frei improvisierte Stücke kreiert, darunter auch solche, die der Konzertreihe Mopomoso¹³ zugeordnet werden können.

Im Bereich Hörspiel und Sounddesign lassen sich keine klassischen Genres definieren. Die in den Interviews genannten Werke sind sowohl realistisch als auch fantastisch.

5.1.4 Motivation und Beweggründe zur Arbeit in Spatial Audio

Bei der Frage nach den Beweggründen, die zur Arbeit an Spatial Audio geführt haben, lassen sich immer wiederkehrende Motive erkennen. Am häufigsten wird der Zugang zur Technik genannt. Erst durch die vor Ort vorhandenen Lautsprechersysteme, die vor allem an Hochschulen und Universitäten installiert ist, wird die Arbeit mit Spatial Audio möglich. Auch Praktika in Tonstudios mit Spatial-Audio-Systemen werden als erster Berührungspunkt genannt. Somit werden das Studium und die vor Ort vorherrschenden technischen Möglichkeiten als Hauptgrund für die Arbeit in Spatial Audio genannt. Ein weiterer erster Kontakt wurde durch Veranstaltungen wie die Tonmeistertagung ermöglicht, bei der neue Technologien und Produktionen vorgestellt werden. Auch die Arbeit in Kinoformaten wie Dolby Surround hat in einigen Fällen dazu geführt, sich im nächsten Schritt mit Dolby Atmos auseinanderzusetzen.

Durch die Markteinführung von Dolby Atmos sehen die produzierenden die Möglichkeit, mit Spatial Audio Geld zu verdienen. Andererseits sehen sich einige gezwungen, in diesem Format zu arbeiten, um wettbewerbsfähig zu bleiben und somit Einnahmen nicht zu verlieren.

Die Möglichkeit, ein neues Format zu erforschen und mitzuentwickeln, ist ein weiterer mehrfach genannter Beweggrund. Technologieinteressierte sehen hier ein weites, noch zu erforschendes Feld und eine neue Herausforderung.

Neben ökonomischen und technischen Aspekten wurden auch neue ästhetische Möglichkeiten, optimierte Workflow-Ansätze sowie Problemlösungen genannt, die im Vergleich zur Stereo-Produktion bislang nicht möglich waren:

- Einfachere Umsetzung immersiver Klangwelten
- Höchst differenzierte Wahrnehmung des Klangraums

¹³ *MODernism, POstMODernism, SO what?* Eine Konzertreihe, die 1991 gegründet wurde und für freie Improvisation steht. Sie widmet sich dem freien musikalischen Ausdruck [Paris Transatlantic Magazine, 2007].

- Möglichkeit, parallele Atmosphären und Geräuschwelten erfahrbar zu machen, ohne die emotionale Wirkung von Sprache und Musik zu beeinträchtigen
- Gestaltung skulpturaler, körperlich erfahrbarer Klänge im 3D-Raum
- Aktive Raumgestaltung: Räume können nicht nur erschaffen, sondern auch bewegt werden
- Erweiterte Möglichkeiten zur Tiefenstaffelung
- Vermeidung des Loudness Wars: mehr Dynamik im Mix
- Erweiterte klangliche Fläche im 360°-Raum: mehr Elemente ohne Kompression oder Frequenzmaskierung

Die Motivation zur Arbeit in Spatial Audio hat somit viele Facetten und gibt vorläufig einen Einblick in die Vorteile der Spatial-Audio-Formate.

5.2 Zusammenfassung der Kernfragen (Schlüsselfragen)

In diesem Kapitel werden die Schlüsselfragen, die die Grundlage zur Beantwortung der drei Thesen bilden zusammengefasst.

5.2.1 Software und Tools

Auf die Frage, welche Software und Plug-ins zur Erstellung von Spatial-Audio-Inhalten benutzt werden, wurde eine Vielzahl an Werkzeugen genannt. Die folgende Aufzählung orientiert sich an der Häufigkeit ihrer Nennung in den Projekten der Befragten. Es ist zu beachten, dass nicht alle genannten Tools ausschließlich für Spatial Audio entwickelt wurden. Teilweise kommen auch stereobasierte Plug-ins zum Einsatz, die in Spatial-Audio-Workflows integriert werden, zum Beispiel Altiverb von Audio Ease, die FabFilter Suite oder die Collections von iZotope und Soundtoys.

Digitale Audio Workstations

1. Reaper – Wird oft für experimentelle Produktionen und Ambisonics-Projekte verwendet.
2. Pro Tools – Insbesondere für Dolby Atmos-Projekte genutzt.
3. Nuendo – Besonders für Dolby Atmos und professionelle Postproduktion verwendet.
4. Ableton Live – Genutzt als Performance-Tool und für die Integration von Spatial Audio.

5. Logic Pro / Cubase – Weniger relevant für Spatial Audio, wird aber gelegentlich genutzt.
6. Digital Performer – In der Vergangenheit häufiger verwendet, heute kaum noch in Gebrauch.

Plug-ins

1. IEM Plug-in Suite (Institute of Electronic Music, Graz) – Sehr beliebt für Ambisonics und HOA.
2. Dolby Atmos Renderer / Atmos Production Suite – Standard für Dolby Atmos-Produktionen.
3. SPARTA (Spatial Audio Plug-ins) – Häufig verwendet für Ambisonics und binaurale Anwendungen.
4. SPAT (IRCAM) – Wird oft in Verbindung mit Max/MSP integriert.
5. XP4L und MAX4L - Als Ableton oder SPAT-Integration verwendet.
6. WigWare (Bruce Wiggins Ambisonics Panner) – Genutzt für Ambisonics-Panning.
7. SSA Plug-ins (Surround Sound Analysis) – Verwendet für Monitoring und Spatial Mixing.
8. Altiverb (Audio Ease) – Reverb-Plug-in für realistische Raumklang-Simulation.
9. Dear Reality / Dear VR – Binaurale und immersive Audio-Plug-ins.
10. Roomenizer - Tool für Raumhall durch hochwertig aufgenommene Impulsantworten.
11. NugenAudio- Upmixing Software.
12. Symphony 3D - Spatial Audio Reverb.
13. FabFilter R2 - Spatial Audio Reverb.
14. God Particle - Optimierung von Mix-Bus-Prozessen.
15. Anaglyph - Panning und Binauralisierung.
16. Kronlacher Suite - Umfasst unterschiedliche Tools zur Arbeit in Ambisonics.

Programmierungsumgebungen

1. Max/MSP – Weit verbreitet für eigene Spatial-Audio-Algorithmen und Live-Processing.
2. SuperCollider – Flexibles, auf Programmierung basierendes Werkzeug.
3. PD (Pure Data) – Selten genutzt, meist für experimentelle Sounddesign-Projekte (kostenfreie Alternative zu MAX/MSP).

5.2.2 Spatial-Audio-Mikrofonierungsansätze

Es gibt viele verschiedene Ansätze, Klang für eine Spatial-Audio-Komposition oder -Produktion aufzunehmen. Generell ist bei den Aufnahmen zu bedenken, dass Direktschall, frühe Reflexionen und der Nachhall eingefangen werden können. Ein beliebtes Instrument für spezielle Mikrofonierungsansätze ist die Orgel, welche für den Raum selbst gebaut wurde und daher mit den akustischen Eigenschaften zusammenarbeitet. Bei der Aufnahme von Orgeln können viele Mikrofone benutzt werden, einige davon können sogar in die Orgel gesetzt werden, um den Direktschall einzufangen. Ein Hauptmikrofon wird für diesen Ansatz für die Aufnahme von frühen Reflexionen und den Nachhall verwendet. Bei Orchestern werden zusätzliche Mikrofone zum Einfangen von Deckenreflexionen und Wandreflexionen am Ende des Raumes eingesetzt.

Bei Hip-Hop werden die meisten Quellen in Mono aufgenommen und im Panorama verteilt. Es werden dafür mehr Aufnahmen gemacht, um mehr Material zur Füllung des Panoramas zu haben, darunter fallen zum Beispiel die Backing-Vocals.

Bei der Aufnahme von Hörspielen werden verschiedene Mikrofonierungsansätze angewendet. Von Ambisonics-Mikrofonen, die als Hauptmikrofon eingesetzt werden, bis zu MS-Einzelmikrofonierungen für die Spielenden richtet sich die Mikrofonierung nach den Gegebenheiten des Projekts. Eine Einzelmikrofonierung kann mit den Hauptmikrofonen kombiniert werden oder als Backup dienen. Durch die Ambisonics-Mikrofonierung können sich die Schauspielenden frei im Raum bewegen und die Choreografie wird akustisch adäquat abgebildet. Tiefenstaffelung kann zum Beispiel durch SM69-Mikrofone erreicht werden, die als Stützen weiter entfernte Klänge aufnehmen. Die Tiefenstaffelung ist durch das Versetzen der Mikrofone variierbar. Besondere Effekte können mit einem Kunstkopf kreiert werden. Aufnahmen mit dem KU100 können beispielsweise als verzerrende Effekte eingesetzt werden.

Ambisonics-Mikrofone haben durch die kompakte Bauweise den Vorteil, schnell und mit wenig Aufwand aufgebaut werden zu können. Des Weiteren sind sie dadurch leichter zu transportieren. Für binaurale Aufnahmen bieten sich In-Ear-Mikrofone an. Aufgrund der eigenen HRTF kann die Aufnahme für andere jedoch deutlich anders klingen. Obwohl Ambisonics-Mikrofone sich durch ihre Funktionalität, kompakte Bauweise und Fähigkeit zur räumlichen Aufnahme auszeichnen, wird laut den Interviewten eine gewisse Unklarheit bei der Aufnahme von Instrumenten bemängelt. Daher wird hierfür der 2L-Cube oder eine PCMA-3D-Mikrofonierung benutzt. Bei diesen Mikrofon-Setups kann durch das Einstellen der Abstände zwischen den Mikrofonen¹⁴ die Aufnahme an die Raumgröße angepasst

¹⁴ Bei dem 2L-Cube werden Kugelmikrofone, bei PCMA-3D vorwiegend Nierenmikrofone eingesetzt.

werden. Diese Hauptmikrofonierung bietet den Vorteil, mit allen Formaten kompatibel zu sein.

Ein Experte vertritt die Auffassung, dass sich der Klang nach naturgetreuen Mustern richten sollte, die sich plausibel anfühlen. Was für das Gehör natürlich klingt, wird subjektiv als positiv wahrgenommen. So können viele verschiedene Mikrofonierungsansätze zu zufriedenstellenden Ergebnissen führen. Durch experimentelle Mikrofonierungen, wie den Einsatz von Grenzflächenmikrofonen an Wänden, können qualitativ hochwertige, diffuse Aufnahmen gemacht werden. Eine weitere Möglichkeit ist es, Ambisonics-Mikrofone wie Tetramic mit MS-Paaren zu ergänzen oder ein Ambisonics-Mikrofon als Rahmen zu benutzen und diesen mit Spots zu füllen. Wichtig scheint es, auf Laufzeitdifferenzen, Auslöschungen und die Nähe zum Instrument zu achten. Es wird mehrfach betont, dass Monostützen bei Spatial-Audio-Aufnahmen unnatürlich wirken und daher eine Stereo-Stütztechnik bevorzugt wird. Sollte man hauptsächlich mit Nierenmikrofonen arbeiten, bietet es sich an, tiefe Frequenzen durch ein Mikrofon mit Kugelcharakteristik aufzunehmen.

Neben 3D-Mikrofonierungen werden aber auch klassische Mono- und Stereo-Aufnahmen benutzt, um daraus einen Spatial-Audio-Mix zu machen. Statt späterem Panning kann Bewegung schon in die Mikrofonierung gebracht werden.

5.2.3 Anzahl der Klangelemente

„Wenn ein Song mit 400 Gitarrenspuren funktioniert, dann hat er eben 400 Gitarrenspuren.“¹⁵

Spatial Audio bietet durch die Erweiterung in den dreidimensionalen Raum deutlich mehr Platz, um Klänge im Panorama zu platzieren. Die Befragten haben zu der Frage, ob es ein Limit gibt wie folgt reagiert:

Einige Komponierende verpacken mehrere Hundert von Klängen in Clustern und lassen diese eine eigene Textur erschaffen. Durch Granulatoren als Multiplier und durch Flocking-Algorithmen wie Boids wird eine Vielzahl von Klängen erstellt und in einer Art Schwarmverhalten bewegt. Im Verlauf des Kompositionsprozess gewöhnt man sich an die Klänge und füllt die Lücken, die man mit der Zeit entdeckt, dadurch entstehen detailreiche Stücke. Die Dopplung von Klängen ist ein beliebtes Stilmittel, um solche Lücken zu füllen. Es wird betont, dass eine Mindestanzahl an Klangelementen, die sich nach der Art der Klänge richtet, erforderlich ist, um den Raum zu füllen. Eine Obergrenze der Anzahl

¹⁵ Interviewte Person D, Interview 4 (eigene Erhebung).

der Elemente nach oben hingegen wurde nicht festgestellt. In der Regel wird mit mindestens 40 Spuren in den Projekten gearbeitet. Die Anzahl der Spuren kann allerdings auch 200 Spuren übersteigen.

Limitiert wird die Anzahl an Klangelementen durch die Rechenleistung des Computers und statischer Limits, da die Lokalisierung mehrerer Klänge auf einmal eine Herausforderung für den menschlichen Hörsinn darstellt. Ein Klangkünstler vermerkt, dass durch eine wissenschaftliche Testreihe herausgefunden wurde, dass von maximal fünf Klangelementen die Lokalisation und Bewegung nach dem Hören reproduziert werden konnten¹⁶. Für die Lokalisation von Klängen bietet es sich an, weniger Elemente zu verwenden, denn eine Überforderung der Hörenden durch die erschwerte räumliche Zuordnung von Klängen ist schnell erreicht. Wirkt das Stück ungewollt verwirrend, so ist die Grenze überschritten – das Gleiche gilt, wenn zu viele ungewollte Maskierungen auftreten. In Genres wie Hip-Hop ist es wichtig, die Anzahl der Elemente gezielt zu begrenzen, um die Energie, Prägnanz und den charakteristischen Punch des Stücks nicht zu beeinträchtigen. Laut eines Produzenten kann es in Dolby Atmos und der Benutzung von Pro Tools aufgrund der Programmstruktur im Projekt unübersichtlich werden – die Systeme geben eigenen Limitierungen vor. Dies ist vor allem in Bezug auf Dolby Atmos der Fall, denn es können maximal 128 Objekte inklusive Bed benutzt werden[Pfanzagl-Cardone, 2023].

5.2.4 Bewegung und Positionierung von Klängen und Instrumenten

Die Befragten gaben umfangreiche Einblicke in die Positionierung und Bewegung von Klangquellen in Spatial-Audio-Stücken. Bei den Dolby-Atmos-Produktionen wurden vor allem Hip-Hop-, Pop- und elektronische Stücke und Hörspiele genannt. Bei den Ambisonic-Produktionen waren es vor allem elektroakustische Stücke und Hörspiele.

Laut den Erfahrungsberichten werden bei Hip-Hop und Pop die Lead-Vocals ausnahmslos nach vorne gesetzt, in der Regel in den Centerlautsprecher. Backing-Vocals, Chöre und Synth-Elemente werden an die Seiten oder nach hinten in den Raum gesetzt. Oben werden lediglich hochfrequente Instrumente wie Flöten, Streicher oder die Okarina platziert. Das Schlagzeug wird wie die Lead-Vocals oft typischerweise nach vorne gesetzt. Dabei werden Snare, Kick und Becken teils mittig oder breit nach links und rechts gepannt. Zusätzlich wird mit der Dopplung der Becken nach hinten experimentiert. Elemente wie Overheads werden bevorzugt nach oben vorne gesetzt, und die Hi-Hats nach links und rechts gepannt, während Lead- und Rhythmusgitarren in den meisten Produktionen auf die Seiten gesetzt werden.

¹⁶ Vgl. Troschka (2021): Listening in 3D [Troschka et al., 2021].

Bei elektronischer Musik lösen sich die Produzierenden stärker vom Anspruch die Realität wiederzugeben und streben keine Bühnenabbildung vorne an, wie es beispielsweise bei einem Orchester der Fall ist. Rhythmische Bewegungen und Synthesizer-Klänge werden frei durch den Raum bewegt. Teilweise sind die Perkussionselemente klassisch nach vorne gesetzt, während die Snaredrum auch mal nach hinten gesetzt wird und mit den anderen Elementen zum Drop des Beats nach vorne verschmilzt. Bass und Lead-Vocals werden nach vorne gesetzt, und ähnlich wie bei Hip-Hop- und Pop-Produktionen werden Keys, Gitarren und Backing-Vocals auf die Seiten und in wenigen Fällen nach hinten verteilt. Der Reverb der Lead-Vocals wird als Effekt genutzt und statt nach vorne auf die Seiten oder nach hinten gepackt.

Ein Interviewter betont, sich generell an der Realität zu orientieren und das Schlagzeug an einer festen Position zu belassen. Einige lösen sich jedoch vollständig von diesem Ansatz, verteilen das Schlagzeug im ganzen Raum und lassen die Snaredrum auch mal als Effekt wandern. Besonders breite Klangflächen lassen sich für langsame Bewegungen einsetzen. Experimentell kann es bei Jazz-Produktionen werden. Hier wird auch gerne mal die Stimme nach oben gesetzt, das Klavier nach hinten und der Bass nach vorne.

Bei Soundscapes bietet sich das Bewegen von Sinustönen, Drone-Sounds und Crackle-Sounds in der Vertikalen an. Soundgruppierungen werden durch den Raum bewegt, und es wird mit den Emotionen der Rezipienten gespielt, indem beispielsweise ein Gefahr suggerierender Klang die Hörenden von hinten überrascht. Klänge werden gezielt positioniert, um ihnen eine zugewiesene Bedeutung zu verleihen und die Aufmerksamkeit des Hörenden zu lenken. Ein Element, das besonders gut bei Ambisonics funktioniert, ist das Bewegung von Klangspiralen, die von oben durch den Raum laufen.

Je nachdem, ob sie diegetisch oder non-diegetisch eingesetzt werden soll, wird bei Hörspielen die Musik mal nach hinten oder realistisch in die Szene gesetzt oder bewusst für Szenenwechsel genutzt. Klangelemente wie ein Hubschrauber werden realitätsgetreu nach oben gesetzt. Die Erzählstimme wird zentral positioniert, und die Sprachverständlichkeit als wichtiger Faktor genannt. Für spezielle Effekte kann die Stimme aber auch nach unten gesetzt werden oder aus allen Richtungen wiedergegeben werden.

Generell eignet sich das Ruf und Antwort Prinzip für Spatial Audio. Auch die kontrapunktische Verteilung wird oft erwähnt. Es wird betont, dass Bewegung nicht ausschließlich um der Bewegung willen eingesetzt werden sollten. Sie kann beispielsweise Spannung erzeugen – so können bestimmte Passagen eines Stückes durch bewegte Elemente hervorgehoben oder zu einem klassischen Drop hingearbeitet werden.

Auch hier stellt sich im Kompositions- und Produktionsprozess die zentrale Frage: Dient die räumliche Bewegung der musikalischen Aussage – oder lenkt sie eher ab?

5.2.5 Frequenzabhängigkeiten im Tieftonbereich

Der Umgang mit dem Tieftonbereich stellt eine besondere Herausforderung in Ambisonics- und binauralen Produktionen dar und ihm wird daher ein eigenes Kapitel gewidmet und in der späteren Analyse erneut aufgegriffen.

Bei Dolby-Atmos-Produktionen wird der LFE nicht immer genutzt, da er bei der Binauralisierung nicht berücksichtigt wird und nicht alle Genres vom LFE profitieren. Der LFE wird bevorzugt leise dazugemischt und selten mit anderen Lautsprechern korreliert. Der LFE wird bei Atmos als Vorteil gesehen, da er den Tieftonbereich ergänzt, auch wenn schwer nachvollziehbare Klangverfärbungen auftreten. Der Tieftonbereich fügt sich in Dolby Atmos besser in den Mix ein. Bei Ambisonics werden oft zusätzliche Subwoofer eingesetzt, auch um Raummoden zu eliminieren. Es hat sich bewährt, Subwoofer in die Ecken des Hörbereichs zu setzen. Stehen keine Subwoofer zur Verfügung, werden in Ambisonics die tieffrequenten Anteile von Klängen oft auf alle Lautsprecher verteilt.

Generell werden Bassanteile nach unten gemischt, da sie in der Höhe schwer ortbar bzw. wahrnehmbar sind. Wie bereits im Kapitel zu Spatial-Audio-Mikrofonierungsansätze (siehe Unterunterabschnitt 5.2.2) erwähnt, werden bei Mikrofonarrays Kugelmikrofone in Kombination mit Nierenmikrofonen eingesetzt, um die tiefen Frequenzen besser einzufangen.

5.2.6 Integration der Hörenden und das Sweet-Spot-Problem

Die Rezipienten in den Produktionsprozess zu integrieren und sich des Sweet-Spot-Problems bewusst zu sein, ist für den Großteil der Befragten ein fester Bestandteil im Spatial-Audio-Workflow. Wie die Befragten damit umgehen, wird in diesem Kapitel zusammengefasst.

Viele Produzierende im Bereich der Musikproduktion betonen, dass kopfbezogene Formate wie Binaural Audio das Sweet-Spot-Problem lösen. Auch wenn das HRTF-Problem noch nicht vollständig gelöst ist, sei dies der beste Weg, Spatial Audio massentauglich zu machen. Lautsprechersysteme seien für den normalen Konsumierenden nicht geeignet, da dieser selten im Sweet Spot sitzt oder die Lautsprecher korrekt aufstellt sind, wie es schon bei Stereo oft der Fall sei.

Als einziges Lautsprechersystem, das einen großen Sweet Spot bietet bzw. in der Lage ist, mehrere optimale Hörpositionen zu erzeugen, gilt die Wellenfeldsynthese. Allerdings ist

sie aufgrund des hohen technischen Aufwands nicht praktikabel. Beim Format Ambisonics lässt sich dem Problem durch gezielten Einsatz des Dopplereffekts, Frequenzwechsel sowie bewusste Pegelunterschiede von Klängen entgegenwirken.

Allgemein wird erwähnt, dass die Software das eigentliche Problem ist, da sie ausnahmslos eine feste Hörerposition in der Mitte vorgibt. Um den Hörenden dennoch in den Workflow zu integrieren, hören die Komponierenden und Produzierenden ihre Stücke aus verschiedenen Positionen, Höhen und Winkeln zu den Lautsprechern an. Es kann auch als Vorteil gesehen werden, an unterschiedlichen Positionen verschiedene Hörerfahrungen eines Stückes zu erleben. Dennoch wird vor allem bei Spatial-Audio-Installationen darauf geachtet, dass aus allen Positionen eine positive Hörerfahrung möglich ist. Daher wird der Raum oft für mehrere Positionen komponiert.

Ein vielversprechender Ansatz ist das Aufnehmen von Raumimpulsantworten des Ausstellungsraumes. Hierdurch kann in der Mischung das Stück aus verschiedenen Positionen angehört werden, ohne vor Ort sein zu müssen. Ein weiterer Ansatz ist die Kombination von kopf- und lautsprecherbezogenen Kompositionen. Durch den Einsatz offener Kopfhörer kann eine zusätzliche Ebene dazugemischt werden, die den Sweet-Spot-Effekt mindert.

Bei Hörspielen kann der Hörer entweder außerhalb des Geschehens bleiben oder in dieses integriert werden. Er kann als eine Art unsichtbarer Mitläufer existieren, ohne selbst Geräusche zu verursachen, oder eine bestimmte Position einnehmen und eine Art Bühne vor sich erleben.

Ein kleinerer Teil der Befragten vernachlässigt das Sweet-Spot-Problem. Es wird betont, dass es sehr viele unterschiedliche Spatial-Audio-Lautsprechersysteme gibt, die sich in Anzahl und Abstand der Lautsprecher untereinander deutlich unterscheiden. Daher lohnt es sich nicht, den Mix in Bezug auf den Sweet Spot zu optimieren. Dennoch sollten alle ästhetischen und technischen Entscheidungen stets in Bezug auf den Hörer getroffen werden, da dieser letztlich das Medium konsumiert und in der kommerziellen Musikproduktion ein optimaler Mix entscheidend für den wirtschaftlichen Erfolg ist.

Die Befragten sind sich einig, dass Spatial-Audio-Mischungen sowohl über Lautsprecher als auch über Kopfhörer abgehört werden sollten, um herauszufinden, ob unerwartete Effekte auftreten.

5.2.7 Räumlichkeit und der Umgang mit Halleffekten

Die Interviewten haben eine Vielzahl von Ansätzen genutzt, um Raum zu erschaffen, ihn zu manipulieren, und Halleffekte einzusetzen.

In Dolby Atmos wird Hall als umhüllender Effekt verstärkt von oben und hinten eingesetzt. Oft wird ein Atmos Reverb Bed gebaut, um den Vorteil zu nutzen, dass der Dolby Renderer die Höhenkanäle spiegelt. In vielen Fällen werden die Halleffekte leise dazugemischt und die Brightness des Halls erhöht. Es wird dabei nicht differenziert, ob sich der Hallklang je nach Position, etwa oben vorne oder oben hinten, unterscheiden sollte. Einige setzen für Dolby- und Multikanal-Setups ausgerichtete 9.1.6-Hall-Plug-ins wie zum Beispiel den R2 von FabFilter ein. Andere wiederum arbeiten mit mehreren Stereo-Hallspuren und bekommen eine dedizierte Hallspur als Stem zugespielt, die im besten Fall mit Atmos-Hall-Plug-ins nachgebaut wird, um Klangfarbe und Stil zu erhalten. Es können auch mehrere Stereo-Hallspuren in ein Bed gesendet werden, um so einen Spatial-Audio-Halleffekt zu erzeugen. Überwiegend werden zusätzlich Delayeffekte und Modulationen in verschiedenen Kombinationen verwendet.

Bei elektronischen Produktionen wird Hall nur dezent eingesetzt, um den Mix nicht diffus und undefiniert klingen zu lassen. Es wird darauf geachtet, mit Klängen Räumlichkeit zu schaffen, ohne direkt auf einen Halleffekt zurückzugreifen, da dieser oft eine eigene Färbung mitbringt. Ein weiterer Ansatz ist es, Stereohall aufzufächern und eine Tiefenstaffelung durch Pegeländerungen zu erreichen. Es wird betont, dass auf einem Atmos-Album keine unterschiedlichen Räume entstehen dürfen, da die Songs stimmig ineinander übergehen müssen.

Ein Produzent erzeugt Tiefenstaffelung mit einem Drop-off in hohen Frequenzen, um durch Luftabsorption Weite zu erzeugen, und hebt tiefe Frequenzen an, um Klang nach vorne zu holen und den Nahbesprechungseffekt nachzuahmen. Ein anderer Produzent wiederum nutzt Impulsantwort-Datenbanken als musikalisches Werkzeug, um beispielsweise ein Schlagzeug anzupassen. Bei Ambient-Stücken wird generell mit mehr Hall gearbeitet.

Bei Hörspielen gibt es Werkzeuge wie *Next Room* von *sinewave lab* [Sinewave Lab, 2024], die verschiedene Mikrofonpositionen im Raum simulieren und so die Hörenden hinter eine Tür oder ein Fenster setzen können. Variierende Halleffekte spielt bei Szenenwechseln eine wichtige Rolle, um den Hörenden einen Umgebungswechsel zu verdeutlichen. Es wird erwähnt, dass Hall bei Ambisonics durch unterschiedliche Lautsprecherabstände zueinander verschieden klingen kann. Auf diese Weise kann sich die subjektiv wahrgenommene Raumgröße von kleinen zu großen Räumen verändern.

Für die Aufnahme von Raumhall gibt es verschiedene Möglichkeiten. Zum Beispiel kann durch die Aufnahme der Reflexionen von der linken und rechten Wand ein transparenter Klangraum erzeugt werden. Die realistischsten Räume entstehen durch die Aufnahme

mehrerer Impulsantworten. Dies erfordert jedoch eine hohe große Anzahl von Messungen und damit einen großen technischen, zeitlichen und finanziellen Aufwand. Die Aufnahme von 11 Lautsprecherpositionen mit 75 Mikrofonen bedeutet einen enormen Aufwand, liefert aber reproduzierbar realistische Hallräume.

Wird diese Technik angewandt, ermöglicht sie neue Möglichkeiten für Live-Aufnahmen, die ohne Störgeräusche wie Klatschen oder Husten sind, da lediglich der Direktschall der Instrumente aufgenommen werden, während der Hall aus den Impulsantworten generiert wird. Somit ist eine nahezu 1:1-Nachbildung des Konzertsaals möglich. Auch mit geringerem Aufwand können Räume durch Sweeps erfasst und die Impulsantworten binaural decodiert werden. Eine einfachere Methode ist es, Mikrofone im Raum zu verteilen und an Orten wie vor einer Tür, an einem Fenster oder von der Decke hängend zu positionieren, um verschiedene Raumanteile aufzunehmen und in die Produktion zu mischen.

Ambisonics erfordert andere Herangehensweisen, um Raum zu kreieren und Halleffekte einzusetzen. Beispielsweise werden algorithmische Halleffekte wie der FDN Reverb von IEM genutzt. Dieser eignet sich vor allem als Kompositionswerkzeug oder für Musikproduktionen, jedoch seltener als Hall-Plug-in, um realistische Räume abzubilden wie zum Beispiel für Hörspiele. Der IEM Room Encoder bietet hierfür bessere Möglichkeiten, da Raumgrößen eingegeben und frühe Reflexionen berechnet werden können. Dies erfordert allerdings eine hohe Rechenleistung.

Reverb wird von Klangkutschaffenden und Sounddesignern vorrangig als Sounddesign-Tool genutzt, seltener zur Erzeugung eines Raums. Einige setzen eher auf Stereo-Hall-Plug-ins und Stereo-Delay-Effekte, statt Ambisonics-spezifische Plug-ins zu verwenden. Beispielsweise werden gerne Hallfahnen als Klangobjekte eingesetzt. Die Kronlachner Suite ermöglicht die Erstellung eigener Presets und die Nutzung eigener Impulsantworten, die häufig in Workflows integriert werden. Ein Komponist unterscheidet zwischen konstruiertem und aufgenommenem Raum. Der Raum kann durch Klangflächen oder Cluster unterschiedlicher Klänge realisiert werden.

Bei Multichannel-Kompositionen wird gerne mit SPAT und Max/MSP gearbeitet. Hier kann der 16-kanalige Reverb von Flux eingesetzt werden. SPAT selbst kann die frühen Reflexionen von Räumen ähnlich wie der Room Encoder bei Ambisonics berechnen. Dies erfordert jedoch eine hohe Rechenleistung. Spannende Effekte entstehen, wenn ein Direktsignal nur aus einem Lautsprecher kommt und der Hallanteil auf andere Lautsprecher verteilt wird. In der Regel werden Hall-Plug-ins als kreatives Tool eingesetzt, um beispielsweise durch Side-Chaining mit einem Kompressor den trockenen oder nassen Teil des Halls zu manipulieren.

Durch das Interpolieren von Hall können verschiedene Phantom-Räume erschaffen werden, die bei Live-Events funktionieren.

Ein besonderer Ansatz ist es, mit Raum zu komponieren, anstatt ihn herzustellen. Werden statische, korrelierte Quellen gegeneinander verschoben, können ebenfalls Raumeffekte entstehen. Ein Komponist betonte, dass sich besonders interessante Erfahrungen durch die Kombination virtueller Räume mit Kopfhörerwiedergabe in Außenumgebungen erzeugen lassen.

5.2.8 Mixing in Spatial Audio

Ein neues Format bedeutet auch neue Mixing-Ansätze, die von den Befragten umfangreich erläutert wurden und Vergleiche zu Stereo-Mixing-Ansätzen ermöglichen. Dabei geht es auch um den Umgang mit Up- und Downmixing in andere Formate.

„Ein Stereo-Downmix ist nur wie ein Bildausschnitt von einem großen Panorama.“¹⁷

Als Ausnahme sticht besonders die Arbeit an einem Ambisonics-Hörspiel hervor. Hier wurde das Mixing über ein Lawo-Pult realisiert, das mit einer DAW kombiniert wurde. Hinzu kamen externe Effektgeräte, und das Panning wurde in 1st-Order-Ambisonics umgesetzt. Dieser Ansatz bildet eine Ausnahme, da er nach bisheriger Erfahrung ungewöhnlich und näher am klassischen Stereo-Workflow angelehnt ist.

Bei Soundscape-Produktionen in Ambisonics wurden EQs eingesetzt, um Maskierungseffekte, die bei der Verwendung von hohen Anzahlen von Klängen auftreten können, zu minimieren. Bewegte Elemente wurden durch EQs hervorgehoben, um die Bewegung präziser wahrnehmbar zu machen. Das Panning wurde frequenzabhängig vorgenommen: Hochfrequente Klänge wurden in der Höhe, tiefe Klänge unten platziert. Oft wurden dafür EQs automatisiert. Es wird erwähnt, dass Panning-Tools teilweise bereits eine Art EQ beinhalten, je nachdem, wo der Klang positioniert wurde. Wichtig sei es, die Dynamik beizubehalten.

Bei experimentellen Ambisonics-Produktionen und installationsbezogenem Mehrkanal-Mixing wurde nicht mehr von einem Mixing-Prozess gesprochen. Der Mix entsteht in der Komposition durch die Klänge. Eventuell werden am Ende noch Rotationen im Panorama eingesetzt oder kleine Anpassungen vorgenommen. Wichtig sei es, sich während der Komposition regelmäßig den Fortschritt anzuhören.

¹⁷ Interviewte Person F, Interview 6 (eigene Erhebung).

Bei Installationen oder beim Abspielen von Stücken über Lautsprechersysteme wird immer an den Abspielraum gedacht. Gegebenenfalls werden EQ oder Kompression eingesetzt, um den Mix an den Raum anzupassen. Es wird betont, dass es schwierig sei, Lautstärkenverhältnisse beizubehalten, da es kaum Metering-Tools für Spatial Audio gibt. Grundsätzlich kann man sich beim Mixing die Frage stellen: Was soll hörbar sein? In vielen Fällen sind Pegelanpassungen und leichte Dynamikprozesse ausreichend. Es wird betont, dass Field Recordings kaum angepasst werden und die Übertragung in Formate wie Mono oder Stereo eine Herausforderung darstellt.

Bei Musikaufnahmen wird verdeutlicht, dass ein optimaler Mix kein EQ oder dynamische Processing-Tools benötigt. Es sei sinnvoller, die Position des Mikrofons zu verändern, anstatt in der Postproduktion die feinen Nuancen der Instrumente zu verlieren. Die Planung ist daher essenziell. Bei Aufnahmen ist die Positionierung und das passende Aussteuern der Pegel der Schlüssel. EQs sollten nur eingesetzt werden, um Störgeräusche zu minimieren oder problematische Raumfrequenzen zu korrigieren, die durch Mikrofonpositionierung nicht gelöst werden können – dies seien lediglich Reparaturmaßnahmen.

Ein Recording-Engineer nimmt in höheren Sampleraten (96 kHz) auf, um Fehler auf größere Frequenzbereiche zu verteilen und sie so weniger ins Gewicht fallen zu lassen. Ein wichtiger Mixing-Prozess ist die Laufzeitkompensierung von Stützmikrofonen zum Hauptmikrofon. Bei Ambisonics-Mikrofonen kann es zu Positions- bzw. Rotationsabweichungen kommen, die in vielen Fällen in der zugehörigen Software korrigiert werden müssen. Ein Produzent erwähnt, dass er nicht unter Berücksichtigung von Codecs mischt, sondern ein freies Mixing durchführt, um den Konsumenten ein transparentes Produkt zu bieten.

Bei Atmos-Produktionen wird mehrfach darauf hingewiesen, Phantomschallquellen zu vermeiden und die Klänge direkt auf die Lautsprecherpositionen zu verteilen. Eine übermäßige Anhebung der tiefen Frequenzen bereits im frühen Mixing-Prozess sollte vermieden werden. Der LFE wird daher in der Regel nur zur Verstärkung tieffrequenter Musikanteile eingesetzt. Auch bei Atmos wird ein frequenzabhängiges Panning durchgeführt. Oft gehen Mix- und Mastering-Prozesse ineinander über. Als erste und in vielen Fällen einzige Schritte werden das Panorama ausgearbeitet und die Lautstärken eingestellt. Frequenzen müssen selten angepasst werden, da der Mix nicht frequenztechnisch bereinigt werden muss, denn es gibt genug Platz im 3D-Panorama. Der Fokus liegt auf dem Panning, nicht auf dynamischem Processing. Der Atmos-Workflow bietet die Möglichkeit, Elemente nach Distanz zu gruppieren. So können nahe und weit entfernte Klänge im Mix strukturiert werden. Es wird mehrfach erwähnt, dass Kompression für einen Spatial-Audio-Mix problematisch sei. Dennoch gibt es einige Ansätze, mit leichter Kompression die den Mix nicht negativ beeinträchtigen. Einige nutzen Bus-Kompressionsmethoden und setzen Kompres-

soren auf allen Spuren ein, um dieses Ziel zu erreichen. Ein weiteres beliebtes Tool im Mixing-Prozess ist der Einsatz von Delay-Effekten. Diese werden häufig zur räumlichen Tiefenstaffelung oder zur Betonung von Bewegung im 3D-Raum eingesetzt. Dies kann zu einer subjektiven Wahrnehmung führen, die einer Kompression ähnelt, ohne dass tatsächlich eine Verringerung des Dynamikumfangs erfolgt.

Generell wird betont, dass sich die Kick im Center durchsetzen kann und dies wichtig sei. Einige bauen eine klassische Frontstage, während andere sich langsam davon abwenden. Klassische Summeneffekte wie bei Stereo gibt es nicht. Diese können nur über einen Workaround, mit einer Vielzahl von Stereo-Masterbussen realisiert werden. Wer sich von der Frontstage entfernt, betont, dass es wichtig sei, Ankerpunkte im Mix zu setzen. Das Panorama wird letztlich zum neuen Mischfeld, da Klänge in Spatial Audio nicht mehr gegeneinander um Platz kämpfen. Elemente sollen bewusst gepannt werden, damit sie einen Zweck erfüllen.

„Musik produzieren heißt, Entscheidungen zu treffen. Die Lead-Vocals als Anker einsetzen – dann ist das die Begründung. Macht das Sinn? So mischt man. Wie transportiere ich die Emotionen?“¹⁸

Allgemein wird EQing in mehreren Fällen mit Halleffekten wie dem Altiverb kombiniert. Das Panning ist sehr zeitaufwendig, insbesondere das Automatisieren der Quellen nimmt den Großteil des Aufwandes einer Mischung ein. Viele wünschen sich daher haptische Lösungen in Form von physischen Controllern, um diesen Prozess zu beschleunigen. Wichtig sei es, gepante Bewegungen mehrfach anzuhören, um sicherzustellen, dass der Klang tatsächlich dort positioniert ist, wo er sein soll. Die Übertragung sei teilweise ungenau.

5.2.9 Mastering in Spatial Audio

Der Mastering-Prozess weicht deutlich von klassischen Stereo-Mastering-Prozessen ab. Hier sind die Antworten zusammengefasst, wie mit der Herausforderung des Masterings in Spatial Audio umgegangen wird.

Bei Dolby-Atmos-Produktionen findet eine Summierung im Renderer statt. Es werden dramaturgisch sinnvolle Lautheiten angestrebt, damit das Album in sich stimmig bleibt. Dies wird durch das Beibehalten der Klangfarbe und Räumlichkeit aufeinanderfolgender Stücke erreicht. Mixing-Engineers wollen jedoch häufig nicht, dass leisere Stücke tatsächlich leiser erscheinen, da die maximale Lautheit des lautesten Stückes in der Regel bei -18 LUFS liegt. Deshalb ist eine kontinuierliche Kommunikation zwischen Mixing- und Mastering-Engineers essenziell. Außerdem wird eine konsistente Wiedergabe der Stücke

¹⁸ Interviewte Person D, Interview 4 (eigene Erhebung).

auf verschiedenen Geräten angestrebt – das heißt, die Musik darf auf Soundbars, Kopfhörern oder Heimkinos nicht signifikant an Qualität verlieren.

Mastering wird als notwendige Qualitätskontrolle in Spatial Audio gesehen, der mehr Gewicht beigemessen wird als bei Stereo, da hier häufiger technische Fehler und ästhetische Probleme auftreten können. Ungewollte Pop- oder Klick-Geräusche im Masterstück werden, in der Regel entfernt. Viele der Spatial-Audio-Stücke werden auf Stereo heruntergemischt, wobei in vielen Fällen Frequenzanpassungen notwendig sind. Klassische Mastering-Chains entfallen weitgehend und mit wenigen Ausnahmen werden keine Master-Busse verwendet.

Ein Produzent verfolgt einen bestimmten Mastering-Prozess: Aufgrund der Limitierung durch Pro Tools wird zunächst ein Stereo-Master erstellt. Dazu werden 118 Objekte auf 59 Masterspuren geroutet und verlinkt, sodass Änderungen direkt auf alle Objekte angewendet werden können. Dieser Workflow ahmt ein klassisches Stereo-Mastering nach, sodass Multiband-Kompressoren, Saturation-Plug-ins, Sidechains, Limiter und das God-Particle-Plug-in genutzt werden können.

Ein Mastering-Prozess bei Hörspielen sieht das Limitieren einzelner Objekte und Busse vor. Ist das Ziel, das Hörspiel im Hörfunk zu senden, sollte es -23 LUFS erreichen. Für Online- bzw. Streamingdienste müssen -18 LUFS erreicht werden, bei Spotify sind es -14 LUFS. Diese Vorgaben sollten eingehalten werden, da andernfalls eine automatische Kompression durch den Streamingdienst vorgenommen wird, die es zu vermeiden gilt. Bei Ambisonics wird das Master-File in der Regel mit 48 kHz/24 Bit im WAV-Format ausgegeben und entweder binauralisiert oder es wird ein WAV-File erstellt, das je nach Ambisonics-Ordnung die jeweilige Anzahl an Spuren enthält.

Eine bereits erwähnte Herausforderung des Masterings ist der Umgang mit der Binauralisierung. Verschiedene Renderer und Plug-ins nutzen unterschiedliche HRTFs, wodurch ein Stück je nach Wiedergabeumgebung unterschiedlich klingen kann. Außerdem werden im Dolby-Atmos-Renderer und bei Apple Music schwer nachvollziehbare Anpassungen vorgenommen, die beim Mix und Mastering berücksichtigt werden müssen. Die Codecs und Binauralisierungsprozesse entwickeln sich ständig weiter. Werden die EQs der Renderer upgedatet, verändert sich auch der Klang der Stücke. Das bedeutet, dass ein Mix von heute in fünf Jahren anders klingen wird. Es bietet sich daher an, verschiedene HRTFs zu testen, um Kompatibilität und Qualität zu gewährleisten. Dies erhöht den Arbeitsaufwand, weshalb ein Produzent betont, dass dieser Aspekt in vielen Fällen nicht berücksichtigt werden kann.

Eine weitere Methode ist die Nutzung der eigenen HRTF. Diese wird teilweise von Apple durch das Scannen der Ohren erzeugt. Die Tools werden regelmäßig optimiert, berücksichtigen jedoch noch nicht die Einflüsse des Torsos. Einige Plug-ins wie Virtuoso oder Harpex bieten eine hochwertige Auswahl an HRTFs, und durch Emulatoren kann im Mastering-Prozess der Einfluss der Apple- und Dolby-Binauralisierung vorgehört werden, sodass Anpassungen mit dem Kunden abgestimmt werden können.

Die Erstellung mehrerer Master für verschiedene Formate oder Codecs ist ein weiterer Ansatz, der jedoch ebenfalls den Arbeitsaufwand erhöht. In vielen Fällen wird das Mischen und Mastern nicht getrennt durchgeführt, während das Mastering bei kommerziellen Produktionen an spezialisierte Mastering-Engineers abgegeben wird. Wichtig ist es, sich an die Vorgaben der Streaming-Plattformen zu halten. Es wird betont, dass sich das Mastering von Spatial-Audio-Stücken oft nur auf ein technisches Mastering beschränkt.

5.2.10 Methoden, Ansätze und Philosophien hinter Spatial-Audio-Produktionen und Kompositionen

Bevor eine Übersicht der verschiedenen Workflows dargelegt wird, wird in diesem Kapitel zusammengefasst, welche Ansätze zu den Arbeitsprozessen gehören. Es soll geklärt werden, ob einer bestimmten Philosophie gefolgt wird und welche Methodik dahintersteckt.

Die Ansätze für die Arbeit in Spatial Audio sind die Grundlage für die Arbeitsprozesse und Workflows, die die Befragten im Laufe der Zeit entwickelt, adaptiert oder weiterentwickelt haben. Es gibt unterschiedliche Herangehensweisen, die jeweils ihre Vor- und Nachteile haben. Dieses Kapitel wird einige Punkte aus den zuvor zusammengefassten Kapiteln zusammenführen. Somit dient es als erste Zusammenfassung der zuvor erwähnten Teilbereiche von Spatial-Audio-Kompositionen bzw. -Produktionen. Dennoch liegt der Fokus in diesem Kapitel auf allgemeinen Herangehensweisen, die sich aus dem Umgang mit Klängen und zum Beispiel dem Mixing-Prozess ableiten lassen. Dieses Kapitel wird nach Format und nach den erörterten Kategorien eingeteilt.

Musik in Dolby Atmos

Der Großteil der Musik in Dolby Atmos wird von den Produzierenden über das sogenannte Stemmixing realisiert. Hierbei werden die Originalspuren oder gemasterten Stereospuren aus einer Stereo-Produktion in einen Dolby-Atmos-Mix umgewandelt. Der am häufigsten genutzte Ansatz ist die Arbeit mit einer Frontbühne, die aufgrund des gewohnten Stereohörens bevorzugt wird. Um einen Ankerpunkt für die Hörenden zu realisieren und den Spatial-Audio-Mix als Erweiterung des Stereo-Tracks zu verstehen, werden oft klassische Elemente wie Lead-Vocals, Bass und Drums in den Centerlautsprecher oder in der Front verteilt gesetzt. Die Seiten, Höhen und der hintere Bereich werden vor allem für Halleffekte genutzt.

fekte oder atmosphärische Klänge verwendet. An den Stereoklang heranzukommen und die Intention sowie die Grundidee in den 3D-Raum zu übertragen ist ein oft angestrebtes Ziel. Laut den Produzierenden sollen die Hörenden nicht überfordert werden, daher wird auf viele bewegte Elemente verzichtet und ein eher statischer Mix angestrebt. Es wird mehrfach betont, dass Spatial Audio kein bloßer Effekt sondern ein eigenständiges, zukunftsweisendes Format ist. Die Übertragung von Stereo auf Spatial Audio stellt damit einen sich kontinuierlich entwickelnden Prozess dar.

Eine Herausforderung besteht in dem Mischen gemasterter Stereospuren, da diese bereits viele Effekte enthalten und den kreativen Prozess stark einschränken. Übermäßiges EQing oder Komprimieren wirkt sich negativ auf die Qualität des Spatial-Audio-Mixes aus. Mit den nicht bearbeiteten Originalspuren zu arbeiten, wird daher bevorzugt. Sie sind jedoch selten verfügbar. Mehrfach wird betont, dass sich die Ansätze der Spatial-Audio-Mischungen nach den Genres richten. Beispielsweise ermöglicht elektronische Musik ein Loslösen von der Realität, weshalb seltener eine Bühnenbildung angestrebt wird. Jedoch bleibt die Stimme ein Ankerpunkt, der auch in diesen Mischungen in den Centerlautsprecher gesetzt wird.

Aus der Studie von Hyunkook Lee¹⁹ geht hervor, dass die linken und rechten Lautsprecher kaum korrelieren – man spricht daher bereits von einem eigenständigen immersiven System innerhalb des 3D-Systems. Der Hauptfokus liegt auf den vorderen Lautsprechern, was den Ansatz der Bühnenbildung bestätigt. Die Analyse von Lees Studie zeigt, dass Surround-Lautsprecher in einer Atmos-Mischung hauptsächlich dazu dienen, das Stereo-Bild zu erweitern. Die hinteren Kanäle werden vor allem für Hüllkurven, die Seiten für rhythmische Elemente mit klaren Transienten verwendet. Darüber hinaus ersetzt der Hallpegel in Atmos-Mischungen die traditionelle Tiefendarstellung, indem der Hall gezielt über der Hörposition platziert wird.

Weitere Überlegungen zu Spatial-Audio-Mischungen betreffen den Konsumkontext: Wann und wo wird die Musik gehört? Die meisten werden Spatial Audio über Kopfhörer konsumieren und sich dabei auch im Freien bewegen. Daher sollte man vorsichtig mit Sounds sein, die die Hörenden irritieren könnten. Insbesondere unerwartete Klänge von hinten sollten vermieden werden, und Bewegungen von Klangquellen überlegt eingesetzt werden. Vor allem sind Bewegungen in der Vertikalen nicht optimal wahrnehmbar. Man kann sich allerdings einiger psychoakustischer Prinzipien bedienen, indem der Frequenzbereich um 5 kHz angehoben wird, um die Bewegung hervorzuheben. Die Horizontale ist die wichtigste Ebene zur Lokalisierung von Klängen. Die vertikale wird hingegen für Effekte und das Spiel mit der Klangfarbe verwendet. Durch die HRTFs sollte man sich der Vorne-

¹⁹ vgl. An Investigation Into Immersive Contemporary Popular Music Mixing Practice

Hinten-Verwechslung bewusst sein – ein Schlagzeug nach hinten zu setzen, funktioniert laut einigen Produzierenden in den wenigsten Fällen. Auffällig ist der Ansatz des frequenzabhängigen Mischens. Auch hier ist eine enge Zusammenarbeit mit den Komponierenden von zentraler Bedeutung.

Musik in Ambisonics

In der Musikproduktion bzw. -Komposition mit Ambisonics wird ebenfalls frequenzabhängig gearbeitet. Hohe Klänge werden nach oben in den Raum gesetzt und tiefe in den unteren Bereich. Klänge mit einem breiten Frequenzspektrum werden für Bewegungen genutzt. Für Ambisonics ist ebenfalls von Bedeutung nicht zu viele bewegte Elemente in den Mix zu setzen, da auch hier zu viele Bewegungen schnell irritierend sein können. Dies kann und wird allerdings auch gezielt als Effekt eingesetzt.

Der klassische Frontbühnenansatz wird seltener verwendet. In Ambisonics wird meist raumbezogen gearbeitet, wodurch mehrere Hörpositionen entstehen. Man soll sich durch den Raum bewegen können und verschiedene Hörerfahrungen erleben. Wird mit einer Front gearbeitet, dann werden Schlagzeug und Lead-Vocals – wie bei Dolby Atmos – nach vorne gesetzt.

Recording von Spatial Audio und Mischung von Livemusik

Bei der Aufnahme von Jazz, Klassik und Orchestern wird mit einer klassischen Bühne vorne gearbeitet. Die Rauminformationen werden nach hinten übertragen. Die Performance wird akustisch eingefangen, während der Hauptfokus auf dem Raum liegt, der aufgenommen und wiedergegeben wird. Von zentraler Bedeutung ist dabei die Dynamik der Aufnahme, denn sie gilt als Schlüssel für eine immersive und lebendige Produktion.

Ein häufig genannter Ansatz ist die Übertragung der aufgenommenen Quellen direkt auf die Lautsprecherpositionen im Raum, ohne Processing. Ziel ist es, kein Panning einzusetzen und somit keine virtuellen oder Phantomschallquellen zu erzeugen, da die Auflösung dadurch reduziert wird. Einen Klang zwischen zwei Lautsprechern zu platzieren, reduziert die Auflösung um die Hälfte. Panning sei dennoch möglich, da unser Gehirn in der Lage ist, ein eigenes Panning durchzuführen. Ein Klang, der erst hinten und dann vorne abgespielt wird, wird automatisch gepannt. Es wird betont, wie wichtig die Auseinandersetzung mit der Psychoakustik des Menschen ist. Eine Möglichkeit in der Software zu pannen, ist ein frequenzbasiertes, laufzeit- oder lautheitsbezogenes Panning.

Kleine Bewegungen eines Instruments können große Veränderungen in der Klangbalance bewirken, weshalb eine bewusste räumliche Platzierung entscheidend ist – unabhängig davon, ob es sich um Stereo- oder Raumklang handelt. Während technische Korrekturen

möglich sind, bleibt die musikalische Relevanz entscheidend, da die Bewegung selbst ein Ausdruck der Musik ist. Das Klangbild entsteht bereits bei der Aufnahme durch gezielte Mikrofonierung, sodass die Komponierenden und Musizierenden die mehrdimensionale Klangskulptur formen. Da das Ziel vieler Produktionen darin besteht, starke emotionale Reaktionen auszulösen, kommt der gezielten Klanginszenierung eine zentrale Bedeutung zu. In der Popmusik beispielsweise basiert ein Großteil des Höreindrucks auf dem Direktschall.

Allgemein sei es wichtig, unkorrelierten Raumklang aufzunehmen, da durch symmetrisches Arbeiten Laufzeitunterschiede, welche das Klangbild zerstören können, vermieden werden, was durch den Fokus auf die Hauptmikrofonierung unterstützt wird. Die Arbeit mit vielen Stützen kann somit vermieden werden. Bewegung von Klang kann bereits bei der Aufnahme durch verschiedene Mikrofonpositionen eines Klanges erzielt werden – dies ist jedoch zeitaufwendig. Auch Tiefenstaffelung kann durch eine ähnliche Methode in die Aufnahme gebracht werden. Hierbei werden die Musizierenden aus verschiedenen Positionen im Raum immer wieder aufgenommen. Eine Herausforderung sind, wie bei allen Aufnahmen, die Raummoden. Diese sind nicht vollständig vermeidbar, daher wird in vielen Fällen mit leichter Kompression gearbeitet und in der Postproduktion nach Bedarf der Frequenzgang angepasst.

Man sollte beachten, dass es zu Übersprechungen kommen kann. In einigen Fällen sei daher eine räumliche Trennung notwendig. Es wird hervorgehoben, in alle Mikrofonpositionen hineinzuhören und eine Stereoaufnahme als Referenz aufzunehmen.

Hörspiel

Im Hörspiel werden viele verschiedene Methoden angewendet, da es sich anbietet, mit verschiedenen Erzählebenen zu arbeiten. Zum Beispiel wird im Spatial-Audio-Mix eine Stereospur integriert, um die Im-Kopf-Lokalisation für Sprecher oder andere Effekte zu nutzen. Räumlichkeit wird am besten bereits in die Aufnahmen eingebracht – durch die Nutzung verschiedener Räume oder die Veränderung der Raumakustik, beispielsweise durch den Einsatz verschiedener Böden im Studio. Wie bei Live-Aufnahmen werden Stereostützen statt Monostützen eingesetzt, um direkt eine räumliche Tiefe in die Mischung zu bringen, da bei strenger Betrachtung streng genommen sind ideale Monoquellen nicht realistisch und kommen in der Natur nicht vor.

Im Hörspiel wird der Ansatz der Verdichtung eingesetzt: Auf realistische Abbildungen wird verzichtet, emotionale Klänge müssen übertrieben werden, um den Spannungsbogen aufrechtzuerhalten. Sprachkunst, atmosphärische Klänge, Geräusche, Sounds und Musik sollen sich gegenseitig kontrapunktisch hervorheben. Darüber hinaus machen beim

Hörspiel eine gute Story und eine optimale Mikrofonierung bereits einen großen Teil der Produktion aus und sind Voraussetzung für die Schaffung einer immersiven Welt. Wichtig ist es, die Zuhörenden aktiv in die Story einzubeziehen: Sind sie Teil des Geschehens oder Beobachtende von außen?

Auch bei Hörspielen wird in der Regel mit einer Frontbühne gearbeitet. Es wird die Arbeit mit Übertreibungen und symbolhaften Klängen empfohlen, anstatt realistische Darstellungen umzusetzen. Es sei wichtig, ihre eigene Interpretation und Vorstellung in die Musik einfließen zu lassen. Zu viele Vorgaben und zu realistische Umsetzungen können dem entgegenwirken. So werden zum Beispiel Laufbewegungen nur angedeutet, statt sie perfekt im Raum zu platzieren. Bewegung sollte generell nur eingesetzt werden, wenn es wirklich um die Bewegung selbst geht, sonst lenkt sie in der Szene vom Hauptgeschehen ab. Bewegung kann zudem die Aufmerksamkeit von einem wichtigen Dialog ablenken. Eine Lösung ist es, dramaturgische Alternativen zu finden. Werden Podcasts produziert, wird ebenfalls empfohlen, eine Bühne als Anker zu nutzen. Spatial Audio bietet die Möglichkeit, Overvoices räumlich zu verteilen und den Hörenden dadurch die Wahl zu geben, sich entweder die Übersetzung oder das Original anzuhören. Ein Interviewter betont, dass nicht ausschließlich kopfbezogen produziert werden sollte, da es durchaus einige Audiophile gibt, die Stücke mit raumbezogenen Lautsprechersystemen wie Soundbars zu Hause anhören.

Experimentelle Musik und Installationen

Ob Fixed Media oder performative Installation – es ist übereinstimmend wichtig, den Raum, in dem etwas Akustisches abgespielt wird, zu verstehen. Verschiedene Perspektiven einzunehmen und verschiedene Hörpositionen durchzugehen, ist essenziell. Es bietet sich an, Raum und Zeit im Kontext zu sehen und Klangquellen nicht als Punktquellen im Raum zu betrachten. Auf diese Weise können unerwünschte Effekte wie Raummoden oder ungeplante Übersprechungen und Auslöschungen vermieden werden. Durch das Verstehen des Raumes können Resonanzen von Klängen betont werden, um verborgene Melodien sichtbar zu machen. Eine Künstlerin betont, auf das Verwenden identischer Wiederholungsschleifen (Loops) zu verzichten, und empfiehlt stattdessen variierende Repetitionen. Das wiederholte Loopen wird als maschineller Vorgang beschrieben und spiegelt nicht die Natur wider, in der kein Klang je exakt gleich ist.

Wiederholt in das Stück hineinzuhören, wird von allen Befragten mehrfach betont. Es kann hilfreich sein, sich Fragen zu stellen wie: Woher kommt der Klang? In welche Richtung bewegt er sich? Und passt er wirklich dorthin, wo er gerade ist? Das Publikum von Anfang an mitzudenken, kann die Qualität des Stückes deutlich verbessern. Es sei wichtig, sich zu überlegen, wer am Ende die Komposition hört: Sind es eher Laien oder Audiophile? Dies kann entscheidend sein für die Frage, wie viele Bewegungen und wie viele Klänge

im Stück eingesetzt werden.

Werden Aufnahmen verwendet, sollten diese so sauber wie möglich sein, denn verschwommene Aufnahmen reduzieren die räumliche und frequenzielle Qualität. Ein Ansatz ist es, aufgenommenen Klang und aufgenommenen Raum in Relation zu betrachten, denn die Kombination bildet einen konstruierten Raum. Auch in diesem Bereich lohnt es sich, Impulsantworten des Raumes aufzunehmen. Diese können beim Mischen mit Kopfhörern helfen, die Kontrolle über den Klang zu behalten. Die Verteilung der Lautsprecher wird gerne von Normen abweichend durchgeführt. Lautsprecher können gezielt positioniert werden, um nahe oder weit entfernte Klänge zu simulieren. Ein mehrfach genanntes Problem ist der Umgang mit tiefen Frequenzen in Ambisonics. Eine Lösung besteht darin, mit separaten Tieftönen-Files zu arbeiten, die direkt auf die Subwoofer geroutet werden, die bevorzugt in allen Ecken des Hörbereichs aufgestellt werden. Wer in 5th-Order arbeitet, nutzt somit 40 Kanäle.

Weitere gern genutzte Effekte sind der Dopplereffekt, generative Prozesse und Kompositionsansätze von John Cage, welche zum Beispiel die Arbeit mit Stille und Zufallsverfahren umfassen. Neben der Entscheidung, mit wenigen Elementen einen klaren Raum zu schaffen, können auch viele Elemente zu Schwärmen kombiniert werden, um organische und vielschichtige Klangmassen zu erschaffen, die sich dynamisch im Raum entfalten.

Ein erwähnter neuer Begriff ist die Multiband-Spatialisierung, welche ein frequenzabhängiges Mischen beschreibt. Dabei geht es darum, tiefe Frequenzen räumlich zu verteilen, während hohe Frequenzen präzise lokalisiert bleiben, um ein eindringlicheres Klangerlebnis zu schaffen. Dieses Prinzip könnte helfen, Proxemik-Grenzen²⁰ zu verwischen und neue immersive Klangräume für Stimmen, Musik und Geräusche zu erschließen.

Um die Klangwiedergabe im Raum zu überprüfen, kann man ein Rauschen in Kreisbewegungen durch den Raum senden. Dies fördert das Verständnis und kann die Spatial-Audio-Darstellung im Mix verbessern. Die optimale Situation ist die Komposition im Abspielraum selbst – jedoch kann es dadurch zu Übertragungsproblemen in andere Räume kommen. Ein weiterer Test ist es, ein Referenzsignal mit einem Ambisonics-Mikrofon zu kreieren. Der Vergleich der Aufnahme mit dem Originalstück kann dazu beitragen, die Natürlichkeit des Klangs zu gewährleisten. Fordernde, komplexe Klänge können zum Verlust der Aufmerksamkeit führen. Im Allgemeinen sind Klänge zu wählen, die mit dem Raum harmonieren oder zumindest kompatibel sind. Letztendlich ist das Ziel kohärente Räume zu erschaffen.

²⁰ Beschreibt die gefühlten Distanzen zwischen Klangquelle und Hörenden.

Die Arbeit mit einer Frontstage wird als Orientierungspunkt für Installationen angesehen. Im Gegensatz zu den bisherigen Formaten ist es nicht das vorrangige Ziel, die Komposition ausschließlich auf die Front auszurichten. Anders ist es der Fall bei Inhalten, die mit visuellen Medien arbeiten.

Formatübergreifende Ansätze

Generell wird ein Upmixing mit entsprechenden Plug-ins nicht empfohlen, da es ausschließlich bei ambienten Sounds oder Stücken den gewünschten Effekt erzielt und sollte daher nicht alleinstehend genutzt werden. Bestimmte Aspekte sind mit Spatial Audio nicht kompatibel. Die Technik kann viele Ansätze nicht vollständig umsetzen. Zum Beispiel kann ein fallendes Objekt in einem Ambisonics-Setup nicht realistisch umgesetzt werden, sondern nur angedeutet werden.

Auffällig ist der Ansatz des frequenzabhängigen Mischens, der sich in allen Herangehensweisen wiederfindet. Sich der Psychoakustik des Menschen bewusst zu sein, wird als unabdingbar beschrieben. Ebenso wird häufig eng mit den Komponierenden zusammengearbeitet – ein Vorgehen, das an die Arbeit in Stereo erinnert. Das wiederholte Hören der eigenen Stücke aus verschiedenen Positionen und Winkeln oder mit unterschiedlichen HRTFs wird immer wieder empfohlen. Darüber hinaus gilt es, eine gewisse Flexibilität gegenüber Hard- und Software zu bewahren: Man sollte sich nicht ausschließlich auf eine bestimmte Arbeitsweise festlegen, um offen für neue und womöglich effektivere Ansätze zu bleiben.

Das bewusste Zurücknehmen von übermäßigen Halleffekten wird als sinnvoller Ansatz bewertet – insbesondere im Hinblick auf die Ortbarkeit von Klängen, die durch übermäßigen Hall beeinträchtigt werden kann. Auch die Arbeit mit einer Vielzahl an Lautsprechern wird als Chance verstanden, mit geringeren Lautstärken zu arbeiten und sich von tendenziell undynamischen, überkomprimierten Mischungen zu distanzieren.

5.2.11 Produktions und -Kompositionsprozesse

Aus den zuvor beschriebenen Methoden und Philosophien lassen sich folgende Workflows und Prozesse ableiten:

Musikproduktion und -Komposition

Die Wahl des Workflows hängt in erster Linie vom Ausgangsmaterial ab. Dieses entscheidet, welche Prozesse angewendet werden können.

- Upmixing – durch spezielle Software werden Stereo- und Surround-Files in Formate wie 9.1 hochgemischt; ist zeitsparend, aber von niedriger Qualität und wird nicht empfohlen.
- Stereomaster-Stemmixing – die häufigst genannte Methode; die Stereospuren werden von den Künstlern exportiert und enthalten zum Teil übermäßiges Processing – liefert deutlich bessere Qualität als Upmixing, jedoch mit erheblichem Einfluss auf kreative Entscheidungen im Mix.
- Originalspuren-Stemmixing – optimale Voraussetzung für einen optimalen Mix
- Direktes Komponieren und Produzieren in Spatial Audio – die beste Qualität kann durch die direkte Arbeit im Format erreicht werden; höchste Flexibilität und freier kreativer Prozess – ausnahmslos von allen Befragten als beste Variante beschrieben.

Je nach Ansatz werden folgende Schritte in der Spatial-Audio-Produktion als unabdingbar gesehen:

- Recording mit verschiedenen Mikrofonierungstechniken.
- Verteilung der Klangquellen im Panorama.
- Lautheitsanpassungen, EQ und – wenn nötig – dynamische Prozesse.
- Bewegungen von Klängen, nach sinnvollen Kriterien, die das Stück qualitativ fördern.
- Technisches oder kreatives Mastering – je nach Format.

Aufnahmen

Das Aufnehmen von Musik in Spatial Audio unterscheidet sich durch die Wahl des Hauptmikrofons, der Anzahl und Art der Mikrofone sowie durch die Wahl der Einzel bzw. Stützmikrofonierung:

- Ambisonics-Hauptmikrofon mit Stützen – gute räumliche Abbildung; durch die Stützen wird die Lokalisation optimiert.
- Mikrofonarray (z. B. 2L-Cube oder PCMA-3D) mit Stützen – flexibel für alle Formate anwendbar.

- Direktmikrofonierung in Kombination mit zuvor aufgenommenen Raumimpulsantworten – für hochwertige Aufnahmen mit wenigen oder ohne Störgeräusche.

Bei Aufnahmen sollten folgende Faktoren bedacht werden:

- Stereostützen statt Mono benutzen – führt zu mehr Räumlichkeit in der Aufnahme.
- Die Aufnahme von Raumimpulsantworten kann mit einem Ambisonics-Mikrofon oder durch die aufwendige Mikrofonierung mit einer hohen Anzahl an Mikrofonen durchgeführt werden – letztere Variante führt zu qualitativ hochwertigeren Impulsantworten.
- Die Verteilung der Mikrofone, zum Beispiel an der Wand oder hinten im Raum, sollte immer überprüft und bei Bedarf angepasst werden.
- Bei der Nutzung von hauptsächlich Nieren- oder Ambisonics-Mikrofonen sollten zusätzliche Kugelmikrofone eingesetzt werden, um ausreichend tieffrequente Anteile aufzunehmen.
- Durch das direkte Panning auf die Lautsprecherpositionen in der Postproduktion wird die höchste Auflösung erreicht.

Hörspiel

Die Produktion eines Hörspiels zeigt Parallelen zur Musikproduktion in Spatial Audio. Ein Hörspiel kann von Stereo nach Spatial Audio hochgemischt werden. Naheliegender wäre in diesem Fall die Verwendung einer der Stemmixing-Varianten. Die von den Experten beschriebene qualitativ beste Variante ist jedoch die Kreation eines Hörspiels direkt im Spatial-Audio-Format. Diese unterscheidet sich durch das gewählte Format – in der Regel Ambisonics oder Dolby Atmos – sowie durch die Mikrofonierungstechniken. Ein weiterer großer Entscheidungspunkt ist die Aufnahme der Foleys bzw. das Sounddesign allgemein. In entsprechenden Studios kann Tiefenstaffelung und Bewegung direkt in die Aufnahme eingebracht oder nachträglich in der Postproduktion erzeugt werden.

Experimentelle Produktionen und Installationen

Experimentelle Produktionen und Installationen werden in der Regel direkt im Spatial-Audio-Format gedacht. Ein Stemmixing ist unüblich und wird seltener angewandt. Die Workflows unterscheiden sich vor allem durch die gewählte Soft- und Hardware. Viele Werke werden mit Ambisonics realisiert, während andere teils komplexe Soft- und Hardware-Kombinationen mit Programmen wie Max/MSP, Pure Data, SPAT und analogen Geräten umfassen.

5.3 Ergänzungen durch Eventualfragen

Die Anzahl der in einem Interview gestellten Fragen war elementar von der zur Verfügung stehenden Zeit abhängig. Viele Interviews boten die Möglichkeit, weitere Fragen zu stellen, die sich in den Gesprächsverlauf integrieren ließen oder als Nachfragen vor der Schlussfrage gestellt werden konnten. Viele Antworten waren bereits so umfangreich, dass bestimmte Fragen automatisch beantwortet wurden. Diese Informationen sind in den folgenden Kapiteln zusammengefasst.

5.3.1 Psychoakustik und Hörgewohnheit

Die Hörgewohnheit und die Psychoakustik des Menschen spielen eine entscheidende Rolle in Workflows und Ansätzen in der Arbeit mit Spatial Audio.

Klänge nach hinten zu platzieren, kann durch unsere Hörerfahrung als Gefahr interpretiert werden und entsprechende Stressreaktionen auslösen. Zudem werden höhere Klänge im Normalfall oben geortet, während tiefere unten wahrgenommen werden. Wird dies nicht berücksichtigt, kann es zu verwirrenden Effekten führen, die den Hörer überfordern können. Unser Gehör dient zum Überleben und nicht, um Musik zu hören – so sind durch die Evolution bestimmte Hörgewohnheiten tief in uns verankert. Einige der Befragten stellten in ihren Forschungen fest, dass vor allem unkorrelierte Klänge entscheidend für die Unterscheidung von Richtungen sind. Durch Head-Tracking kann den Hörenden die Gewinnung zusätzlicher Richtungsinformationen ermöglicht werden, da durch das Bewegen des Kopfes Klänge besser lokalisiert werden – insbesondere in der Vorne-Hinten-Ortung.

Generell ist es wichtig, Klänge so einzusetzen, dass sie in sich schlüssig wirken und mit den Prinzipien der menschlichen Wahrnehmung übereinstimmen. Besonders für die Wahrnehmung von Entfernung spielen Veränderungen der Lautstärke und der Dopplereffekt eine zentrale Rolle. Weder Lautsprecher noch Kopfhörer geben Schallquellen naturgetreu wieder – ebenso wenig wie Phantomquellen, die in der realen Welt nicht vorkommen. Damit wir dennoch nicht die Orientierung verlieren, interpretiert unser Gehirn die Höreindrücke so, dass sie sinnvoll erscheinen. Diese Fähigkeit ist evolutionär von Bedeutung, da eine fehlerhafte Wahrnehmung potenziell lebensbedrohliche Situationen zur Folge haben könnte.

Neben Lautstärkeveränderungen und dem Dopplereffekt lassen sich weitere psychoakustische Prinzipien einsetzen, wie etwa frequenzabhängige Klanggestaltung, um die subjektive Raumwahrnehmung zu beeinflussen. Diese ist nicht von kartesischen Koordinaten abhängig, sondern entsteht durch unsere rationale Wahrnehmung und kognitive Verarbeitung.

Besonders deutlich wird dies bei Ambisonics, wo tiefe Frequenzen für den Gesamteindruck wesentlich sind, während feine räumliche Strukturen in höheren Lagen oft unbemerkt bleiben.

Ein besonderes Phänomen ist die Fähigkeit unseres Gehirns, eine Klangbewegung zu simulieren, auch wenn technisch gesehen kein Panning stattfindet.²¹ Wird ein Klang nacheinander über zwei unterschiedliche Lautsprecher wiedergegeben, entsteht im Kopf automatisch der Eindruck einer Bewegung. Auch visuelle Einflüsse können die auditive Wahrnehmung beeinflussen: So führen visuelle Reize, zum Beispiel ein Raum voller Lautsprecher, dazu, dass wir erwarten, aus allen Richtungen Klang zu hören – unser Gehirn passt die Wahrnehmung entsprechend an. Diese visuellen Biases können gezielt genutzt werden, um das Hörerlebnis zu verstärken oder zu verändern.

Ein weiterer zu beachtender Bereich ist die Proxemik. Diese untersucht, wie Abstände zwischen Menschen soziale Interaktionen beeinflussen, wobei verschiedene Distanzzonen von Intimität bis Öffentlichkeit unterschieden werden. Dieses Konzept wurde in räumlichen Klangsystemen genutzt, um durch gezielte Lautsprecherplatzierung die Wahrnehmung von Nähe oder Distanz akustisch zu manipulieren. Bei Spatial Audio kann das Publikum bewusst entscheiden, ob es sich vollständig auf den Klangraum einlässt und den realen Raum ausblendet – ähnlich wie im Theater oder Kino, wo eine innere Übereinkunft entsteht, die Vorstellungskraft zu aktivieren. Wenn es keine visuellen Vorgaben gibt, entstehen individuelle, mentale Klangbilder, die sich aus Erinnerungen und Erfahrungen ergeben und so eine tiefere immersive Wirkung entfalten können.

5.3.2 Probleme und Herausforderungen

Die Nutzung und Entwicklung neuer Technologien geht in der Regel mit Herausforderungen und Problemen einher. Welche das sind, wird in diesem Kapitel anhand der Primär- und Sekundärdaten zusammengefasst.

Ein grundlegendes Problem ist die Unbekanntheit der Spatial-Audio-Formate und wie sie zu verstehen sind. Was Spatial Audio ist, wird meistens erst durch Demonstrationen nachvollziehbar. Bewusste A-B-Vergleiche zeigen das Potenzial der Technik, die neue künstlerische Freiheit und ihren Mehrwert. Auch die Musikschaaffenden müssen erst von Spatial-Audio überzeugt werden oder sind sich der Formate gar nicht bewusst.

Viele auf Plattformen wie Apple Music veröffentlichte Mischungen schöpfen das räumliche Potenzial von Spatial Audio nur unzureichend aus. Häufig werden sie fälschlich

²¹ Vgl. Person N, Interview 14 (eigene Erhebung).

als herkömmliche Stereomischungen wahrgenommen – etwa aufgrund fehlender psychoakustischer Kenntnisse oder begrenzter Ressourcen bei der Produktion. Ein Großteil der Dolby-Atmos-Mischungen gilt als qualitativ unzureichend, da oft nur bestehende Stereo-Versionen rekonstruiert oder automatisiert hochgemischt werden. Häufig fehlt das für 3D-Audio notwendige Originalmaterial, was die Umsetzung zusätzlich erschwert.

Die kopfbezogene Wiedergabe löst das mehrfach erwähnte Sweet-Spot-Problem, bringt aber auch eigene Herausforderungen mit sich. Ein bekanntes Problem ist das falsche Platzieren der Lautsprecher. Viele Konsumierende halten sich nicht an die vorgegebene Aufstellung, was den Stereo-Effekt und schließlich bei Spatial Audio die räumlichen Effekte und das Panorama zerstört. Bei der kopfbezogenen Wiedergabe funktionieren die generischen HRTFs nicht für jede Person gleichermaßen. Anbieter wie Dolby und Apple erschweren den Mixing- und Mastering-Prozess, durch intransparente Prozesse im Hintergrund. Dadurch kann zum Beispiel das Head-Tracking von Apple bei den neuen AirPods Pro, das die Metadaten von Dolby Atmos umgeht, einen dedizierten Atmos-Mix gravierend beeinträchtigen, da dies in der Mischung nicht berücksichtigt werden konnte. Die Tiefenstaffelung und Räumlichkeit gehen dadurch verloren. Hinzu kommt das bereits diskutierte Problem der sich laufend erneuernden Softwareprozesse, die in der Zukunft die bereits produzierten Mixe immer wieder verändern werden. Eine weitere Herausforderung in den Workflows ist der Umgang mit tiefen Frequenzen bei Binauralmischungen und Ambisonics. Bei Dolby Atmos wird der LFE nicht auf die Kopfhörermischung übertragen, weshalb ein alternativer Ansatz gefunden werden muss.

Die Hörgewohnheit und die Psychoakustik des Menschen werden immer wieder als Schlüsselfaktoren genannt. Die Stereowiedergabe über Kopfhörer wird traditionell falsch wiedergegeben, da eine für 60°-Lautsprecherwinkel optimierte Mischung auf eine 180°-Kopfhörerwiedergabe übertragen wird, was den Raumklang verzerrt. Die Binauralisierung kann dies korrigieren, indem sie die ursprüngliche Stereobasis auf den Kopfhörer überträgt. Doch dies wird für viele Hörenden ungewohnt klingen, da es von ihren Hörgewohnheiten abweicht. Daher ist bei Spatial-Audio- und binauralen Mischungen ein klares Erwartungsmanagement entscheidend. Komponierende und Produzierende sollten frühzeitig auf Unterschiede in Dynamik, Frequenzgang und Räumlichkeit vorbereitet werden, um Missverständnisse zu vermeiden.

Viele der erarbeiteten Workflows stoßen durch die Software und Hardware an ihre Grenzen. Vor allem für Installationen und Live-Events werden häufig Programme miteinander verkettet, um die gewünschten Ideen umzusetzen. Beliebt sind hierfür Programme wie Spat Revolution oder Max/MSP. Dolby Atmos wird durch die Begrenzung auf 128 Objekte in einem Projekt eingeschränkt. Die Software und der Workflow von Dolby Atmos

sind auf die Arbeit mit einer Front ausgelegt, was die Kreativität einschränkt. Die hohen Lizenzkosten, mit welchen Dolby Atmos verbunden ist, stellen zudem für viele eine Barriere dar. Außerdem ist der Ansatz der Kombination von Binaural und Stereo, um eine weitere Ebene in den Mix zu bringen nicht möglich. Zum Beispiel kann keine Im-Kopf-Lokalisation, für besondere Effekte erzeugt werden.

Unterschiedliche Räume besitzen eigene akustische Eigenschaften. Das Aufnehmen von Raumimpulsantworten kann helfen, das Mischen zu optimieren. Eine weitere aktuelle Herausforderung ist die konsistente Wiedergabe von Spatial-Audio-Mischungen auf unterschiedlichen Systemen, da nicht jede Plattform alle Formate unterstützt. Auch die Kombination von DAWs mit externen Hardwaremodulen stellt die Audioschaffenden vor Herausforderungen, da speziell entwickelte Hardware-Komponenten fehlen. Besonders haptisches Panning könnte viel Zeit beim Mischen einsparen.

Die Genres bringen jeweils eigene Herausforderungen mit sich. So ist es nicht einfach, die Durchsetzungskraft minimalistischer Hip-Hop-Tracks über Kopfhörer überzeugend zu reproduzieren. Generell werden viele Klangelemente benötigt, um das Panorama sinnvoll zu füllen. Beim Stemmixing mit wenig Ausgangsmaterial muss häufig auf Dopplungen von Spuren zurückgegriffen werden, was die kreativen Entscheidungen einschränken kann. Eine wirklichkeitsgetreue Abbildung realer Klangbilder ist technisch nicht vollständig möglich – daher sind gezielte gestalterische Überlegungen erforderlich. Klänge, die sich von oben nach unten bewegen sollen, werden in den meisten Fällen nicht als solche wahrgenommen.

In vielen Produktionen wird das Potenzial von Spatial Audio bislang nur teilweise genutzt. Das liegt an fehlendem Wissen, zu wenig Budget und einer mangelnden Auseinandersetzung mit der Psychoakustik. Viele Labels und Produzierende sehen Spatial-Audio lediglich als Add-on. Einige Produktionen basieren auf einer schnellen Übertragung bestehender Stereo-Stücke in das Spatial-Audio-Format, um wirtschaftlich auf den aktuellen Trend zu reagieren. Außerdem werden teils unrealistische Erwartungen geweckt und Versprechen gemacht, die weder technisch noch künstlerisch vollständig einlösbar sind. Es kursieren zahlreiche irreführende Vorstellungen über die Möglichkeiten von Spatial Audio. Ein treffendes Zitat fasst diese Problematik gut zusammen, denn solche Aussagen sind bedingt nur mit der Wellenfeldsynthese möglich und somit für den normalen Hörenden nicht zugänglich:

„Stell dir vor, du sitzt in einem Raum und da ist ein Fenster offen. Dann kommt eine Fliege reingeflogen, fliegt durch den Raum, um deinen Kopf, dann in den Nebenraum und wieder raus. Das machen wir mit Spatial Audio. Weißt du, wie oft ich diese Geschichte gehört habe? Einfach unfassbar.“²²

Es besteht Konsens, wenn es um die Qualität der Spatial-Mischungen geht: Wir brauchen hochwertige Produktionen, sonst nimmt das Interesse an Spatial Audio wieder ab und wir haben die gleiche Situation wie bei der Quadrophonie in den 70er/80er-Jahren. Derzeit lässt sich bei einem Großteil der Spatial-Audio-Mischungen diskutieren, inwiefern sie das Potenzial des Formats sinnvoll ausschöpfen.

5.3.3 Faszination und Vorteile von Spatial Audio

Die Befragten sind sich einig: Ein durchdachter Spatial-Audio-Mix klingt besser und sorgt bei Hörenden für einen Wow-Effekt, der die eigene Motivation, in diesem Format zu arbeiten, steigert. Es ist faszinierend, ein neues Format und die dazugehörigen Themen rund um das Hören und die Klangforschung in Verbindung zu bringen und Teil des Entwicklungsprozesses zu sein. Spatial Audio weckt die Neugier und bietet die Chance, eigene Workflows in einem noch nicht durchstandardisierten Format zu finden.

Den Komponierenden und Kunschtchaffenden bietet Spatial Audio eine größere Leinwand und somit viele neue Freiheiten im Umgang mit Klängen sowie die Möglichkeit, Konflikte aus dem Stereoformat, wie etwa den Loudness War, hinter sich zu lassen. Mit einem höheren Dynamikumfang zu arbeiten, eröffnet viele kreative Möglichkeiten. Orte sind neu erfahrbar, Produkte können auf innovative Weise präsentiert werden.

Durch Spatial Audio wist es deutlich einfacher, immersive Welten zu erschaffen. Es ermöglicht, Hörende mit Klängen zu umhüllen und gleichzeitig verschiedene Dimensionen zu erzählen, während sie selbst entscheiden können, worauf sie achten möchten. Dies bietet neue Chancen für das Storytelling und beeinflusst, wie Hörspiele und Podcasts gestaltet werden können. Es ist möglich, parallele Klangwelten oder Klänge zu erzeugen, die eine hochdifferenzierte Wahrnehmung des Klangraums ermöglichen. Dadurch kann eine verbesserte Raumtiefe geschaffen werden. Zudem kann Spatial Audio als eigenständiges Medium funktionieren und ist nicht auf visuelle Medien angewiesen.

Die neu gewonnenen Freiheiten, die Erforschung und Bewegung in die Musik zu bringen, fasziniert die Menschen. Eine Klangkünstlerin betont ihre Faszination für Noise-Musik: Spatial Audio bietet die Möglichkeit, Hörende auf ganz andere Weise zu fordern und

²² Interviewte Person L, Interview 12 (eigene Erhebung).

generell auf einer tieferen emotionalen Ebene zu erreichen²³. Klangobjekten kann in Spatial Audio eine umfangreichere Identität verliehen werden, die den Hörenden auf neuen, unterschiedlichen Wegen vermittelt werden können.

5.4 Zusammenfassung der Abschlussfragen

In diesem Kapitel werden die Antworten der Abschlussfragen zusammengefasst. Welche Tipps geben die Befragten Lernenden, bzw. welche hätten sie selbst gerne gehabt und gibt es Hard- oder Software, die sich die Befragten für ihre Workflows wünschen? Wie denken die Befragten über die Zukunft von 3D-Audio, und würden sie die Arbeit in diesem Format empfehlen?

5.4.1 Zukunft von 3D-Audio

Die Industrie hat sich bereits dafür entschieden, Spatial Audio weiterzuentwickeln und dauerhaft zu etablieren. Für Audioschaffende bedeutet das, sich zwangsläufig mit dem Format auseinanderzusetzen. Zwar wird Stereo nicht vollständig verdrängt, doch wer auch in Zukunft relevant bleiben möchte, sollte in der Lage sein, mit Spatial Audio zu arbeiten.

Dolby Atmos ist im Mainstream etabliert. Konzerne wie Google arbeiten an eigenen Formaten, beispielsweise für YouTube, ohne Lizenzierungsverfahren, was von vielen als großes Potenzial gesehen wird. Das Hören über Kopfhörer ist der Schlüssel für die Verbreitung von Spatial Audio. Dolby Atmos-fähige Smart Speaker kommen auf den Markt, und Dolby Atmos-Systeme werden in neue Autos integriert. Durch Binaural Audio ist 3D-Sound für jeden zugänglich, denn es bedarf keiner vier Lautsprecher und teurer Endstufe wie bei der Quadrophonie in den 70er-Jahren. Interaktive Anwendungen haben enormes Potenzial, unter anderem weil Hersteller mittlerweile Spatial-Audio-Tools direkt in DAWs integrieren.

Spatial Audio wird bleiben – darin sind sich die Befragten sicher. Dennoch besteht große Unsicherheit über die Art und Weise, wie es in Zukunft genutzt wird. Der Musikbereich entwickelt sich noch, während Spatial Audio in Film, Games, Planetarien, Museen oder Events bereits etabliert ist. Objektbasierte Formate sind flexibler als klassische Surround-Formate. Dolby Atmos und andere objektbasierte Workflows versprechen, dass Mischungen in verschiedenen Abhörsituationen immer gleich klingen können, auch wenn die Umsetzung noch nicht der Realität entspricht.

Spatial Audio wird als ästhetische Bereicherung für die Audiobranche wahrgenommen. Es fördert bei Anwendenden nicht nur das räumliche Denken, sondern vermittelt auch

²³ Interviewte B, Interview 2 (eigene Erhebung).

neue Fähigkeiten, die sich positiv auf Stereo-Produktionen auswirken können. Die Arbeit mit Spatial Audio wird als inspirierend empfunden und steigert die Experimentierfreude. Auf die Frage, ob die Arbeit mit Spatial Audio zu empfehlen sei, antworteten alle Befragten einstimmig mit *Ja*. Ebenso herrscht Einigkeit darüber, dass es zukünftig zu einer Koexistenz von Stereo und Spatial Audio kommen wird.

5.4.2 Wünsche

Wenn sie einen Wunsch frei hätten und sich eine noch nicht existierende Hard- oder Software wünschen könnten, was wäre diese? Diese Frage wurde mit sehr diversen Antworten entgegnet – eine sticht besonders hervor: ein haptisches Werkzeug, um das Panning im 3D-Raum zu erleichtern und somit viel Zeit einzusparen. Weitere Wünsche sind:

- Kopfhörer mit Bone-Conduction-Sound und Spatialisation – ähnlich einem Exoskelett, das es ermöglicht, die Räumlichkeit physisch zu spüren.
- Günstige und natürlich klingende Acht-Mikrofone oder ein neuer Mikrofon-Ansatz: ein leichtes, fast masseloses Material, das durch Licht abgetastet wird.
- Panning durch Gestik.
- Qualitativ hochwertige Upmixer.
- Source Management wie bei Spitfire, um Klangquellen direkt zu adressieren, ohne Trajektorien definieren zu müssen.
- Eine Zusammenarbeit zwischen Dolby und Apple, um gemeinsam einen Binauralisierungsprozess zu entwickeln.
- Ein Konvertierungstool für alle Spatial-Audio-Formate, das die Beibehaltung der Distanzinformationen und Lautstärkeanpassungen gewährleistet.
- Umwandlungstools, um Ambisonics in Dolby Atmos zu nutzen.
- Eine gut klingende Ambisonics-Hallsimulation als Software.
- Bessere Delay-Plug-ins mit inspirierenden Presets.
- Ein rotierender Tisch für die Arbeit in Ambisonics-Domes.
- Ein gut klingender De-Mixer für besseres Stemmixing – um bestimmte Passagen, wie Vocals, aus einem Beat zu extrahieren und sie besser im Raum verteilen zu können.
- Mehrfach wurde eine Vereinfachung der Bus-Prozessierung in Dolby Atmos gewünscht.
- Generell sollen Stereo-Plug-ins für die Nutzung in Spatial Audio optimiert werden.

5.4.3 Tipps

Welche Tipps würden die Fachkundigen Lernenden geben oder hätten sie gerne selbst bekommen? Ein oft genannter Tipp ist es, bewusst im Alltag zu hören und sich folgende Fragen zu stellen: Was bedeutet es, einen Klang zu lokalisieren? Wie hören sich die Räume an, in denen ich mich befinde? Das Experimentieren mit Räumen und Klängen ist essenziell für die Arbeit in Spatial Audio. Mehrfach wird auch betont, die Arbeit in Mono und Stereo nicht zu vernachlässigen. Die Fähigkeit, zum Beispiel mit Tiefenstaffelung umzugehen, sollte ebenso in Mono und Stereo erlernt werden.

Nicht jeder Track und jedes Konzert bietet sich an, in 3D aufgenommen zu werden. Diese Unterscheidung treffen zu können, ist eine wichtige Grundlage für die Arbeit mit Spatial Audio. Wer in einem Ambisonics Dom arbeitet, sollte sich immer wieder bewegen, den Kopf drehen und nicht im optimalen Sweet Spot verharren.

„Bleib flexibel, arbeite an deinen Workflows, werde dir der Routingstrukturen bewusst und probiere alles aus – auch wenn es Stereo-Typen sind.“²⁴

Den Austausch über Spatial-Audio-Inhalte und Workflows zu fördern und bereits vorhandenes Wissen zu Nutzen und zu erweitern ist für die weitere Entwicklung notwendig. Die Dokumentation der Plug-ins zu lesen, kann das Verständnis für die Technik fördern. Wer viel experimentiert, macht Fehler – und das ist gut so. In der Lage zu sein, Entscheidungen zu begründen und selbst die Grenzen der Formate zu testen sind wichtige Empfehlungen. Neue Dinge zu erlernen und anzuwenden, ist eine essenzielle Fähigkeit. Nicht nur für die Arbeit in Spatial Audio.

5.5 Meinung zu Standardisierten Spatial Audio Workflows

Bei der Frage, ob es bereits Standards in Bezug auf Technik und Ästhetik in Spatial Audio gibt und ob solche Standards sinnvoll sind, gehen die Meinungen der Befragten auseinander. Das Erkennen von Gemeinsamkeiten und Mustern kann sich jedoch als hilfreich erweisen. Es gibt gewisse Aspekte, die in Spatial Audio einfach nicht umgesetzt werden können – zu wissen, was ästhetisch möglich ist und was nicht, wird als hilfreich eingeschätzt. Häufig sind sich die Kunstschaaffenden nicht bewusst, was technisch realisierbar ist. Für den Einstieg kann es daher sinnvoll sein, sich an grundlegende Guidelines zu orientieren. Dennoch sollten eigene kreative Ansätze entwickelt werden, um authentisch mit Spatial Audio zu arbeiten.

²⁴ Interviewte Person L, Interview 12 (eigene Erhebung).

Technische Standards existieren bereits. Dolby Atmos gibt einen standardisierter Workflow mit zugehöriger Software vor. Auch Streamingdienste geben technische Anforderungen vor, insbesondere im Hinblick auf die Lautheit von Mischungen. Spätestens im Rundfunk wird es Standards geben, um eine gleichbleibender Qualität sicherzustellen. Früher oder später werden sich Schemata entwickeln, die bei Spatial-Audio-Produktionen angewendet werden – denn am Ende muss Zeit und Geld gespart werden.

Aus ästhetischer und psychoakustischer Sicht scheint es eigene Standardansätze etabliert zu haben. Die Stimme wird in der Regel im Centerlautsprecher platziert. Zudem wird in den meisten Fällen bevorzugt, mit einer Bühne in der Front zu arbeiten – auch, weil man sich an die noch vorherrschenden Hörgewohnheiten der Hörenden anpassen möchte und die Idee der Stereo-Stücke, die oft Grundlage für Spatial Audio sind, nicht verlieren möchte. Viele der Befragten sind sich einig, und sprechen sich für Standardmethoden zur Konvertierung der Formate aus, um zum Beispiel Lautheiten auf verschiedenen Wiedergabegeräte konsistent zu halten.

Gleichzeitig warnt der Großteil der Befragten vor einer Standardisierung der kreativen und gestalterischen Aspekte. Die künstlerische Freiheit darf nicht eingeschränkt werden. Dennoch kann es hilfreich sein, bestimmte Grundlagen mit auf den Weg zu geben. Technische Standards sind sinnvoll, während künstlerische und kreative Standards, die die Ästhetik betreffen, kritisch betrachtet werden – auch wenn sie Prozesse vereinfachen können.

6 Analyse

In diesem Kapitel werden die Primär- und Sekundärdaten anhand der aufgestellten Thesen ausgewertet. Zunächst werden die Erkenntnisse aus den Spatial Audio Listening Sessions und der vorausgegangenen experimentellen Forschung zur Wahrnehmung komplexer Klänge in Spatial Audio Formaten erläutert. Nach der methodischen Analyse der Thesen werden diese in der Diskussion kritisch hinterfragt.

6.1 Erkenntnisse aus den Spatial Audio Listening Sessions und der experimentellen Forschungsarbeit

Die Grundlage und Idee dieser Arbeit entwickelte sich aus der praktischen Auseinandersetzung mit Spatial Audio über einen Zeitraum von circa sechs Jahren. Die experimentelle Forschungsarbeit *Sound in Space*, die von den Experten Stefan Troschka und Prof. Thomas Görne begleitet wurde, ermöglichte umfassende Einblicke in die subjektive Wahrnehmung der Hörenden und den Einfluss psychoakustischer Effekte. Ziel der Arbeit war es, herauszufinden, wie Zuhörende komplexe Klänge anhand der elektroakustischen Kompositionen von Erik Nyström wahrnehmen und beschreiben. Die Daten wurden im Rahmen fokussierter Gruppeninterviews im Format der Spatial Audio Listening Sessions (SALS) erhoben. Im Folgenden werden die wichtigsten Ergebnisse, die für diese Arbeit relevant sind, vorgestellt:

- Einen großen Einfluss auf die Wahrnehmung hatte der Sweet Spot, der maßgeblich für die unterschiedlichen Eindrücke der Räumlichkeit und der komplexen Klänge verantwortlich war.
- Überraschend war die Wahrnehmung von Klängen in der vertikalen Ebene: Obwohl es sich um ein Surround-Setup ohne aktive Höhenlautsprecher handelte, wurden von allen Teilnehmenden Klänge in der Höhenebene wahrgenommen. Es kann angenommen werden, dass die Hörerfahrung, das frequenzabhängige Hören sowie das bloße Vorhandensein der Höhenlautsprecher, trotz Inaktivität, dabei eine Rolle spielten. Diese Erkenntnisse wurde in den SALS bestätigt (siehe nächster Abschnitt).

Nach Abschluss der experimentellen Forschungsarbeit wurden die SALS pausiert, bis sie im August 2024 mit Unterstützung von Lucca Riitano, Jan Feldmann und David Stolz aus dem Masterstudiengang Sound/Vision der HAW Hamburg mit einem neuen Konzept einen größeren Zulauf fanden und seitdem regelmäßig stattfinden. Während der SALS entstehen zahlreiche Diskurse über die Wahrnehmung von Raum, Ästhetik und Lokalisation in Spatial-Audio-Produktionen. Folgende Grundlagen und relevante Erkenntnisse für diese Arbeit konnten dabei gewonnen werden:

- Die Relative Position der Hörenden bezüglich des Sweet Spots zeigt sich erneut als wesentlicher Einflussfaktor auf die Wahrnehmung von Spatial-Audio-Stücken. Die Abhängigkeit vom gewählten Format ist dabei deutlich erkennbar: Kanalbasierte Stücke weisen einen empfindlicheren Sweet Spot auf als szenenbasierte Formate wie Ambisonics.
- Das Sweet-Spot-Problem wird durch die Abschattung anderer Teilnehmender verstärkt. Sitzt oder steht eine Person im Schallweg eines Lautsprechers, kann die Wahrnehmung des Stückes stark beeinträchtigt werden. Infolgedessen fällt das Spatial-Audio-Bild teilweise zusammen, denn es werden nur die anderen Schallquellen wahrgenommen. Auch hier zeigen sich deutliche Unterschiede zwischen kanalbasierten, szenenbasierten und objektbasierten Formaten. Im Allgemeinen werden Klänge im Bereich oberhalb der Horizontalebene, also in Richtung der Höhenlautsprecher, seltener durch Objekte oder andere Personen verdeckt. Formate, die auf die Verwendung von Höhenlautsprechern zurückgreifen, können folglich zu einer stabileren räumlichen Wahrnehmung beitragen, da Verzerrungen durch akustische Abschattung in diesen Bereichen seltener auftreten.
- Spatial-Audio-Stücke, die explizit im Zielformat produziert wurden und nicht lediglich durch Stem-Mixing oder Upmixing übertragen, wurden von den Hörenden als qualitativ hochwertiger empfunden.
- Die Qualität in Spatial Audio wurde als Genre abhängig empfunden. Genres wie Rock und Metal werden seltener produziert und tendenziell schlechter abgemischt, während elektroakustische Stücke, elektronische Musik und Klanglandschaften deutlich mehr Zustimmung erhalten.
- Visuelle Cues beeinflussen die auditive Wahrnehmung. Dieses Phänomen zeigte sich bereits in der experimentellen Forschungsarbeit: Dort, wo Lautsprecher sichtbar sind, werden häufig Klänge wahrgenommen - auch wenn diese Lautsprecher in der Realität nicht bespielt werden.

Jedes dieser Phänomene spielt eine entscheidende Rolle bei der Wahl des Formats und erfordert eine Anpassung der Workflows, um den damit verbundene Herausforderungen zu begegnen. Diese Ansätze werden in der Diskussion (siehe Abschnitt 7) erneut aufgegriffen.

6.2 Analyse der Thesen

Für die qualitative Inhaltsanalyse wurden drei Thesen formuliert. Diese entstanden aus der jahrelangen Forschung und Arbeit in Spatial Audio und aus einer Vielzahl von Gesprächen mit Spatial-Audio-Schaffenden. Die Analyse orientiert sich an den drei Thesen.

6.2.1 Hypothese 1

Unabhängig von Genre, Format oder beruflichem Hintergrund zeigen sich in der Produktion und Komposition von Spatial Audio wiederkehrende Muster und Gemeinsamkeiten in kreativen und technischen Herangehensweisen.

Im Regelfall werden innerhalb von Praktika und in der Ausbildung an Universitäten oder Hochschulen mehrere Formate – in der Regel Ambisonics, Dolby Atmos und kanalbasierte Setups wie zum Beispiel das Würfelsetup mit acht Lautsprechern – gelehrt bzw. stehen den Auszubildenden zur Verfügung. Aufgrund der geringen Nutzung und Verbreitung ist die Wellenfeldsynthese kein fester Bestandteil der Lehre und es liegen daher nur wenige Daten vor. Im schulischen Kontext werden Stücke, die in Ambisonics produziert wurden, gerne in Dolby Atmos reproduziert, und umgekehrt. Dies ermöglicht direkte Vergleiche und fördert ein umfassendes Verständnis der Auszubildenden für beide Formate. Oft wird versucht, dieselbe Idee in verschiedenen Formaten umzusetzen. Dabei werden die technischen Voraussetzungen weniger berücksichtigt, was dazu führt, dass einige Elemente der Stücke nicht wie gewünscht im 3D-Raum funktionieren.

Aufseiten der Soft- und Hardware gibt es vor allem Überschneidungen bei Plug-ins, die im Normalfall für die Bearbeitung von Mono und Stereo eingesetzt werden. Dies ist in der Regel auf die zum aktuellen Zeitpunkt nicht vorhandene Verfügbarkeit entsprechender Versionen für das jeweilige Spatial-Audio-Format zurückzuführen. Die meisten Plug-ins sind formatspezifisch – insbesondere, wenn es um die IEM Plug-in Suite, die Dolby-Atmos-Suite, Reverbs für 9.1.6 oder spezielle Reverbs für 3rd-Order-Ambisonics geht. Der Großteil der Software ist untereinander nicht kompatibel. Ein selten eingesetzter Workaround ist die Nutzung von Spat Revolution oder Programmen wie Max/MSP. Mithilfe dieser Tools können zumindest Einschränkungen der DAWs umgangen werden. In Nuendo/Cubase ist es beispielsweise nicht möglich, über 3rd-Order-Ambisonics hinauszuarbeiten, während Reaper insbesondere für Ambisonics-Produktionen genutzt wird, um Inhalte in 7th-Order-Ambisonics zu realisieren. Es gibt deutlich mehr Plug-ins, die auf binaurale Arbeit ausgelegt sind und in verschiedenen Formaten verwendet werden können. Zum Beispiel die Plug-ins von DearVR.

Im Vergleich zur Software in Spatial Audio sind Mikrofoniertechniken formatübergreifend einsetzbar. Der 2L-Cube von Morten Lindberg kann sowohl für Dolby-Atmos-Aufnahmen als auch für Ambisonics-Produktionen verwendet werden. Genre- und formatübergreifend wird die Einzelmikrofonierung genutzt. Vor allem bei Musikproduktionen und Hörspielen wird in vielen Fällen auf eine Einzelmikrofonierung zurückgegriffen, die mittlerweile vorwiegend mit Stereomikrofonierungen, um eine räumliche Erweiterung des Klangbildes zu erreichen, kombiniert wird. Eine Ausnahme bildet Hip-Hop – hier werden vor allem Mo-

noquellen aufgenommen. Generell gilt es, auf Laufzeitdifferenzen, Phasenauslöschungen und die Nähe zum aufgenommenen Klang zu achten. Unterschiedliche Mikrofoncharakteristiken führen zu unterschiedlichen Aufnahmen – dies gilt für alle Genres und Formate. Somit werden Ambisonics-Mikrofone bevorzugt zur Aufnahme von Klanglandschaften und Atmosphären genutzt, während die Verwendung von Mikrofonarrays zur Aufnahme von Musik präferiert wird. Wenn es um die Aufnahme für kopfbezogene Formate geht, sind sich die Interviewten einig: Der KU100 ist hierfür das geeignetste Mikrofon. Auch sind sich alle dessen bewusst, dass der Umgang mit Direktschall, frühen Reflexionen und dem Nachhall entscheidend ist – ob direkt in der Aufnahme oder nachträglich durch eine Software realisiert.

Die Arbeit in Spatial Audio erfordert eine Menge an Ausgangsmaterial, um das große Panorama zu füllen. Limitierungen gibt es lediglich durch die Rechenleistung oder durch Formatbeschränkungen – wie bei Dolby Atmos, das auf 128 Objekte begrenzt ist. Bei Genres wie Hip-Hop muss die Balance zwischen einem gefüllten Panorama und Effekten, wie dem gewünschten Punch, des Stückes gefunden werden. Die Benutzung von bewegten Elementen sollte bewusst eingesetzt werden, um eine Verwirrung oder unangenehmen Momenten zu vermeiden. Dies kann allerdings auch als gewollter Effekt eingesetzt werden – wird jedoch vor allem bei experimentellen Produktionen, Kompositionen und Installationen genutzt. Die Lenkung der Aufmerksamkeit der Hörenden wird hierfür gerne mit bewegten Klangelementen realisiert. Bewegte Klänge in der Vertikalen können nur schwer umgesetzt werden. Aus diesem Grund wird die Bewegung in der Regel auf die horizontale Ebene beschränkt.

Wenn es um die Positionierung der Klänge im Panorama geht, scheinen die Lead-Vocals und Stimmen generell in der Front, vor allem in der Center-Position, als Anker übergreifend umgesetzt zu werden. Ausnahmen bilden Kompositionen im experimentellen Bereich, bei Jazz, Installationen und Hörspielen. Hier wird oft ein Überraschungseffekt eingesetzt, indem Stimmen nach hinten oder oben gesetzt werden. Die Dopplung von Klängen und das Verwenden eines Ruf und Antwort Prinzips, der Einsatz des Dopplereffekts und eine kontrapunktische Verteilung von Klängen werden übergreifend eingesetzt.

Die Positionierung der Klänge im Tieftonbereich stellt generell eine Herausforderung in Spatial Audio dar. Lediglich durch den Einsatz des LFEs, der jedoch nur bei der Hälfte der Dolby-Atmos-Produktionen eingesetzt wird, die nicht binauralisiert wiedergegeben werden, ist ein einfacheres Bassmanagement gegeben. Die Erreichung idealer Bassverhältnisse bei Ambisonics-Produktionen und -Installationen erfolgt durch das gezielte Platzieren mehrerer Subwoofer an den akustisch geeigneten Stellen des Raumes. Diese Positionierung erfolgt häufig in den Ecken oder entlang der Raumgrenzen, wobei die konkrete Ausrich-

tung von der Geometrie des Abhörraums abhängt. Generell werden die Bassanteile der Klänge nach unten gemischt, um der Hörgewohnheit und der Psychoakustik des Menschen gerecht zu werden.

Zur Verbreitung und um Spatial-Audio-Kompositionen zugänglich zu machen oder sie zu archivieren, wird die Binauralisierung bevorzugt. Bei Musikproduktionen erfolgt die Binauralisierung über Software, während installative Kompositionen oder Performances zur realitätsnahen Archivierung mit einem Kunstkopf – vor allem dem KU100 – aufgezeichnet werden. Um die Hörenden in den Workflow zu integrieren, wird die Komposition und Produktion aus verschiedenen Positionen und binaural abgehört. Den Befragten ist das Sweet-Spot-Problem und das Problem durch die HRTFs bewusst. Beides wird jedoch auch als Chance wahrgenommen, denn so können aus unterschiedlichen Hörpositionen verschiedene Hörerlebnisse entstehen. Ein Spatial-Audio-Stück kann somit mehrfach gehört werden und bietet dennoch die Möglichkeit, neue Klänge wahrzunehmen.

Dem Einbeziehen von Raum, der Räumlichkeit und Hall wird eine wichtige Bedeutung zugesprochen. Der Großteil verwendet Stereo- oder 3D-Hall-Plug-ins, um Raum zu erschaffen. Ein verbreitetes Mittel ist vor allem der Einsatz von Delays und das Spiel mit Frequenzen. So können durch die Anhebung hoher oder Absenkung tiefer Frequenzen der Nahbesprechungseffekt oder Weite erzeugt und somit Raumtiefe suggeriert werden. Reverbs werden von vielen Komponierenden als Sounddesign-Tool eingesetzt. Das Verwenden aufgenommener Impulsantworten wird aufgrund der hohen Qualität und der realitätsnahen Abbildung des Raumes zu einem zunehmend beliebten Ansatz.

Durch das große Panorama bei Spatial Audio wird der Einsatz von dynamischen Prozessen wie das Benutzen von Kompression in den meisten Fällen überflüssig und bewirkt somit einen größeren Dynamikumfang bei Spatial-Audio-Stücken. Die Befragten sind sich einig, dass dies dem Loudness War positiv entgegenwirkt. Beim Mischen in Spatial Audio findet der Panning-Prozess in der Regel als erstes statt. Hochfrequente Klänge werden in der Höhe platziert, tiefe unten. Dann folgt das EQing. Im Allgemeinen werden EQs eingesetzt, um Klänge hervorzuheben und die Position im Raum zu verdeutlichen, den Klang des Instruments klarer erscheinen zu lassen oder Störgeräusche zu minimieren. Das frequenzabhängige Mischen ist in jedem Fall die bevorzugte Methode in Spatial Audio. Ein Ansatz, der sich langsam durchsetzt, ist das Routen von Klängen direkt auf die Lautsprecherpositionen, da virtuelle Quellen oder Phantomschallquellen die Klangqualität reduzieren. Um einen Panning-Effekt zu realisieren, werden weitere Ansätze angewendet – wie zum Beispiel frequenzbasiertes, laufzeitbezogenes oder lautheitsbezogenes Panning.

Der Mastering-Prozess ist in vielen Fällen im Mixing-Prozess enthalten. Klassische Mastering-Chains entfallen weitgehend, stattdessen liegt der Fokus darauf, die Mischung an die jeweilige Abhörsituation anzupassen. Genaue Lautheitswerte werden nur bei Stücken für Streaming oder den Rundfunk eingehalten, um nachträglich durch den Anbieter vorgenommene Dynamikprozesse zu vermeiden. Bei Formaten wie Dolby Atmos wird außerdem darauf geachtet, dass die Mixe auf den verschiedenen Abspielgeräten die gleiche Qualität haben. Ein wichtiger Mastering-Schritt ist das Hören von binauralisierten Inhalten mit verschiedenen HRTFs. Dieser Vorgang kann Rückschlüsse darauf zulassen, ob der Mix in seiner Gesamtheit eine zufriedenstellende Klangqualität aufweist.

Ein weit verbreiteter Ansatz ist das Mischen nach Ankerpunkten, die wichtig sind, um den Hörenden Orientierung und Fokus zu ermöglichen. Vor allem wird – mit Ausnahme der experimentellen Produktionen und Installationen, bei Musikproduktionen, Hörspielen und teilweise Klanglandschaften mit einer Front, einer klassischen Bühne, gearbeitet. Der hintere und obere Bereich wird für Effekte oder als Erweiterung des Stereobildes verwendet. Bei der Aufnahme und Mischung von Livemusik ist die Ausrichtung auf eine Frontbühne gängige Praxis. Der Fokus liegt auf der Aufnahme des Raums und der Dynamik, welche Schlüssel für eine immersive Produktion ist. Diese Arbeitsweise ist allerdings im Wandel. Im Rahmen der Analyse der zweiten Hypothese wird dieses Thema ausführlich diskutiert.

Die Befragten sind sich einig, dass das Komponieren und Produzieren direkt in Spatial Audio in vielen Aspekten die höchste Qualität liefert. Besonders profitieren dabei die Lokalisierung, der räumliche Klang und die Umhüllung, die zentrale Faktoren für die Erzeugung eines optimalen Hörerlebnisses sind. Bei den Befragten besteht ein Konsens darüber, dass es eine Limitierung in der Umsetzung von Bewegung im Raum gibt. Vor allem in Bezug auf Bewegung in der vertikalen und der Anzahl der bewegten Klangelemente. Ein Objekt, das von oben durch den Raum fällt, kann nicht realitätsnah abgebildet werden – lediglich eine Andeutung ist möglich. Wichtig ist es, die Psychoakustik des Menschen zu beachten, da Spatial Audio eine Annäherung an die auditive Wahrnehmung der Realität darstellt und anstrebt diese hinsichtlich der Umhüllung zu imitieren.

Die Herangehensweisen der Befragten weisen – sowohl aus den Primär- als auch aus den Sekundärdaten – Gemeinsamkeiten auf, die format- und genreübergreifend in Produktions- und Kompositionsansätzen deutlich werden. Das frequenzabhängige Mischen, das von einem Klangkünstler gezielt als *Multiband-Spatialisierung* bezeichnet wird, findet sowohl bewusst als definierte Methode Anwendung als auch unbewusst als Bestandteil des gesamten Mixing-Prozesses. In der Regel erfolgt die Orientierung an der Psychoakustik des Menschen. So werden beispielsweise bewegte Klänge bevorzugt in der Horizontalen eingesetzt. Aus den Daten ist ersichtlich, dass durch technische Voraussetzungen Limitierungen

in der Kreativität entstehen und sich dadurch Workflows und Muster voneinander unterscheiden. Bereits die Tatsache, dass bei Ambisonics mit einer kugelförmigen Sphäre und bei Dolby Atmos mit einem quaderförmigen Setup mit festgelegter Frontrichtung gearbeitet wird, schränkt die gestalterischen Möglichkeiten grundlegend ein, was sich wiederum direkt auf die Arbeitsweise und Philosophie auswirkt.

Die Hörenden in die Workflows zu integrieren, ist ebenfalls ein wichtiger Faktor. Durch den Wechsel der Hörposition im Abhörraum, das Testen verschiedener HRTFs und die Berücksichtigung der Hörgewohnheiten, kann die Hörerfahrung optimiert werden. Es zeichnen sich wiederkehrende Muster ab, wenn es um die Arbeit mit dem Panorama geht. Für Spatial Audio wird je nach Genre eine hohe Anzahl an Klänge gebraucht, um das Panorama zu füllen. Entscheidend ist das Arbeiten mit Ankerpunkten, um den Hörenden Orientierung zu ermöglichen und Irritation zu vermeiden – außer, diese wird bewusst eingesetzt. Das Arbeiten mit einer Front bzw. einer Bühne wurde in den meisten Workflows wiedererkannt. Ein wichtiger Bestandteil aller Arbeitsweisen ist das wiederholte bewusste Hören im Produktionsprozess. Mehrfach in den Mix hineinzuhören, sich nicht ausschließlich auf visuelle Panning-Fahrten zu verlassen, sondern durch das Hören selbst wahrzunehmen, ob der Klang wirklich dort ist, wo er sein soll ist ein fundamentaler Prozess. Die Empfehlung ist klar: Das bewusste Hören im Alltag sollte gefördert werden. Wie klingt echtes Spatial Audio in alltäglichen Hörsituationen, insbesondere außerhalb des Studios?

Es scheint ein Prozess im Gange zu sein, bei dem wir uns langsam von der klassischen Frontbühne entfernen und experimentierfreudiger werden. Wenn es um kommerzielle Musikproduktionen geht, wird hier jedoch bisher vorsichtig gearbeitet. Die Herangehensweise erinnert an den Übergang von Mono auf Stereo. Das Gehör musste erst an Stereo gewöhnt werden – und genauso ist es jetzt mit Spatial Audio: Die Menschen müssen lernen, dieses neue Hören zu verstehen und einzuordnen.

6.2.2 Hypothese 2

Die Produktion und Komposition von Spatial Audio erfordert angepasste Ansätze und Workflows, um den Übergang von Stereo zu Spatial Audio hinsichtlich unserer Hörgewohnheit zu erleichtern und in der Zukunft das umfassende Potenzial von Spatial Audio nutzen zu können.

Bei der Einführung von Stereo gab es eine Übergangszeit, in der sich die Menschen an das neue Format gewöhnen mussten. Stereo war vielversprechend und bot mit den Panoramaeffekten von links nach rechts ein neues Hörerlebnis. Mit der Einführung von Stereoanlagen für Privathaushalte in den 1960er Jahren wurde Stereo zum Mainstream, die Einführung von Walkmans und das Hören über Kopfhörer verstärkten den Trend. Dennoch wurden

weiterhin viele Schallplatten in Mono produziert. Der Übergang von Mono zu Stereo dauerte Jahrzehnte [Roginska and Geluso, 2018]. Diese Parallele lässt sich heute bei der Einführung von Spatial Audio erkennen.

Der Großteil der Spatial-Audio-Werke – und somit auch der Workflows – ist auf eine Front bzw. eine Bühne ausgerichtet. Diese Herangehensweise ist deutlich an das Stereohören und das realistische Abbilden von Konzerten oder Performances angelehnt, bei denen Hörende in der Regel nach vorne auf die Bühne sehen. Es gibt verschiedene Abstufungen in der Art und Weise, wie diese Front mit Hinblick auf das benutzte Panorama umgesetzt wird. Ein Teil der Mischungen klingt lediglich nach einem einfachen LCR-Setup oder wie ein sehr breiter Stereomix. Diese Ansätze werden jedoch kritisiert, da sie das Potenzial von Spatial Audio nicht ausschöpfen und oft nicht die Qualität eines hochwertigen Stereomixes erreichen. Bei hochwertigeren Spatial-Audio-Mischungen werden räumliche Effekte durch Klänge, die nach oben und hinten in den Raum gesetzt werden, realisiert. Außerdem werden Instrumente breiter in der Horizontalen verteilt. In einer weiteren Herangehensweise dient lediglich die Stimme als Ankerpunkt im Centerlautsprecher, während der gesamte übrige Raum mit Instrumenten gefüllt wird. Schließlich gibt es als letzte Variante das freie Experimentieren – hier wird nicht länger mit einer festen Front gearbeitet, und das volle Potenzial von Spatial Audio wird vollständig genutzt.

Vor allem in der Musikproduktion und im Hörspiel ergaben die erhobenen Daten, dass das Arbeiten mit einer Bühne den Standard darstellt. Die Hörenden langsam an Spatial Audio heranzuführen, erscheint als logischer Schritt und vereinfacht den Einstieg in das räumliche Hören über Kopfhörer- und Lautsprechersysteme. Ein wiederkehrender Faktor ist die Beachtung der Psychoakustik des Menschen. Workflow-übergreifend wird frequenzabhängig gearbeitet, und je nach gewünschtem Effekt wird darauf geachtet, nicht zu viele bewegte Klangelemente einzusetzen und generell die Anzahl gleichzeitig abgespielter Klänge zu begrenzen, um die Hörenden nicht ungewollt zu irritieren und ein angenehmes Hörerlebnis zu erzeugen.

Insgesamt lässt sich ableiten, dass wir uns in einem Wandel der Hörgewohnheiten befinden, der in der Spatial-Audio-Komposition und -Produktion nicht vernachlässigt werden darf – ohne dabei jedoch das Potenzial des Formats gravierend einzuschränken.

6.2.3 Hypothese 3

Für die Produktion und Komposition von Spatial Audio lassen sich Guidelines formulieren, die sowohl Anfängern als auch erfahrenen Produzierenden Orientierung bieten, ohne die künstlerische Freiheit einzuschränken.

Grundlegend wird sich für eine Standardisierung der Technik rund um Spatial Audio ausgesprochen, vor allem wenn es um die Kompatibilität unterschiedlicher Formate geht. Die Einführung von Guidelines zur Standardisierung wird hingegen kritischer betrachtet, da insbesondere der experimentelle und forschende Ansatz als essenziell für Spatial Audio angesehen wird. Gleichzeitig gibt es jedoch Aspekte, die in Spatial Audio nicht oder nur eingeschränkt umsetzbar sind. Dies führt dazu, dass viele aktuelle Spatial-Audio-Mischungen, insbesondere in Dolby Atmos, das vorhandene Potenzial nicht ausschöpfen. Soll sich die Spatial-Audio-Branche langfristig etablieren, braucht es bessere Hörerlebnisse durch qualitativ hochwertige Produktionen.

Auf Grundlage der Interviews und der Erfahrung durch die Spatial Audio Listening Sessions wurde eine Lösung gefunden: ein Guide für optimierte Spatial-Audio-Produktionen, der an den aktuellen Stand angepasst ist, dabei aber die künstlerische Freiheit nicht einschränkt. Stattdessen regt er gezielt zum Testen an und dient als Orientierungshilfe. Einige Tipps sind allgemeingültig, andere beziehen sich auf spezifische Formate.

Formatübergreifende Guidelines

Der optimale Ansatz für die Erstellung von Spatial-Audio-Stücken ist das Komponieren und Produzieren direkt in dafür vorgesehenen Spatial-Audio-Formaten. Im Idealfall steht bereits von Anfang an fest, dass es sich um eine Spatial-Audio-Mischung handelt, sodass vom ersten Entwurf bis zum finalen Mix, dieser Ansatz verfolgt werden kann. Hierdurch können alle notwendigen Klänge gezielt aufgenommen und optimiert werden.

Das Erweitern oder Neudenken von Stereomischungen für Spatial Audio kann zwar funktionieren, hängt jedoch vom vorliegenden Material ab, das die künstlerische Freiheit einschränken kann. Werden für ein Spatial-Audio-Stück eigene Aufnahmen erstellt, bietet es sich an, auf folgende Erkenntnisse zu achten:

Mikrofonierung

Das Aufnehmen eines Stückes oder einzelner Klänge erfordert die Berücksichtigung von Direktschall, frühen Reflexionen und dem Nachhall. Diese variieren durch die Eigenschaften des Raumes und können mit verschiedenen Ansätzen eingefangen werden. Die Ausgangslage ist dabei entscheidend: Soll mit digitalen Spatial-Audio-Hall-plugin-ins oder mit aufgenommenen Impulsantworten gearbeitet werden? Bei Letzterem reicht die Aufnahme des Direktschalls mithilfe von Einzelmikrofonierungen aus. Der Grad der aufgenommenen Räumlichkeit kann durch den Einsatz von Stereopaaren oder das Aufnehmen der Wandreflektionen mit nah an der Wand positionierten Mikrofonen deutlich erhöht werden. Es bietet sich an, Bewegungen bereits bei der Aufnahme einzufangen und auf klassisches

Panning (siehe Abschnitt *Positionierung und Bewegung von Soundquellen*) wenn möglich zu verzichten. Dies kann allerdings die Flexibilität in der Nachbearbeitung einschränken und sollte entsprechend bedacht werden.

Eine Hauptmikrofonierung sollte je nach Bedarf und gewähltem Format erfolgen. Mikrofonarrays wie PCMA-3D oder der 2L-Cube bieten sich für klare und richtungstreue Aufnahmen an, während Ambisonics-Mikrofonierungen mehr Räumlichkeit, aber weniger Lokalisationsschärfe liefern. Es sollte bedacht werden, dass sich Aufnahmen mit Ambisonics-Mikrofonen nur bedingt für Atmos-Mischungen eignen, da sie nicht dem objektbasierten Ansatz von Dolby Atmos entsprechen und eine nachträgliche Dekodierung sowie Umwandlung erforderlich machen, welche in der Regel zu Qualitätsverlusten führt.

Ist der Platz und die notwendige Technik vorhanden, bietet es sich an, Tiefenstaffelung durch das Positionieren verschiedener Mikrofone im Raum zur Klangquelle zu realisieren. Die Aufnahme mit einem Kunstkopf eignet sich ideal zur Archivierung und Verbreitung von Spatial-Audio-Stücken in binauraler Form. Durch die korrekte Mikrofonierung lassen sich viele der gewünschten räumlichen Effekte bereits in der Aufnahme realisieren.

Positionierung und Bewegung von Soundquellen

Grundsätzlich sollte man sich bei der Produktion und Komposition von auditiven Werken der Psychoakustik und den menschlichen Hörgewohnheiten bewusst sein. Besonders fällt dies bei Spatial Audio ins Gewicht, da das Hörerlebnis eine räumliche Umhüllung ermöglicht, die dem realen Hören deutlich näherkommt. Es ist einfach, die Hörenden durch Klänge von hinten zu irritieren oder durch zu viele Klänge zu überfordern. Je nach Stück und Genre sollte dies bedacht und entsprechend vermieden werden.

Die Wahrnehmung von Klang ist in hohem Maße frequenzabhängig, sodass das frequenzabhängige Mischen von entscheidender Bedeutung ist, um einen Mix zu erstellen, der an die Hörgewohnheiten des Menschen angepasst ist. Darüber hinaus sollte bedacht werden, dass softwarebasiertes Panning zu Qualitätsverlusten führen kann. In vielen Fällen ist es sinnvoll, auf alternative Panningmethoden zurückzugreifen, zum Beispiel auf laufzeit- oder lautheitsbasiertes Panning oder frequenzabhängige Ansätze.

Es gibt kein festes Limit in Bezug auf die Anzahl der Klänge, die eingesetzt oder bewegt werden können. Lediglich Formatvorgaben, wie zum Beispiel das 128-Objekte-Limit bei Dolby Atmos, oder technische Faktoren wie die Rechenleistung können einschränkend wirken. Generell ist jedoch eine höhere Anzahl an Klängen als bei Stereoproduktionen notwendig, was bereits bei der Aufnahme berücksichtigt werden muss. Letztlich muss immer gegengehört werden, ob der Klang, der positioniert und optional bewegt wird,

wirklich dort wahrgenommen werden kann. Hierbei wird empfohlen, Stücke sowohl über Kopfhörer (binaural) als auch auf einer Lautsprecheranlage abzuhören. Bewährte Ansätze, wenn es um Klangposition und Bewegung geht, sind folgende:

- Lead-Vocals in den Center positionieren, Halleffekte auf Raumlautsprecher (Seiten/hinten/oben) verteilen.
- Weitere Ankerpunkte wie Bass und Schlagzeug auf der Front verteilen, aber frequenzabhängig mischen.
- Breite Klangflächen sind mit langsamer Bewegung deutlich ortbar. Bewegung kann zum Spannungsaufbau eingesetzt werden.
- Bei Hörspielen empfiehlt es sich, mit symbolischen oder überzeichneten Klangelementen zu arbeiten, anstatt die akustische Realität 1:1 abzubilden, da dies oft nicht überzeugend wirkt oder nicht funktioniert.

Generell bietet es sich an, Ankerpunkte zur Orientierung zu setzen. Dies kann vom Ansatz der Bühne in der Front über das Setzen der Lead-Vocals im Center bis hin zu bühnenunabhängigen Ankerpunkten realisiert werden. Hier sollte zu Beginn die Frage gestellt werden: Für welches Publikum und welchen Raum wird das Stück produziert bzw. komponiert? Wird es ein experimentelles Stück für eine Installation oder ein Dolby-Atmos-Mix, der auf Apple-Music veröffentlicht werden soll und mit Stereotracks konkurrieren wird?

Hörerintegration

Die Hörenden mit in den Workflow zu integrieren, kann das Hörerlebnis signifikant steigern. Das Abhören des Mixes aus verschiedenen Positionen oder mit verschiedenen HRTFs steigert das Verständnis über den Mix, kann aber zu längeren Arbeitsprozessen führen. Denn die Abhörbedingungen sind durch das HRTF- und das Sweet-Spot-Problem oder durch das Abspielen der Stücke in verschiedenen Räumen mit unterschiedlichen Systemen nie identisch. Spannend bleibt die Integration des Hörenden bei Hörspielen. Hier kann je nach gewähltem Format der Hörende in das Geschehen gebracht oder als außenstehende Person positioniert werden. Umfassende Möglichkeiten bieten installative Spatial-Audio-Produktionen. Bei diesen können mehrere Erlebnisse realisiert werden, indem nach bestimmten Bereichen im Raum gemischt wird.

Raum und Nachhall

Das Arbeiten mit Raum und der Räumlichkeit mithilfe von beispielsweise Halleffekten sollte mit Vorsicht behandelt werden. Es gibt viele Methoden, um ein Gefühl von Räumlichkeit zu erzeugen. Spatial-Audio-Halleffekte können die Grundlage bilden oder

als Sounddesign-Tool gesehen werden. Durch den Einsatz von Effekten wie dem Doppellereffekt, Delays oder Volumeshifts lassen sich komplexe Räume konstruieren. Diese Herangehensweise ist essenziell für Hip-Hop und elektronische Stücke, die auf den Punch angewiesen sind, um zu funktionieren. Der Ansatz, den Raum in aufgenommenen Raum und konstruierten Raum zu unterscheiden, kann ein hilfreicher Aspekt zum Verständnis von Raum und der empfundenen Räumlichkeit sein.

Mixing und Mastering

Der Mixing-Prozess besteht bei Spatial Audio zum Großteil aus dem Panning-Prozess der Klänge im Panorama. Durch das 360°-Panorama sind umfangreiche Dynamikprozesse nicht länger notwendig. Kompression kann und sollte zur Beibehaltung der feinen Klangnuancen nur geringfügig eingesetzt werden. Der Mastering-Prozess bei Spatial-Audio-Produktionen ist je nach Format nur bedingt existent. Mastering kann als wichtige Qualitätskontrolle bei der Veröffentlichung binauraler Inhalte oder Dolby-Atmos-Produktionen dienen oder ein rein technischer Arbeitsschritt sein. Die Workflows sind derzeit limitiert, da bisher keine Software verfügbar ist, um klassische Mastering-Chains zu etablieren. Entscheidend ist letztlich, einen passenden Dynamikumfang zu erreichen. Wichtig zu betonen ist, dass – wenn das Stück auf Streamingplattformen oder im Rundfunk veröffentlicht werden soll – bestimmte Lautheitsvorgaben erreicht und eingehalten werden müssen, um eine unerwünschte Komprimierung zu vermeiden.

Allgemeine Ergänzungen

Die Planung bei der Komposition und Produktion von Spatial-Audio-Stücken ist essenziell. Das bewusste Hören in den Alltag zu integrieren, ist für das Verständnis von Klang im Raum von großer Bedeutung und kann die Mixing-Fähigkeiten positiv beeinflussen. Außerdem wird empfohlen, den Mix wiederholt abzuhören – vor allem, wenn es um bewegte Klänge geht. Wie bei Stereoproduktionen gilt: Nach visuellen Cues platzierte oder bewegte Klänge werden nicht zwangsläufig an der intendierten Position wahrgenommen.

Die wichtigste Empfehlung ist das eigenständige Experimentieren mit Spatial Audio. Dies umfasst auch die Überprüfung dieser Guidelines. Den größten Lernerfolg erzielen wir in der Audiobranche selten durch rein theoretische Überlegungen, sondern durch praktische Erfahrungen. Inwiefern diese Guidelines die Kreativität einschränken oder nicht, wird in der folgenden Diskussion aufgegriffen.

7 Diskussion und Ausblick

Die Analyse bestätigt, dass die zuvor aufgestellten Hypothesen zutreffen. Es zeigen sich – unabhängig von verwendetem Genre, Format oder beruflichem Hintergrund – wiederkehrende Muster in Produktion und Komposition, sowohl in kreativen als auch technischen Herangehensweisen. Neben diesen gemeinsamen Mustern treten jedoch auch deutliche Unterschiede auf, die durch die jeweiligen Formate und die Technik selbst vorgegeben werden. Besonders auffällig ist dies bei der Arbeit mit einer Frontbühne in Dolby Atmos sowie der Art der Spatialisierung, die beispielsweise bei Ambisonics das Arbeiten in einer kugelförmigen Anordnung vorgibt.

Die Datenlage legt nahe, dass wir uns derzeit in einer Übergangsphase der Hörgewohnheiten befinden – von Stereo und Surround zu vollständig dreidimensionalen Spatial-Audio-Werken. Parallelen zu historischen Übergangsphasen, wie etwa dem Wandel von Mono zu Stereo, sind deutlich erkennbar.

Wenn es um die Frage geht, ob sich Guidelines formulieren lassen, die unterstützend wirken, ohne kreative Prozesse einzuschränken, kann keine pauschale Antwort gegeben werden. Solche Guidelines müssen in der Praxis getestet werden, um die Hypothese letztlich zu bestätigen. Dennoch ließen sich aus den analysierten Ansätzen Schemata ableiten, auf deren Grundlage ein Guideline-Entwurf entstanden ist – mit dem Ziel, folgende Fragen zu beantworten: Was lässt sich in Spatial Audio konkret umsetzen, und welche Faktoren sind entscheidend, um ein positives Hörerlebnis zu schaffen? Angesichts der aktuellen Situation wird deutlich zu wenig über die Ästhetik von Spatial Audio gesprochen. Ein solcher Guide könnte der Produktion und Komposition zukünftiger Spatial-Audio-Mischungen geringer Qualität entgegenwirken oder zumindest eine Diskussion darüber anstoßen.

Aus den Ergebnissen geht hervor, dass die Forschung über Ästhetik und Philosophie in Bezug auf Spatial-Audio-Herangehensweisen auf ein großes Potenzial hindeutet. Sie ist nicht annähernd abgeschlossen, sondern wird sich über einen längeren Zeitraum hinweg weiterentwickeln. Während Stereo weitgehend erforscht ist, bleibt offen, ob dies auch jemals in gleichem Umfang für Spatial Audio gelten wird – nicht zuletzt, weil auf technischer Ebene kontinuierlich neue Entwicklungen in den Bereichen Hardware, Codecs und Spatialisierungsmethoden hinzukommen.

Die Interviews wurden nicht nur zur Überprüfung der Thesen geführt, sondern auch genutzt, um darüber hinausgehende Fragestellungen und Themenkomplexe zu erschließen. Die freie Interviewführung und das kontinuierliche Anpassen der Fragen und Codes erwiesen sich als angemessene Methode, um ein wenig erforschtes Thema im kreativen und

technischen Bereich zu untersuchen. Eine rein objektive Perspektive aus forschungspraktischer Sicht einzunehmen, war jedoch nicht vollständig möglich. Da der Autor selbst in Spatial Audio komponiert und produziert, kann eine Tendenz zu positiven Ergebnissen nicht ausgeschlossen werden. Entscheidend ist, dass künftig auch andere Forschende unabhängig Daten erheben, um Vergleichswerte zu generieren und Erkenntnisse bestätigen oder widerlegen zu können.

Die erhobene Anzahl qualitativer Experteninterviews, ergänzt durch fundierte Sekundärdaten, ermöglichen zwar aussagekräftige Schlüsse, jedoch bedarf es zur wissenschaftlich fundierten Validierung größerer Studienreihen und einer breit aufgestellten Metaanalyse. Des Weiteren wäre eine Ausweitung der weiteren Fachkundigen, ein überarbeiteter Fragebogen und eine Durchführung der Interviews durch verschiedene Forschende wünschenswert. Dies könnte mögliche Biases reduzieren und gezieltere Aussagen ermöglichen. Daraus ergibt sich die Chance, auf den gewonnenen Ergebnissen aufzubauen, um einen fundierten, praxisorientierten Guide zu erstellen und diesen in konkreten Anwendungsfeldern zu testen, ohne dabei die kreativen und forschenden Prozesse einzuschränken. Angesichts der dynamischen Weiterentwicklung von Technologien rund um Spatial Audio bleibt die Arbeit in diesem Bereich ein fortlaufender Prozess, der kontinuierlich begleitet, diskutiert und analysiert werden sollte.

Ein Faktor, der die Tiefe und Länge der Interviews beeinflusst hat, war die Erfahrung der Befragten. Personen, die bereits in mehreren Formaten gearbeitet und eine größere Zahl an Projekten realisiert haben, konnten die Fragen meist deutlich ausführlicher beantworten. Die Herausforderung für den Interviewenden bestand darin, die Beantwortung der Schlüsselfragen zu ermöglichen und bei Zeitmangel auf weniger relevante Themenbereiche zu verzichten.

Die intensive Auseinandersetzung mit Spatial-Audio-Workflows und der Ästhetik der Werke warf zudem viele weitere Fragen auf. Aufgrund des Umfangs dieser Arbeit konnten diese nicht weiter untersucht werden. Die erhobenen Daten bieten jedoch weiteres Analysepotenzial, das in zukünftigen Arbeiten erschlossen werden kann. Diese Arbeit wurde bewusst offen gestaltet, was eine flexible Bearbeitung ermöglichte, jedoch auch Einschränkungen mit sich brachte, wenn es um das Formulieren weitreichender, generalisierbarer Aussagen geht. Trotz des Umfangs wirkt sie für den Autor noch oberflächlich und lässt viel Raum zur vertiefenden Bearbeitung ästhetischer und philosophischer Fragen.

8 Fazit

In dieser Arbeit wurden Kompositions- und Produktionsansätze in Spatial Audio aus ästhetischer und philosophischer Sicht betrachtet. Grund hierfür ist die vermehrte Arbeit mit Formaten wie Ambisonics und Higher Order Ambisonics, eine steigende Anzahl immersiver Installationen sowie Dolby Atmos, das in den letzten Jahren den Schritt in die kommerzielle Nutzung von 3D-Audio vorgenommen hat. Aus eigener Erfahrung mit der Thematik wurden drei Hypothesen aufgestellt, um gemeinsame Muster in der Arbeit mit Spatial Audio bei den Audioschaffenden zu erkennen, die Hörgewohnheiten des Menschen in Bezug auf Spatial Audio zu beleuchten und einen vorläufigen Guide zu entwickeln, der es Audioschaffenden ermöglichen soll, qualitativ hochwertige Spatial-Audio-Werke zu erarbeiten, ohne in den kreativen Prozess einzugreifen oder bereits im Vorfeld einzuschränken.

Für die Durchführung der Arbeit wurde eine erweiterte Form des fokussierten Interviews entwickelt. Diese Form der qualitativen Interviewführung ist eine Kombination aus dem Experten-, fokussierten und problemzentrierten Interview und bietet eine hohe Flexibilität sowie die Chance zur Reflexivität. Dies ermöglichte einen offenen Diskurs, der die Grundlage bildete, um aussagekräftige Daten zu sammeln. Zur Erweiterung des Datensatzes wurden Sekundärquellen in Form von Podcasts, wissenschaftlichen Papern und weiteren Interviews herangezogen, sofern diese mindestens die Kernfragen beantworteten. Dies stellt einen klassischen Mixed-Methods-Ansatz dar, der selbst erhobene und vorhandene Datenquellen miteinander verbindet. Die Daten wurden schließlich nach der strukturierenden qualitativen Inhaltsanalyse nach Kuckartz ausgewertet. Hierzu wurde die Software MAXQDA eingesetzt, die für Analysen dieser Art ausgelegt ist. Besonders die Transkription und Kodierung der Texte wurde dadurch erheblich erleichtert.

Durch die gewählte Methodik und die Analyse anhand der formulierten Thesen konnten umfangreiche Erkenntnisse über die Workflows in Bezug auf Ästhetik und Philosophie hinter Spatial Audio gewonnen werden. Im Umgang mit Klang und Raum lässt sich eine frequenzbasierte Arbeitsweise erkennen, die sich anhand der Psychoakustik des Menschen erklären lässt. Ebenso ist ein Übergang vom stereobasierten Arbeiten im Sinne einer klassischen Frontbühne zu freieren Verteilungen von Klangelementen im gesamten Raum erkennbar. Durch die ausführlichen Interviews lassen sich Schemata erkennen, die es ermöglichten, den vorläufigen Guide zu formulieren, der helfen kann, qualitativ hochwertigere Spatial-Audio-Werke zu schaffen. Die Hypothesen konnten grundsätzlich bestätigt werden.

Die Ergebnisse zeichnen ein differenziertes Bild der Arbeitsweise mit Spatial Audio, sollten jedoch mit Bedacht betrachtet werden. Der Umfang umfasste lediglich 25 Primär- und Sekundärquellen die analysiert wurden und es besteht ein gewisser Bias auf Seiten des Verfassers durch der eigenen Tätigkeit im Bereich Spatial Audio. Aufgrund der geringen Datenlage zu Technologien bzw. Formaten wie der Wellenfeldsynthese konnten hierzu keine fundierten Aussagen über Ansätze Workflows getroffen werden.

Diese Arbeit kann als Grundlage oder als ein weiterer Schritt in der Erforschung der Ästhetik und der Arbeitsweisen in Spatial Audio gesehen werden. Sie kann einen Beitrag zur künstlerischen Entwicklung und Diskussion leisten. Für weiterführende Aussagen sind zusätzliche Arbeiten notwendig, die sich intensiver mit diesem Thema auseinandersetzen beispielsweise im Rahmen einer zukünftigen Metastudie. Darüber hinaus sollten die gewonnenen Erkenntnisse in der Praxis überprüft und evaluiert werden.

Diese Arbeit wird sowohl in theoretischer als auch in praktischer Hinsicht fortgeführt. In einer folgenden praktischen Arbeit werden einige der analysierten Arbeitsweisen untersucht und in eigenen Kompositionen zusammengeführt. Die Arbeit mit Spatial Audio ist ein dynamischer Entwicklungsprozess, der im Rahmen der Spatial Audio Listening Sessions weiterhin begleitet, beobachtet und analysiert wird.

A Digitaler Anhang

Die codierten Transkripte der Interviews befinden sich als PDF-Dateien im digitalen Anhang. Aufgrund des Datenschutzes sind sie mit Interview 1-17 und den Bezeichnungen Person A bis Q gekennzeichnet. Außerdem sind der Fragebogen in deutscher und englischer Sprache sowie die Liste der Befragten für die Prüfenden beigefügt.

Inhalt des digitalen Anhangs:

- Fragebogen (DE/EN)
- Interviews 1–17 (anonymisiert)
- Liste der Befragten (Nur für die Prüfenden)

Literatur

- [Bagenal, 1951] Bagenal, H. (1951). Musical taste and concert hall design. *Proceedings of the Royal Musical Association*, 78(1):11–29.
- [Berkhout et al., 1993] Berkhout, A. J., de Vries, D., and Vogel, P. (1993). Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5):2764–2778.
- [Berliner Festspiele, 2024] Berliner Festspiele (2024). Maerzmusik 2024 – festival für zeitfragen. Zugriff am 29. Januar 2025.
- [Blauert, 1997] Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press.
- [Blessner and Salter, 2007] Blessner, B. and Salter, L.-R. (2007). *Spaces Speak, Are You Listening?* The MIT Press.
- [Blumlein, 1931] Blumlein, A. D. (1931). Improvements in and relating to sound-transmission, sound-recording, and sound reproducing systems.
- [Bregman, 1990] Bregman, A. S. (1990). *Auditory Scene Analysis*. MIT Press, Cambridge, MA.
- [Broderick, 2012] Broderick, A. E. (2012). Grand messe des morts: Hector berlioz’s romantic interpretation of the roman catholic requiem tradition. Master’s thesis, Bowling Green State University.
- [Chourmouziadou and Kang, 2008] Chourmouziadou, K. and Kang, J. (2008). Acoustic evolution of ancient greek and roman theatres. *Applied Acoustics*, 69(6):514–529.
- [Collins, 2008] Collins, P. (2008). Theatrophone: The 19th-century ipod. *New Scientist*, pages 44–45.
- [Deutschlandfunk Kultur, 2025] Deutschlandfunk Kultur (2025). Musikarchäologie: Die melodien der steinzeit. Zugriff am 12. Januar 2025.
- [Dickreiter, 1997] Dickreiter, M. (1997). *Handbuch der Tonstudioteknik, Band 1 (6. verbesserte Auflage)*. De Gruyter Saur.
- [Emmerson, 2007] Emmerson, S. (2007). *Living Electronic Music*. Ashgate Publishing, Aldershot, Hants, England; Burlington, VT.
- [Frisecke, 2014] Frisecke, A. (2014). *Die Audioenzyklopädie, Ein Nachschlagewerk für Tontechniker (2. Auflage)*. De Gruyter Saur.

- [Gerzon, 1973] Gerzon, M. A. (1973). Periphony: With-height sound reproduction. *Journal of the Audio Engineering Society*, 21(1):2–10.
- [Gullö et al., 2024] Gullö, J.-O., Hepworth-Sawyer, R., Paterson, J., Toulson, R., and Marrington, M., editors (2024). *Innovation in Music: Cultures and Contexts*. Perspectives on Music Production. Routledge.
- [Görne, 2008] Görne, T. (2008). *Tontechnik: Hören, Schallwandler, Impulsantwort und Faltung, digitale Signale, Mehrkanaltechnik, tontechnische Praxis*. Hanser Fachbuchverlag, München.
- [Görne, 2015] Görne, T. (2015). *Tontechnik. Hören, Schallwandler, Impulsantwort und Faltung, digitale Signale, Mehrkanaltechnik, tontechnische Praxis; mit 33 Tabellen (Medien, 4., aktualisierte Aufl.)*. Hanser.
- [Hess, 2012] Hess, W. (2012). Head-tracking techniques for virtual acoustics applications. In *Proceedings of the 133rd Convention of the Audio Engineering Society*, number 8782, San Francisco, CA, USA. Audio Engineering Society.
- [Huber and Runstein, 2017] Huber, D. M. and Runstein, R. E. (2017). *Modern Recording Techniques*. Routledge, New York, 9 edition.
- [Kendall, 2014] Kendall, G. S. (2014). Spatial perception and cognition in multichannel audio for electroacousticmusic.
- [Kirchmair, 2022] Kirchmair, R. (2022). *Qualitative Forschungsmethoden: Anwendungsorientiert: vom Insider aus der Marktforschung lernen*. Textbook, Berlin, Heidelberg. © 2022.
- [Kuckartz, 2012] Kuckartz, U. (2012). Qualitative inhaltsanalyse: Methoden, praxis, computerunterstützung. In *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Beltz Juventa, Weinheim und Basel, 2 edition.
- [Lossius et al., 2009] Lossius, T., Baltazar, P., and de la Hogue, T. (2009). Dbap—distance-based amplitude panning. In *Proceedings of the International Computer Music Conference (ICMC)*.
- [Mikos, 2017] Mikos, L. (2017). *QualitativeMedienforschung*. UVK Verlagsgesellschaft mbH.
- [Paris Transatlantic Magazine, 2007] Paris Transatlantic Magazine (2007). June 2007 issue. Zugriff am 12. April 2025.
- [Pfanzagl-Cardone, 2023] Pfanzagl-Cardone, E. (2023). *The Art and Science of 3D Audio Recording*. Springer Cham. Buch.

- [Roads, 2015] Roads, C. (2015). *Composing Electronic Music: A New Aesthetic*. Oxford University Press, New York, NY.
- [Roginska and Geluso, 2018] Roginska, A. and Geluso, P., editors (2018). *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Routledge, New York, NY.
- [Rumsey, 2001] Rumsey, F. (2001). *Spatial Audio*. Focal Press, Oxford.
- [Rädiker and Kuckartz, 2019] Rädiker, S. and Kuckartz, U. (2019). *Analyse qualitativer Daten mit MAXQDA*. Springer VS, Wiesbaden.
- [Schreier, 2023] Schreier, M. (2023). *Qualitative Erhebungsmethoden*, pages 247–279. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Sinewave Lab, 2024] Sinewave Lab (2024). Next room – post production reverb for film. Zugriff am 11. April 2025.
- [Slotki, 1936] Slotki, I. W. (1936). Antiphony in ancient hebrew poetry. *The Jewish Quarterly Review*, 26(3):199–219.
- [Smalley, 1997] Smalley, D. (1997). *Organised Sound: Spectromorphology: explaining sound-shapes*. Cambridge University Press.
- [Sudahl, 2023] Sudahl, L. (2023). Sound in space: Die wahrnehmung von komplexen klängen. Experimentelle Forschungsarbeit, internes Paper der HAW Hamburg: Sound/Vision, Master of Arts.
- [Troschka et al., 2021] Troschka, S., Stephan, M., Wong, Y.-F., and Görne, T. (2021). Listening experiment on the perception of spatial sound configurations. In *Proceedings of the International Conference on Spatial Audio (ICSA)*.
- [Weinzierl, 2008] Weinzierl, S., editor (2008). *Handbuch der Audiotechnik*. Springer, Berlin, Heidelberg.
- [Ziemer, 2023] Ziemer, T. (2023). *Psychoakustische Schallfeldsynthese für Musik: Raum schaffen für Komposition, Aufführung, Akustik und Wahrnehmung*. Springer.
- [Zotter and Frank, 2019] Zotter, F. and Frank, M. (2019). *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, volume 19 of *Springer Topics in Signal Processing*. Springer.
- [Zwicker and Fastl, 1999] Zwicker, E. and Fastl, H. (1999). *Psychoacoustics: Facts and Models*. Springer, Berlin, Heidelberg, 2nd edition. Hörschwelle und Schmerzgrenze.

Erklärung zur selbstständigen Bearbeitung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit eigenständig und ohne unerlaubte Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel verwendet und alle wörtlich oder sinngemäß übernommenen Inhalte als solche kenntlich gemacht. Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Hamburg, den _____

(Ort, Datum)

(Unterschrift)