

Entwicklung und Evaluation eines Prototyps zur Integration von 3D-Audio in einen virtuellen Wavetable Synthesizer

Masterarbeit

zur Erlangung des akademischen Grades

Master of Engineering (M.Eng.)

im Studiengang Audiovisuelle Medien

an der Hochschule der Medien Stuttgart

Vorgelegt von: Jonas Kieser
Matr.-Nr.: 37428
E-Mail: jonaskieser@outlook.com
Datum: 13.08.2021
Erstprüfer: Prof. Dr. Frank Melchior
Hochschule der Medien, Stuttgart
Zweitprüfer: Prof. Oliver Curdt
Hochschule der Medien, Stuttgart
Betreuer: Julian Stuchlik (M.Eng.)
Klangerfinder, Stuttgart
Daniel Rudrich (M.Sc.)
atmoky, Graz

Ehrenwörtliche Erklärung

Hiermit versichere ich, Jonas Kieser, ehrenwörtlich, dass ich die vorliegende Masterarbeit mit dem Titel: „Entwicklung und Evaluation eines Prototyps zur Integration von 3D-Audio in einen virtuellen Wavetable Synthesizer“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§23 Abs.2 Master-SPO (3 Semester) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Jonas Kieser

Stuttgart, den 13.08.2021

Kurzfassung

3D-Audio-Formate haben Einzug in die Produktion populärer Musik erhalten. In aktuellen Workflows ist die Spatialisierung für 3D-Audio zumeist Teil des Postproduktionsprozesses. Um die gestalterischen Potentiale von 3D-Audio voll auszuschöpfen, ist möglicherweise ein Umdenken in bestehenden Workflows nötig. Grundlegender Vorschlag dieser Arbeit ist es, die Gestaltung der räumlichen Attribute als Teil der Komposition und des Sounddesigns zu betrachten und diese zum Komponisten zu verlagern. Ein Ansatz für diese Verlagerung kann die Integration von 3D-Audio in virtuelle Instrumente zur Klangerzeugung darstellen – wofür aktuell kaum Beispiele bestehen. Im Rahmen dieser Arbeit wird ein Prototyp implementiert, der mehrere Vorschläge für die Integration von 3D-Audio in einen virtuellen Wavetable Synthesizer macht. Der Synthesizer erzeugt objektbasierten, formatagnostischen 3D-Audio Output mit einer Vielzahl an Audioobjekten. Mit polyphonen Voices und unabhängigen Wavetable Oszillatoren können zwei verschiedene Arten von Klangkomponenten auf die Audioobjekte verteilt werden. Zudem besteht die Möglichkeit diese über Hüllkurven und LFOs sowie MIDI-Befehle dynamisch zu panoramieren. Darüber hinaus wird eine makroskopische Steuerung über einen Ankerpunkt vorgeschlagen, der das einfache Kontrastieren der räumlichen Ausdehnung ermöglicht.

In der Evaluation des Prototyps mit Experten aus den Bereichen Komposition & Sounddesign sowie der Postproduktion wurde die Verlagerung der Spatialisierung in Richtung der Komponisten begrüßt. Die einzelnen Vorschläge der Implementierung wurden überwiegend positiv bewertet. Der Ansatz der räumlichen Verteilung polyphoner Voices sowie die Steuerung der Bewegung mehrere Audioobjekte über einen Ankerpunkt erhielten großen Zuspruch. Zudem konnten in der Evaluation wichtige Entwicklungspotentiale, wie beispielsweise die Reduktion auf wirkungsorientierte Parameter, erkannt und weiterführende Ideen generiert werden. Es wurde deutlich, dass künstlerische Forschung gepaart mit Erkenntnissen der Psychoakustik und innovativen Designs virtueller Instrumente notwendig ist, um die Entwicklungen an der Schnittstelle von Klangerzeugung und 3D-Audio weiter voranzutreiben.

Schlagwörter: 3D-Audio, Wavetable Synthese, Max for Live, objektbasiertes Audio, Musikproduktion, Sounddesign, Spatialisierung

Abstract

Spatial audio formats have made their way into popular music production. In current workflows, spatialization is mostly part of the post-production process. To unlock the full creative potential of spatial audio, a rethinking of existing workflows may be necessary. The basic proposal of this thesis is to consider the design of spatial attributes as part of the composition and sound design stage. One approach for the shift to the composers can be the integration of spatial audio into virtual instruments for sound generation - for which there are currently hardly any examples. In this work, a prototype is implemented that makes several proposals for the integration of spatial audio into a virtual wavetable synthesizer. The synthesizer generates object-based, format-agnostic spatial audio output with a variety of audio objects. The sound components polyphonic voices and independent wavetable oscillators can be distributed to the audio objects. In addition, there is the possibility to move them via envelopes and LFOs as well as MIDI messages. Furthermore, macroscopic control via an anchor point is proposed, which allows easy contrasting of the spatial extent.

The evaluation of the prototype with experts from the fields of composition & sound design as well as post-production resulted in approval for shifting the spatialization towards the composer. The individual proposals of the implementation were predominantly evaluated positively. Especially the approaches of the spatial distribution of polyphonic voices as well as the control of the movement of several audio objects via an anchor point received approval. In addition, important development potentials, such as the reduction to effect-oriented parameters, could be identified in the evaluation and further ideas could be generated. It became clear that artistic research together with insights from psychoacoustics and innovative designs of virtual instruments are necessary to advance further developments in the field of sound synthesis and spatial audio.

Keywords: Spatial audio, wavetable synthesis, Max for Live, object-based audio, music production, sounddesign, spatialization

Danksagung

An dieser Stelle möchte ich all jenen danken, die durch ihre fachliche und persönliche Unterstützung zum Gelingen dieser Arbeit beigetragen haben.

Ein besonderer Dank gilt Professor Dr. Frank Melchior für die hervorragende Betreuung meiner Arbeit, vor allem die umfangreichen Fachgespräche waren mir eine große Hilfe und haben mich zu der einen oder anderen Extrameile motiviert. Ihr Engagement hinsichtlich der thematischen Unterstützung über den Rahmen dieser Arbeit hinaus weiß ich zu schätzen.

Professor Oliver Curdt danke ich herzlich für die Begleitung meines gesamten Studiums. Von ihrem Wissen, ihrem Netzwerk und ihrer Unterstützung meiner Projekte habe ich all die Jahre sehr profitiert.

Weiterer Dank gilt meinen Praxisbetreuern Daniel Rudrich und Julian Stuchlik für die fachliche Beratung, das große Interesse an meinem Thema und den inspirierenden Austausch.

Ich danke Jan Langhammer, Andy Hoppe, Dustin Zorn, Martin Rieger, Michael Romanov, Johannes Regnier, Steffen Günther, Ulli Scuda und Thomas Koch für ihre Zeit und die Bereitschaft als Experten zur Verfügung zu stehen und ihr umfangreiches Wissen mit mir zu teilen.

Ein herzlicher Dank geht an meine Kommilitonen für eine großartige Studienzeit, aus der spannende Projekte und gute Freundschaften hervorgegangen sind. Im Speziellen danke ich Marcel Remy, Daniela Rieger und Jona Eisele für das Korrekturlesen dieser Arbeit.

Ein Dank geht zudem an meine Eltern und Geschwister für ihre Unterstützung in der Phase der Masterarbeit im Speziellen und der Studienzeit im Allgemeinen.

Milena danke ich herzlich für ihre umfassende Unterstützung, die wissenschaftliche Beratung und ihre aufmunternden Worte.

Inhaltsverzeichnis

Abbildungsverzeichnis	VII
Formelverzeichnis.....	VIII
Tabellenverzeichnis.....	VIII
Abkürzungsverzeichnis.....	IX
1 Einleitung.....	1
2 Definitionen und Eingrenzungen	4
2.1 Wavetable Synthese	4
2.2 3D-Audio	9
3 Forschungsstand	15
3.1 Raum als kompositorische Dimension im musikgeschichtlichen Kontext.....	15
3.2 Projekte an der Schnittstelle von Klangsynthese und Spatialisierung.....	20
3.3 Forschungsfragen und methodisches Vorgehen.....	23
4 Prototypische Implementierung	25
4.1 Systemintegration.....	25
4.1.1 MIDI-Befehle	27
4.1.2 Audioobjekte	28
4.1.3 3D-Audio Rendering.....	29
4.2 Spatial Wavetable Synthesizer	32
4.2.1 Modulatoren.....	33
4.2.2 Klangsynthese.....	36
4.2.3 Richtungsberechnung	42
4.2.4 Benutzeroberfläche.....	48
5 Experten-Evaluation.....	52
5.1 Methodisches Vorgehen	52
5.1.1 Expertensampling.....	53
5.1.2 Themenbereiche und Leitfaden.....	57
5.2 Auswertung.....	60
5.2.1 Integration von 3D-Audio in das virtuelle Instrument.....	61
5.2.2 Objektifizierung von Klangkomponenten und objektbasierter Output	64
5.2.3 ‚Layer-Mode‘ und ‚Voice-Mode‘.....	66
5.2.4 Bewegung der Audioobjekte	69
5.3 Diskussion.....	72
6 Schlussbetrachtung.....	77
6.1 Kritische Beleuchtung	77
6.2 Fazit.....	79

6.3	Ausblick	82
	Literaturverzeichnis	84
	Anhang.....	92
	A1 Leitfaden.....	92
	A2 Lautsprecherlayout ‚Produktionsstudio‘	94
	A3 Digitaler Anhang	95

Abbildungsverzeichnis

Abbildung 1:	Wellenform eines 2D-Wavetable WAV-Files mit sieben Frames und Framesize von 1024 Samples.	5
Abbildung 2:	Blockschaltbild zum Aufbau eines 2D-Wavetable Oszillator-Blocks, Quelle: schematisiert nach Puckette, 2007, S. 32.	6
Abbildung 3:	Blockschaltbild zum Aufbau eines Klangsyntheseblocks.	7
Abbildung 4:	Blockschaltbild zum Voice Management am Beispiel vier polyphoner Voices, Quelle: schematisiert nach Puckette, 2007, S. 100.	8
Abbildung 5:	Schematische Funktionsweise des Voice Managements mit aktivem Voice Stealing, Quelle: schematisiert nach Puckette, 2007, S. 100.	8
Abbildung 6:	Workflow-Übersicht zur Betrachtungsweise der Panoramisierung als Teil des virtuellen Instruments.	13
Abbildung 7:	Kopfbezogenes Kugelkoordinatensystem, Quelle: Blauert & Braasch, 2008, S. 88.	14
Abbildung 8:	Blockschaltbild zum schematischen Aufbau eines virtuellen Instruments mit objektbasiertem 3D-Audio Output.	25
Abbildung 9:	Blockschaltbild zur Integration des SWS in seine Peripherie.	26
Abbildung 10:	Blockschaltbild des implementierten 3D-Audio Renderings mit Higher Order Ambisonics Workflow.	30
Abbildung 11:	Blockschaltbild des schematischen Aufbaus des SWS mit Klangsynthese und Richtungsberechnung.	33
Abbildung 12:	Verlauf ADSR-Hüllkurve, Quelle: schematisiert nach Lazzarini, 2019, S. 220.	34
Abbildung 13:	Blockschaltbild zum Aufbau des Klangsyntheseblocks im SWS.	36
Abbildung 14:	Blockschaltbild zur Tonhöhenberechnung im Klangsyntheseblock des SWS.	37
Abbildung 15:	Blockschaltbild zum Routing der Klangkomponenten im ‚Voice-Mode‘ am Beispiel von drei Audioobjekten.	39
Abbildung 16:	Blockschaltbild zum Routing der Klangkomponenten im ‚Layer-Mode‘ am Beispiel von fünf Audioobjekten.	40
Abbildung 17:	Blockschaltbild zum Klangsyntheseblock des SWS mit Modulatoren und MIDI-Befehlen unter exemplarischer Darstellung einer Oszillatorkette.	41
Abbildung 18:	Blockschaltbild zum schematischen Ablauf der Richtungsberechnung eines Audioobjekts.	43
Abbildung 19:	Equirektangulare Kugelflächenprojektion und kartesisches 2D-Koordinatensystem, Quelle: schematisiert nach Rudrich, 2021.	45
Abbildung 20:	Benutzeroberflächen-Ausschnitt des SWS zur Klangsynthese.	49

Abbildung 21:	Benutzeroberflächen-Ausschnitt des SWS zur Richtungsberechnung	50
Abbildung 22:	Benutzeroberflächen-Ausschnitt des SWS zum Routing und der Preset- Verwaltung.....	51
Abbildung 23:	Matrix zur Einordnung der Experten in ihre Berufsfelder.	54

Formelverzeichnis

(1)	Allgemeine Funktion der ADSR-Hüllkurve	34
(2)	Berechnung der LFO-Frequenz aus dem Notenwert.....	35
(3)	Berechnung der Phasor-Frequenz aus der MIDI Note Number.....	38
(4)	Skalierung der LFO 1 Signale im ‚Voice-Mode‘	43
(5)	Umrechnung der Koordinaten der Kugelflächenprojektion in Kugelflächenkoordinaten	45
(6)	Modulation der Azimut- und Elevationswinkel über Steuersignale.....	46
(7)	Umrechnung der Kugelflächenkoordinaten in kartesische Koordinaten.....	46
(8)	Drehmatrix aus drei elementaren Rotationsmatrizen.....	46
(9)	Modulation der Kardanwinkel über Steuersignale	47
(10)	Vektor-Matrix-Multiplikation zur Drehung um die Z -, Y- und X-Achse.....	47
(11)	Translation zum Ankerpunkt.....	47
(12)	Modulation des Translationsfaktors.....	47
(13)	Umrechnung der kartesischen Koordinaten in Kugelflächenkoordinaten	48

Tabellenverzeichnis

Tabelle 1:	Gewichtenden prozentuale Faktoren für die Modulationsoptionen der Klangsynthese.....	42
Tabelle 2:	Themenbereiche und Vorschläge der Implementierung und Evaluation.	58
Tabelle 3:	Lautsprecherlayout ‚Produktionsstudio‘.....	94

Abkürzungsverzeichnis

ILD	Interaural Level Difference
ITD	Interaural Time Difference
LFO	Low Frequency Oscillator
ITU	International Telecommunication Union
MIDI	Musical Instrument Digital Interface
DAW	Digital Audio Workstation
VBAP	Vector Base Amplitude Panning
LEV	Listener Envelopment
IRCAM	Institut de Recherche et Coordination Acoustique/Musique
CCRMA	Center for Computer Research in Music and Acoustics
GRM	Groupe de Recherches Musicales
BEAST	Birmingham ElectroAcoustic Sound Theatre
ASDF	Audio Scene Description Format
M4L	Max for Live
SWS	Spatial Wavetable Synthesizer
OSC	Open Sound Control
MTC	MIDI Timecode
LTI	Linear time-invariant System
BRIR	Binaural Room Impulse Response
A.-H.	Amplituden-Hüllkurve
F.-H.	Filter-Hüllkurve
CPU	Central Processing Unit
DoF	Degrees of Freedom
VST	Virtual Studio Technology

Aus Gründen der besseren Lesbarkeit wird in dieser Masterarbeit die Sprachform des generischen Maskulinums verwendet. Es wird an dieser Stelle ausdrücklich darauf hingewiesen, dass sämtliche Personenbezeichnungen gleichermaßen für alle Geschlechter gelten.

Zudem wird darauf hingewiesen, dass umgangssprachliche Formulierungen sowie eigens vergebene Begriffe mit ‚einfachen Anführungszeichen‘, direkte Zitate mit „doppelten Anführungszeichen“, Marken-, Produkt und Werknamen *kursiv* und Vektoren und Matrizen **fett** kenntlich gemacht sind.

1 Einleitung

Musikproduktion in 3D-Audio gewinnt zunehmend an Relevanz: In den Jahren 2020 und 2021 nahmen mehrere Musik-Streaming-Dienste¹ das Streaming von Musik in 3D-Audio in ihr Angebot auf. Bei den in 3D-Audio verfügbaren Titeln handelt es sich weitestgehend um ursprünglich in Stereo produzierte und veröffentlichte Produktionen, die nun ein ‚3D-Upmixing‘ erfahren haben. Die Reaktionen waren mitunter kontrovers (Welch, 2021).

Der Musikproduzent und Toningenieur Hans-Martin Buff setzt sich seit mehreren Jahren mit dem Thema Musikproduktion in 3D-Audio auseinander und zieht den Vergleich zur Umbruchphase von Mono zu Stereo in den sechziger Jahren: „In Stereo verpflanzte Monowerke wirken fast so gut wie das Mono-Original, in 3D-Audio umgetopfte Stereowerke wirken fast so gut wie das Stereo-Original [...]“ (Buff, 2020, S. 7). Dabei hat 3D-Audio das Potential „eine neue, größere, mächtigere, spannendere, mehr Spaß machende Version eurer Kreativität [entstehen zu lassen]. Es geht um die Wirkung eurer Musik“. Buff ist sich sicher: „Um den Hörer Teil einer Welt werden zu lassen, muss man diese Welt bauen, nicht nur eine schon gebaute Welt in die neue Welt umbiegen“ (ebd., S. 7).

In aktuellen Workflows der Musikproduktion ist die Spatialisierung für 3D-Audio zumeist Teil der Postproduktion – möglicherweise ist eine Anpassung bestehender Workflows nötig, um die Potentiale von 3D-Audio vollumfänglich zu nutzen. Für Komponist, Sounddesigner und Entwickler Johannes Regnier darf die Spatialisierung nicht nur einen zusätzlichen Schritt am Ende der Produktionskette darstellen, sondern muss tief in den kompositorischen Prozess verankert sein: „Can can we consider it (spatialization) as part of the composition, integrated? And can we justify spatialization as part of the sound and not just at a mixing stage at the end?“ (Regnier, 2021, ab 13:10). Dass der Einsatz von 3D-Audio inhaltlich begründet sein muss, sieht auch Komponist Dustin Zorn so – man setzt schließlich „einen ästhetischen Marker, [wenn man] räumlich arbeitet“ (Zorn, 2021, ab 22:17).

Speziell für die Produktion populärer Musik kann 3D-Audio Potentiale hinsichtlich der Weiterentwicklung von Produktionsästhetiken darstellen. Denn „populäre Musik [nimmt] zumeist nicht auf vorbestehende oder klanglich klar definierbare Aufführungssituationen Bezug. Sie entsteht [...] in dem technischen Medium“ (Weinzierl, 2008, S. 779). Eine Verlagerung der Spatialisierung zu den Komponisten und Sounddesignern könnte dabei helfen bereits in der Konzeptionsphase die erweiterten räumlichen Gestaltungsmöglichkeiten von 3D-Audio zu berücksichtigen und die

¹ Nach Stand der Arbeit *Apple Music, Amazon Music, Deezer, Tidal, nugs.net, Hungama Music* und *Anghami* (Rieger, 2021a).

Möglichkeiten des „technischen Mediums“ 3D-Audio für die „wirksame Unterstützung der Wahrnehmung musikalischer Strukturen“ (ebd., S. 779) verstärkt zu nutzen.

Die Komposition populärer Musik findet überwiegend in Digital Audio Workstations (DAW) und unter der Verwendung virtueller Instrumente² zur Klangerzeugung statt. Um die Spatialisierung Teil des Kompositionsprozesses werden zu lassen, wird eine Integration von 3D-Audio in die virtuellen Instrumente für sinnvoll erachtet. Ein weit verbreitetes Verfahren zu Klangerzeugung, speziell in der Produktion populärer Musik, ist die Wavetable Synthese. Mit dieser ist die Simulation traditioneller Instrumente ebenso möglich, wie das Sounddesign artifizierlicher, neuartiger Klänge. Allerdings bestehen aktuell keine in DAWs nutzbaren Wavetable Synthesizer, die Möglichkeiten zur Spatialisierung für 3D-Audio beinhalten.

Das Forschungsinteresse dieser Arbeit liegt somit in der Verzahnung eines Wavetable Synthesizers mit 3D-Audio. Daraus leitet sich die erste Forschungsfrage ab:

Wie kann 3D-Audio in ein virtuelles Instrument der Wavetable Synthese integriert werden?

Für die Beantwortung der Forschungsfrage wurde ein exploratives Vorgehen gewählt, aus dem sich mehrere Ansätze für die Integration von 3D-Audio in ein virtuelles Instrument der Wavetable Synthese entwickelt haben. Diese Ansätze wurden in einen Prototyp implementiert und sind als Vorschläge zu verstehen. Diese Vorschläge werden durch Experten evaluiert, um eine Einschätzung hinsichtlich ihrer Sinnhaftigkeit und Relevanz zu erhalten. Daraus ergibt sich die zweite Forschungsfrage:

Wie bewerten Experten aus den Bereichen Klangerzeugung und Postproduktion die Vorschläge der prototypischen Implementierung?

Zu deren Beantwortung wurden neun Anwender und Entwickler aus den Bereichen Komposition & Sounddesign und der Postproduktion im Rahmen von Experteninterviews befragt. Die Ergebnisse der Evaluation vermitteln ein Stimmungsbild hinsichtlich der implementierten Vorschläge und stellen wichtige Erkenntnisse zur Weiterentwicklung des Prototyps sowie für weitere Untersuchungen in diesem Feld dar.

Die persönliche Motivation zur Arbeit an der beschriebenen Forschungslücke resultiert aus der produktionspraktischen Arbeit. Im Rahmen einer 3D-Audio-Popmusikproduktion für die *Student*

² Unter einem virtuellen Instrument wird ein Software-Synthesizer verstanden, der sich in eine DAW einbinden lässt, über MIDI-Befehle gesteuert werden kann und die erzeugten Klänge in Form von Audiosignalen in die DAW zurückführt (Raffaseder, 2010, S. 228).

*3D Audio Production Competition*³ 2020, entstand der kreative Wunsch, bei der Arbeit mit einem Synthesizer einzelne Töne arpeggiert gespielter Harmonien aus unterschiedlichen Richtungen auf der oberen Hemisphäre er- und ausklingen zu lassen. Jeder Ton ist dabei, unabhängig von seiner Dauer, ortsfest. Dies führt zu einem klanglichen ‚Aufspannen‘ des Raums über tonale Direktsignalanteile. Die Realisierung gelang schlussendlich über das manuelle Exportieren der einzelnen Töne der Akkorde und anschließender Panoramisierung. Der kreative Arbeitsfluss war allerdings gestört und die Flexibilität dieser Vorgehensweise gering. Daraus ergab sich aber die Motivation, 3D-Audio in ein virtuelles Instrument zu integrieren, um neue Möglichkeiten zu schaffen und 3D-Audio Workflows zu beschleunigen.

Aufbau dieser Arbeit

In **Kapitel 2** erfolgt eine implementierungsspezifische Definition und Eingrenzung von Wave-table Synthese sowie von 3D-Audio, um die Perspektive dieser Arbeit auf diese Themen zu umreißen.

Kapitel 3 gibt aus Werkperspektive einen Überblick darüber, wie die Gestaltung räumlicher Attribute seit jeher Teil der kompositorischen Praxis ist. Zudem werden exemplarisch Projekte vorgestellt, die sich bereits mit der Schnittstelle zwischen Klangsynthese und 3D-Audio beschäftigt haben. Abschließend werden die Vorschläge des implementierten Prototyps dargelegt.

Kapitel 4 dokumentiert den implementierten Prototyp. Zunächst wird die Systemintegration des Prototyps mit ihren Schnittstellen und dem sich anschließenden 3D-Audio Rendering vorgestellt. Anschließend wird dessen technische Funktionsweise überwiegend auf Basis von Blockschaltbildern beschrieben und die Intentionen sowie partiell die Klanggestaltungsmöglichkeiten der implementierten Funktionen erläutert.

Kapitel 5 umfasst die Darlegung des methodischen Vorgehens der Experteninterviews, die Vorstellung der Experten sowie die Auswertung der Experteninterviews und zeigt ein Stimmungsbild hinsichtlich der im Prototyp implementierten Vorschläge auf.

Kapitel 6 beleuchtet das Vorgehen in dieser Arbeit kritisch und fasst in einem abschließenden Fazit die Ergebnisse im Sinne der Beantwortung der Forschungsfragen zusammen. Zuletzt wird ein Ausblick auf sich anschließende Forschungsbereiche gegeben.

³ Ein jährlich vom Institut für elektronische Musik (IEM) in Graz ausgerichteter internationaler Wettbewerb für studentische Produktionen in High Order Ambisonics in den drei Kategorien Contemporary/ Computer Music, Audio Drama/ Documentary/ Soundscapes und Music Recording/ Studioproduction.

2 Definitionen und Eingrenzungen

Im Rahmen dieser Arbeit wird die Integration von 3D-Audio in ein virtuelles Instrument zur Wavetable Synthese untersucht. Dieses Kapitel dient zur Eingrenzung und Vorstellung der Betrachtungsweise von 3D-Audio im Kontext dieser Arbeit (Kapitel 2.2) sowie der Definition verwendeter Begrifflichkeiten. Zudem wird der Einsatz von digitaler Wavetable Synthese begründet und die für die Implementierung relevanten Aspekte dieses Syntheseverfahrens vorgestellt (Kapitel 2.1).

Die ersten Experimente zur computergestützten Synthese von Klängen begannen in den Bell Telephone Laboratories unter der Leitung von Max V. Mathews in den späten 1950er Jahren (Ruschkowski, 2019, S. 300ff.). Die Entwicklungen basierten auf der Lookuptable Synthese (auch direkte Synthese, engl. direct digital synthesis, DDS), die bis heute Kern vieler Klangsyntheseverfahren ist – so auch der Wavetable Synthese (Roads, 1996, S. 87). Viele der auf Lookuptable Synthese basierenden Klangsyntheseverfahren sind nicht trennscharf kategorisierbar. So besteht ein fließender Übergang von Sampling über Wavetable Synthese bis hin zu Granularsynthese (Puckette, 2007, S. 34). Die Basis dieser Verfahren stellt ein Audio-Sample dar. Der Ursprung des Samples spielt rein technisch keine Rolle. Beim Sampling wird die Makrostruktur des Samples verarbeitet und das Audio hinsichtlich seiner klanglichen Gestalt neu interpretiert. Granularsynthese hingegen verarbeitet Mikrostrukturen des Audio-Samples und erschafft über Grains⁴ eine neue klangliche Gestalt. Dieses Konzept ist in der Vielzahl seiner Umsetzungsmöglichkeiten eine eigene Syntheseklasse (Bilbao, 2009, S. 8). Es lässt sich festhalten, dass eine Vielzahl an verschiedenen Verfahren zur Klangsynthese besteht, und alle Ansätze jeweils ihre spezifischen klanglichen Charakteristika mit sich bringen (Ruschkowski, 2019, S. 298ff.). Im Kontext von Klangsynthese und Räumlichkeit wurden bereits experimentelle Projekte realisiert, die auf Granularsynthese und additiver Synthese basieren (vgl. Kapitel 3.2). Wavetable Synthese ist im Zusammenspiel mit 3D-Audio bisher wenig erforscht. Die Grundlagen digitaler Wavetable Synthese werden im Folgenden beschrieben.

2.1 Wavetable Synthese

Mit Wavetable Synthese lässt sich eine große Bandbreite verschiedenster Klangfarben auf einfache und effiziente Art und Weise erzeugen. Im Vergleich zu der klanglich limitierten subtraktiven Synthese, bietet Wavetable Synthese eine größere Flexibilität hinsichtlich erzeugbarer Wellenformen und ist der additiven Synthese hinsichtlich Effizienz klar überlegen (Bristow-Johnson, 1996, S. 1f.). Verglichen mit der klanglich ausgereizten und dem Sound der 1980er Jahre

⁴ Ein Grain ist ein mikroakustisches Ereignis, mit einer typischen Länge zwischen einer und hundert Millisekunden. In der Granularsynthese ergeben viele solcher Grains einen Klang (Roads, 2001, S. 96).

assoziierten FM-Synthese, ist mit Wavetable Synthese die Simulation traditioneller Instrumente ebenso möglich, wie auch das Sounddesign artifizierlicher, neuartiger Klänge (Dean, 2009, S. 91f.). Klanglich interessant macht die Wavetable Synthese die Fähigkeit zur Veränderung des Obertonspektrums über die Zeit und den daraus resultierenden komplexen Möglichkeiten zur Klangfarbenmodulation. Diese Variante wird als time-varying Wavetable Synthese oder 2D-Wavetable Synthese betitelt (Roads, 1996, S. 159). Die in der Implementierung eingesetzte 2D-Wavetable Synthese ist ein weit verbreitetes Verfahren zur Erzeugung elektronischer Klänge⁵ und Bestandteil kommerzieller virtueller Synthesizer Instrumente wie z.B. *Serum* oder *Pigments*, die Einsatz in zahlreichen Produktionen elektronischer und populärer Musik finden (Duda, 2021; Arturia, 2021).

Ein Wavetable⁶ enthält für die 2D-Wavetable Synthese mehrere berechnete Wellenformen, auch Frames genannt. Die Frames eines Wavetables haben eine feste Länge. Abbildung 1 zeigt die Wellenformen eines 2D-Wavetables mit sieben Frames und 1024 Samples pro Frame, das auch Verwendung in der prototypischen Implementierung dieser Arbeit findet.

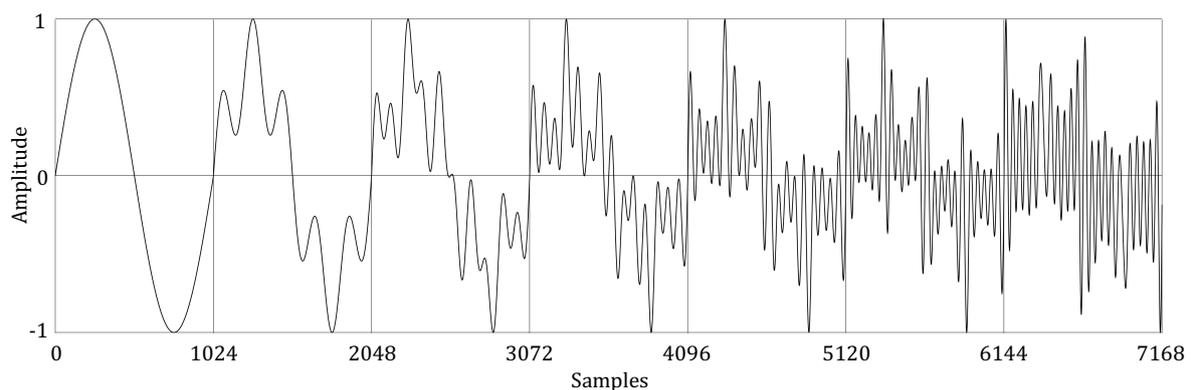


Abbildung 1: Wellenform eines 2D-Wavetable WAV-Files mit sieben Frames und Framesize von 1024 Samples.

Im einfachsten Fall der 2D-Wavetable Synthese, bei zwei Wellenformen, reicht eine lineare Überblendung der beiden Signale aus (Bristow-Johnson, 1996, S. 4). Zu beachten ist, dass die korrespondierenden Teiltöne der überblendeten Wellenformen zueinander in Phase sind, sodass sich ihre Amplituden konstruktiv überlagern. Das Prinzip der 2D Wavetable Synthese kann auf beliebig viele Wellenformen ausgeweitet werden (Puckette, 2007, S. 31; Roads, 1996, S. 95).

⁵ Der Terminus ‚elektronische Klänge‘ bezeichnet im Allgemeinen jene Klänge, die elektronisch oder über digitale Verfahren erzeugt werden und zu deren Wiedergabe Lautsprecher benötigt werden. Die Geschichte der elektronischen Klänge beginnt Anfang des 20. Jahrhunderts und ist eng mit dem technischen Fortschritt verknüpft. Bei den frühen elektronischen Instrumenten bestand neben dem Wunsch nach unverbrauchten Klangfarben auch die Bestrebung, neue Spielweisen zu entwickeln, um zu neuen künstlerischen Ausdruckssphären zu gelangen (Essl, 2019; Ruschkowski, 2019).

⁶ Meist eine Audiodatei im WAV-Format.

Puckette (2007) beschreibt den grundlegenden Aufbau eines Wavetable-Oszillators: Die Wellenformen sind im Wavetable, einer Sequenz von Samples, gespeichert. Im Syntheseprozess dienen die Werte eines sich periodisch wiederholenden Phasor-Signals⁷ als Indizes zum ‚Nachschlagen‘ (engl. lookup) der Werte aus der Sequenz. Dieser Vorgang ergibt das Ausgangssignal des Wavetable-Oszillators. Abbildung 2 zeigt diesen Block schematisch unter Verwendung von zwei Frames. Die Signale der beiden Frames werden in Abhängigkeit des Steuersignals (Wavetable-Position) mit Wertebereich [0;1] linear überblendet (Puckette, 2007, S. 27ff.).

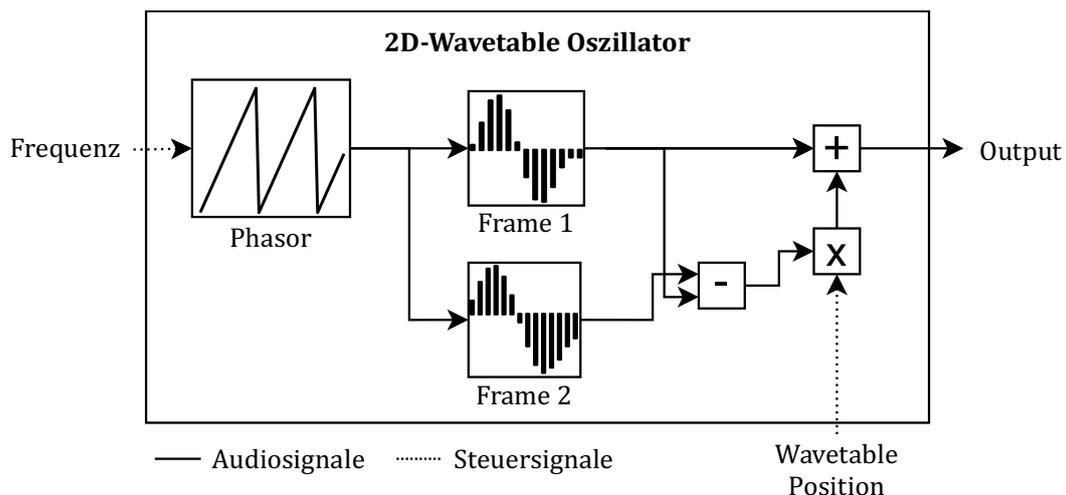


Abbildung 2: Blockschaltbild zum Aufbau eines 2D-Wavetable Oszillator-Blocks, Quelle: schematisiert nach Puckette, 2007, S. 32.

Die Werte des Phasor-Signales fallen in der Regel nicht exakt mit den Indizes des Wavetables zusammen. Für die Interpolation der erforderlichen Zwischenwerte des Wavetables bestehen mehrere Methoden. Neben linearer Interpolation gibt es die Methode der fraktionierten Sample Interpolation (Bristow-Johnson, 1996, S. 4) sowie alternative elaborierte Interpolationsverfahren, für die auf Roads (1996) verwiesen wird (Roads, 1996, S. 95).

Im virtuellen Instrument bettet sich der Wavetable Oszillator-Block in einen Klangsyntheseblock ein, dessen wichtigste Komponenten Abbildung 3 zeigt. Die Frequenz des Phasor-Signals wird im Block Tonhöhenberechnung aus den eingehenden MIDI⁸-Befehlen errechnet. Das Ausgangssignal des Wavetable-Oszillators kann tiefpassgefiltert werden und wird anschließend mit einem Hüllkurvensignal multipliziert. Allgemein gesprochen erzeugen Hüllkurven-Generatoren wie auch LFOs (low frequency oscillators) Modulationssignale, die im Syntheseprozess zur Modulation klangformender Parameter, wie beispielsweise der Tiefpassfilter-Grenzfrequenz, dienen (vgl. Kapitel 4.2.1). Dem Hüllkurvensignal zur Amplitudenmodulation des Ausgangssignals am Ende des

⁷ Phasor-Signal bezeichnet hier ein rampenförmiges Signal mit einem Wertebereich [0;1].

⁸ Ein weit verbreiteter Standard in der Ansteuerung elektronischer und digitaler Instrumente (Lehrman & Tully, 2017, S. 4).

Klangsyntheseblock kommt eine besondere Funktion zu, denn es definiert Anfang und Ende des synthetisierten Klangs einer gespielten Note.

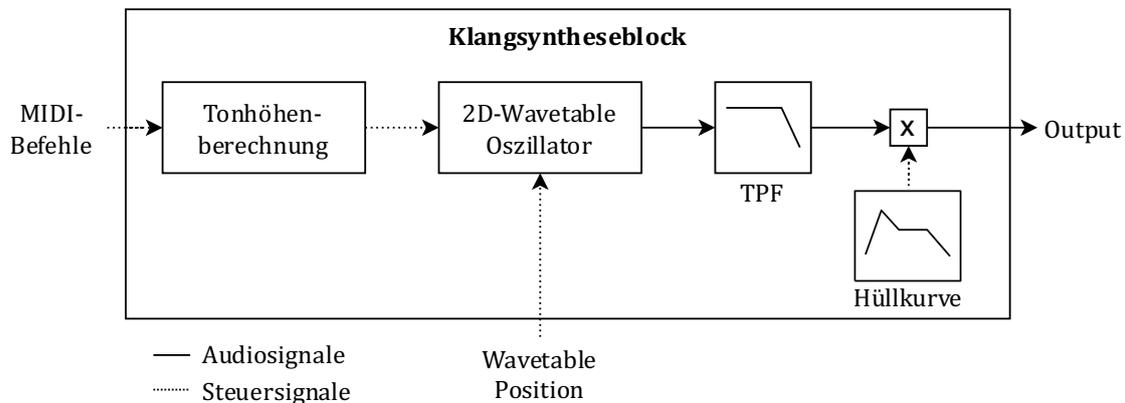


Abbildung 3: Blockschaltbild zum Aufbau eines Klangsyntheseblocks.

Neben der Verwendung von 2D-Wavetables, kann zudem über Wavestacking (auch Wavetable Stacking) für eine zusätzliche Variation des Timbres gesorgt werden. Dabei werden mehrere parallele Wavetable Oszillatoren mit unterschiedlichen Wavetables zusammengemischt (engl. layering) um komplexe Obertonspektren zu erzeugen. Jeder Wavetable Oszillator verfügt über eine eigene Hüllkurve, worüber der zeitliche Verlauf des Klangs gesteuert und somit beispielsweise Einschwing-, stationäre und Ausschwingphasen eines Klangs simuliert werden können (Roads, 1996, S. 160ff.). Der Klangsyntheseblock in Abbildung 3 beinhaltet nur einen 2D-Wavetable Oszillator. Im Sinne des Wavestacking kann dieser um mehrere Einheiten erweitert werden.

Ein weiterer Aspekt des Wavetable Synthesizers ist die Polyphonie. In der Musik ist Polyphonie ein Begriff für verschiedene Arten von Mehrstimmigkeit. Neben beispielsweise einem mehrstimmigen Chor kann auch ein einzelnes Instrument mehrstimmig sein. Das Klavier ist ein polyphones Instrument, auf dem sich theoretisch bis zu 88 Töne gleichzeitig spielen lassen (Puckette, 2007, S. 98). Integriert man eine Instanz des Klangsyntheseblocks in ein virtuelles Instrument, ließe sich lediglich eine einstimmige Melodie spielen. Um eine Mehrstimmigkeit des virtuellen Instruments zu erreichen, braucht es mehrere Instanzen des Klangsyntheseblocks. In diesem Zusammenhang spricht man von polyphonen Voices (Stimmen). Eine Voice enthält dabei eine Kopie des Klangsyntheseblocks. Während beim Klavier für jede Taste die Mechanik zur Erzeugung des Tons zu jeder Zeit verfügbar ist, steuert im virtuellen Instrument das Voice Management (auch Voice Allocation Algorithmus) die Verteilung der eingehenden MIDI-Noten auf die begrenzte Anzahl polyphoner Voices. Die Voices sind aus Gründen der Performance auf eine endliche Zahl festgelegt. (Puckette, 2007, S. 98). Über das Voice Management wird die Reihenfolge der Instanziierung der Voices bestimmt.

Abbildung 4 zeigt schematisch den Aufbau des Voice Managements mit vier Voices. Die Ausgangssignale der Voices werden in Mono-Synthesizern auf einen Kanal summiert. Bei manchen Stereo-Synthesizern besteht die Option, die Ausgangssignale der Voices mit Pegel- und Laufzeitdifferenzen, auf die zwei Kanäle zu verteilen und durch dieses Stereo-Panning eine räumliche Auffächerung im Stereobild zu erzielen.

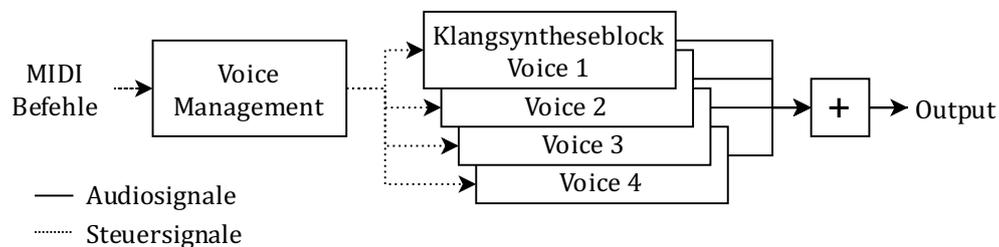


Abbildung 4: Blockschaltbild zum Voice Management am Beispiel vier polyphoner Voices, Quelle: schematisiert nach Puckette, 2007, S. 100.

Der Note On MIDI-Befehl ist Trigger für die Instanziierung einer Voice und startet deren Amplituden-Hüllkurve. Ist die Amplituden-Hüllkurve nach Beenden der Release-Phase wieder auf 0 gefallen, wird eine Mute Nachricht an das Voice Management gesandt und signalisiert somit die erneute Verfügbarkeit dieser Voice. Jede Voice ist über eine ID identifiziert (Puckette, 2007, S. 100).

Abbildung 5 stellt schematisch die Funktionsweise eines einfachen Voice Managements dar. Dabei werden die eingehenden MIDI-Noten a, b, c und d auf zwei Voices verteilt. Zu dem Zeitpunkt an dem d einsetzt, ist keine Voice frei. Im Voice Management ist festgelegt, ob d ignoriert oder Voice 2 von b ‚gestohlen‘ wird (engl. Voice Stealing) (Puckette, 2007, S. 100). Abbildung 5 verdeutlicht das Voice Stealing, bei dem die am längsten aktive Voice ‚gestohlen‘ wird. Voice Stealing ist auch in der Implementierung dieser Arbeit aktiviert.

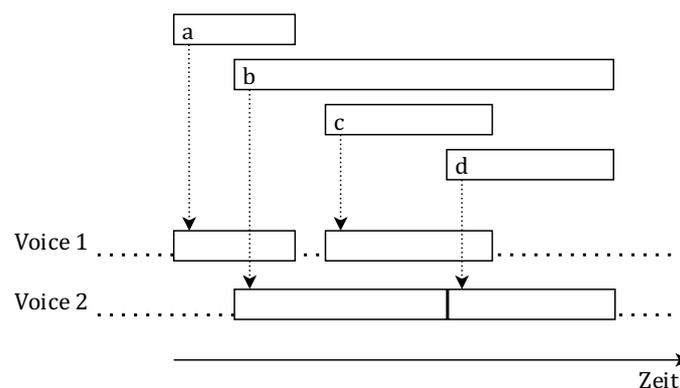


Abbildung 5: Schematische Funktionsweise des Voice Managements mit aktivem Voice Stealing, Quelle: schematisiert nach Puckette, 2007, S. 100.

Zusammenfassend betrachtet sind 2D-Wavetable Oszillatoren der Kern des Klangsyntheseblocks. Bei einem polyphonen Synthesizer werden mehrere Instanzen des Klangsyntheseblocks über ein

Voice Management verwaltet. Die Implementierung dieser Arbeit baut auf diesen Grundlagen der digitalen Wavetable Synthese auf.

2.2 3D-Audio

Bei der Frage nach der Integration von 3D-Audio in einen Wavetable Synthesizer schließt sich die Frage an, wie 3D-Audio zu verstehen ist und was davon seinen Weg in die Implementierung finden soll. Bei primitiver Betrachtung und einem stark vereinfachten Szenario mit festgelegter Hörerausrichtung⁹, lässt sich festhalten, dass Stereo eine Schallreproduktion von vorne ermöglicht. Surround erweitert dies um die Schallreproduktion von hinten und von den Seiten – ein 3D-Audiosystem ermöglicht darüber hinaus die Schallreproduktion von oben und teilweise von unten (Roginska & Geluso, 2018, S. 50f.; Görne, 2015, S. 161). Die Reproduktion von 3D-Audioinhalten ist sowohl über kopfbezogene Wiedergabeverfahren, meist unter dem Begriff Binauraltechnik, als auch über raumbezogene Wiedergabeverfahren, sprich mehrkanaliger Lautsprechersysteme, möglich (Roginska & Geluso, 2018, S. 221ff.). Dabei besteht eine Vielzahl an verschiedenen Lautsprecheranordnungen, die den Anforderungen zur Wiedergabe von 3D-Audio entsprechen. Einen Überblick über mögliche Anordnungen gibt beispielsweise die ITU in der Empfehlung ITU-R BS.2051-2 (ITU-R, 2018). Die exakte Ausgestaltung des 3D-Audio Wiedergabesystems soll für die weitere Betrachtung ausgeklammert werden. Es wird davon ausgegangen, dass das eingesetzte 3D-Audio Wiedergabesystem ausreichend gut funktioniert und in die Lage ist, die im Folgenden beschriebenen Ziele hinsichtlich der Hörereignisbildung umzusetzen.

Die Intention eines 3D-Audiosystems und die sich daraus ableitenden gestalterischen Potentiale werden von Melchior (2020) resümiert: „Ein 3D-Audiosystem hat das Ziel Hörereignisse¹⁰ zu erzeugen, die in Richtung, Entfernung, Ausdehnung und ihrer akustischen Umgebung frei manipulierbar sind und für möglichst viele Hörer mit möglichst gleichen Eigenschaften wahrnehmbar sind“ (Melchior, 2020, S. 3). Diese Kategorisierung deckt sich weitestgehend mit den von Zacharov et al. (2015) vorgeschlagenen Begrifflichkeiten zur Beschreibung und Klassifizierung von Attributen der Audioqualität. Entfernung wird hier als eine Differenzierungsebene von räumlicher Lokalisation betrachtet und darüber hinaus die akustische Umgebung sowie Ausdehnung in mehrere Kategorien untergegliedert (Zacharov et al., 2016, S. 4). Die Betrachtungsweise von Melchior (2020) wird für diese Arbeit als passend bewertet.

⁹ Wie beispielsweise im Kinosaal, mit fester Sitz- und Blickrichtung zur Leinwand.

¹⁰ Das Hörereignis repräsentiert die Wahrnehmung, das Schallereignis bezieht sich auf die physikalische Welt (Spors et al., 2013, S. 3).

Der Prozess zur Gestaltung der räumlichen Attribute Richtung, Entfernung, Ausdehnung und akustische Umgebung¹¹ wird in dieser Arbeit als Spatialisierung¹² bezeichnet. Die Attribute beschreiben die räumliche Gestalt von 3D-Audioinhalten und werden anhand eines simplen idealisierten Beispielszenarios näher beleuchtet. Dabei wird angenommen, dass ein Komponist in einer 3D-Audio fähigen DAW arbeitet und über ein MIDI-Keyboard ein virtuelles Instrument zur Wave-table Synthese mit Mono-Output ansteuert. Der Komponist spielt eine Melodie und gestaltet simultan dazu die beschriebenen räumlichen Attribute. Die Wiedergabe erfolgt über ein lautsprecherbasiertes 3D-Audio Wiedergabesystem.

Richtung

In einem 3D-Audio-Panning-Tool legt der Komponist fest, aus welcher Richtung die Melodie während der Wiedergabe erklingen soll. Über ein Panning-Verfahren, wie beispielsweise Vector Base Amplitude Panning (VBAP), werden für das Audiosignal der Melodie über den im Panning-Tool festgelegten Vektor die Lautsprecher-signale für die Wiedergabe über eine spezifische Lautsprecheranordnung berechnet (Weinzierl, 2008, S. 657). Der Prozess der Richtungsgestaltung eines Audiosignals wird im Folgenden Panoramisierung genannt. Die Panoramisierung eines Klangs kann sowohl statisch als auch dynamisch, mit einer über der Zeit variierenden Richtung, erfolgen. Die Berechnung zur Umsetzung der Panoramisierung wird als 3D-Audio Rendering bezeichnet. Es sei darauf hingewiesen, dass die unterschiedlichen Verfahren des 3D-Audio Renderings jeweils ihre eigenen klanglichen Qualitäten und Eigenschaften mit sich bringen. Um es mit einem vorgezogenen Zitat aus den Experteninterviews des Entwicklers Jan Langhammer zu sagen: "[...], Artefakte im Rendering sind [...] Teil des Sounddesigns. Entweder es passt dann oder es passt halt nicht" (Langhammer, 2021, ab 43:45). Für die konzeptuelle Betrachtung dieser Arbeit wird das eingesetzte 3D-Audio Rendering-Verfahren als gegeben betrachtet und mögliche klangliche Färbungen ausgeklammert.

Nach Blauert & Braasch (2008) ist die Richtungswahrnehmung Teil des räumlichen Hörens¹³ und beruht auf der Auswertung von interauralen Laufzeitdifferenzen (engl. interaural time difference, ITD), interauralen Pegeldifferenzen (engl. interaural level difference, ILD) und spektralen Unterschieden zwischen beiden Ohren. ITD und ILD werden zusammen mit spektralen Unterschieden

¹¹ Unter der akustischen Umgebung wird, der die Schallquellen umgebende (physische oder künstlich erzeugte) Raum verstanden. Dessen Beschaffenheit sorgt über Reflexionen und Nachhall für eine klangliche Färbung. Auch die Abwesenheit eines umgebenden Raums ist eine Form der akustischen Umgebung und setzt klar einen ästhetischen Marker. Darüber hinaus können auch Atmosphären als akustische Umgebung betrachtet werden, die eine spezifische perzeptive Kontextualisierung hervorrufen (Zacharov et al., 2016, S. 6).

¹² Auch in Stereo können räumliche Attribute gestaltet werden. Spatialisierung bezieht sich aber im Rahmen dieser Arbeit, sofern nicht anders geschrieben, auf 3D-Audio-Formate.

¹³ Für Details zum räumlichen Hören wird auf (Blauert, 2005; Blauert & Braasch, 2008; Görne, 2015; Howard & Angus, 2009) verwiesen.

beim natürlichen Hören gleichzeitig ausgewertet (Blauert & Braasch, 2008, S. 89). Der Zusammenhang von ITD, ILD und der wahrgenommenen Schalleinfallrichtung hängt zudem stark von der Beschaffenheit des Schallsignals ab. Zu einer präzisen Lokalisation tragen im Allgemeinen breitbandige und transientenreiche Signalanteile bei (Görne, 2015, S. 126f.). In dem Beispielszenario wird der Klang des Synthesizers einen Einfluss auf die Präzision der Lokalisation der Melodie haben. Der Klang eines Wavetables mit obertonreicher Wellenform gepaart mit kurzen Attack- und Decay-Zeiten der Amplituden Hüllkurve wird besser lokalisierbar sein als der eines obertonarmen Wavetables mit langen Attack-Zeiten.

Die Lokalisationsunschärfe¹⁴ hängt zudem von der Schalleinfallrichtung ab. Ihr Maß variiert im dreidimensionalen Raum – wobei sie besonders gering im horizontalen Bereich der Blickrichtung ist. Bei frontalem Schalleinfall beträgt die Lokalisationsunschärfe maximal 2° und nimmt mit seitlicher Schalleinfallrichtung zu (Dickreiter & Goeres-Petri, 2013, S. 131). Allerdings liefert ein exakt frontales Schallereignis keinerlei ITDs und unterscheidet sich von einem Schallereignis von hinten marginal. Fehlt zudem eine klare visuelle Information, kann es zur Vorne-Hinten-Vertauschung (engl. front-back confusion) kommen (Weinzierl, 2008, S. 675f.). Abhilfe schaffen hier sogenannte ‚Peilbewegungen‘ – kleine Kopfbewegungen auf Horizontalebene, die zur Gewinnung zusätzlicher Richtungsinformation beitragen. Diese Änderung der Kopfhaltung und die damit einhergehende Änderung der Ohrsignale ist auch in der kopfbezogenen 3D-Audiowiedergabe über Binauraltechnik ein wichtiger Faktor – die dynamisch gewonnene auditive Information dominiert bei der Bildung der Hörereignisorte in der Regel über die statisch empfangene (Blauert & Braasch, 2008, S. 93f.). Ein Grund, warum die binaurale Wiedergabe in der Implementierung dieser Arbeit durch das Erfassen der Kopfposition über einen Head-Tracker dynamisch gestaltet wird.

Wie erwähnt kann die visuelle Information, im Szenario beispielsweise die graphische Oberfläche des 3D-Audio-Panning-Tools, einen Einfluss auf die Richtungswahrnehmung haben – Hören ist ein multimodales Phänomen (Blauert & Braasch, 2008, S. 89).

Ausdehnung

Neben der Richtung ist die Ausdehnung ein weiteres räumliches Attribut. Der Komponist kann eine Ausdehnung (auch apparent source width) des Klangs der Melodie über die Dekorrelation einer Gruppe von Duplikaten des Signals der Melodie erreichen. Der Grad der Dekorrelation sowie die Ausdehnung der räumlichen Verteilung der dekorrelierten Signale bestimmen dabei das

¹⁴ Die Lokalisationsunschärfe beschreibt die kleinste Änderung des Schallereignisortes, die in der Wahrnehmung des Hörers zu einer Änderung des Hörereignisortes führt (Blauert & Braasch, 2008, S. 95).

Maß an wahrgenommener Ausdehnung (Potard & Burnett, 2004, S. 2f.). Die Dekorrelation kann in der Produktionspraxis auf unterschiedliche Art und Weise, beispielsweise über Delay-Effekte, Allpass-Filter (Kermit-Canfield & Abel, 2016) oder Tonhöhenmodulationseffekte realisiert werden.

Elektronische Klänge bringen keine natürliche Ausdehnung mit sich, wie das beispielsweise bei einer akustischen Gitarre der Fall ist. Dies eröffnet Gestaltungspotentiale hinsichtlich einer artifizialen, übernatürlichen Ausdehnung ihres Klangs. Die Ausdehnung des elektronischen Klangs könnte sich potenziell über den gesamten reproduzierbaren Hörereignisbereich des 3D-Audiosystems erstrecken. Heißt, das Instrument ist nicht eine den Raum anregende Schallquelle, sondern spannt vielmehr einen Raum um den Hörer herum auf.

In diesem Zusammenhang spielt der Begriff der Umhüllung eine Rolle, der mitunter synonym zu Immersion verwendet wird (Berg, 2009, S. 4f.). Immersion ist allerdings ein kognitiver Zustand und Umhüllung ein perzeptives Phänomen (Agrawal et al., 2020, S. 409). Im Kontext der Konzertsaalakustik wird das Phänomen Umhüllung häufig als listener envelopment (LEV) beschrieben und bezieht sich auf die frühen Reflexionen und den Nachhall (Griesinger, 1997, S. 137). Bei der Reproduktion von Klang über ein 3D-Audiosystem kann Umhüllung aber auch über mehrere, räumlich verteilte Direktschallquellen oder einen Klang mit großer räumlicher Ausdehnung erzeugt werden (Berg, 2009, S. 2f.; Roginska & Geluso, 2018). Umhüllung beschreibt in dieser Arbeit allgemein den perzeptiven Zustand, von Klang umhüllt zu sein. Die Umhüllung ist ein wichtiger Parameter in der räumlichen Gestaltung (Spors et al., 2013, S. 1923f.), wird allgemein als positiv bewertet und stellt ein klangliches Qualitätsmerkmal dar (Weinzierl, 2008, S. 74f.).

Eine wichtige kompositorische Praxis ist das Kontrastieren, das Spiel mit der Gegensätzlichkeit. Die präzise Lokalisation einer punktförmigen Schallquelle kann einem weit ausgedehnten, umhüllend wirkenden Klang gegenübergestellt werden. Dazu wird in der prototypischen Implementierung ein Vorschlag für eine einfache Umsetzung dieser Kontrastierung vorgestellt.

Entfernung

Um mit der wahrgenommenen Entfernung des Klangs der Melodie zu spielen, müsste der Komponist im Beispielszenario mehrere Parameter koppeln. Nach Dickreiter & Goeres-Petri (2013) wird das Entfernungshören vorrangig über die Auswertung von Lautstärke, spektralen Veränderungen (u.a. natürliche Dissipation) und in geschlossenen Räumen dem Verhältnis von Direkt- und Diffusschall realisiert. Mit zunehmender Entfernung der Schallquelle wird der Direktschallpegel geringer, wobei der Pegel des Diffusschalls konstant bleibt. Auch beim Entfernungshören spielen Hörerfahrung, Bekanntheit des Klangs und sofern vorhanden, eine visuelle Information

über die Entfernung der Schallquelle eine Rolle. Die Richtungswahrnehmung einer Schallquelle ist grundsätzlich präziser als die Wahrnehmung ihrer Entfernung (Dickreiter & Goeres-Petri, 2013, S. 133f.).

Um die Entfernung des Klangs seiner Melodie gestalten zu können, müsste der Komponist auf ein 3D-Hallgerät zurückgreifen. Dieses bringt neue Fragestellungen mit sich und erhöht den Komplexitätsgrad hinsichtlich der Implementierung. Für die weiteren Betrachtungen dieser Arbeit soll daher der Aspekt der Entfernungsgestaltung ausgeklammert werden.

Akustische Umgebung

Für die Einflussnahme auf die Gestaltung der akustischen Umgebung, in die sich der Klang der Melodie betten soll, wäre wie auch bei der Gestaltung der Entfernung der Einsatz eines 3D-Hallgeräts notwendig. Auch wenn dessen Einsatz aus klanggestalterischer Sicht reiz- und sinnvoll erscheint und die diffusen Signalanteile des Nachhalls zur klanglichen Umhüllung beitragen, würde dieser eine zusätzliche Ebene eröffnen und vom Fokus dieser Arbeit ablenken, welcher der Panoramisierung von Klangkomponenten des virtuellen Instruments und der darüber erzeugbaren Ausdehnung des Instruments im Raum gilt.

Darüber hinaus gilt festzuhalten, dass die raumakustische Beschaffenheit des Wiedergaberaums, in dem sich das 3D-Audiosystem befindet, auch einen Einfluss auf die Wahrnehmung der akustischen Umgebung hat und mit ihren raumakustischen Eigenschaften für eine klangliche Färbung sorgt. Es wird an dieser Stelle darauf hingewiesen, dass die Raumanteile des Wiedergaberaums eine wichtige Rolle für Externalisierung des Klangbilds bei der binauralen Virtualisierung eines raumbezogenen 3D-Audiosystems spielen (Li et al., 2019).

Die Spatialisierung für 3D-Audio beschränkt sich im Rahmen dieser Arbeit auf die Panoramisierung von Audiosignalen. Im Sinne einer Integration von 3D-Audio in das virtuelle Instrument wird die Verlagerung der Panoramisierung in das virtuelle Instrument für sinnvoll erachtet. Die Umsetzung der Panoramisierung, das 3D-Audio Rendering, geschieht in einem separaten, ausgelagerten Schritt (vgl. Abbildung 6).

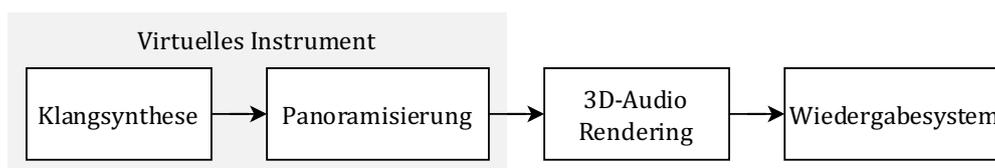


Abbildung 6: Workflow-Übersicht zur Betrachtungsweise der Panoramisierung als Teil des virtuellen Instruments.

Die Schnittstelle zwischen dem virtuellen Instrument und dem 3D-Audio Rendering erfolgt auf Basis des objektbasierten Paradigmas. Nach Herre et al. (2015) werden bei objektbasiertem Audio die Audioinhalte durch Audioobjekte beschrieben. Ein Audioobjekt ist eine Art virtuelle Klangquelle bestehend aus einem Audiosignal und zugehörigen Metadaten. Die Metadaten können unterschiedliche Informationen zum Audioobjekt enthalten wie z.B. Lautheit, Position oder Inhalt des Audiosignals (Herre et al., 2015, S. 770f.). Objektbasiertes Audio ist somit mehr als nur eine Repräsentationsform von 3D-Audio und stellt im Zusammenhang mit Next Generation Audio (NGA) für den Broadcast Bereich Funktionen für die Übertragung interaktiver auditiver Inhalte zur Verfügung. Zwingend notwendig für die Wiedergabe ist ein Renderer, der die Audioinhalte für das vorhandene Wiedergabesetup decodiert (Olivieri et al., 2019, S. 9).

Das Konzept von objektbasiertem Audio ist dabei nicht an ein spezifisches Format (wie z.B. MPEG-H) gekoppelt. Bezogen auf 3D-Audio bietet der objektbasierte Ansatz die Möglichkeit, die Panoramisierung eines Audiosignals unabhängig von einem spezifischen 3D-Audiosystem zu beschreiben. Die Idee der Formatagnostik ist die Basis für den Output des implementierten Prototyps dieser Arbeit.

Im Rahmen dieser Arbeit erfolgt die Beschreibung der Richtungen der Audioobjekte über kopfbezogene Kugelkoordinaten (Polarkoordinaten) (vgl. Abbildung 7). Der Koordinatenursprung befindet sich auf der interauralen Achse zwischen den beiden Ohren, auf Höhe der Ohrkanäleingänge. Die frontale Blickrichtung ist als 0° -Achse definiert.

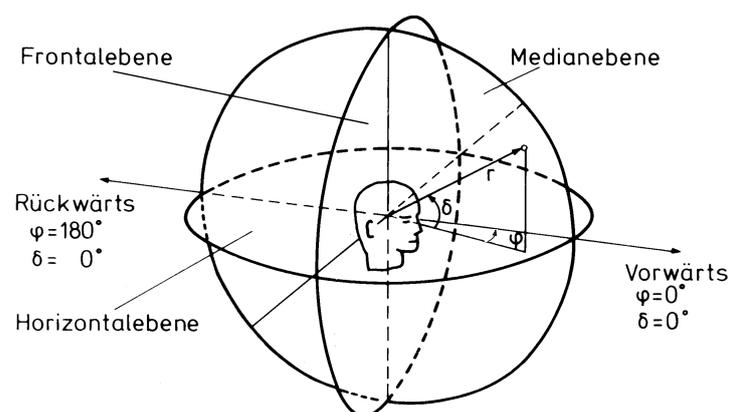


Abbildung 7: Kopfbezogenes Kugelkoordinatensystem, Quelle: Blauert & Braasch, 2008, S. 88.

Wie beschrieben wird auf eine Entfernungsdarstellung und damit auch auf Radius r verzichtet, beziehungsweise $r = 1$ konstant gesetzt. Wird der Abstand zum Ursprung konstant gehalten, spricht man auch von Kugelflächenkoordinaten. Die Richtung eines Audiosignals wird somit über die Winkel Azimut φ und Elevation δ beschrieben (Blauert & Braasch, 2008, S. 88). Der Azimut-Winkel wird im mathematisch positiven Sinne angegeben.

3 Forschungsstand

Schon lange vor der elektroakustischen Reproduktionstechnik und 3D-Audiosystemen haben sich Komponisten verschiedener Epochen mit den räumlichen Attributen kompositorisch auseinandergesetzt. Aufführungsräume stellen nicht nur einen akustischen und visuellen Rahmen für die Musik dar, sondern dienen als Klangverstärker, und beeinflussen Klangstruktur sowie Klangfarbe. Ihre Akustik wirkt sich auf die räumliche, zeitliche, dynamische und klangfarbliche Dimension von Musik aus (Reuter et al., 2014, S. 148). Eine fragmentarische, weit gefasste Zusammenstellung zur Gestaltung räumlicher Attribute im geschichtlichen Kontext und mit Werkperspektive dient dazu, ein Bewusstsein für die Vielschichtigkeit und die Möglichkeit der künstlerischen Interpretation zu schaffen (Kapitel 3.1). Daran anschließend werden in Kapitel 3.2 ausgewählte Projekte vorgestellt, die Untersuchungen an der Schnittstelle von Klangsynthese und Spatialisierung zum Thema hatten. In Kapitel 3.3 wird abschließend die Forschungslücke aufgezeigt und die beiden Forschungsfragen vorgestellt. Darüber hinaus werden die im Implementierungsteil umgesetzten Vorschläge dargestellt.

3.1 Raum als kompositorische Dimension im musikgeschichtlichen Kontext

Allgemein gesprochen sind räumliche Beziehungen zwischen Klangkörpern seit jeher fester Bestandteil der Aufführungspraxis. Volkslieder wechseln zwischen Stimmgruppen, Militärkapellen marschieren über den Marktplatz, Alphörner übermitteln Botschaften über Kilometer hinweg. Musikgeschichtliche Beispiele, die sich in besonderer Art und Weise der räumlichen Dimension der Musik widmen, können einen Anhaltspunkt für Entwicklungen an der Schnittstelle von Klangzeugung und Spatialisierung darstellen.

Traditionelle akustische Musik

Die frühesten bekannten und dokumentierten Werke, die Raum als kompositorisches Element nutzen, stammen aus der Mitte des 16. Jahrhunderts aus der Basilica San Marco in Venedig. Eine Besonderheit dieser Kirche waren zwei sich gegenüberstehende Orgeln. Der Kapellmeister und Komponist Adrian Willaert machte sich diese Anordnung zunutze und komponierte antiphonale¹⁵ Werke für zwei räumlich getrennte Chöre und Instrumentalgruppen. Willaerts 8-stimmige *Vesper* (1550) ist das früheste bekannte Werk, in dem räumliche Dialogform und Echoeffekte eingesetzt wurden. Die Praxis verbreitete sich anschließend in ganz Europa (Zvonar, 2005).

¹⁵ Antiphonie bedeutet Gegengesang oder Wechselgesang (auch ‚call and response‘) und beschreibt eine Art des Musizierens, bei der vorgetragene musikalische Elemente von anderen Stimmen oder Instrumenten beantwortet werden.

Auch in der Romantik finden sich Beispiele für die effektvolle räumliche Verteilung von Musikern. Ein Paradebeispiel dafür ist Hector Berlioz' *Requiem* (1837). Vier separate Blechbläserensembles sind als Fernorchester zur Darstellung des Jüngsten Gerichts räumlich getrennt in die vier Himmelsrichtungen platziert. In ähnlicher Weise wurden ebenfalls außerhalb der Bühne Blechbläserensembles in Giuseppe Verdis *Requiem* (1874) und in Gustav Mahlers 2. *Symphonie* (1894) eingesetzt (Scherliess, 2020).

Mit dem Beginn des 20. Jahrhunderts fanden sich neue Herangehensweisen an die räumliche kompositorische Dimension. Der amerikanische Experimentalist Charles Ives begann mit unabhängigen Ensembles zu arbeiten, wie z.B. in *The Unanswered Question* (1908). In diesem Stück befinden sich die im Off stehenden Streicher als eine Art Gegenpart zu dem Solo-Trompeter und dem Holzbläserensemble auf der Bühne. Ives' Experimente von Klang und Raum wurden später in der Musik von Henry Brant weitergeführt. *Voyage Four* (1963) erfordert drei Dirigenten, die Schlagzeug und Blechbläser auf der Bühne dirigieren, Geigen auf einem Seitenbalkon, Bratschen und Celli auf der gegenüberliegenden Seite, Bässe auf dem Boden im hinteren Bereich, Holzbläser und einige Streicher in den hinteren Balkonen sowie mehrere Einzelspieler im Publikum (Zvonar, 2005). *Windjammer* (1969) hatte einen stationären Hornsolisten und mehrere Holzbläser, die sich während des Spiels auf vorgeschriebenen Routen bewegen (Harley, 1997, S. 71 ff.). Sozusagen ein Vorläufer von Trajektorien: Bewegungspfade, wie sie mitunter bei der dynamischen Panoramisierung elektronischer Klänge zum Einsatz kommen.

Die wenigen Beispiele sollen genügen, um zu zeigen, dass eine bewusste Gestaltung der wahrgenommenen Richtung des Klangs sowie dessen Bewegung, Ausdehnungen, Entfernungen und vor allem auch der gezielte Einsatz akustischer Umgebungen ebenfalls Teil der Musik ist, die ohne den Einsatz elektroakustischer Reproduktionstechnik auskommt.

Die Anfänge der elektroakustischen Musik

Mit der Erfindung der Tonaufzeichnung und Übertragung wurde eine neue musikalische Ära eingeleitet, in der Klang erstmals losgelöst von seiner Quelle hörbar war. 1923 erfand Leon Theremin das nach ihm benannte Instrument Theremin, das durch die Bewegung der Hände in der Nähe zweier Antennen gesteuert wurde. Der Klang des Instruments kam aus einem Lautsprecher (Essl, 2019, S. 2f.). Theremin-Ensemble Aufführungen könnten somit das erste Beispiel für mehrkanalige Lautsprechermusik gewesen sein.

1948 präsentierte Pierre Schaeffer die ersten musikalischen Werke, die mit Hilfe von Schallplatten-Recordern entstanden. Diese, als *Musique Concrète* betitelte Kompositionstechnik, entstand aus Aufnahmen von Alltagsgeräuschen. In Zusammenarbeit mit Pierre Henry schuf Schaeffer ein

Repertoire an Werken für Tonband, darunter *Symphonie pour un Homme seul* (1950). Aufgrund fehlender Mehrspurrecorder behelfen sich die Komponisten mit mehreren Mono-Tonbandgeräten, wobei mehrere Tonbandsignale über ein 4-Kanal Lautsprechersystem tetraedischer Anordnung wiedergegeben wurden. Schaeffer entwickelte für die Verteilung des Klangs auf dem Wiedergabesystem einen Mechanismus namens *Potentiomètre d'espace* (1951), der Induktionsspulen zur Steuerung der Signalführung verwendete. Die Steuerung bestand aus vier, den Interpret umgebenden großen Reifen, welcher mit seinen Armbewegungen die räumliche Positionierung des Klangs steuerte (Ruschkowski, 2019, S. 248 ff.). Dieses System kann im Entferntesten als ein historischer Vorläufer heutiger 3D-Audio-Panning-Tools angesehen werden.

In New York gründeten John Cage und eine Gruppe von Kollegen das *Project for Music for Magnetic Tape*. Zu den Werken, die durch das Verarbeiten von oft winzigen Bandschnipseln entstanden, gehörte Cages *Williams Mix* (1952) für 8 Mono-Bänder, die jeweils über einen eigenen Lautsprecher abgespielt werden. Die Lautsprecher waren in gleichen Abständen um die Hörer verteilt (Zvonar, 2005). Dies weckt Assoziationen zu Surround-Lautsprecherlayouts.

1956 vollendete Karlheinz Stockhausen im Studio für Elektronische Musik des Westdeutschen Rundfunks (WDR) das Werk *Gesang der Jünglinge* (Bates, 2009, S. 130 ff.). Das aus elektronischen Klängen und der Stimme eines Knabensoprans bestehende Stück, gilt als das erste Stück für Mehrspurband. Die wahrscheinlich erste echte quadrophonische Komposition war Stockhausens *Kontakte* (1960) (Roads, 1996, S. 453). Die vier Spuren waren für Lautsprecherpositionen vorne links und rechts sowie hinten links und rechts ausgelegt. Für die Erzeugung von Kreisbewegungen befestigte Stockhausen einen Lautsprecher auf einem von vier Mikrofonen umgebenen Drehtischsystem, mit denen er Teile der Komposition wiederaufnahm, um bei der erneuten Wiedergabe über die quadrophone Lautsprecheranordnung eine Bewegung des Klangs zu erzielen (Roads, 1996, S. 471). Die räumliche Bewegung des Klangs ist auch in Stockhausens späterem Schaffen ein zentrales Merkmal geblieben und nicht nur auf seine elektronischen Werke beschränkt. Das Werk *Stimmung* (1968) für sechs Vokalistinnen, wurde durch sechs das Publikum umgebende Lautsprecher verstärkt, um die Zuhörer klanglich zu umhüllen (Kurtz, 1993).

Klanginstallationen

Klanginstallationen sind mit ihren aufwendigen und Raum-schaffenden Aufbauten prädestiniert für den Einsatz räumlicher Wiedergabesysteme. Als historisch relevantes Beispiel kann die Multimedia-Installation der Brüsseler Weltausstellung des Philips-Pavillons angeführt werden, in dem die Tonbandkomposition *Poème Electronique* (1958) von Edgard Varèse zu hören war. Die Musik wurde durch ein Mehrspurband-System mit den visuellen Effekten synchronisiert. Für

Bewegungen des Klangs auf dem 425 Lautsprecher umfassenden Wiedergabesystem wurden neun verschiedene ‚Routen‘ vergleichbar mit Trajektorien, programmiert (Zvonar, 2005).

Auf der EXPO 70 in Osaka (Japan) gab es mehrere Mehrkanal-Klanginstallationen mit der Musik der Avantgarde-Komponisten. Der Komponist und Architekt Iannis Xenakis, präsentierte seine 12-kanalige Tonbandkomposition *Hibiki Hana Ma*, die über ein System von 800 Lautsprechern rund um das Publikum, über und unter den Sitzen spatialisiert wurde. Zur gleichen Zeit spielten im Deutschen Pavillon Karlheinz Stockhausen und eine Gruppe von 20 Solisten Konzerte in einem kugelförmigen Auditorium welches 600 Zuhörer fasste. Stockhausen steuerte die Spatialisierung aus der Mitte der Kugel heraus und verteilte den Klang in kreis- und spiralförmigen Bahnen auf 55 Lautsprechern. Die Solisten selbst befanden sich auf sechs kleinen Balkonen hoch an den Wänden, und das Publikum saß auf einem schalltransparenten Gitter, so dass es von oben und unten vollständig von der Musik umhüllt wurde (Stockhausen, 1978).

Fortschritte der Spatialisierung

In den 1960ern erweiterten einige Studios für elektronische Musik ihre Wiedergabesysteme auf eine quadrofone Lautsprecheranordnung. Diese Entwicklung erfolgte aus der Erweiterung von Stereo und fällt zeitlich mit der Verfügbarkeit von Vierspur-Tonbandgeräten zusammen. Obwohl Quad im audiophilen Endkunden Markt eine kurze und kommerziell wenig erfolgreiche Episode darstellt, hat Vierspur-Tonbandmusik die kommerzielle Quad-Technik um einige Jahre überlebt. Die ersten Anwender von Quad-Panning waren noch dazu gezwungen, ihre eigenen Panning-Tools zu bauen oder spezielle Fader- und Pan-Pot Verschaltungen zu entwickeln. Mit den modularen Analogsynthesizern Mitte der 1960er Jahre konnte auch räumliche Modulation mit spannungsgesteuerten Verstärkern (VCAs) durchgeführt werden (Zvonar, 2005).

John Chowning stellte 1966 ein Programm fertig, mit dem er eine Trajektorie zeichnen konnte und die dafür notwendige Energieverteilung zwischen vier Lautsprechern vom Computer berechnet wurde (Zvonar, 2005). Chowning wandte sich anschließend der digitalen Klangsynthese zu und kehrte mit einer neuen Klangpalette zu seiner Arbeit an der räumlichen Klanggestaltung zurück. Er vollendete das quadrophonische Werk *Sabelith* (1966) und hielt 1970 auf dem Kongress der Audio Engineering Society seinen Vortrag *The Simulation of Moving Sound Sources* - eine Analyse der Auswirkungen des Dopplereffekts auf die Wahrnehmung der Geschwindigkeit eines sich bewegenden Klangs sowie der lokalen und globalen Nachhalleffekte (Chowning, 1971).

Chowning führte seine Forschung am CCRMA¹⁶ in Stanford durch – einer renommierten Einrichtung für Computermusik und Gastgeber für Komponisten aus aller Welt. Dort experimentierte zudem Roger Reynolds mit Hallgeräten und einem spannungsgesteuerten räumlichen Lokalisationsystem. *Eclipse (Voicespace III)* (1979) und *The Palace (Voicespace IV)* (1980) nutzten die damals neuen Möglichkeiten der digitalen Klangbearbeitung und die Fortschritte digitaler Hallalgorithmen. Die kollaborativen Experimente hatten weiterhin Bestand. Reynolds *Watershed IV* (1996) ist ein Stück für Solo-Schlagzeug und Surround-Sound Spatialisierung auf einem sechskanaligen Lautsprechersystem. Dieses basierte auf TRAnSiT (Toward Real-time Audio Spatialization Tools) – ein Projekt, das aus dem Zusammenschluss führender Forschungsinstitute, Komponisten und Entwickler resultierte (Bohn, 2002). Die Bearbeitung des Stücks wurde am IRCAM¹⁷ in Paris durchgeführt. Ebenso wie das CCRMA ist auch das IRCAM Gastgeber für Komponisten elektronischer Musik und Entwickler innovativer musikalischer Technologie. Ergebnisse dieser Forschungsarbeit werden heute überwiegend in Form von Software-Suites den Kunstschaffenden zur Verfügung gestellt. So wurde die Multimedia-Entwicklungsumgebung *Max* in ihrer ersten Version von Miller Puckette Mitte der 1980er Jahre am IRCAM entwickelt. Für den Bereich 3D-Audio steht mit *Spat* eine umfangreiche Echtzeit Spatialisierung für die Kreation, Postproduktion und Aufführung von Musik zur Verfügung, die sich in *Max* integrieren lässt.

Nach Essl (2019) gab es neben diesen diskreten Vorgehensweisen zur Spatialisierung auch andere Ansätze, die auf Diffusion und artifizielle Ausdehnung der Klänge abzielten. Dabei wurde vorproduziertes Mono-oder Stereomaterial durch einen ‚Diffusion Desk‘ über ein ganzes ‚Lautsprecherorchester‘ abgespielt. Dabei ist es üblich, die Lautsprecher sowohl auf der Bühne als auch an den Seiten, hinter sowie über dem Publikum zu platzieren. Auf gestalterischer Ebene liegt dabei das Hauptinteresse nicht in der Bewegung des Klangs im Raum und der damit verbundenen Lokalisation, sondern in der Artikulation der Musik durch die Wiedergabe verschiedener Passagen über unterschiedlich klingende Lautsprecheranordnungen (Essl, 2019, S. 14). Eine Form dieser räumlichen Interpretation ist das 1974 von Francois Bayle für die Groupe de Recherches Musicales (GRM) entwickelte Acousmonium - ein 50-80teiliges Ensemble unterschiedlichster Lautsprechertypen. Mit dem Birmingham ElectroAcoustic Sound Theatre (BEAST) entwickelte sich auch in England ein solches akusmatisches¹⁸ System (Zvonar, 2005).

Es lässt sich zusammenfassend festhalten, dass in der elektroakustischen Musik die Spatialisierung für räumliche Wiedergabesysteme seit Jahrzehnten eine Rolle spielt. Komponisten

¹⁶ CCRMA: Center for Computer Research in Music and Acoustics

¹⁷ IRCAM: Institut de recherche et coordination acoustique/musique.

¹⁸ Der Begriff Akusmatik bezeichnet Musik, bei deren Aufführung das Hören ohne visuellen Bezug zur Klangquelle geschieht – die Musik ohne Performance über Lautsprecher wiedergegeben wird.

elektroakustischer Musik sind meist auch Entwickler ihrer Instrumente und Spatialisierungs-Tools. Die Arbeiten erfolgen häufig in Entwicklungsumgebungen wie beispielsweise *Max*, *Pure Data* und *OpenMusic*. Die im Rahmen der Experteninterviews befragten Komponisten elektroakustischer Musik Johannes Regnier und Dustin Zorn konnten diese Arbeitsweise bestätigen.

Die Produktion populärer Musik wird aber überwiegend in DAWs realisiert. Diese Softwareumgebungen weisen einen hohen Standardisierungsgrad auf und bieten in der Regel umfangreiche Funktionen für die Aufnahme, Bearbeitung und Mischung von Audio. Trotz langer Zeit geringer Verbreitung von 3D-Audio in den Produktions-Workflows populärer Musik, gibt es Beispiele von Produktionen, die in 3D-Audio-Formaten entstanden sind, bzw. dafür gemischt wurden. So wurde seit 2005 ein *Grammy Award* für das *Best Surround Sound Album* verliehen, der 2019 in *Best Immersive Audio Album* umbenannt wurde. Mit *3-D The Catalogue* waren mit Kraftwerk auch Vertreter populärer elektronischer Musik in der offiziellen Auswahl der Kategorie vertreten (Recording Academy, 2017). Dieses Kraftwerk Album besteht aus 3D-Audio Mischungen bereits veröffentlichter Titel der Gruppe. Die Kompositionen erfolgten somit nicht in Hinblick auf die gestalterischen Potentiale eines 3D-Audiosystems, sondern wurde im Nachgang spatialisiert (Wendt, 2018). Die positive Resonanz auf das Album verdeutlicht, dass sich mit dieser Vorgehensweise überzeugende Musik in 3D-Audio produzieren lässt.

3.2 Projekte an der Schnittstelle von Klangsynthese und Spatialisierung

Nachdem die Werkperspektive einen Eindruck von der Bedeutung der räumlichen Dimension im musikgeschichtlichen Kontext vermittelt hat, werden im Folgenden exemplarisch ausgewählte Projekte vorgestellt, die sich mit ähnlichen wie den dieser Arbeit zugrunde liegenden Anforderungen beschäftigt haben.

Spatialisierung additiver Synthese

Ein mehrfach implementiertes Konzept ist die Spatialisierung additiver Synthese. Die additive Synthese ist eine der ältesten und meist untersuchtesten Verfahren. Sie stellt eine Klasse von Klangsynthesetechniken dar, die auf der Addition elementarer Wellenformen basieren und darüber komplexe Wellenformen erzeugen (Roads, 1996, S. 134ff.). Die Teiltöne können dabei ganzzahlige Vielfache der Grundschwingung sein und ein harmonisches Obertonspektrum erzeugen. Aber auch arbiträre Frequenzverhältnisse sind möglich und somit unharmonische bis geräuschte Spektren synthetisierbar. Es besteht die Möglichkeit, einzelne Teiltöne über die Zeit in Frequenz und Amplitude zu ändern und darüber lebhaftere, sich entwickelnde Timbres zu erzeugen. Die vorgestellten Ansätze basieren im Wesentlichen auf dem Ansatz, die über Teiltonoszillatoren entstehenden Klangkomponenten unabhängig zu panoramieren.

Spatio-Operational Spectral (S.O.S.) Synthesis ist ein Projekt von Topper et al. (2002) aus den frühen 2000er Jahren. Hierbei werden die spektralen Komponenten additiver Synthese und physikalischer Modellierung auf einzelne oder Gruppen von Lautsprechern eines achtkanaligen Surround Systems geroutet. Dieses Projekt zur Echtzeitspatialisierung wurde in *Max* implementiert (Topper et al., 2002).

Verron et al. (2008) stellen mit ihrem Projekt *A 3D Immersive Synthesizer for Environmental Sounds* das Design eines 3D-Synthesizers für Umgebungsgeräusche vor, das explizit den Anspruch hat, eine Alternative zur Spatialisierung als nachgelagerten Schritt zu präsentieren. Die Architektur der Synthese-Engine ermöglicht die Kombination des additiven Synthesemodell mit der Spatialisierung auf einer Ebene. Der Algorithmus basiert auf amplitudenbasierter Klangpositionierung. Dies ermöglicht eine Synthese und Spatialisierung von deterministischen und stochastischen Signalen, um punktförmige und ausgedehnte Schallquellen zu simulieren. Jede Komponente kann individuell im Raum positioniert werden. Die Panoramisierung erfolgt zunächst formatagnostisch und kann für unterschiedliche Wiedergabeformate gerendert werden (Verron et al., 2008).

Eine weitere Implementierung ist das Projekt *A Jack-based application for spectro-spatial additive synthesis* von von Coler (2019). Die Synthese fokussiert sich auf die Nachbildung eines Violinklangs in Legatotechnik. Dabei werden Violinensamples analysiert und mit Hilfe von Sinusschwingungen und gefilterten Rauschbändern re-synthetisiert. Die einzelnen Klangkomponenten gehen über die virtuelle Soundkarte *JACK* in einen Renderer und werden dort räumlich auf ein Surround-Lautsprecherlayout verteilt. Über OSC¹⁹ erhält der Renderer zusätzliche Information aus der in *Pure Data* realisierten Synthese für die räumliche Verteilung der Klangkomponenten (von Coler, 2019). Das Projekt ist Teil von Colers Dissertationsschrift *A System for Expressive Spectro-Spatial Sound Synthesis*, die darüber hinaus eine aufschlussreiche Referenz für die Arbeit an der Schnittstelle von Klangsynthese und Spatialisierung darstellt (von Coler, 2020).

Spatialisierung von Granularsynthese

Einen weiteren, mehrfach implementierten Ansatz stellt die Kombination von Granularsynthese und Partikel-Simulationen, meist unter der Verwendung von Flocking-Algorithmen, dar. Flocking-Algorithmen, wie Craig Reynolds *Boids*, modellieren die Gesamtbewegung von Gruppen autonomer Partikel, beispielsweise eines Vogelschwarms und werden häufig in der Computergrafik und Animation eingesetzt (Reynolds, 1987). Speziell mit Granularsynthese scheinen sich Flocking-Algorithmen viele Eigenschaften zu teilen. Beide verwenden eine große Anzahl

¹⁹ Open Sound Control (vgl. Kapitel 4.1.2).

individueller Partikel oder Grains, erzeugen so Komplexität und lassen sich durch wenige Parameter auf Makroebene steuern. Flocking-Algorithmen legen eine Spatialisierung nahe, bei der jedes Grain entsprechend der Position eines einzelnen Partikel in der algorithmisch erzeugten Wolke im Raum positioniert wird (Bates, 2009).

Eine auf diesem Ansatz basierende Arbeit ist Kim-Boyles *Spectral and Granular Spatialization with Boids*. Kim-Boyle implementiert dabei eine Methode, die Reynolds Flocking-Algorithmus für die Verteilung von Grains auf ein quadrofonen Lautsprechersetup nutzt. Die Position der Boids wird dabei im zweidimensionalen Raum auf die Lautsprecher abgebildet. (Kim-Boyle, 2006).

Auch das Projekt *Creating Timbre in Space: Granular Synthesis and Ambisonics Spatialization* von Rossetti & Manzolli untersucht die Beziehung zwischen granular synthetisierten Klängen und Spatialisierung. Allerdings werden hier die Grains nicht diskret in Abhängigkeit der Boids im Raum positioniert, sondern mehr von einer Boid gesteuerten Diffusion Gebrauch gemacht (Rossetti & Manzolli, 2019).

Virtuelle Synthesizer mit räumlichem Bezug

In DAWs integrierbare virtuelle Instrumente zur nativen Klangerzeugung in 3D-Audio-Formaten sind nicht verbreitet. Zwei Beispiele für ansatzweise Surround-fähige Synthesizer werden in Kurzform vorgestellt. Mit dem *Subtrakt 5.1* von Fabian Grote besteht ein Konzept eines Surround fähigen Software Synthesizer. Dieser war allerdings zum Stand dieser Arbeit nicht mehr verfügbar. Er basiert auf der Idee des ‚Layering‘²⁰ und bietet sechs Oszillatoren an, welche diskret auf die sechs Lautsprecherkanäle des 5.1 Surround Formates geroutet werden können (Grote, o. J.). Der Synthesizer *Absynth* von Native Instruments ist ebenfalls Surround fähig. Dekorrelierte Surround Signale werden dabei über integrierte Effekte, wie beispielsweise Multitap Delays und Resonatoren, erzeugt (Sidi, 2009, S. 90ff.).

Objektbasierte Formate in der Produktion

Nicht unerwähnt sollen Projekte bleiben, die den Versuch der Etablierung eines objektbasierten Zwischenformats für die Produktion von 3D-Audio gewagt haben, sich aber nicht durchsetzen konnten. Ein Grund, warum für die Implementierung dieser Arbeit kein standardisiertes objektbasiertes Format zum Einsatz kommt und die erzeugten Audiosignale und Metadaten in roher Form und in Anlehnung an das objektbasierte Paradigma erfolgen. Ein Ansatz war das Audio Scene Description Format (ADSF) – ein Austausch-Format für objektbasierte Audioszenen und in

²⁰ Layering meint im Kontext Sounddesign das Erzeugen komplexer Timbres über das ‚Schichten‘ mehrerer Klänge, welche in der Wahrnehmung zu einem Hörereignis verschmelzen.

Zusammenspiel mit dem SoundScapeRenderer konzipiert (Geier et al., 2010). Auch *SpatDIF* war ein Ansatz für ein Zwischen-, bzw. Austausch-Format für 3D-Audio Szenen, dessen Entwicklung seit Mitte der 2010er Jahre nicht weiter verfolgt wurde (Miyama et al., 2013; Schacher et al., 2014).

Eine Untersuchung von Coleman et al. (2018) im Bereich der objektbasierten Audioproduktion kommt zu dem Ergebnis, dass ein Bedarf an objektbasierten Produktions-Tools besteht (Coleman et al., 2018). Obwohl der Fokus dabei auf der Aufnahme akustischer Klänge liegt, können auch virtuelle Synthesizer zu Produktions-Tools gezählt werden. Im Artikel wird dabei ein objektbasierter Workflow von der Klangerzeugung über Repräsentation des 3D-Audio-Inhalts bis hin zum Rendering vorgeschlagen.

3.3 Forschungsfragen und methodisches Vorgehen

Die Aufarbeitung des Forschungsstands, sowohl aus Werkperspektive als auch aus projektbezogener Perspektive hat gezeigt: Räumlichkeit wird seit Jahrhunderten als kompositorische Dimension der Musik bewusst gestaltet. Elektronische Klänge und elektroakustische Reproduktionssysteme bringen bezüglich der räumlichen Gestaltung eigene Möglichkeiten mit sich. Mit diesen wurde in der zweiten Hälfte des vergangenen Jahrhunderts umfassend experimentiert, wobei Instrumente an der Schnittstelle von Klangsynthese und Spatialisierung entstanden sind. Bisherige Projekte haben sich dabei vor allem auf additive Synthese und Granularsynthese konzentriert. Für die Wavetable Synthese konnte kein vergleichbares Projekt ausfindig gemacht werden – die Verzahnung dieses Klangsyntheseverfahrens mit 3D-Audio im Rahmen dieser Arbeit stellt somit einen neuen Ansatz dar. In Hinblick auf Produktionsumgebungen populärer Musik wurde ein Bedarf an virtuellen Klangerzeugern festgestellt, welche gestalterische Potentiale von 3D-Audio mit dem Klangerzeugungsprozess verzahnen und dabei in DAW-basierte Produktions-Workflows integrierbar sind. Daraus leitet sich die erste Forschungsfrage ab:

Wie kann 3D-Audio in ein virtuelles Instrument der Wavetable Synthese integriert werden?

In einem explorativen Prozess wurde ein Prototyp entwickelt und implementiert. Dieser basiert auf einer Reihe von Schlussfolgerungen des Forschungsstands und besteht aus mehreren konkreten Umsetzungsvorschlägen, aus welchen sich die Integration von 3D-Audio in das virtuelle Instrument zusammensetzt.

In einem ersten Schritt werden Vorschläge zum generellen Ansatz des virtuellen Instruments und zu dem System, in welches sich dieses integriert, gemacht: Ein Vorschlag besteht in der Objektivierung von Klangkomponenten des Syntheseprozesses und deren Panoramisierung innerhalb

des virtuellen Instruments. Zudem wird ein objektbasierter, formatagnostischer Output des Instruments vorgeschlagen, an den sich ein ausgelagertes 3D-Audio Rendering anschließt.

In einem zweiten Schritt werden Vorschläge zu konkreten Funktionen dieses Wavetable Synthesizers gemacht: Es werden polyphone Voices und mehrere unabhängigen Oszillatoren als Klangkomponenten für die Objektifizierung vorgeschlagen. Diese Klangkomponenten ergeben sich aus dem in Kapitel 2.1 beschriebenen Aufbau des Klangsyntheseblocks. Zudem erfolgt der Vorschlag zur dynamischen Panoramisierung. Die Bewegung der Audioobjekte wird über Modulationssignale des Syntheseprozesses und durch die Eingabe des Musikers gesteuert. Darüber hinaus wird eine Form der Makrobildung für die einfache Zugänglichkeit des Kontrastierens zwischen einem gering ausgedehnten, klar lokalisierbaren Klang und einer breiten Ausdehnung dieses Klangs vorgeschlagen.

Diese Vorschläge werden als eine Diskussionsgrundlage begriffen und sollen zum Zwecke ihrer Bewertung mit Hilfe von Experteninterviews evaluiert werden. Daraus resultiert die zweite Forschungsfrage:

Wie bewerten Experten aus den Bereichen Klangerzeugung und Postproduktion die Vorschläge der prototypischen Implementierung?

Im Anschluss an die Auswertung der Experteninterviews wird in der Diskussion ein Stimmungsbild zu den Vorschlägen auf Basis der Aussagen der Experten gezeichnet. Die Erkenntnisse könnten zudem für eine mögliche Weiterführung des Entwicklungsprozesses dienen. Als Experten kommen Anwender und Entwickler aus den Bereichen Klangerzeugung und Postproduktion zu Wort: Dabei handelt es sich um Komponisten & Sounddesigner, Synthesizer-Programmierer, Toningenieure und Postproduktionstool-Entwickler.

4 Prototypische Implementierung

Wie im vorangehenden Kapitel beschrieben, fanden bereits diverse Experimente an der Schnittstelle von Klangsynthese und räumlicher Tonwiedergabe statt. Dabei handelte es sich meist um komplexe Anwendungen, die für Kompositionen im Bereich experimenteller elektroakustischer und akusmatischer Musik entstanden sind. Die Fragestellung dieser Arbeit zielt darauf ab, wie sich 3D-Audio in ein virtuelles Instrument zur digitalen Wavetable Synthese integrieren lässt. Durch die Integration in eine DAW wird das virtuelle Instrument Workflows zur Produktion populärer Musik zugänglich gemacht. Abbildung 8 zeigt schematisch den vorgeschlagenen Ansatz zur Integration von 3D-Audio in das virtuelle Instrument. Dieser sieht die Erzeugung von Audioobjekten im virtuellen Instrument und das 3D-Audio Rendering in einem separaten Plug-In innerhalb der DAW vor.

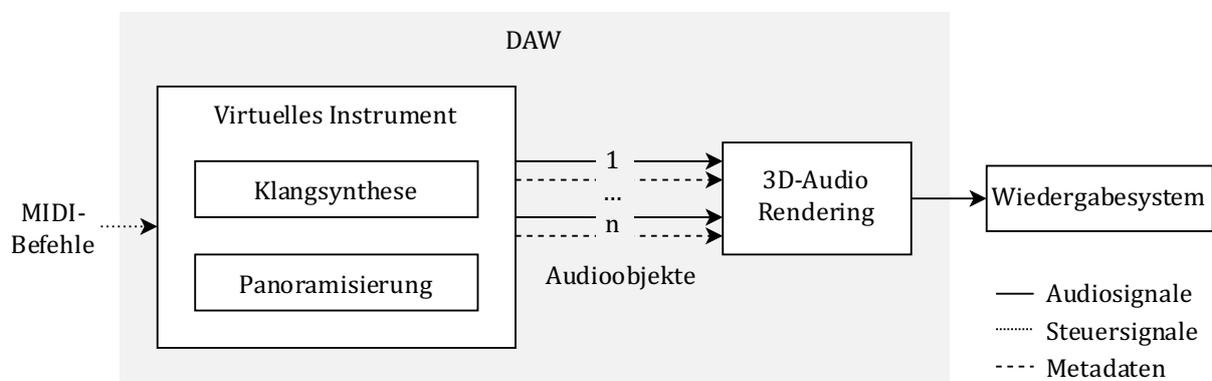


Abbildung 8: Blockschaltbild zum schematischen Aufbau eines virtuellen Instruments mit objektbasiertem 3D-Audio Output.

Den Input des Instruments stellen MIDI-Befehle dar. Das virtuelle Instrument erzeugt Audiosignale über die Klangsynthese und die zugehörigen Metadaten in der Panoramisierung. Diese stellen den objektbasierten Output des virtuellen Instruments dar. Die Metadaten bestehen aus Richtungsinformationen der Audiosignale und repräsentieren deren Panoramisierung auf der Kugeloberfläche. Durch den objektbasierten Output wird eine weitestgehend formatagnostische ausgangsseitige Schnittstelle vorgeschlagen. In der Theorie können die Daten der Panoramisierung von unterschiedlichen 3D-Audio-Panning-Tools umgesetzt werden – in der Praxis muss dieser einseitig über eine offene Schnittstelle verfügen. Der genaue Ablauf der Umsetzung der Panoramisierung im 3D-Audio Rendering ist abhängig vom verwendeten Format. Das bei der Implementierung dieser Arbeit verwendete Verfahren wird in Kapitel 4.1.3 beschrieben.

4.1 Systemintegration

Die Umsetzung des virtuellen Instruments erfolgt in der grafischen Programmierumgebung *Max 8*. Die Entwicklung mit *Max 8* hat in der frühen, prototypischen Entwicklungsphase den

Vorteil, dass auf bestehende Blöcke zur Signalverarbeitung zurückgegriffen werden kann und in der begrenzten Zeit die prototypische Entwicklung eines virtuellen Instruments mit erforderlichem Komplexitätsgrad möglich ist. Das in *Max 8* entwickelte virtuelle Instrument kann als *Max for Live* (M4L) Device in die DAW *Ableton Live 10* inseriert werden. Das finale M4L-Device befindet sich im digitalen Anhang dieser Arbeit (vgl. A3) und kann auf Anfrage zur Verfügung gestellt werden. Da *Ableton Live 10* ausschließlich über Stereo-Audiospuren verfügt und damit nicht 3D-Audio fähig ist, wird das 3D-Audio auf eine zweite DAW, *Cockos Reaper*, ausgelagert. *Reaper* unterstützt Spuren mit einer Bussbreite von bis zu 64 Kanälen und ist damit 3D-Audio fähig. Abbildung 9 zeigt die spezifische Integration des virtuellen Instruments, im Folgenden ‚Spatial Wavetable Synthesizer‘ (SWS), in seine Peripherie.

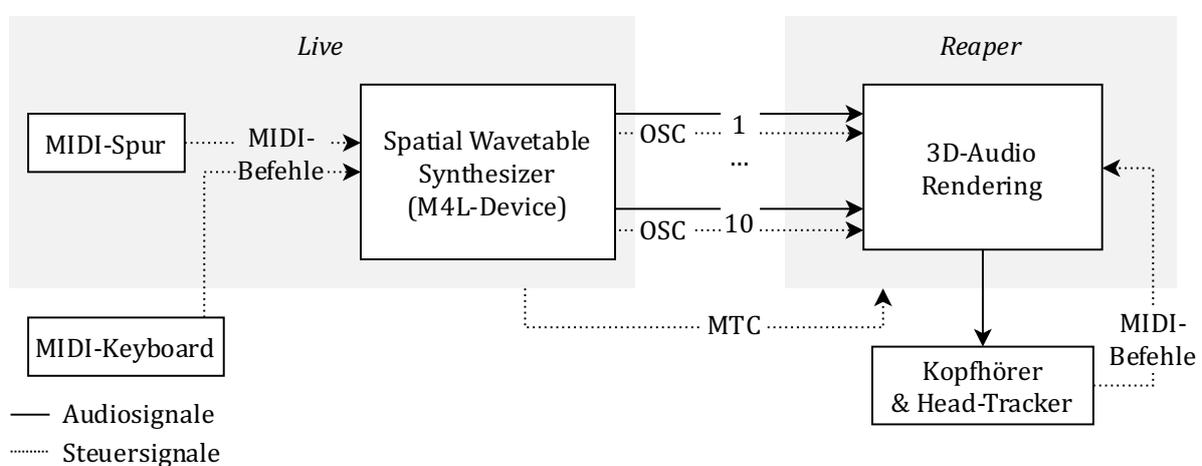


Abbildung 9: Blockschaltbild zur Integration des SWS in seine Peripherie.

Die eingangsseitigen MIDI-Befehle (Kapitel 4.1.1) können von einem externen MIDI-Keyboard oder einer MIDI-Spur in *Live* an den SWS gesandt werden. Die im SWS erzeugten Audioobjekte (Kapitel 4.1.2) werden für das 3D-Audio Rendering in *Reaper* weiterverarbeitet. Das 3D-Audio Rendering sieht eine kopfbezogene Wiedergabe vor und erfolgt über das szenenbasierte Ambisonics Format unter Verwendung mehrerer Plug-Ins. Der Ablauf wird in Kapitel 4.1.3 näher beschrieben. Die DAWs *Live* und *Reaper* sind über MIDI Timecode (MTC)²¹ synchronisiert. *Live* dient dabei als Master, *Reaper* als Slave.

Als Zeitvariable wird t definiert. Aus Gründen der einfacheren Darstellung werden im Folgenden alle Signale kontinuierlich dargestellt. In der Implementierung handelt es sich um diskrete Signale, bei denen verschiedene Abtastraten zum Einsatz kommen. Die in den Blockschaltbildern als Audiosignale gekennzeichneten Signale, werden immer mit ‚Audio Rate‘ übertragen. Die Steuersignale können in Abhängigkeit ihrer spezifischen Verwendung mit unterschiedlichen Raten

²¹ MIDI Timecode (MTC): Über MIDI-Befehle austauschbare Zeitinformationen für die Synchronisation mehrere MIDI-Geräte oder Anwendungen (MMA, 1996, S. 27f.).

übertragen werden. Für die Erläuterung der Funktionsweise anhand der Blockschaltbilder ist die exakte Rate der Steuersignale zweitrangig. In *Max 8* werden Signale entweder mit einer für Audio typischen Abtastfrequenz von 48kHz („Audio Rate“) zeitlich erfasst oder mit einer Abtastfrequenz von 1kHz („Control Rate“) (Cycling '74, 2020). Für *Max*-spezifische Details wird auf die Patches im digitalen Anhang verwiesen. Steuersignale wie MIDI-Befehle und OSC-Daten werden ebenfalls mit niedrigeren Abtastraten übertragen.

4.1.1 MIDI-Befehle

Das MIDI (Musical Instrument Digital Interface) Protokoll wurde 1983 als Spezifikation für die Ansteuerung und Verbindung von elektronischen Musikinstrumenten veröffentlicht. Trotz der stark begrenzten Auflösung ist es immer noch ein Standard für konventionelle Anwendungen (Lehrman & Tully, 2017, S. 4). Die Spezifikationen des Protokolls sind in *MIDI 1.0 Detailed Specification* ausführlich beschrieben (MMA, 1996).

Von acht verfügbaren Voice Messages, bestehend aus jeweils zwei Daten Bytes, finden Note On, Note Off, Control Change und Pitch Bend im SWS Einsatz. Der Note On Befehl wird beispielsweise beim Drücken einer Taste auf dem MIDI-Keyboard erzeugt und gibt Auskunft über die gespielte Note. Dessen erstes Daten Byte, die Note Number, beschreibt die Tonhöhe. Dieser Wert wird im Folgenden durch $n \in [0; 127]$ repräsentiert. Das zweite Daten Byte, die Note On Velocity, beschreibt die Geschwindigkeit, mit der die Taste gedrückt wurde. Dieser Wert wird im Folgenden durch $v \in [0; 127]$ repräsentiert. Beim Loslassen der Taste wird ein Note Off Befehl mit der zum Note On Befehl identischer Note Number, aber mit Note Off Velocity $v = 0$ gesandt (MMA, 1996, S. 10).

Das MIDI-Protokoll lässt bis zu 120 verschiedene Control Change Befehle zu. Diese bestehen aus der Controller Number und dem dazugehörigen Controller Value. Beim SWS findet das Modulation Wheel Einsatz, dem in den MIDI-Spezifikationen die feste Controller Number 1 (CC01) zugewiesen ist (Lehrman & Tully, 2017, S. 14). Die Modulation Wheel Werte werden im Folgenden durch $m \in [0; 127]$ repräsentiert.

Pitch Bend dient zur globalen Tonhöhenvariation der aktiven MIDI-Noten. Bezüglich der Auflösung stellt Pitch Bend eine Ausnahme dar und wird mit Hilfe beider Daten Bytes über least significant bit (LSB) und most significant bit (MSB) dargestellt. Die $2^{14} = 16384$ Pitch Bend Werte ermöglichen auch bei großen Tonumfängen ein stufenloses Glissando (Lehrman & Tully, 2017, S. 15).

4.1.2 Audioobjekte

Der Output des SWS basiert auf dem objektbasierten Paradigma. Die entstehenden Audioobjekte werden dem prototypischen Charakter der Umsetzung entsprechend in roher Form weiterverarbeitet und nicht in ein standardisiertes objektbasiertes Audioformat wie beispielsweise SpatDIF encodiert. Die im SWS erzeugten Audioobjekte bestehen aus Audiosignalen $s_i(t)$ und Metadaten. i ist der Index des Audioobjekts mit $i \in [1; 10]$ bei zehn möglichen Audioobjekten. Die Metadaten sind simpel strukturiert und enthalten pro Objekt die Richtung in Form von kopfbezogenen Kugelflächenkoordinaten des zugehörigen Audiosignals, welche über die Panoramisierung im SWS berechnet werden. Die Richtung der Objekte wird über den Vektor $\mathbf{p}_i(t) = \begin{pmatrix} \varphi_i(t) \\ \delta_i(t) \end{pmatrix}$ beschrieben, bestehend aus dem Azimut-Winkel $\varphi_i(t)$ und Elevationswinkel $\delta_i(t)$. Bei den Audioobjekten handelt es sich um dynamische Objekte, die ihre Positionen über die Zeit verändern können.

Die im SWS berechneten Richtungsdaten werden in Open Sound Control (OSC) Nachrichten encodiert. OSC ist eine Datenprotokoll Spezifikation für die Echtzeitkommunikation zwischen verschiedenen Anwendungen über Netzwerk. Ursprünglich wurde es als akkurate und vielseitige Methode für die Steuerung musikalischer Echtzeit Performances und die Kommunikation zwischen Computern, Synthesizern und anderen Multimedia Devices entwickelt. OSC kann als flexiblere MIDI-Alternative verstanden werden: es bietet einen offenen, benutzerdefinierten Adressraums, der eine beliebige parametrische Steuerung über konventionelle Netzwerktechnik ermöglicht. (Matt, 2021)

Der OSC Stream ist eine Sequenz von Frames, die in Bezug auf einen als Timetag bezeichneten Zeitpunkt definiert sind. Innerhalb der als Bundles bezeichneten Frames, befinden sich eine bestimmte Anzahl von Nachrichten, die jeweils den Zustand eines Sub-Streams zum letzten Referenzzeitpunkt repräsentiert. Die Sub-Streams sind mit einer von Menschen lesbaren Adresse gekennzeichnet. Im Gegensatz zu Audiodaten, die in einem festen zeitlichen Raster abgetastet werden, bauen sich die Bundles flexibel aus den Sub-Streams auf, die mit variablen Raten abgetastet werden (Schmeder et al., 2010, S. 1f.).

Die Syntax eines ausgehenden OSC Sub-Streams besteht aus einem Adress- und einem Datenteil und baut sich für das SWS spezifische 3D-Audio Rendering wie folgt auf:

`/<PluginName>/<command>/ <arg1> <arg2> ...`

Für das Audioobjekt 1 in Blickrichtung mit $\mathbf{p}_1 \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ werden folgende OSC-Nachrichten vom SWS an das *MultiEncoder* Plug-In, das die erste Instanz der 3D-Audio Rendering-Kette darstellt, gesandt.

```
/MultiEncoder/azimut1 0.
```

```
/MultiEncoder/elevation1 0.
```

Der OSC Stream wird über das User Datagram Protocol (UDP) übertragen. Dafür wird im SWS die IP-Adresse des Zielrechners (in diesem Fall der Host Rechner mit Loopback Adresse 127.0.0.1) und der Port für den Empfang der Datenpakete spezifiziert. Im SWS sind Send Intervalle zwischen *1ms* und *200ms* einstellbar, die damit die Abtastrate für diese Steuersignale festlegt.

Die im SWS erzeugten Audiosignale müssen von *Ableton Live* nach *Reaper* geroutet werden. Dabei kommt die virtuelle Soundkarte *Blackhole* zum Einsatz worüber sich Audiosignale über bis zu 16 separate Kanäle zwischen zwei Anwendungen austauschen lassen (Roth, 2021). In *Reaper* wird eingangsseitig die virtuelle Soundkarte und ausgangsseitig für den Anschluss der Kopfhörer eine Soundkarte mit physischen Ausgängen benötigt. Dafür wird auf die *MacOS* spezifische Funktion der Bildung eines ‚Aggregate Device‘ im Audio MIDI Setup zurückgegriffen. Es lassen sich hier zwei oder mehr Soundkarten zu einer aggregierten Soundkarte virtuell vereinen. Die Ein- und Ausgänge der beiden aggregierten Soundkarten sind dann simultan nutzbar. Eine der Soundkarten wird dabei als Clock Source bestimmt. Die aggregierte Soundkarte kann nun in *Live* und *Reaper* ausgewählt und sowohl die Ein- und Ausgänge von *Blackhole* als auch die physischen Ausgänge der verwendeten Hardware Soundkarte genutzt werden.

Die Audiosignale und Richtungsdaten können in *Reaper* aufgezeichnet werden. Die eingehenden Audiosignale werden dabei als WAV-Dateien auf einer Spur aufgenommen. Die Richtungsdaten werden als Automationsdaten in die korrespondierende Audiospur geschrieben. So könnten die Metadaten der Audioobjekte auch nach der Erzeugung noch bearbeitet werden. Allgemein ist eine zwischen den DAWs bestehende Latenz zu beachten, die sich auf die Audiopuffer-Größe der Hardware-Soundkarte im ‚Aggregate Device‘ zurückführen lässt.

4.1.3 3D-Audio Rendering

Das 3D-Audio Rendering wird mit Plug-Ins der *IEM Plug-In Suite* realisiert. Dabei handelt es sich um Open-Source-Software für die Arbeit mit Ambisonics bis zu siebter Ordnung (Rudrich, 2021). Die Plug-Ins sind frei verfügbar und vollständig über OSC steuerbar. Sie eignen sich daher explizit für die Anbindung an den SWS, der auf eine offene Schnittstelle dieser Art angewiesen ist.

Abbildung 10 zeigt schematisch den Signalfluss des implementierten 3D-Audio Renderings. Dieses hat zum Ziel, die Abhörsituation des 3D-Audio-Studios ‚Produktionsstudio‘ über Faltung mit Hilfe dort gemessener binauraler Raumimpulsantworten zu virtualisieren.

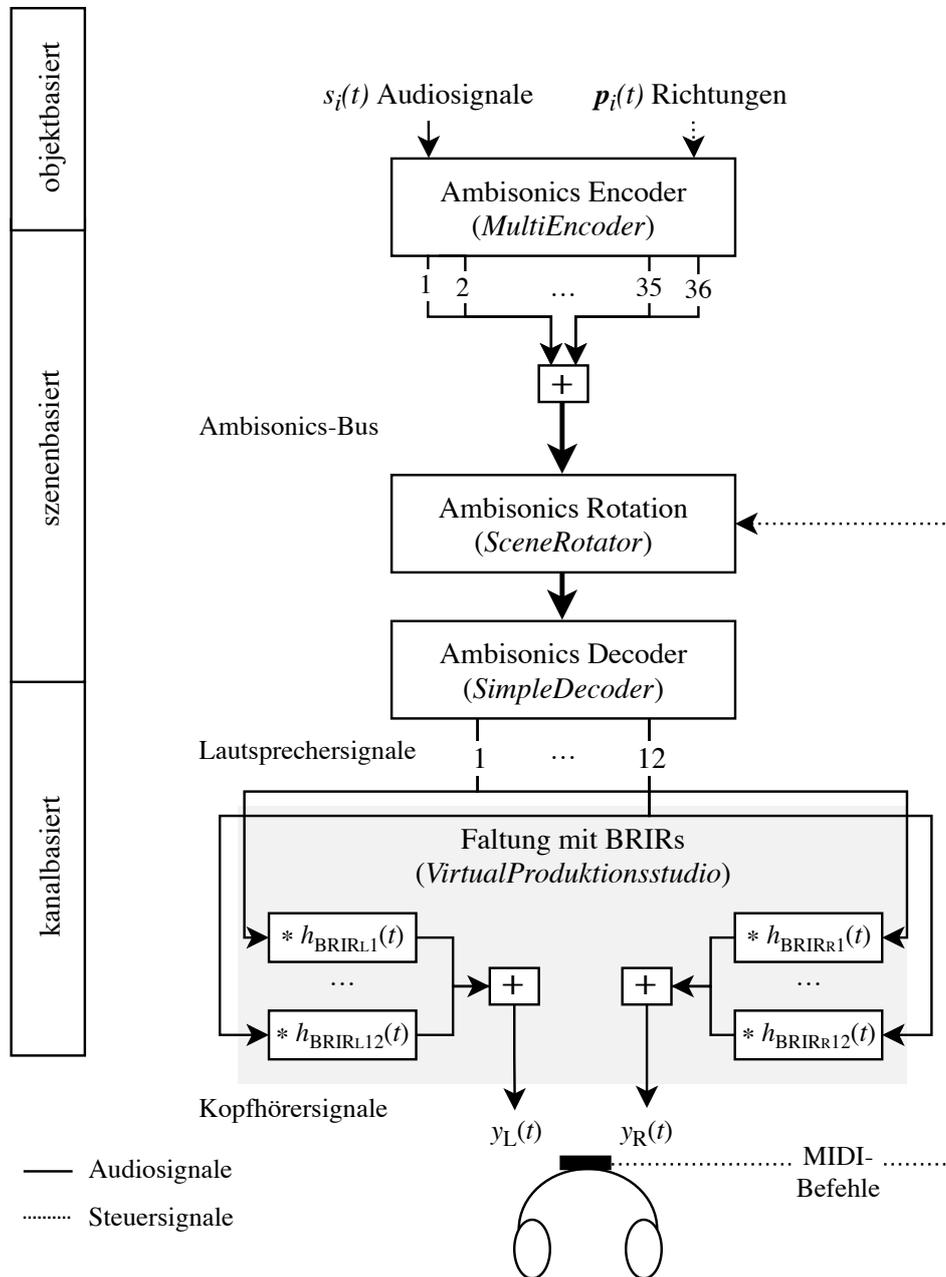


Abbildung 10: Blockschaltbild des implementierten 3D-Audio Renderings mit Higher Order Ambisonics Workflow.

Das 3D-Audio Rendering erhält dem objektbasierten Output des SWS entsprechend Audiosignale und Richtungsdaten. In einem ersten Schritt werden die im SWS erzeugten Audiosignale $s_i(t)$ in Abhängigkeit der errechneten und über OSC gesandten Richtungen $p_i(t)$ mit Hilfe des *MultiEncoder* Plug-Ins in eine Ambisonics-Szene encodiert. Alle richtungencodierten Signale werden auf einen mehrkanaligen Ambisonics-Bus summiert, der die gesamte Klangszene beinhaltet. Die Breite des mehrkanaligen Busses wird durch die Ordnung N über $(N + 1)^2$ bestimmt und steht

in Abhängigkeit zur gewünschten Richtungsauflösung (Frank et al., 2015, S. 2). Die Encodierung erfolgt mit $N = 5$ entsprechend in 36 Kanäle, was einer ungefähren Richtungsauflösung von 30 Grad entspricht (Frank & Zotter, 2020, S. 28). Für Grundlagen und Details zum Thema Ambisonics wird auf Zotter & Frank 2019 verwiesen.

Die Realisierung der dynamischen Variante der Binauralsynthese erfolgt über eine Rotation der Ambisonics-Szene invers zur Kopffrotation. Das zur Kopfbewegung inverse Nachführen der Ambisonics-Szene ist gegenüber einer statischen Binauralisierung wahrnehmungstechnisch ein Zugewinn (Roginska & Geluso, 2018, S. 101ff.). Dies lässt sich unter anderem über die in Kapitel 2.2 beschriebenen Peilbewegungen begründen und kann zudem auch zu einer verbesserten Externalisierung führen (ebd.). Die Kopfbewegung wird über den am Kopfhörer montierten DIY-Head-Tracker *MrHeadTracker* erfasst. Der Head-Tracker besteht aus dem 9-DOF Sensor *BNO055* und einem *Arduino Micro* und entstammt einem Entwurf von Michael Romanov. Mit Hilfe des im Sensor vorhandenen Gyroskops und Beschleunigungsmessers kann die Kopffrotation relativ zur Ruhelage berechnet werden. Der Head-Tracker wird vom *SceneRotator* Plug-In als MIDI-Gerät erkannt. Quaternionen oder Euler-Winkel werden in 14-bit MIDI-Befehlen, analog zu LSB/MSB Pitch Bend, für eine feinere Auflösung der Rotationsdaten, übertragen (Romanov, 2017).

Anschließend wird ein Decoder eingesetzt, der aus dem szenenbasierten Ambisonics-Format die Lautsprecher-signale für das konkrete Lautsprecherlayout erzeugt (Frank et al., 2015, S. 5f.). Bei dem Lautsprecherlayout handelt es sich um ein 13-kanaliges Lautsprechersystem in dem als ‚Produktionsstudio‘ bezeichneten Abhör-raum. Eine Übersicht über das Lautsprecherlayout des Studios findet sich im Anhang A2. Das *SimpleDecoder* Plug-In kommt hier als Decoder zum Einsatz und basiert auf dem All-Round Ambisonic Decoding Ansatz (AllRAD) (Zotter & Frank, 2019, S. 92f.). Die Koordinaten der Lautsprecher sowie deren Kanalnummern werden in das *SimpleDecoder* Plug-In eingelesen. Hinsichtlich des Bass-Managements ist im *SimpleDecoder* Plug-In eine Trennfrequenz und ein Gain-Wert für das Tieftonlautsprecher-Signal einstellbar. Für den Tieftonlautsprecher wird der tiefpassgefilterte erste Kanal des Ambisonics-Busses, der omnidirektionale Kanal, verwendet (Zotter & Frank, 2019, S. 111).

Der Übertragungsweg von den Lautsprechern im 3D-Audio-Studio zum Hörer kann als ein lineares zeitinvariantes System (LTI-System) betrachtet werden und lässt sich daher mit einer binauralen Raumimpulsantwort (BRIR) abbilden (Roginska & Geluso, 2018, S. 92f.). Die Impulsantworten der dreizehn Lautsprecher wurden über den Kunstkopf *Neumann KU100* im Sweet Spot des 3D-Audio Studios ‚Produktionsstudio‘ gemessen. Für die Virtualisierung dieses Wiedergabesystems über Kopfhörer werden die zwölf Lautsprecher-signale mit den entsprechenden BRIRs

gefaltet und ergeben die binauralen Kopfhörersignale. Die Faltung erfolgt innerhalb des in *Reaper* insertierten Plug-Ins *VirtualProduktionsstudio*.

Für die Binauralisierung hätte man nicht zwangsläufig den Weg über die Decodierung und Faltung gehen müssen. Es besteht auch die Möglichkeit aus dem Ambisonics-Format direkt binaurale Signale zu berechnen, wie es beispielsweise im *BinauralDecoder* Plug-In implementiert ist (Schoerhuber et al., 2018). Da während der Umsetzungsphase aus Gründen der Machbarkeit nicht in einem 3D-Audio Studio abgehört werden konnte, wurde dieses über die Faltung mit den BRIRs virtualisiert. Die Virtualisierung des ‚Produktionsstudios‘ bildet auch dessen Räumlichkeit, sprich dessen akustische Umgebung ab. Die Akustik des ‚Produktionsstudios‘ ist einem professionellen Studio entsprechend optimiert, mit kurzer Nachhallzeit und nach subjektiver Beurteilung wenig färbend. Die Reflexionen und der dezente Nachhall des Raums sorgen für eine bessere Externalisierung bei der binauralen Wiedergabe und können einer In-Kopf-Lokalisation entgegenwirken (Li et al., 2019).

Abschließend wird darauf hingewiesen, dass die Rotation der Ambisonics-Szene nicht einer Kopfrotation im ‚Produktionsstudio‘ gleichkommt. Mit einer Kopfdrehung im ‚Produktionsstudio‘ verändert sich der Bezug zwischen Lautsprechern und Hörer und damit auch das LTI-System. Die verwendeten BRIRs stellen einen statischen Bezug zwischen Lautsprechern und Kunstkopfmikrofon her. Die Rotation dreht vielmehr die Abbildung der Ambisonics-Szene auf dem Lautsprecher-Wiedergabesystem. Dies ist allerdings nicht weiter störend, da das Verfahren wahrnehmungstechnisch überzeugt.

4.2 Spatial Wavetable Synthesizer

Nachdem die Systemintegration des SWS beschrieben wurde, dokumentiert dieses Kapitel die Implementierung des ‚Spatial Wavetable Synthesizers‘. Zwei wesentliche Vorschläge charakterisieren den SWS und unterscheiden ihn von konventionellen Mono- oder Stereo-Synthesizern:

- Zum einen die Objektivierung von Klangkomponenten des Klangsyntheseblocks in Form polyphoner Voices und unabhängiger Oszillatoren (Layer). Die zwei Varianten werden fortan ‚Voice-Mode‘ und ‚Layer-Mode‘ genannt.
- Zum anderen die Möglichkeit zur dynamischen Panoramisierung der Audioobjekte durch Modulationssignale und MIDI-Befehle sowie eine Form der Makrobildung für das einfache Steuern räumlicher Ausdehnung. Die Makrobildung wird im Folgenden als ‚Anchor-Mode‘ bezeichnet.

Der Aufbau des SWS gliedert sich in zwei Blöcke: die Klangsynthese und die Richtungsberechnung (vgl. Abbildung 11). Zwei Hüllkurven sowie ein LFO sind als Modulatoren Teil des Klangsyntheseblocks. Die Modulatoren werden in Kapitel 4.2.1 näher beschrieben. In Abhängigkeit des gewählten Modus werden entweder polyphone Voices oder unabhängige Oszillatoren der über das Voice Management instanziierten Klangsyntheseblöcke als Klangkomponenten objektifiziert. Unabhängig vom Modus sind jeweils bis zu zehn Audioobjekte möglich. Die Umsetzung des Klangsyntheseblocks sowie die Objektifizierung der Klangkomponenten wird in Kapitel 4.2.2 näher beschrieben.

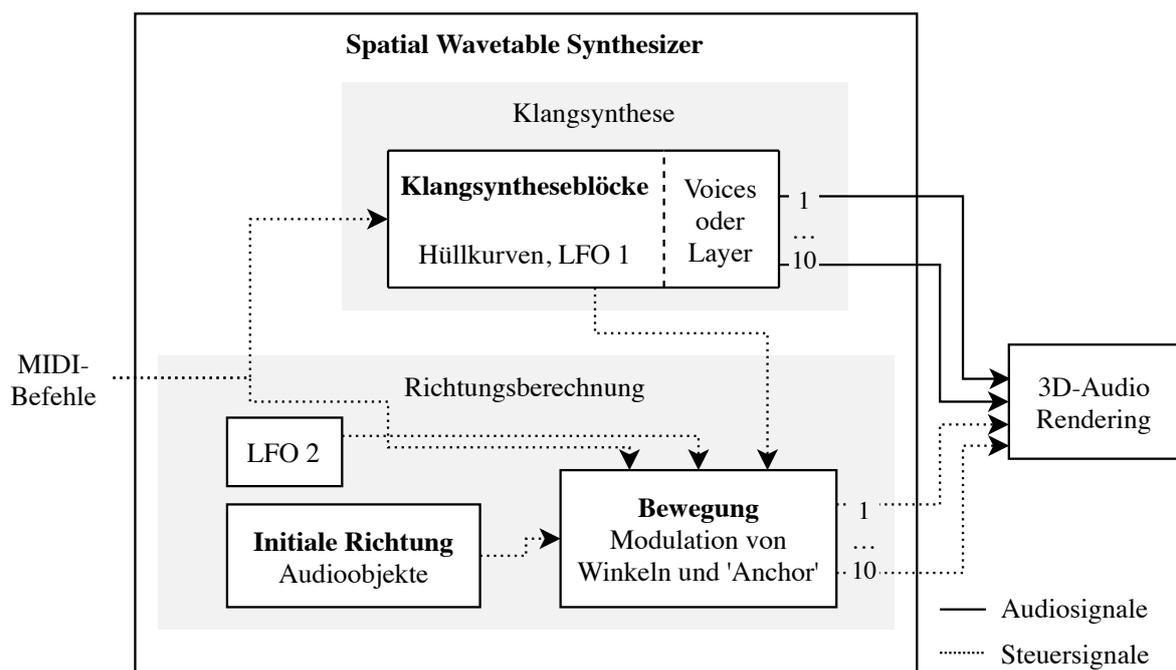


Abbildung 11: Blockschaltbild des schematischen Aufbaus des SWS mit Klangsynthese und Richtungsberechnung.

Den Audioobjekten wird eine initiale Richtung zugewiesen. Die Bewegung der Audioobjekte wird über Modulation ihrer richtungsbeschreibenden Winkel durch Hüllkurven- und LFO-Signale sowie MIDI-Befehle und die Makroform 'Anchor-Mode' realisiert. Die Umsetzung der Richtungsberechnung wird in Kapitel 4.2.3 näher beschrieben. In Kapitel 4.2.4 wird abschließend der Aufbau der graphischen Benutzeroberfläche gezeigt und eine Verbindung zu den zuvor beschriebenen Funktionen hergestellt.

4.2.1 Modulatoren

Zwei Hüllkurven-Generatoren und zwei LFOs sind die Modulatoren im SWS. Deren Signale dienen zum einen zur Modulation klangformender Parameter im Klangsyntheseblock (z.B. Grenzfrequenz des Tiefpassfilters), zum anderen zur Modulation der Richtung der Audioobjekte (z.B. über Azimut-Winkel).

Hüllkurven

Die zwei Hüllkurven-Generatoren sind Teil des Klangsyntheseblocks. Deren Instanziierung erfolgt somit dynamisch über das Voice Management. Der zeitliche Ablauf des Hüllkurven-Signals gliedert sich in die vier Segmente Attack, Decay, Sustain und Release (vgl. Abbildung 12). Der Note On MIDI-Befehl ist Trigger für die Attack-Phase, deren Dauer, wie auch die Zeiten von Decay- und Release-Phase, über die Benutzeroberfläche festgelegt wird.

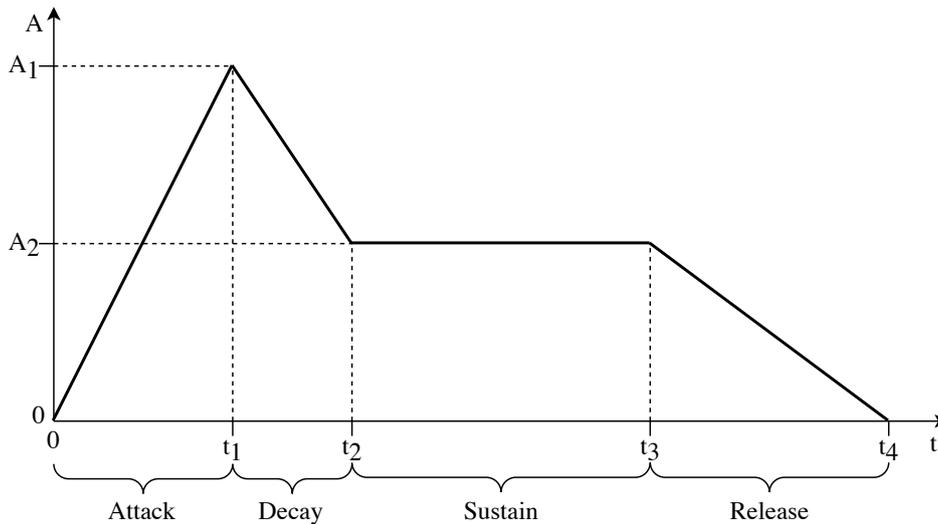


Abbildung 12: Verlauf ADSR-Hüllkurve, Quelle: schematisiert nach Lazzarini, 2019, S. 220.

Das Hüllkurven-Signal kann allgemein über

$$e(t) = \begin{cases} \frac{A_1}{t_1} t, & 0 < t \leq t_1 \\ A_1 - \frac{A_1 - A_2}{(t_2 - t_1)} (t - t_1), & t_1 < t \leq t_2 \\ A_2, & t_2 < t \leq t_3 \\ A_2 - \frac{A_2}{t_4 - t_3} (t - t_3), & t_3 < t \leq t_4 \end{cases} \quad (1)$$

beschrieben werden.

Der maximale Anstieg A_1 der Funktion ergibt sich aus der Velocity v einer Voice mit $A_1 = 1/(v/127)$. Das Sustain-Level A_2 ist relativ zu A_1 und ergibt sich über den in der Benutzeroberfläche festgelegten Sustain-Wert $b_s \in [0; 1]$ mit $A_2 = A_1 b_s$. Der Note Off MIDI-Befehl startet die Release-Phase und das Hüllkurvensignal sinkt über die Release-Zeit zurück auf 0 (Lazzarini, 2019, S. 220; Puckette, 2007, S. 89ff.).

Innerhalb jedes Klangsyntheseblocks wird eine Amplituden-Hüllkurve $e_a(t)$ und eine Filter Hüllkurve $e_f(t)$ erzeugt. Wie in Kapitel 2.1 beschrieben, hat die Amplituden-Hüllkurve eine besondere Funktion, da sie über die Multiplikation mit dem Ausgangssignal des Klangsyntheseblocks die

zeitliche Dauer einer Voice begrenzt. Nach Ende der Release-Phase sendet die Amplituden-Hüllkurve eine Mute Nachricht an das Voice-Management und signalisiert die Verfügbarkeit der Voice. Über die Amplituden-Hüllkurve lässt sich der Lautstärkeverlauf einer Note über der Zeit gestalten. Kurze Attack- und Decay-Zeiten mit anschließenden niedrigen Sustain-Werten resultieren in transienten Klängen. Lange Attack Zeiten simulieren lange Einschwingphasen und resultieren tendenziell in sphärischen Klängen.

Die Filter-Hüllkurven sind optionale Modulationssignale für klangformenden Parameter der Klangsynthese und modulieren nicht zwangsläufig die Grenzfrequenz des Tiefpassfilters.

LFOs

Die zwei LFOs im SWS erzeugen periodische Modulationssignale unterschiedlicher Wellenformen im subauditiven Frequenzbereich mit Amplitudenwerten im Bereich [-1;1]. Ein Beispiel für deren Verwendung ist die Erzeugung eines Tonhöhenvibratos über die Modulation des Wertes der MIDI Note Number im Klangsyntheseblock mit einem sinusförmigen LFO-Signal. Ein Beispiel für die dynamische Panoramisierung ist die Modulation des Azimut-Winkels mit einem sägezahnförmigen LFO-Signal, was in einer Rotation um die z-Achse resultiert.

Die Frequenz der LFOs wird in der Benutzeroberfläche entweder als Frequenzwert oder als Tempo relativer Notenwert $q \in \{\frac{1}{48}, \frac{1}{32}, \frac{1}{24}, \frac{1}{16}, \frac{1}{12}, \frac{1}{8}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, 4, 6, 8, 16\}$ angegeben. Letzteres ist im musikalischen Kontext besonders interessant, da sich LFO gesteuerte Modulationen auf das in der DAW vorherrschende Metrum²² beziehen lassen. Dazu fragt das M4L-Device den Tempowert $c \in [30; 220]$ der DAW in ‚Schlägen pro Minute‘ (engl. beats per minute, bpm) ab. Ein ‚Schlag‘ ist dabei als Viertelnote definiert (Raffaseder, 2010, S. 209). Die Frequenz des LFOs errechnet sich aus

$$f_{lfo} = \frac{c}{4 q 60s} . \quad (2)$$

Im SWS gibt es zwei LFOs, deren Signale über $l_1(t)$ für LFO 1 und $l_2(t)$ für LFO 2 beschrieben werden. LFO 1 ist Teil des Klangsyntheseblocks und dessen Instanziierung erfolgt damit dynamisch über das Voice Management. Es besteht eine Retrigger-Option, bei deren Aktivierung die Phasen aller aktiver LFO 1 Instanzen mit jedem neu eingehenden Note On Befehl zurückgesetzt und die LFO-Signale darüber synchronisiert werden. LFO 2 befindet sich außerhalb des Klangsyntheseblocks und ist somit unabhängig vom Voice Management. Während LFO 1 als

²² Das Metrum ist ein zeitlich basiertes Ordnungsprinzip, innerhalb dessen sich ein Rhythmus bewegt (Görne, 2017, S. 80).

Modulationsquelle sowohl für Parameter des Klangsyntheseblocks als auch für die Winkelmodulation der Audioobjekte verfügbar ist, steht LFO 2 nur für die Modulation der Audioobjekt-Koordinaten zur Verfügung.

Für jeden der beiden LFOs kann aus sechs verschiedenen Wellenformen gewählt werden. Die wählbaren Wellenformen der LFOs sind Sinus-, Sägezahn-, invertierter Sägezahn-, Dreieck-, Rechteck-, sowie eine ‚Sample & Hold‘-Wellenform, bei der ein zufälliger Wert für eine arbiträre Zeit gehalten wird.

4.2.2 Klangsynthese

Die Objektifizierung von Klangkomponenten des Syntheseprozesses ist ein grundlegender, in den SWS implementierter Vorschlag. Dabei werden im ‚Voice-Mode‘ polyphone Voices und im ‚Layer-Mode‘ die unabhängigen Wavetable Oszillatoren des Klangsyntheseblocks zu den Audiosignalen der Audioobjekte. Im Folgenden wird zunächst der Aufbau des Klangsyntheseblock des SWS beschrieben und im Anschluss näher auf die Funktionsweise der beiden Modi eingegangen. Abbildung 13 zeigt dessen schematischen Aufbau.

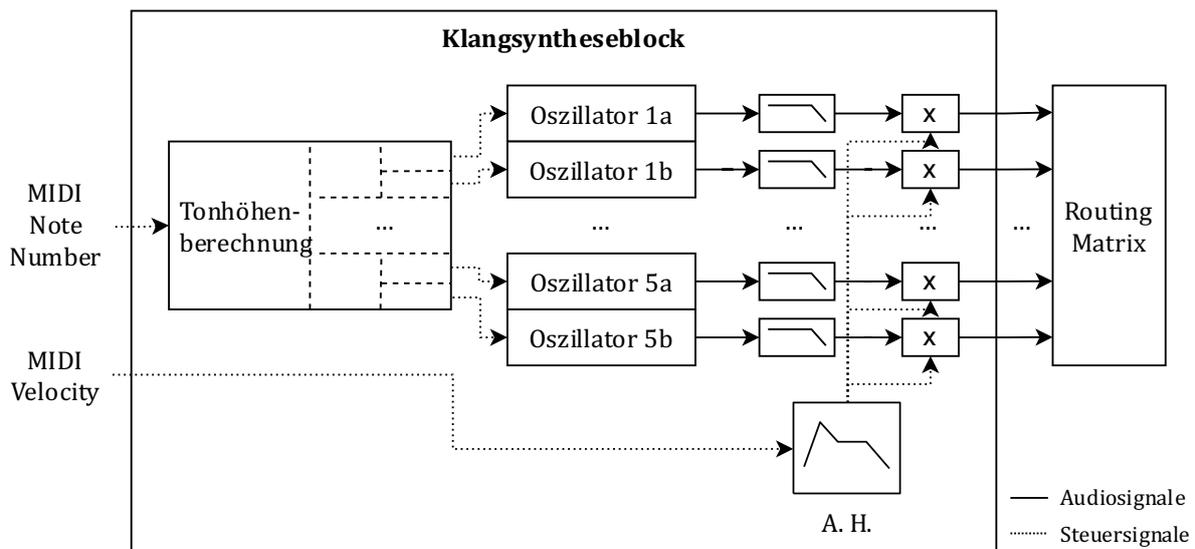


Abbildung 13: Blockschaltbild zum Aufbau des Klangsyntheseblocks im SWS.

Bei den Oszillatoren handelt es sich um 2D-Wavetable Oszillatoren wie in Kapitel 2.1 beschrieben. Die im SWS auswählbaren 2D-Wavetables stammen vom *Live* eigenen virtuellen Instrument *Wavetable* (Ableton AG, 2021). Die WAV-Dateien sind eine Sammlung von über 100 2D-Wavetables verschiedenster klanglicher Ausprägungen, kategorisiert in 11 Gruppen. Die verwendeten 2D-Wavetables haben eine feste Framesize von 1024 Samples und bestehen aus zwischen 4 und 256 Frames pro 2D-Wavetable.

In Anlehnung an den Ansatz des Wavestacking (vgl. Kapitel 2.1) sind in einem Klangsyntheseblock des SWS zehn Oszillatoren enthalten. Diese sind in Zweierblöcke gruppiert (1 bis 5). Jede Oszillator-Gruppe kann ein eigenes 2D-Wavetable laden. Durch das Schichten (engl. layering) unabhängiger Wavetable-Oszillatoren mit unterschiedlichen Wellenformen können komplexe klangliche Texturen erschaffen werden. Jede Oszillator-Gruppe kann zudem in Oktavabständen gegen die globale Tonhöhe transponiert werden. Die zwei Oszillatoren einer Gruppe (a und b) können über eine Tonhöhenverschiebung im Cent-Bereich²³ (im Folgenden auch ‚Detune‘ genannt) zudem leicht gegeneinander verstimmt werden, um eine Dekorrelation der beiden Signale zu erreichen.

Bei den Wavetables handelt es sich mitunter um obertonreiche Wellenformen. Im Rahmen des Sounddesigns ist es wünschenswert, den synthetisierten Klang zu filtern. Die Ausgangssignale der Oszillatoren können daher über ein Tiefpass-Filter spektral geformt werden. Da die einzelnen Signale für die anschließende Objektifizierung separat vorliegen müssen, sind zehn parallele Tiefpassfilter-Instanzen pro Klangsyntheseblock implementiert.

Die Amplituden-Hüllkurve des Klangsyntheseblocks formt alle Oszillatorsignale gleichermaßen. Der Ausgang eines Klangsyntheseblocks besteht somit aus zehn Audiosignalen. Die sich anschließende Routing-Matrix wird im Zusammenhang mit den zwei Modi erläutert. Im Folgenden wird zunächst auf die Tonhöhenberechnung des Klangsyntheseblocks eingegangen.

Tonhöhenberechnung

Die Tonhöhenberechnung (Abbildung 14) folgt in einer Baumstruktur vom Groben ins Feine.

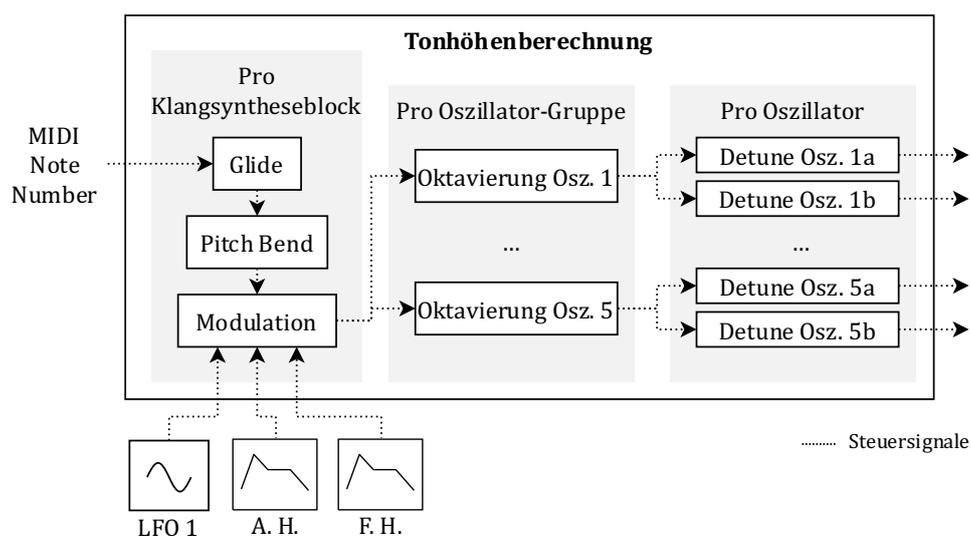


Abbildung 14: Blockschaltbild zur Tonhöhenberechnung im Klangsyntheseblock des SWS.

²³ Ein Cent ist ein Hundertstel eines Halbtonschriffs in der gleichschwebend-temperierten Stimmung. Ein Halbtonschritt hat demnach 100 Cent und eine Oktave 1200 Cent (Sengpiel, 2008).

Zunächst wird eine globale Tonhöhe pro Klangsyntheseblock berechnet. Ausgangspunkt ist die vom Voice Management zugewiesene Note Number. Bei aktivem ‚Glide‘-Mode wird in einem ersten Schritt ein Glissando ausgehend von der zuletzt gespielten Note Number über eine vom Nutzer festgelegte Dauer berechnet. Dazu addiert sich der Pitch Bend Wert, der zuvor auf einen vom Nutzer festgelegten Tonumfang in Halbtonschritten skaliert wurde. Anschließend werden die Modulationssignale in Abhängigkeit ihrer in der Modulationsmatrix festgelegten Gewichtungen mit dem Tonhöhenwert verrechnet. Die Modulation mit den Hüllkurven- und LFO-Signalen wird später näher beleuchtet. Das Zwischenergebnis kann pro Oszillatorgruppe um bis zu ± 3 Oktaven transponiert werden, um bei komplex geschichteten Klängen Differenzierungsmöglichkeiten für das spektrale Sounddesign der Oszillatorgruppen zu geben. Anschließend besteht die Option zur Dekorrelation der Oszillatorsignale einer Oszillatorgruppe über einen leichten ‚Detune‘. Diese ‚Verstimmung‘ der beiden Oszillatoren gegeneinander erfolgt statisch mit bis zu $\pm 20\text{Cent}$ und kann über den ‚Spread‘-Parameter in der Benutzeroberfläche eingestellt werden.

Die gesamte Tonhöhenberechnung erfolgt in der Note Number Domäne mit Fließkommazahlen. Der über die Zeit modulierte Tonhöhenwert $n_{\text{mod}}(t)$ wird mit

$$f_p(t) = 2^{\frac{n_{\text{mod}}(t)-69}{12}} 440 \frac{1}{\text{s}} \quad (3)$$

in die Frequenz für die Ansteuerung des Phasors umgerechnet (Lazzarini, 2019, S. 170). Die MIDI Note Number 69 repräsentiert die Note A_4 (auch Kammerton a') und ergibt $f_p(t) = 440\text{Hz}$. Die Tonhöhenberechnung in der Note Number Domäne hat den Vorteil der linearen Verteilung der Notenwerte über die musikalischen Oktaven hinweg.

Oszillator-Gruppen

Eine Oszillator-Gruppe besteht aus zwei Wavetable Oszillatoren mit gleichem 2D-Wavetable. Wie bereits beschrieben, können die Signale der beiden Oszillatoren über leichten ‚Detune‘ dekorreliert werden. Darüber hinaus bestehen weitere Operationen, die auf die fünf Oszillator-Gruppen des SWS ausgeführt werden können, deren Processing nach den Klangsyntheseblöcken und noch vor dem Modus abhängigen Routing auf die Audioobjekte stattfindet. Bei den Operationen handelt es sich um simple Synthesizer typische Parameter, die dem Sounddesign dienen.

Jede Oszillator-Gruppe kann ein- und ausgeschaltet sowie über einen Gain in der Lautstärke verändert werden. Zudem besteht die Möglichkeit, jede Oszillator-Gruppe um bis zu 50ms zu verzögern, um sich den Präzedenzeffekt im Sounddesign der innerklanglichen Struktur zu Nutze machen zu können (Weinzierl, 2008, S. 103). Darüber hinaus gibt es sogenannte ‚MW Zones‘, die pro Oszillator-Gruppe definiert werden können. Innerhalb einer definierten ‚Zone‘ steigt die

Lautstärke dieser an. Über die Stellung des Modulation Wheel werden die ‚MW Zones‘ ausgelesen. Ist die Modulation Wheel Position unterhalb der Zone einer Oszillator-Gruppe, ist diese nicht hörbar. Im Verlauf der Zone wird die Oszillator-Gruppe lauter. Befindet sich die Modulation Wheel Position oberhalb der Zone einer Oszillator-Gruppe, hat diese ihre maximale Lautstärke erreicht. Die ‚MW Zones‘ sind somit eine Art makroskopisches Bedienelement zur Steuerung der Lautstärkeverläufe einzelner ‚Layer‘.

Im Folgenden wird das Modus abhängige Routing der Klangkomponenten auf die Audioobjekte dargestellt.

‚Voice-Mode‘

Im ‚Voice-Mode‘ werden die Ausgangssignale eines Klangsyntheseblocks zusammengemischt und ergeben das Signal einer polyphonen Voice, die sich aus den Signalen aller potenziell aktiven Oszillatoren zusammensetzt. Die Voices werden in der Reihenfolge der Voice IDs über eine Routing Matrix auf die Audioobjekte verteilt. Abbildung 15 zeigt das Routing exemplarisch mit drei Audioobjekten. Die Anzahl der Audioobjekte kann über die Benutzeroberfläche des SWS verändert werden.

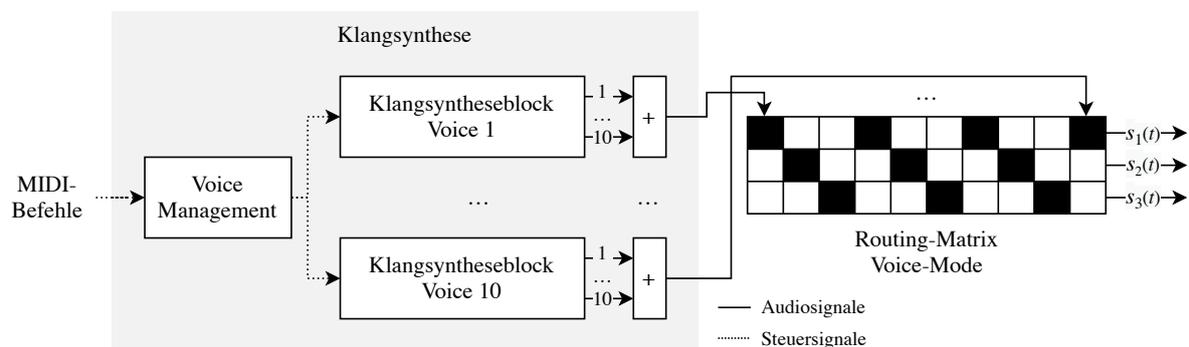


Abbildung 15: Blockschaltbild zum Routing der Klangkomponenten im ‚Voice-Mode‘ am Beispiel von drei Audioobjekten.

Aus der Verteilung einzelner polyphoner Voices im Raum resultiert eine räumliche Verteilung der harmonischen Struktur des Klangs und lässt eine Form der Raumharmonik entstehen (Stockhausen, 1978, S. 381).

‚Layer-Mode‘

Während in den meisten Synthesizern die einzelnen Oszillatoren des Klangsyntheseblocks auf einen Mono- oder Stereoausgang zusammengemischt werden, können sie beim SWS im ‚Layer-Mode‘ in der Benutzeroberfläche frei auf Audioobjekte geroutet werden. Dabei werden im Gegensatz zum ‚Voice-Mode‘ nicht die Ausgangssignale eines Klangsyntheseblocks summiert, sondern

die jeweils korrespondierenden Oszillatoren über die zehn Instanzen des Klangsyntheseblocks hinweg aufsummiert. Ein Audioobjekt im Layer-Mode enthält somit alle gespielten musikalischen Noten, aber nur die Audiosignale eines ‚Oszillator-Typs‘ mit gleichem Wavetable (z.B. des Oszillators a1 vgl. Abbildung 13). Abbildung 16 zeigt schematisch die Verteilung der über Voices hinweg summierten Wavetable Oszillatoren exemplarisch auf fünf Audioobjekte. In dem Beispiel werden die Signale der Oszillator-Gruppe 1 auf die Audioobjekte 1 und 2, die der Gruppe 3 auf die Audioobjekte 3 und 4 und ein Oszillator der Gruppe 5 auf Audioobjekt fünf geroutet.

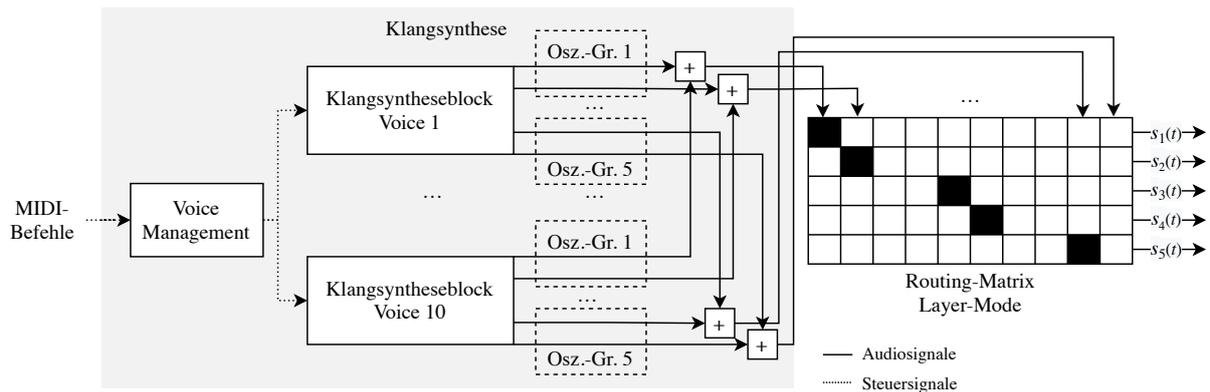


Abbildung 16: Blockschaltbild zum Routing der Klangkomponenten im ‚Layer-Mode‘ am Beispiel von fünf Audioobjekten.

Die Verteilung einzelner Oszillator-Signale mit unterschiedlichen Wavatables kann zu einer räumlichen Verteilung der innerklanglichen Struktur des komplex gelayerten Klangs und damit zu einer Art dessen räumlichen Auffächerns auf Timbre-Ebene führen.

Modulation klangformender Parameter

Aus Gründen der Übersichtlichkeit wurde in den bisherigen Darstellungen des Klangsyntheseblocks weitestgehend auf die Modulatoren verzichtet und der Fokus auf die Objektifizierung der Klangkomponenten gelegt. Für das Sounddesign mit einem Synthesizer ist die Modulation klangformender Parameter der Synthese ein wichtiger Bestandteil. Darüber lässt sich Bewegung, Dynamik und Lebendigkeit im Klang erzeugen. Modulationen sind in unterschiedlichen Ausprägungen fester Bestandteil diverser virtueller Instrumente. Die Umsetzung im SWS ist in Anlehnung an das M4L-Device *Poli* entstanden (Ableton & Amazing Noises, 2020).

Abbildung 17 zeigt den Klangsyntheseblock des SWS ergänzt um die Modulatoren und aus Gründen der Übersicht mit der Reduktion auf eine exemplarische Oszillatorkette 1a. Alle Modulationen werden in der Implementierung gleichermaßen auf alle Oszillatorketten angewandt. Als Quellensignale der Modulation dienen die in Kapitel 4.2.1 beschriebenen Amplituden-Hüllkurve, Filter-Hüllkurve und LFO 1 sowie die MIDI-Befehle Note Number, Velocity und

Modulation Wheel. Modulationsziele sind die Tonhöhe, die Wavetable Position, die Tiefpassfilter-Grenzfrequenz sowie die Amplitude des Ausgangssignals eines Klangsyntheseblocks.

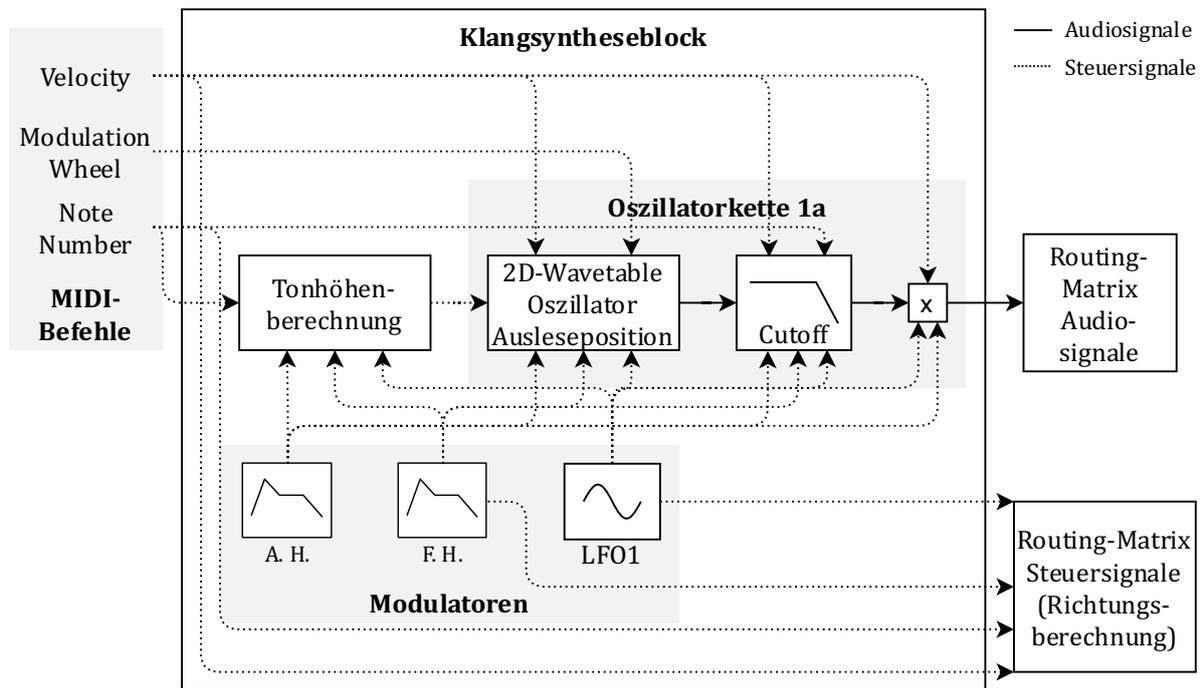


Abbildung 17: Blockschaltbild zum Klangsyntheseblock des SWS mit Modulatoren und MIDI-Befehlen unter exemplarischer Darstellung einer Oszillatorkette.

Jede Instanz des Klangsyntheseblocks gibt neben dem synthetisierten Audiosignal die Signale ihrer Filter-Hüllkurve, des LFO 1 sowie den Velocity und Note Number Wert aus. Diese werden zur Modulation in der Richtungsberechnung des SWS weiterverarbeitet und analog zu den Audiosignalen in Abhängigkeit des gewählten Modus über eine Routing-Matrix dem entsprechenden Audioobjekt zugeteilt (vgl. Abbildung 15, Abbildung 16). Velocity und Note Number sind zwar keine im Klangsyntheseblock erzeugten Werte, werden aber über das Voice Management einem spezifischen Klangsyntheseblock (einer Voice) zugewiesen und daher für die Zuordnung zum Audioobjekt ebenfalls mit aus dem Klangsyntheseblock ausgegeben.

Alle Modulationen der klangformenden Parameter sind Teil des Klangsyntheseblocks. Allerdings sind nicht alle Modulationsquellen mit allen Modulationszielen kombinierbar. Für die Auswahl sind sowohl die klanggestalterische Sinnhaftigkeit einer Modulationsoption sowie deren technische Machbarkeit ausschlaggebend. Zu jeder verfügbaren Modulation kann über die Modulationsmatrix auf der Benutzeroberfläche eine Gewichtung in Prozent angegeben werden. Im Klangsyntheseblock werden die entstehenden Wertebereich in Abhängigkeit des Modulationsziel auf sinnvolle, explorativ festgelegt Werte übersetzt. Auf die Darstellung der gewichtenden Faktoren wurde in Abbildung 17 aus Gründen der Übersichtlichkeit verzichtet. Tabelle 1 zeigt die möglichen Modulationsoptionen mit ihren prozentualen Wertebereichen.

	Tonhöhe	2D-W. Osz. Pos.	TPF Grenzfrequenz	Amplitude
Amplituden H.	-100% – 100%	0% – 100%	-100% – 100%	-
Filter H.	-100% – 100%	0% – 100%	-100% – 100%	-
LFO 1	-100% – 100%	0% – 100%	0% – 100%	0% – 100%
Note Number	-	-	0% – 200%	-
Velocity	-	0% – 100%	0% – 100%	0% – 100%
Mod. Wheel	-	0% – 100%	0% – 100%	-

Tabelle 1: Gewichtende prozentuale Faktoren für die Modulationsoptionen der Klangsynthese.

Die Modulationsstärke der Amplituden-Hüllkurve auf die Amplitude des Ausgangssignals kann nicht eingestellt werden, da hier eine harte Verbindung besteht, worüber die Dauer eine Voice zeitlich definiert ist. Darüber hinaus ist anzumerken, dass die Filter-Hüllkurve standardmäßig zu 50% die Tiefpassfilter-Grenzfrequenz moduliert. Die Velocity moduliert zudem standardmäßig zu 100% die Amplitude des Ausgangssignals. Das ist die technische Umsetzung der Anschlagstärke einer MIDI Keyboard Taste auf die Lautstärke des Tons, was für ein organisches Spielgefühl essenziell ist. Bei Änderung des gewichtenden Faktors von Velocity auf Amplitude zu 0%, erklingt unabhängig von der Anschlagstärke jede Note in gleicher Lautstärke.

4.2.3 Richtungsberechnung

Zu jedem Audiosignal $s_i(t)$ eines Audioobjekts wird im SWS über die Zeit eine Richtung auf der Kugeloberfläche berechnet, welche die Panoramisierung des Audioobjekts beschreibt. Die errechneten Richtungen werden zu den Metadaten der Audioobjekte. Die Richtung eines Audioobjekts wird über $\mathbf{p}_i(t)$ beschrieben (vgl. Kapitel 4.1.2). Wie in Kapitel 2.2 begründet, wird auf die Berechnung eines Radius und einer Entfernungsdarstellung verzichtet.

Im Folgenden wird die Richtungsberechnung exemplarisch anhand eines einzelnen Audioobjekts dargelegt und auf den Index zur Nummerierung des Audioobjekts aus Gründen der einfacheren Darstellung verzichtet. Die Berechnung erfolgt für alle Audioobjekte gleichermaßen. Die verwendeten Indizes im Richtungsberechnungsprozess dienen zur Identifikation von Zwischenergebnissen der Berechnung.

Abbildung 18 zeigt den schematischen Ablauf der Richtungsberechnung. Jedem aktiven Audioobjekt wird in der Benutzeroberfläche des SWS zunächst eine initiale Richtung \mathbf{p}_{init} zugewiesen. Daran anschließend erfolgen verschiedene Modulationen und die Translation zu einem Ankerpunkt. Die Koordinaten des resultierenden, über die Zeit veränderlichen Richtungsvektors $\mathbf{p}_{\text{mrt}}(t)$, werden abschließend in OSC-Nachrichten formatiert (vgl. Kapitel 4.1.2).

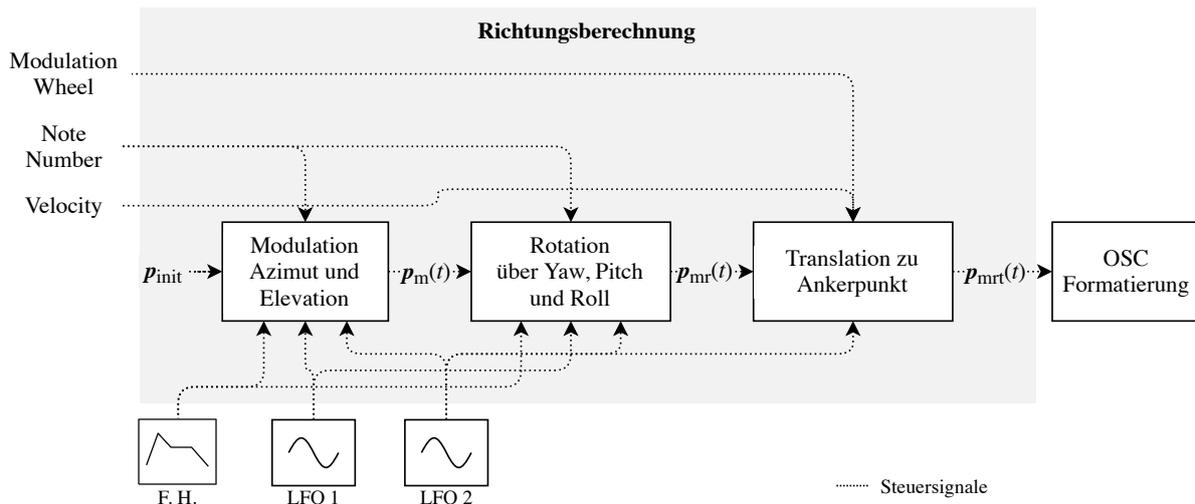


Abbildung 18: Blockschaltbild zum schematischen Ablauf der Richtungsberechnung eines Audioobjekts.

Für die Modulation stehen unter anderem Modulationssignale aus der Klangsynthese zur Verfügung. Dies ist einer der Vorschläge dieser Arbeit zur Verzahnung von Klangsynthese und 3D-Audio. Filter-Hüllkurve, LFO 1, Velocity und Note Number sind Teil des Klangsyntheseblocks und werden von diesem ausgegeben (vgl. Abbildung 17). Velocity, Note Number und Modulation Wheel werden auf den Wertebereich $[0;1]$ skaliert. Die Steuersignale aus den Klangsyntheseblöcken werden analog zu den Audiosignalen über Routing-Matrizen den entsprechenden Audioobjekten zugeordnet. Für den Einsatz in der Richtungsberechnung müssen die Steuersignale in Abhängigkeit des gewählten Modus („Voice-Mode“, „Layer-Mode“) aufbereitet werden.

Im „Voice-Mode“ werden $e_f(t)$, $l_1(t)$ und n pro Audioobjekt summiert und durch die Anzahl der aktiven Voices des zugehörigen Audioobjekts geteilt. Die auf ein Audioobjekt gerouteten Voices werden mit $k \in [1; 10]$ dargestellt. Die Anzahl aktiver Voices pro Audioobjekt wird über $k_{on}(t)$ mit $k_{on} \in [1; k]$ repräsentiert und ist in Abhängigkeit der gespielten Noten über die Zeit veränderlich. Am Beispiel der Abbildung 15 mit drei Audioobjekten, würden für das erste Audioobjekt die Modulationssignale der Voices 1, 4, 7 und 10 summiert und durch die Anzahl der zu einem Zeitpunkt aktiven Voices geteilt werden. Für den LFO 1 ergibt sich das skalierte LFO 1 Signal für ein Audioobjekt

$$l_{1,scaled}(t) = \frac{1}{k_{on}(t)} \sum_{a=1}^k l_1^a(t), \quad (4)$$

wobei $l_1^a(t)$ die LFO 1 Signale eines Audioobjekts beschreibt. Auf den Index des Audioobjekts wird wie eingangs erwähnt aus Gründen der einfacheren Darstellung verzichtet.

Die Berechnung von $e_{f,scaled}(t)$ und n_{scaled} erfolgt analog dazu. Dadurch werden die Modulationssignale für die Bewegung der Audioobjekte in ähnlichen Wertebereichen gehalten,

unabhängig davon, aus wie vielen Voices sich ein Audioobjekt zusammensetzt. Dieses Verfahren hat sich im explorativen Vorgehen dieser Arbeit als sinnvoll erwiesen.

Das Skalierungsverfahren im ‚Layer-Mode‘ unterscheidet sich dahingehend, dass die Modulationssignale aller Voices aufaddiert und durch die Anzahl aller aktiver Voices geteilt werden. Das liegt darin begründet, dass im ‚Layer-Mode‘ jedes Audioobjekt alle Voices enthält, da nicht die Voices, sondern die unterschiedlichen Oszillatorketten auf die Audioobjekte geroutet werden. Da alle Oszillatorketten eines Klangsyntheseblocks dieselben Modulationssignale verarbeiten, ergeben sich daraus für die Richtungsmodulation aller Audioobjekte ebenfalls identische Modulationssignale.

Die Velocity Werte werden ausschließlich für die Translation zum ‚Anchor‘ verwendet, errechnen sich ‚Mode‘ unabhängig über die Mittelwertbildung der Velocity Werte aller aktiven Voices und ergeben v_{scaled} . LFO 2 erzeugt das Modulationssignal $l_2(t)$ außerhalb der Klangsyntheseblöcke. Dieses bedarf somit keiner Skalierung.

Im Folgenden werden die einzelnen Schritte der Richtungsberechnung für die Audioobjekte näher beschrieben.

Initiale Richtung

Mit der Auswahl der Anzahl an Audioobjekten wird deren initiales Verteilungsmuster auf der Kugeloberfläche festgelegt. Die Richtungen können zum einen über die numerische Eingabe der Azimut- und Elevationswinkel, zum anderen in der graphischen Repräsentation in Form einer equirektangularen Kugelflächenprojektion (Abbildung 19) auf der Benutzeroberfläche festgelegt werden. Die equirektangularen Kugelflächenprojektion (auch Rektangulärprojektion) ist die einfachste Form der Kugelflächenprojektion. Die Verzerrung nimmt ausgehend von der Horizontalebene hin zu den beiden Polen zu (Snyder, 1993, S. 5–8). Bei der graphischen Darstellung der Kugelflächenprojektion wird das Gradmaß verwendet, wohingegen in den folgenden mathematischen Darstellungen auf das Bogenmaß zurückgegriffen wird.

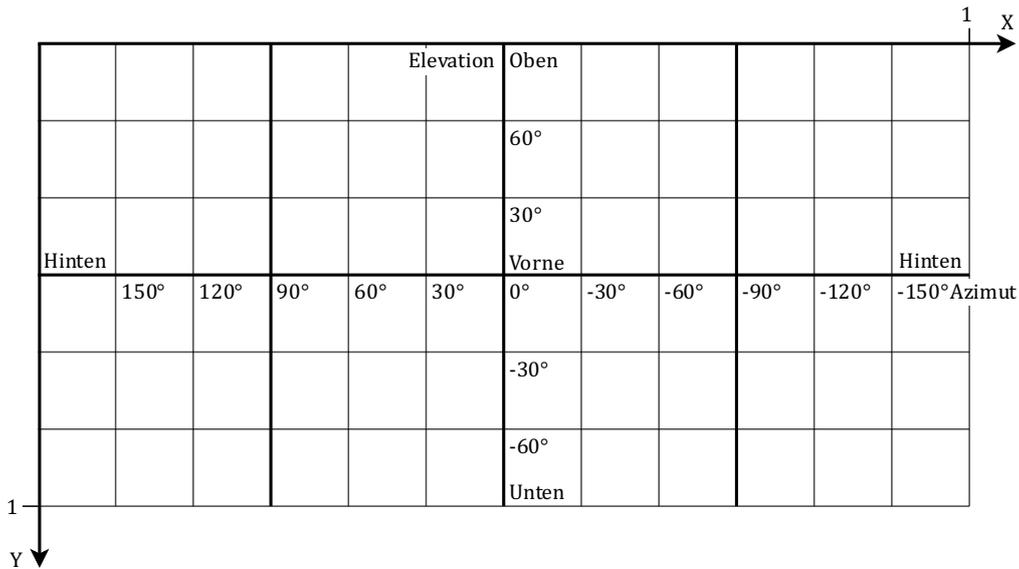


Abbildung 19: Equirektangulare Kugelflächenprojektion und kartesisches 2D-Koordinatensystem, Quelle: schematisiert nach Rudrich, 2021.

Über diese Form der Projektion lässt sich die Kugelfläche auf einfache Art und Weise in einem kartesischen 2D-Koordinatensystem abbilden. Die kartesischen Koordinaten x, y werden über

$$\begin{aligned} \varphi_{\text{init}} &= -2\pi (x - 0.5) \\ \delta_{\text{init}} &= -\pi (y - 0.5) \end{aligned} \tag{5}$$

in die beiden Winkel der Kugelkoordinaten umgerechnet. Daran schließt sich die dynamische Richtungsberechnung über Modulationssignale an.

Wie auch bei der Modulation der klangformenden Parameter im Klangsyntheseprozess, wird die Stärke einer Modulation über Faktoren gewichtet, die sich in einer Modulationsmatrix auf der Benutzeroberfläche einstellen lassen. Die gewichtenden Faktoren werden im Folgenden über $w_{i,j}$ mit $i \in \{f, 1, 2, n, v\}$ für die Modulationsquellen skalierte Filter-Hüllkurve $e_{f,\text{scaled}}(t)$, skaliertes LFO 1 Signal $l_{1,\text{scaled}}(t)$, LFO 2 Signal $l_2(t)$, skalierte Note Number n_{scaled} , skalierte Velocity Werte v_{scaled} und $j \in \{a, e, y, p, r, u\}$ für die Modulationsziele Azimut-Winkel φ , Elevationswinkel δ , die Kardanwinkel Yaw ψ , Pitch θ , Roll ϕ und den Translationsfaktor u beschrieben.

Modulation Azimut und Elevation

In einem ersten Schritt werden Azimut- und Elevationswinkel moduliert und dadurch Bewegungen parallel zur Horizontalebene und vertikale Bewegungen der Audioobjekte durch Zenit und Nadir möglich. Das Zwischenergebnis $\mathbf{p}_m(t)$ errechnet sich über

$$\begin{aligned}\varphi_m(t) &= \varphi_{\text{init}} - \pi \left(e_{f,\text{scaled}}(t) w_{f,a} + l_{1,\text{scaled}}(t) w_{1,a} + l_2(t) w_{2,a} + n_{\text{scaled}} w_{n,a} \right), \\ \delta_m(t) &= \delta_{\text{init}} - \frac{\pi}{2} \left(e_{f,\text{scaled}}(t) w_{f,e} + l_{1,\text{scaled}}(t) w_{1,e} + l_2(t) w_{2,e} + n_{\text{scaled}} w_{n,e} \right).\end{aligned}\tag{6}$$

Um die Bewegungsoptionen zu erweitern, werden die Kugelkoordinaten über

$$\mathbf{p}_m(t) = \begin{pmatrix} \cos(\delta_m(t)) \cos(\varphi_m(t)) \\ \cos(\delta_m(t)) \sin(\varphi_m(t)) \\ \sin(\delta_m(t)) \end{pmatrix}\tag{7}$$

in kartesische Koordinaten transformiert (Papula, 2016, S. 127). Zu beachten ist die verwendete Konvention der 0°-Achse auf Horizontalebene (Polrichtung in Blickrichtung des kopfbezogenen Kugelkoordinatensystems), sprich als geographische Breite mit $\delta \in [-\pi/2; \pi/2]$.

Achsenrotationen mit Kardanwinkeln

Anschließend werden weitere Bewegungen über die Vektor-Matrix-Multiplikation mit dynamischen Rotationsmatrizen realisiert. Dynamisch sind die Rotationsmatrizen aufgrund der Modulation der enthaltenen Kardanwinkeln. Die Motivation für die Rotationen mit Kardanwinkeln liegt darin begründet, dass darüber auch Bewegungen parallel zur Median- und Vertikalebene möglich sind und vertikale Bewegungen nicht zwangsläufig über Zenit und Nadir verlaufen.

$\mathbf{p}_m(t)$ wird nun über die Kardanwinkel (auch Tait-Bryan-Winkel) Ψ , θ und ϕ um die Z-, Y- und X-Achse gedreht. Die Drehung mit den Kardanwinkeln wird in der Benutzeroberfläche als Yaw, Pitch und Roll (Gier-, Nick- und Rollwinkel) aufgrund der Geläufigkeit dieser Begriffe angegeben. Es wird darauf hingewiesen, dass die Modulation von Ψ (Yaw) in einer zur Modulation des Azimut-Winkels redundanten Bewegung resultiert.

Die Drehung um eine Achse wird elementare Rotation genannt und erfolgt in Richtung der Drehachse im mathematisch positiven Sinne. Für die elementaren Drehmatrizen, deren Elemente Sinus- und Cosinus-Werte der Kardanwinkel sind, wird auf Rill & Schaeffer, 2014, S. 7ff. verwiesen. Aus den drei elementaren Matrizen setzt sich die kombinierte Matrix

$$\mathbf{M}(\psi(t), \theta(t), \phi(t)) = \mathbf{M}_z(\psi(t)) \cdot \mathbf{M}_y(\theta(t)) \cdot \mathbf{M}_x(\phi(t))\tag{8}$$

zusammen (Rill & Schaeffer, 2014, S. 5–9).

Die Modulation der Kardanwinkel über die Zeit erfolgt mit

$$\begin{aligned}
\psi(t) &= -\pi(e_{f,\text{scaled}}(t) w_{f,y} + l_{1,\text{scaled}}(t) w_{1,y} + l_2(t) w_{2,y} + n_{\text{scaled}} w_{n,y}), \\
\theta(t) &= -\frac{\pi}{2}(e_{f,\text{scaled}}(t) w_{f,p} + l_{1,\text{scaled}}(t) w_{1,p} + l_2(t) w_{2,p} + n_{\text{scaled}} w_{n,p}), \\
\phi(t) &= -\pi(e_{f,\text{scaled}}(t) w_{f,r} + l_{1,\text{scaled}}(t) w_{1,r} + l_2(t) w_{2,r} + n_{\text{scaled}} w_{n,r}).
\end{aligned} \tag{9}$$

Für die Rotation von $\mathbf{p}_m(t)$ um die Achsen des Koordinatensystems wird dieser über die Vektor-Matrix-Multiplikation zu

$$\mathbf{p}_{mr}(t) = \mathbf{M}(\psi(t), \theta(t), \phi(t)) \mathbf{p}_m(t). \tag{10}$$

Translation

Den letzten Schritt der Richtungsberechnung stellt die Translation von $\mathbf{p}_{mr}(t)$ zu einem Ankerpunkt (engl. anchor), beschrieben durch \mathbf{p}_{anc} in Abhängigkeit eines über der Zeit modulierbaren Faktors $u(t)$, $u \in [0; 1]$ über

$$\mathbf{p}_{mrt}(t) = \mathbf{p}_{mr}(t) + (\mathbf{p}_{\text{anc}} - \mathbf{p}_{mr}(t)) u(t) \tag{11}$$

dar. Die zugrundeliegende Idee ist die Bildung eines makroskopischen Effekts für die Bewegung der Audioobjekte. Diese können aus ihrer Richtung zu einem auf der Kugelfläche festgelegten Ankerpunkt hin- und wieder zurückbewegt werden. Darüber ist beispielsweise eine Kontrastierung zwischen der Konzentration aller Audioobjekte in eine Richtung und deren Verteilung und Ausdehnung über einen größeren Bereich kompakt steuerbar.

In der Benutzeroberfläche wird ein Wert u_0 als initialer Translationsfaktor festgelegt. Modulationsquellen für $u(t)$ sind v_{scaled} , m und $l_2(t)$. Es gilt

$$\begin{aligned}
u(t) &= u_1(t) + (1 - u_1(t)) v_{\text{scaled}} w_{v,u}, \\
u_1(t) &= u_m(t) + (1 - u_m(t)) l_2(t) w_{2,u}, \\
u_m(t) &= u_0 + (1 - u_0) m w_{m,u}.
\end{aligned} \tag{12}$$

Für die Rücktransformation in Kugelkoordinaten mit der Konvention δ als geographische Breite und $r = 1$ gilt:

$$\delta_{\text{mrt}}(t) = \arcsin\left(\frac{z_{\text{mrt}}(t)}{r}\right), \quad (13)$$

$$\varphi_{\text{mrt}}(t) = \arctan2(y_{\text{mrt}}(t), x_{\text{mrt}}(t)).$$

Die modulierten Koordinaten werden abschließend als OSC-Nachrichten formatiert (vgl. Kapitel 4.1.2) und stellen die Metadaten der Audioobjekte dar.

4.2.4 Benutzeroberfläche

Der SWS wird als virtuelles Instrument in Form eines M4L-Device in *Live* integriert. Die technisch beschriebenen Funktionen zur Klangsynthese und Richtungsberechnung werden über eine Benutzeroberfläche (engl. user interface, UI) für die klanggestalterische Arbeit in der DAW zugänglich gemacht. Diese versteht sich wie auch die technische Umsetzung als ein prototypischer Entwurf um die konzeptuellen Vorschläge zu evaluieren. Die Gestaltung der Benutzeroberfläche zielt daher primär auf pragmatische Funktionalität ab. Für die UI-Objekte wie z.B. Eingabefelder und Drehregler wurde bei der Umsetzung in *Max* soweit möglich auf die *Live* spezifischen Ausführungen der Objekte zurückgegriffen, was zu einem *Live* typischen Erscheinungsbild der UI führt. Eine Gestaltungsrichtlinie bieten zudem die *Max for Live Production Guidelines* (Ableton AG, 2020). Der SWS orientiert sich zudem bei der Darstellung der Modulationsmatrizen und Bedienelemente für Hüllkurven und LFOs am M4L-Instrument *Poli* (Ableton & Amazing Noises, 2020). Die Beschriftungen der Bedienelemente wurden in englischer Sprache vorgenommen, da alle gängigen virtuellen Instrumente eine englische Benutzeroberfläche haben und die englischen Begriffe somit weit verbreitet und geläufig sind.

Die Benutzeroberfläche gliedert sich in die drei Bereiche Klangsynthese, Richtungsberechnung und Routing. Diese sind durch die in *Live* bedingte Darstellung mit großer horizontaler Ausdehnung von links nach rechts angeordnet. Im Folgenden soll eine Verknüpfung zwischen der beschriebenen technischen Umsetzung und den Bedienelementen hergestellt werden. Für eine Demonstration der Funktionalität (inklusive Hörbeispiele) wird auf die Demo-Videos im digitalen Anhang (vgl. A3) dieser Arbeit verwiesen, die für die Experteninterviews produziert wurden.

Klangerzeugung

Abbildung 20 zeigt den Teil der UI zur Steuerung der Klangsynthese. Der SWS befindet sich im ‚Layer-Mode‘. Unterhalb des ‚Mode‘ Toggles können für die Oszillator-Gruppen die 2D-Wavetables ausgewählt und mit dem türkisfarbenen Slider eine initiale Wavetable-Position bestimmt werden. Unterhalb lassen sich Pitch Bend in Halbtonschritten und ‚Glide‘ in Millisekunden für die Tonhöhenberechnung einstellen.

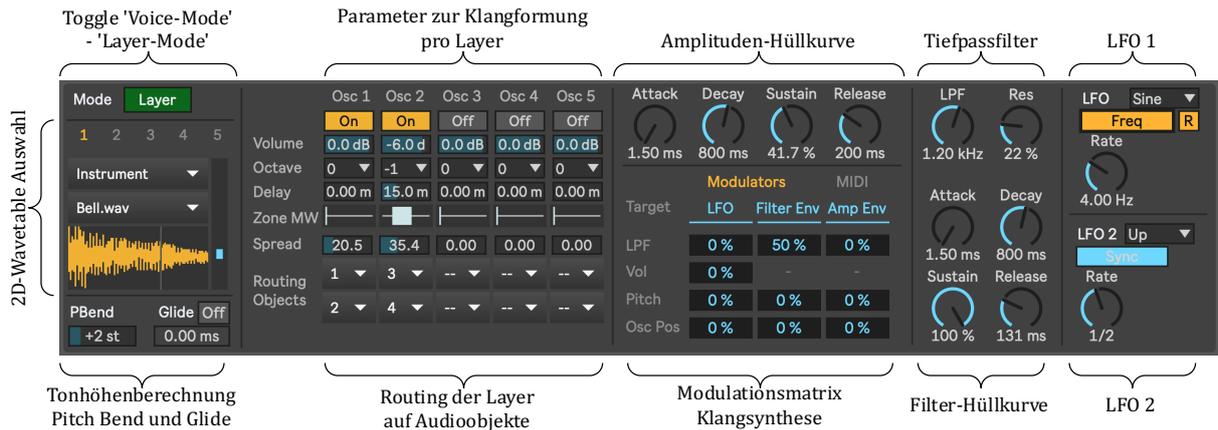


Abbildung 20: Benutzeroberflächen-Ausschnitt des SWS zur Klangsynthese.

Rechterhand können für die Oszillator-Gruppen, aus Gründen der Übersicht in der UI, 'Osc'²⁴ benannt, Einstellungen getroffen werden. Layer 1 und 2 sind aktiv geschaltet. Zudem ist Layer 2 relativ um 6dB Gain reduziert, eine Oktave nach unten transponiert, um 15ms verzögert und ist durch die definierte ‚MW Zone‘ in der Lautstärke von der Modulation Wheel Position abhängig. Layer 1 und 2 werden auf jeweils zwei unterschiedliche Audioobjekte geroutet und die beiden zugrundeliegenden Oszillatoren der Oszillatorgruppen über den ‚Spread‘-Parameter dekorreliert. Dieser steuert das Maß der Tonhöhenvariation (‚Detune‘). Bei aktivem ‚Voice-Mode‘ entfallen die Bedienelemente ‚Spread‘ und ‚Routing Objects‘ aufgrund der über das Voice Management festgelegten Verteilung der polyphonen Voices auf die Audioobjekte.

Im rechten Teil befinden sich die Bedienelemente zur Parametrisierung der Amplituden- und Filter-Hüllkurve sowie zur Festlegung von Grenzfrequenz und Resonanz des Tiefpassfilters. Unterhalb der Amplituden-Hüllkurve befindet sich die Modulationsmatrix zur Bestimmung der gewichtenden Faktoren für die Modulation klangformender Parameter. Die Filter-Hüllkurve moduliert mit 50% eines explorativ festgelegten Wertes die Grenzfrequenz des Tiefpassfilters. Darüber hinaus ist in Abbildung 20 keine weitere Modulation aktiv. Die Modulationsmatrix der Klangsynthese besteht aus zwei Tabs für Modulatoren und MIDI-Befehle, wobei der ‚Modulators‘ Tab aktiv ist. Auf der rechten Seite werden LFO 1 und LFO 2 parametrisiert. Obwohl LFO 2 nicht für die Modulation klangformender Parameter zur Verfügung steht, befinden sich dessen Bedienelemente aus Gründen einer kompakten UI-Gestaltung im Block zur Klangsynthese. Für LFO 1 ist eine sinusförmige Wellenform gewählt, die Frequenz über einen Frequenzwert festgelegt und die Retrigger-Option, zum Zurücksetzen der Phase bei jedem eintreffenden MIDI Note On Befehl, aktiviert. Für LFO 2 ist eine aufsteigende sägezahnförmige Wellenform gewählt und die Frequenz

²⁴ Synonym zu dem was hier unter ‚Layer‘ verstanden wird. In der Implementierungslogik bezeichnet ein ‚Layer‘ eine Oszillator-Gruppe, bestehend aus zwei Oszillatoren mit gleichem Wavetable und ‚Detune‘ Option. Aus Anwendersicht wird eine Oszillator-Gruppe in der Regel nur als Oszillator oder ‚Layer‘ bezeichnet.

des LFO über den temporelativen Notenwert ‚Halbe‘ definiert. LFO 2 ist nicht Teil des Klangsyntheseblocks und verfügt daher über keine Retrigger-Option.

Richtungsberechnung

Abbildung 21 zeigt den Bereich der UI zur Richtungsberechnung der Audioobjekte. Im oberen Teil ist hier die Anzahl der Objekte auf vier festgelegt und ein initiales Verteilungsmuster gewählt, das die vier Objekte in gleichen Abständen auf der Horizontalebene verteilt. Die Option zur Bewegung der Audioobjekte ist aktiviert. Die initialen Richtungen der Audioobjekte lassen sich über numerische Eingabe auf der linken Seite oder über die Repositionierung der nummerierten weißen Kreise auf der Rektangularprojektion der Kugelfläche ändern. Die blauen Kreise repräsentieren die bewegten Pendants der Audioobjekte.

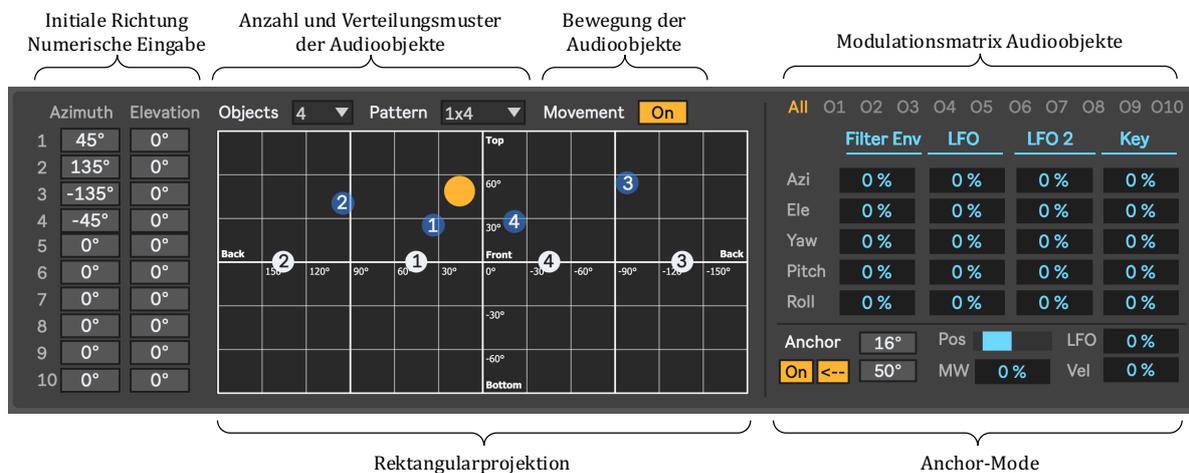


Abbildung 21: Benutzeroberflächen-Ausschnitt des SWS zur Richtungsberechnung.

Rechterhand befindet sich die Modulationsmatrix zur Festlegung der gewichtenden Faktoren für die Modulation von Azimut- und Elevationswinkeln sowie Yaw-, Pitch- und Roll-Winkeln. Diese können sowohl für alle Objekte (Tab ‚All‘) als auch jedes Objekt einzeln (Tab O1 bis O10) eingestellt werden. In Abbildung 21 ist keine Modulation aktiv. Aktiv ist hingegen der ‚Anchor-Mode‘. Der Ankerpunkt wird durch den orangenen Kreis auf der Rektangularprojektion repräsentiert, dient als ‚Zielpunkt‘ für die Translation. Über ‚Pos‘ wird ein initialer Faktor für die Translation festgelegt (vgl. u_0), der hier etwa mittig eingestellt ist. Der Translationsfaktor kann über das Modulation Wheel, das LFO 2 Signal (in der UI nur mit LFO angegeben) und über skalierte Velocity Werte moduliert werden (vgl. Formel (12)).

Output & Presets

Rechterhand der Richtungsberechnung lässt sich ein Block für das Routing der rohen Audioobjekt-Daten und zur Preset-Verwaltung öffnen (Abbildung 22). Die Audiosignale werden

paarweise auf die virtuelle Soundkarte *Blackhole* geroutet. Für die OSC formatierten Metadaten wird Host-Adresse und Port spezifiziert, sowie ein Send-Intervall angegeben.

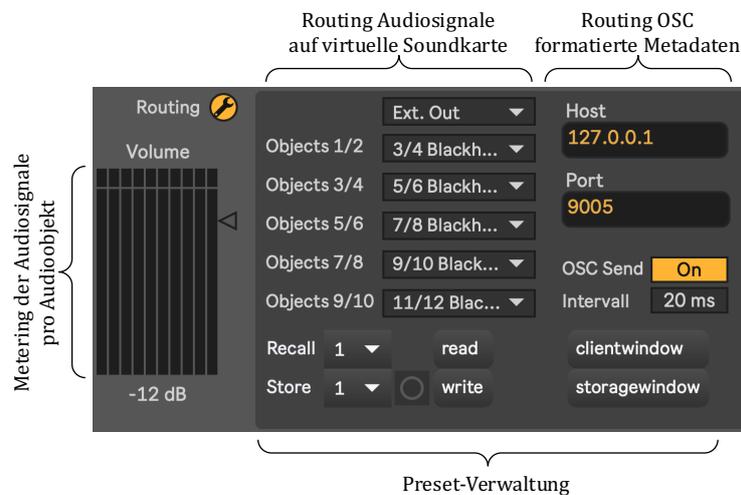


Abbildung 22: Benutzeroberflächen-Ausschnitt des SWS zum Routing und der Preset-Verwaltung.

Im SWS ist zudem ein System zur Preset-Bildung integriert, das für den Sounddesign-Prozess eine unerlässliche Funktion darstellt. Der Zustand aller Parameter des M4L-Devices kann darüber gespeichert und zu einem späteren Zeitpunkt wiederhergestellt werden. Die Einstellungen werden in eine XML-Datei geschrieben.

5 Experten-Evaluation

Der ‚Spatial Wavetable Synthesizer‘ und seine System-Integration sind als Vorschläge zur Beantwortung der Forschungsfrage 1 (*Wie kann 3D-Audio in ein virtuelles Instrument der Wavetable Synthese integriert werden?*) zu verstehen. Diese werden im Folgenden über Experteninterviews evaluiert. Die Evaluation ist in Forschungsfrage 2 (*Wie bewerten Experten aus den Bereichen Klangerzeugung und Postproduktion die Vorschläge der prototypischen Implementierung?*) verankert und bildet zusammen mit der Implementierung des SWS den zentralen praktischen Teil dieser Arbeit.

Die Implementierung des SWS war geprägt von einem explorativen Prozess, dessen letzter Stand in Kapitel 4 dokumentiert ist. Perspektivisch kann dieser Stand einen Ausgangspunkt für weitere Entwicklungen darstellen. Dafür ist es sinnvoll, die umgesetzten Vorschläge mit Experten tangierender Themenfelder zu evaluieren. Dabei wird das frühe Entwicklungsstadium des Prototyps als Vorteil erachtet, um in Abhängigkeit des durch die Experten vermittelten Stimmungsbilds den richtigen Fokus für weitere Entwicklungen zu finden.

In Kapitel 5.1 wird das methodische Vorgehen der Experteninterviews beschrieben, die Experten vorgestellt und die leitfadengestützte Struktur der Interviews dargelegt. In Kapitel 5.2 folgt die deskriptive Auswertung, in der die wichtigsten Aussagen der Experten zusammengefasst werden. Abschließend werden in Kapitel 5.3 die Aussagen der Experten zu den im SWS umgesetzten Vorschlägen eingeordnet und diskutiert.

5.1 Methodisches Vorgehen

Die durchgeführten Experteninterviews sind an das theoriegenerierende Experteninterview aus dem Bereich der qualitativen Sozialforschung angelehnt, dessen Ausrichtung auf die Rekonstruktion subjektiver Deutungen und Interpretationen der Experten abzielt (Bogner et al., 2014, S. 2).

Den Experten zeichnen seine Sachkenntnis und sein Spezialwissen aus: „Das Besondere am Expertenwissen besteht nicht nur in dessen besonderer Reflexivität, Kohärenz oder Gewissheit, sondern auch insbesondere darin, dass dieses Wissen in besonderer Weise praxiswirksam und damit orientierungs- und handlungsleitend für andere Akteure wird“ (Bogner et al., 2014, S. 14). Speziell erfahrungsbasiertes und anwendungsbezogenes Wissen in den relevanten Themenbereichen ist für die Evaluation des SWS von Interesse. Der Experte ist bis zu einem gewissen Grad das Konstrukt des Forschungsinteresses (Meuser & Nagel, 2009). Damit zeichnet den Experten nicht nur sein Wissen in Form fachspezifischer Kompetenz aus, sondern die Fähigkeit, Verbindungen zu anderem Wissen herzustellen und die Relevanz des eigenen Wissens zu reflektieren

(Hitzler, 1994, S. 21f.). Speziell in ihrer Ausgestaltung als qualitative Interviews liegt die Stärke von Experteninterviews in der Erhebung von Deutungswissen: „Deutungswissen ist gewissermaßen ‚immer wahr‘. So widersprüchlich und inkonsistent es erscheinen mag – es ist ein ‚soziales Faktum‘, das unabhängig von seinem Wirklichkeitsbezug sozialwissenschaftlich relevant ist“ (Bogner et al., 2014, S. 20).

Bei theoriegenerierenden Experteninterviews wird ein Experte meist als Vertreter einer bestimmten Gruppe angesprochen. Es steht „[...] die subjektive Dimension im Vordergrund: implizite Entscheidungsmaximen, handlungsanleitende Wahrnehmungsmuster, Routinen, Weltbilder, etc. [...]“ (Bogner et al., 2014, S. 25) – der Fokus liegt auf dem Deutungswissen der Befragten. Das Interview trägt in der Hinsicht zur Theoriegenerierung bei, in dem es sich analytisch und interpretativ mit dem empirischen Material auseinandersetzt, um Zusammenhänge aufzuzeigen und theoretische Ansätze zu entwickeln (ebd.). Die Zusammenstellung von Experten aus mehreren relevanten Gruppen wurde im folgenden Expertensampling berücksichtigt.

5.1.1 Expertensampling

Die Auswahl der Experten basiert auf den thematischen Schwerpunktfeldern dieser Arbeit. Aus einem traditionellen Workflow-Verständnis kann die Wavetable Synthese dem Bereich der Klangzerzeugung und die Spatialisierung für 3D-Audio der Postproduktion zugeordnet werden. Beide Bereiche sollen sowohl von Anwender-, als auch Entwicklerperspektive betrachtet werden, um einen einseitigen Blickwinkel auf die Vorschläge zu vermeiden. Somit wurde nach Experten gesucht, die neben Erfahrungswerten in der Arbeit mit 3D-Audio einen beruflichen Hintergrund in mindestens einem der Berufsfelder Komponist & Sounddesigner, Toningenieur, Synthesizer-Programmierer oder Postproduktionstool-Entwickler aufweisen.

Die Resonanz der Experten auf die Interviewanfragen war wie folgt: Von 12 Anfragen resultierten neun in einem Interview, zwei Anfragen wurden aus zeitlichen Gründen abgelehnt, eine weitere Anfrage nicht beantwortet. Es wurde darauf geachtet, dass die Experten unterschiedlichen Netzwerken angehören. Dies wurde durch die Streuung über unterschiedliche Institute, Unternehmen und Arbeitsbereiche der Experten sichergestellt. Zudem wurde die Auswahl auf deutschsprachige Experten beschränkt, um den Komplexitätsgrad der Durchführung im Machbarkeitsrahmen dieser Arbeit zu halten.

Die befragten Experten sind in der Reihenfolge der durchgeführten Interviews: Jan Langhammer, Andy Hoppe, Dustin Zorn, Martin Rieger, Michael Romanov, Johannes Regnier, Steffen Günther, Ulli Scuda und Thomas Koch. Abbildung 23 gibt einen Überblick über die berufliche Einordnung der Experten. Die Experten nahmen diese Einordnung während des Interviews selbst vor. Es sei

darauf hingewiesen, dass die Berufsfelder keine trennscharfen Kategorien darstellen und es sich in der Praxis vielmehr um ein Spektrum mit fließenden Übergängen zwischen den Bereichen handelt. Dennoch gibt diese Einordnung einen guten Überblick über die fachlichen Konstellationen. Alle gewünschten Berufsfelder konnten mehrfach abgedeckt werden, wobei es zu einer Schwerpunktbildung bei den Toningenieuren gekommen ist. Ulli Scuda und Michael Romanov haben ihre Berufsfelder absteigend priorisiert (1 und 2).

	Klangerzeugung	Postproduktion
Anwendung	Komponist & Sounddesigner <div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Hoppe</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Regnier</div> </div> <div style="display: flex; justify-content: center; align-items: center; margin: 2px;"> <div style="border: 1px solid black; padding: 2px; margin: 0 10px;">Rieger</div> </div> <div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Günther</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Zorn</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Romanov (2)</div> </div>	Toningenieur <div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Hoppe</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Scuda (1)</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Regnier</div> </div> <div style="display: flex; justify-content: center; align-items: center; margin: 2px;"> <div style="border: 1px solid black; padding: 2px; margin: 0 10px;">Koch</div> <div style="border: 1px solid black; padding: 2px; margin: 0 10px;">Rieger</div> </div> <div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Günther</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Zorn</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Romanov (1)</div> </div>
Entwicklung	Synthesizer-Programmierer <div style="display: flex; justify-content: center; align-items: center; margin: 2px;"> <div style="border: 1px solid black; padding: 2px; margin: 0 10px;">Regnier</div> </div> <div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Günther</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Zorn</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Romanov (1)</div> </div>	Postproduktionstool-Entwickler <div style="display: flex; justify-content: center; align-items: center; margin: 2px;"> <div style="border: 1px solid black; padding: 2px; margin: 0 10px;">Scuda (2)</div> <div style="border: 1px solid black; padding: 2px; margin: 0 10px;">Regnier</div> </div> <div style="display: flex; justify-content: center; align-items: center; margin: 2px;"> <div style="border: 1px solid black; padding: 2px; margin: 0 10px;">Koch</div> <div style="border: 1px solid black; padding: 2px; margin: 0 10px;">Langhammer</div> </div> <div style="display: flex; justify-content: space-around; align-items: flex-start;"> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Günther</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">Zorn</div> </div>

Abbildung 23: Matrix zur Einordnung der Experten in ihre Berufsfelder.

Im Folgenden werden die einzelnen Experten in Hinblick auf ihre berufliche Tätigkeit und ihrem Bezug zu 3D-Audio vorgestellt, ein kurzer Überblick über die von ihnen verwendeten Workflows zur Produktion von 3D-Audio Inhalten gegeben und – sofern verbalisiert – ihre Haltungen bezüglich Musikproduktion in 3D-Audio dargelegt.

Anzumerken ist, dass Johannes Regnier mitunter auf Englisch geantwortet hat. Direkte Zitate wurden aufgrund einer unverfälschten Darstellung nicht übersetzt. Darüber hinaus wird festgehalten, dass Andy Hoppe und Michael Romanov in Bezug auf 3D-Audio mit ihrer Tätigkeit bei *z-Achse* einen gemeinsamen Hintergrund haben. Dies trifft auch Thomas Koch, Johannes Regnier und Steffen Günther zu, deren Arbeiten mit 3D-Audio im Netzwerk der Universität der Künste (UdK) Berlin eine Schnittmenge aufweisen.

Thomas Koch ist in der Forschung und Entwicklung am Fraunhofer Heinrich-Hertz-Institut und in der Lehre an der UdK Berlin zum Thema 3D-Audio tätig. Bei seinen Arbeiten mit 3D-Audio Bezug handelt es sich häufig um Browser basierte audiovisuelle Werke: 360°-Videos, 3DoF- und 6DoF-Videos im Kontext Web-Audio. Bezüglich des Workflows überprüft Koch zunächst das Endformat, da dieses nach seiner Aussage Auswirkungen auf die Produktions-Tool-Auswahl hat. Die

Eigenschaften des Formats fließen in seine Konzeptionen ein. Räumlichkeit begreift Koch als kompositorische Dimension und das Mastering in 3D-Audio als einen rein technischen Prozess. Für den Produktionsprozess von 3D-Audio Inhalten bestätigt Koch eine Trennung zwischen Klangzeugung und Spatialisierung, weist aber auf die Lernfähigkeit und das Abstraktionsvermögen der Künstler hin.

Jan Langhammer ist als freier Ingenieur und Berater tätig. Er hat einen Hintergrund in der Entwicklung von Produktions-Tools für Wellenfeldsynthese und zudem die *Spatial Audio Workstation*²⁵ mitentwickelt. Häufig stellt er die Schnittstelle zwischen den Kreativen und den Labels her. Bei *Encircled Audio* arbeitet Langhammer mit unterschiedlichen 3D-Audio Formaten, deren Produktion meist mit der Anlieferung verschiedener Audio-Assets in Mono und Stereo durch den Künstler beginnt. In den seltensten Fällen könne der Künstler die erzeugten Klänge in seiner Produktionsumgebung über ein 3D-Audio Wiedergabesystem abhören. Die Produktion der Assets auf Künstlerseite erfolgt mit der Intention der Spatialisierung; die Abstraktion auf Künstlerseite funktioniere gut. Während der Spatialisierung priorisiert Langhammer: Assets mit hoher Priorität definiert er als Audioobjekt mit eigenem Panning-Pfad, andere Assets mischt er auf einen Kanal-Bus.

Johannes Regnier entwickelt digitale Werkzeuge und Algorithmen für Musik, ist als Komponist und Tonmeister aktiv und in der Lehre von 3D-Audio an der UdK Berlin tätig. Seine Arbeiten mit 3D-Audio umfassen Rauminstallation mit eigenen Stücken und Echtzeit-Performances, Arbeiten mit Wellenfeldsynthese-Systemen und unter anderem die Entwicklung der MMT Toolbox – einer Sammlung von Tools für die Kontrolle von Echtzeit-Spatialisierung für Wellenfeldsynthese. In seiner künstlerischen Forschung setzt sich Regnier intensiv mit dem Zusammenspiel von Klangsynthese und Spatialisierung für 3D-Audio auseinander. Die Arbeiten zur Spatialisierung von Granularsynthese mit Partikel-Simulationen nennt er einen „Reflex“ (Regnier, 2021, ab 17:05). Bei seinen Arbeiten kommen mit *IRCAM Spat*, *ICST Toolbox* und *IEM Plug-Ins* unterschiedliche 3D-Audio Rendering Software zum Einsatz.

Michael Romanov programmiert Synthesizer und ist Komponist elektronischer Musik. Als Entwickler hat er unter anderem den DIY Head-Tracker *MrHeadTracker* entwickelt, der Einsatz in der Rendering-Kette der SWS-Implementierung findet. Er war Entwickler im Kollektiv *z-Achse*, welches sich der Etablierung von 3D-Audio in den Clubs der Techno-Szene widmete. Darüber hinaus hat er den *Klangdome*, eine mobile 3D-Audio Wiedergabelösung, mitentwickelt. Den Workflow für die Produktion von 3D-Audio Inhalten unterscheidet er in zwei Ansätze: Zum einen

²⁵ Die *Spatial Audio Workstation* ist ein Mixing-Tool für die Arbeit mit objektbasiertem 3D-Audio und als Plug-In für die DAW *Nuendo* verfügbar (Encircled Audio, 2020).

der Spatialisierung fertig produzierter Klänge in einem nachgelagerten Schritt. Diesen Ansatz bezeichnet Romanov als eine Form des Upmixings. Wie auch Langhammer unterscheidet er zwischen Audioquellen im Fokus, die über Bewegung hervorgehoben werden können und Klänge, die sich in ein ‚Bett‘ oder Soundscape mischen. Er weist zudem auf die Abhängigkeit der Spatialisierung von der Signalbeschaffenheit hin. Zum anderen könne die Produktion von 3D-Audio Inhalten über Tools erfolgen, die Klangerzeugung und Spatialisierung gleichzeitig machen. Seine Entwicklungen bei *z-Achse* konzentrierten sich auf Gestik gesteuerte Tools, wie beispielsweise ein modifizierter *Wii-Controller* für die Klangerzeugung und Bewegung im Raum, auch ‚Techno-Peitsche‘ genannt. Bei so einer Vorgehensweise kämen „ganz andere Sachen raus“ (Romanov, 2021, ab 21:50).

Andy Hoppe ist aktuell primär in der Beratung und Konzeption sowie als Creative Director bei *Klangerfinder* tätig. Seine Erfahrung mit 3D-Audio sammelte er unter anderem mit Romanov im Kollektiv *z-Achse*. Darüber hinaus arbeitete er am *Klangdome* mit. Seine aktuellen Arbeiten mit 3D-Audio konzentrieren sich auf auditives Feedback im Automotiv-Bereich. Bezüglich des Workflows zur Produktion von 3D-Inhalten berichtet Hoppe ähnliches wie Romanov. Speziell im Kontext Musik werde häufig im Nachgang spatialisiert – als Beispiel nennt er *Kraftwerk – 3-D The Catalogue*.

Martin Rieger ist als Toningenieur und Sounddesigner in den Bereichen Virtual Reality, 360°-Videos, Hörspiel und Filmton tätig. Nahezu alle seine Produktionen haben einen 3D-Audio Bezug. Endformat ist häufig die binaurale Wiedergabe über Kopfhörer mit und ohne Head-Tracking. Darüber hinaus ist Rieger als Blogger und Berater für 3D-Audio tätig. Musik in 3D-Audio bezeichnet er als „Wundertüte“ (Rieger, 2021b, ab 17:27), weil nicht absehbar sei, wo die Reise hingehet. Die Musik für seinen Workflow wird in der Regel als separate Stems oder Stereo-Mix angeliefert, die er diskret im Raum platziert oder „durch einen Spatializer jagt“ (ebd. 25:42).

Dustin Zorn ist primär Komponist elektroakustischer Musik, programmiert Synthesizer und entwickelt Audioeffekte für eigene Kompositionen. Bei seinen Arbeiten mit 3D-Audio handelt es sich um Kompositionen, Installationen und Live-Setups mit vielen Lautsprechern im Raum. Seine Arbeitsweise ist stark abhängig vom Aufführungsort, denn Räume seien nicht neutral und speziell in der elektroakustischen Musik würde für einen spezifischen Raum komponiert. Er sagt: "Meistens lohnt es sich dann eher über Funktionen von Lautsprechern nachzudenken. Das heißt im Prinzip genau das Gegenteil von neutral, sondern jeder Lautsprecher hat durch seine Position im Raum `ne Färbung und auch `ne musikalische Rolle. Und das hat manchmal mehr mit der musikalischen Realität zu tun [...]. In dem Sinn, dass man wirklich für die Rollen komponiert – oder

doch irgendwo probiert ein neutrales Setup in 'nem Raum hinzubiegen. Das sind so die beiden Perspektiven" (Zorn, 2021, ab 05:40).

Ulli Scuda ist Toningenieur bzw. Tonmeister, an der Entwicklung von Postproduktions-Tools beteiligt und hat einen Background als Komponist und Sounddesigner. Er bezeichnet sich selbst als Skeptiker elektronischer Klänge und ist nach eigenen Angaben kein Synthesizer-Kenner. Seine Arbeiten beschäftigen sich mit der Aufnahme, Übertragung und Wiedergabe von 3D-Audio, schwerpunktmäßig objektbasiertes Audio mit MPEG-H. Den Workflow einer 3D-Audio-Produktion beschreibt er als geteilt: Zunächst die Generierung des Klangmaterials, beispielsweise die Aufnahme von Mono- und Stereospuren einer Band, daran anschließend die Mischung, bei der die Signale klanglich und räumlich zusammengebracht werden. Das 3D-Panning, das räumliche Arrangieren, sei dabei meist ein zusätzlicher separater Schritt. Beim Erzeugen der Signale spielt bei ihm die Räumlichkeit selten eine Rolle. Bei Filmmusik, Popmusik und elektronischer Musik sei der Raum, den man am Ende will, schließlich nicht der, der bei der Erzeugung da ist. Den Sonderfall stelle die Abbildung eines natürlichen Raums dar, beispielsweise eines Ensembles in einer Kirche.

Steffen Günther ist Sounddesigner und Komponist elektronischer Musik in 3D-Audio. Er entwickelt Synthesizer und Audioeffekte in *Max* und *Unity*. Mit 3D-Audio ist er darüber hinaus am Heinrich-Hertz-Institut im Kontext Klanginstallationen tätig und assistiert Thomas Koch bei der Vorlesung – wobei er primär den anwendungsorientierten Part übernimmt. Sein 3D-Audio Produktions-Workflow baut in der Regel auf synthetisierten Mono- und Stereo-Klängen auf, die gelegentlich um Soundfield-Mikrofonaufnahmen ergänzt werden. Seine Kompositionen sind modular als Live-Set aufgebaut. Presets in der DAW sind meist Basis für die räumliche Struktur. Game Engines sieht er auch aufgrund der Reflexionsberechnung über ray tracing und die etablierte dynamische Binauralsynthese als starkes Tool für die 3D-Audio Produktion der Zukunft.

5.1.2 Themenbereiche und Leitfaden

Nach der Expertenvorstellung wird nun der Aufbau der Interviews erläutert. Die Vorschläge zur Implementierung des Prototyps werden im Sinne einer besseren Handhabung für die Auswertung der Interviews in vier Themenbereiche gegliedert (vgl. Tabelle 2). Die Themenbereiche geben die Struktur für den Interviewleitfaden vor. Darüber hinaus wird durch die inhaltliche Strukturierung eine vergleichende Auswertung der Experteninterviews möglich.

Themenbereiche	Vorschläge
Konzept	Integration von 3D-Audio in das virtuelle Instrument
Workflow	Objektifizierung Klangkomponenten, objektbasierter Audiooutput, und separates 3D-Audio Rendering
Klangkomponenten	„Layer-Mode“, „Voice-Mode“
Bewegung	Modulationssignale & MIDI-Befehle, Makrobildung „Anchor-Mode“

Tabelle 2: Themenbereiche und Vorschläge der Implementierung und Evaluation.

Unter ‚Konzept‘ wird der generelle Ansatz dieser Arbeit verstanden, 3D-Audio in das virtuelle Instrument zu integrieren und damit die Spatialisierung zum Komponisten zu verlagern. Dies stellt den grundlegenden Gedanken dieser Arbeit dar. Da die Spatialisierung auch an anderer Stelle im Musikproduktionsprozess geschehen kann, wird zu Beginn der Evaluation der Experteninterviews ein Stimmungsbild zu dieser grundlegenden Herangehensweise gezeichnet. Unter dem Themenbereich ‚Workflow‘ werden der Vorschlag zur Objektifizierung von Klangkomponenten des Syntheseprozesses sowie der Vorschlag eines objektbasierten, formatagnostischen Outputs des Instruments mit sich anschließendem ausgelagerten 3D-Audio Rendering, zusammengefasst. Im Themenbereich ‚Klangkomponenten‘ werden die zwei vorgeschlagenen Modi ‚Voice-Mode‘ und ‚Layer-Mode‘ besprochen. Der Themenbereich ‚Bewegung‘ beinhaltet den Vorschlag die Bewegung der Audioobjekte durch Modulationssignale des Syntheseprozesses und MIDI-Befehle zu steuern sowie den vorgeschlagenen ‚Anchor-Mode‘.

Leitfaden

Ebenfalls auf die vier Themenbereiche bezieht sich der Leitfaden, der die Grundlage für die Experteninterviews gebildet hat. Die Varianz dessen, was in der Literatur unter einem Leitfaden verstanden wird ist recht groß. Es reicht von allgemein gehaltenen Themen-Richtlinien (Arthur & Nazroo, 2003) bis hin zu standardisierten Leitfäden mit konkret formulierten Fragen (Kaufmann, 1999, S. 65ff.). Der Leitfaden dieser Arbeit beruht auf einem Mittelweg, in dem er auf feste Formulierungen zurückgreift, um eine vergleichende Auswertung zu ermöglichen. Er bietet aber auch genug Raum für spontane Reaktionen und Nachfragen, was dem momentanen Status des SWS als Prototyp gerecht wird. Der Leitfaden beinhaltet eine thematische Strukturierung in Anlehnung an die vier Themenbereiche (vgl. Tabelle 2), ist dabei aber kein engmaschiges Netz zur Informationsgewinnung, sondern offener und lockerer. Die Interviewsituation lässt sich nicht exakt planen: „Das Interviewen bleibt immer eine ‚art of science‘, eine Art Kunstfertigkeit, die sich nicht vollständig durch methodische Regeln vorbestimmen lässt“ (Fontana & Frey, 1998). Der Leitfaden wurde bei Pre-Tests mit den Praxisbetreuern dieser Arbeit einem Realitätscheck unterzogen und iterativ optimiert (Kaiser, 2014, S. 69). Die finale Version des Leitfadens, der auch zur Protokollierung der Interviews diente, findet sich im Anhang A1. Der erarbeitete Leitfaden sieht folgenden Ablauf des Interviews vor:

Bereits vor dem Interview hat der Experte sich fünf ein-minütige, vorproduzierte Videos angesehen. Diese erklären die grundlegenden Konzepte und Funktionalitäten des SWS und geben über statisch binauralisierte Hörbeispiele einen klanglichen Eindruck. Die fünf Demo-Videos finden sich im digitalen Anhang dieser Arbeit und können auf Anfrage zur Verfügung gestellt werden.

Zu Beginn des Gesprächs stellt der Interviewer sich und das Thema dieser Arbeit vor. Daran schließt sich die fachliche Verortung des Experten sowie die Beschreibung des beruflichen Hintergrunds und gegebenenfalls der zugehörigen Institution, beziehungsweise des Unternehmens an. Als Einstieg wird nach Erfahrung mit 3D-Audio und dem aktuellen Workflow bei der Produktion von Inhalten in 3D-Audio gefragt. Dabei handelt es sich um erzählungsgenerierende Fragen, die auf eine längere, eigenständig strukturierte Schilderung abzielen (Bogner et al., 2014, S. 62). Diese dienen dazu den fachlichen Hintergrund des Experten in Bezug auf 3D-Audio besser einordnen zu können.

Vor dem Einstieg in die Fragen zu den vier Themenbereichen, werden die wichtigsten Funktionalitäten des SWS auf acht Folien zusammengefasst, um eventuellen Verständnisschwierigkeiten vorzubeugen. Die Folien befinden sich im digitalen Anhang (vgl. A3).

Daran anschließend folgt der Hauptteil des Interviews. Dabei dienen folgende sinngemäße Fragen als Hauptfragen, die in Abhängigkeit des Umfangs der Antwort um spezifische Nachfragen ergänzt werden. Es handelt sich hierbei um Stellungnahmen und Bewertungsfragen, die weniger auf Beschreibungen und mehr auf Einschätzungen und normative Beurteilungen abzielen (Bogner et al., 2014, S. 64).

- Wie bewerten Sie allgemein das Konzept der Verzahnung von Klangerzeugung und Spatialisierung?
- Wie bewerten Sie das vorgeschlagene Konzept der Objektivierung einzelner Klangkomponenten mit objektbasiertem Output und ausgelagertem 3D-Audio Rendering?
- Was ist Ihr Eindruck von der räumlichen Verteilung mehrerer Layer? Was ist Ihr Eindruck von der räumlichen Verteilung polyphoner Voices?
- Was ist ihr Eindruck von der Bewegung der Audioobjekte über Modulationssignale der Klangsynthese? Was ist Ihr Eindruck von der Makrobildung ‚Anchor-Mode‘?

Darüber hinaus wurden im Leitfaden einige Zusatzfragen festgelegt, die zur zusätzlichen Informationsgenerierung gedacht waren, allerdings aufgrund des Umfangs nicht in der Auswertung der Interviews in dieser Arbeit berücksichtigt werden.

In der Interviewdurchführung kamen häufig Zurückspiegelungen zum Einsatz: kurze bilanzierende Zusammenfassung des Interviewers, vor allem für die Ordnung von Aussagen verteilt über das Interview zu einzelnen Themengebieten (Bogner et al., 2014, S. 67). Die Zurückspiegelungen waren meist gepaart mit einem ‚Verständnisfragen‘-Charakter und kamen auch zur thematischen Steuerung des Interviews zum Einsatz.

Die Interviews fanden in Form einer Videokonferenz im Zeitraum vom 18.05.2021 bis 28.05.2021 statt. Die Dauer der Interviews lag zwischen 51 Minuten und 87 Minuten. Besondere Vorkommnisse traten nicht auf. Die Dokumentation der Interviews erfolgte über eine Tonaufnahme der Videokonferenz und stichpunktartige handschriftliche Protokollierung während des Interviews sowie dessen Ergänzung um ein Gedächtnisprotokoll im Anschluss an das Interview. Die Tonaufnahmen der Experteninterviews befinden sich im digitalen Anhang (vgl. A3) dieser Arbeit.

5.2 Auswertung

Die Auswertung gliedert sich in einen deskriptiven Teil auf Basis der vorhandenen Dokumentation der Interviews (Kapitel 5.2) und einer anschließenden Diskussion der gewonnenen Information (Kapitel 5.3).

Der deskriptive Teil der Auswertung basiert auf den Antworten der Experten zu den Hauptfragen. Diese gingen zumeist wie gewünscht über den Kern der eigentlichen Frage hinaus und resultierten in erzählenden Antworten. Aus Gründen der Machbarkeit wurde auf eine vollständige Transkription aller neun Interviews verzichtet. Eine partielle, bzw. selektive Transkription direkter Zitate, die Kernaussagen der Experten zu den Vorschlägen widerspiegeln, wird als ausreichend erachtet. Die Datenbasis der Auswertung sind die Tonaufnahmen der Interviews in Abgleich mit den angefertigten handschriftlichen Protokollen. In einem Zwischenschritt wurden die Aussagen der Experten, in der Struktur der definierten Themengebiete, paraphrasiert. Das Ziel ist die Komplexitätsreduktion, sprich die Herstellung übersichtlicher und gut handhabbarer Textpassagen, in denen die besonders relevanten Themen ausführlicher und weniger relevante Themen gerafft dargestellt werden (Bogner et al., 2014, S. 80). Es soll darauf hingewiesen sein, dass die Paraphrasierung bereits eine Art interpretative Auseinandersetzung mit dem Material darstellt. Die Bewertung eines Vorschlags bestehen meist aus Aussagen an verschiedenen Stellen in der Tonaufnahme, die in einer Paraphrase zusammengefasst wurden. Daher wird bei Paraphrasen auf einen Kurzbeleg verzichtet, die Quelle ist die Tonaufnahme des Interviews. Darüber hinaus werden markante Aussagen direkt zitiert. Bei den direkten Zitaten wird auf den Startzeitpunkt des Zitats in der Tonaufnahme referenziert.

5.2.1 Integration von 3D-Audio in das virtuelle Instrument

In einem ersten Schritt werden die Aussagen²⁶ zum Vorschlag der Integration von 3D-Audio in das virtuelle Instrument und der damit einhergehenden Verlagerung der Spatialisierung zum Komponisten gegenübergestellt.

Für Koch muss die Gestaltung räumlicher Attribute Teil der Komposition sein. Eine Verlagerung der Spatialisierung hin zum Komponisten hält er für wichtig. Er gibt aber zu bedenken, dass auch in der Postproduktion eine Kontrolle über räumliche Aspekte wichtig ist und stellt den Vergleich zum externen Mastering her; ein Beispiel für eine sinnhafte Arbeitsteilung. Wo genau die Trennung bei Produktionen von 3D-Audio Musik erfolgen sollte, lässt er offen: „Ich glaube [...], dass man das zum jetzigen Zeitpunkt noch gar nicht abschließend beantworten kann, welche Aspekte das sind. Aber grob passiert im Moment noch zu viel in der Postproduktion und zu wenig im künstlerischen Prozess - das kann man schon mal sagen.“ (T. Koch, persönliche Kommunikation, 28. Mai 2021, ab 22:12). Den SWS begreift Koch auch als eine Chance zum Lernen, Experimentieren und für künstlerische Forschung: „Ich finde dein Projekt total super und ich habe in verschiedenen Kontexten schon darüber nachgedacht, dass räumliche Positionen verschmelzen mit dem klanglichen Inhalt, um eben auch primär den Komponisten, Sounddesignern und Postproduktion Toningenieuren den Zahn zu ziehen, dass sie alles irgendwie verteilen - was Nonsense ist, oder im schlimmsten Fall ständige Kreisbewegungen zu machen, weil das wirklich unerträglich ist. Also dann höre ich mir lieber einen Mono-Mix an, tatsächlich. Solche Werkzeuge sind unglaublich hilfreich, zum einen um zu Gestalten und zum anderen aber auch um zu verstehen was verschiedene Bewegungen in verschiedenen musikalischen Kontexten überhaupt mit uns machen.“ (ebd., ab 09:10).

Langhammer findet die Verlagerung der Spatialisierung zum Komponisten ebenfalls gut. Aber auch für ihn sollte die räumliche Einflussnahme in der Postproduktion selbstverständlich bleiben. Ein Tool wie der SWS helfe in der Postproduktion einen Anhaltspunkt zu haben, was sich der Künstler bezüglich Positionen und Bewegungen der Klänge vorgestellt hat. Wörtlich sagt er: "Der Schritt zu den eigentlichen Kreativen, den Komponisten oder Sounddesigner, [sie] in die Lage zu versetzen während der Produktion schon zu sondieren wie er oder sie sich die räumliche Verteilung oder die räumliche Ausbreitung eines gewissen Sounds vorstellt finde ich super, weil wie ich vorhin erklärt hab', wir gehen jetzt davon aus, dass wir immer nur einen Haufen an Spuren

²⁶ Es wird auf die unterschiedlichen Abstraktionsebenen hingewiesen. In der technischen Beschreibung der Implementierung wird mitunter bewusst zwischen Begrifflichkeiten differenziert. Im Evaluationsteil werden Begriffe in Abhängigkeit des Experten mitunter unterschiedlich verwendet, meist umgangssprachlich verbalisiert, mehr aus der Anwenderperspektive gedacht und weniger aus der Logik des technischen Prozesses heraus begriffen. Die Formulierungen werden in der Auswertung weitestgehend übernommen, um das interpretative Maß speziell im deskriptiven Teil möglichst gering zu halten.

bekommen und dann können wir diese Spuren auf Objekte legen und damit halt die Ausdehnung, die Verräumlichung sozusagen halt erst im Mix [machen]." (Langhammer, 2021, ab 28:49).

Für Regnier ist die Spatialisierung fester Teil der Komposition, bestehend aus mehreren Parametern und mit hoher Komplexität. Für ihn stellt sich die Frage nach der Verzahnung nicht, sondern ausschließlich nach dem ‚Wie‘ der Implementierung. Er bezeichnet es als „crazy difficult task. It's where you can explore things, what people didn't do yet" (Regnier, 2021, ab 29:35). Er sieht aber durchaus die Problematik im Status quo: "Can we consider spatialization of sound not at last stage of the work, not additional stage of the work? You compose, then you spatialize. Can we consider it as part of the composition, integrated and can we justify spatialization as part of the sound and not just as a mixing stage at the end? Very often it's difficult." (Regnier, 2021, ab 13:10). Ferner beschreibt er den aktuellen Stand von Spatialisierung im Produktionsprozess populärer Musik anhand einer Analogie: „Many times, I see people doing a composition and seeing spatialization as an additional layer. You have the same problem with spatialization audio as you had before and still today with electronic music with people working with acoustic instrument [...] they want electronics as well in their piece and very often they add electronics as sort of additional layer. But you don't see a real reason, a real connection between the composition with acoustic instruments and the electronic parts. And with spatilization I have the feeling, there is often this problem as well. People are making sounddesign, composition, sound synthesis AND ok now let's spatialize. But what about considering that the sound synthesis [...] is already integration spatialization deeply within the synthesis method." (ebd., ab 13:48). Für Regnier stellt das eine große Herausforderung dar; Klangexperimente, die nur über die Durchführung und empirische Überprüfung gelingen können. „And this is a difficult question, because it raises many, many issues, I think. It raises question about relationship sound localization and spectrum. Also, the conflict between what you hear and what you see on your display of spatialization and many things in my opinion you cannot anticipate in terms of will this strategy be successful: you have an idea, you say ok this sounds like a cool idea for spatialization. What about I spread my polyphonic voices all around the space. And then you try and it's not as nice as it felt. And sometimes you have an idea that seems stupid and then it sounds nice. There is a lot with spatial audio and with audio in general, that [...] you cannot know if the idea is good until you try it in the real space with the system and you see aaah, what was looking good and what seemed like a good idea is actually not working" (ebd., ab 15:04). Darüber hinaus sieht er die Integration von 3D-Audio in ein Tool zur Klangerzeugung als eine Spezialisierung des Tools, das zu einer Einschränkung des Nutzerkreises führt. Diesem Risiko steht die erhöhte Chance gegenüber, etwas Neues, Beeindruckendes zu machen.

Romanov hält es für wichtig, speziell für 3D-Audio komponieren zu können. Er schätzt die vorge-schlagene Integration als einen guten Ansatz für ein Kompositionstool für 3D-Audio ein. Er gibt

zu bedenken, dass eine Integration auch zu einem schnelleren Workflow führen kann, da „Sound in 3D [...] schon da ist“ (Romanov, 2021, ab 34:19). Komplexe Bewegungen vieler Audioobjekte einzeln zu Pannen könne einen erheblichen Mehraufwand darstellen.

Auch Hoppe hält die Integration für einen richtigen Vorschlag. Bei Trennung von Komposition und Spatialisierung lande man immer bei ähnlichen Ergebnissen. „Dieses Verfahren [...] jemand macht ein Musikstück und jemand anderes spatialisiert das. Das ist das was man aufbrechen muss, weil es immer in ähnlichen Ergebnissen landet. Und richtig emotional wird's ja nur wenn man's verzahnt und `ne Sängerin z.B. mit dem Raum mal richtig spielen kann aus der Emotion heraus. Das ist aber immer nur ein Teil der Wahrheit“ (Hoppe, 2021, ab 20:05). Hoppe denkt das Thema weiter und macht deutlich, dass es auch auf die Wirkung beim Publikum ankommt: „Weil selbst das muss bei den Leuten ankommen. Das heißt dann haben wir weitergeguckt, kann man das alles in eine sehr plakative performante Bewegung packen, dass das Publikum das auch noch versteht. Das ist ,nen weiter Weg von deinen Plug-Ins Azimut-Wert Pitch hoch runter zu nem Publikum das nen flashigen 3D-Effekt abbekommt. Das ist mitunter ein schwieriger und barrierevoller Weg“ (ebd., ab 20:30).

Rieger findet den Vorschlag gut und schätzt solche Tools als hilfreich ein, um zu überprüfen, ob native 3D-Audio-Musik einen Mehrwert bietet: „Also ich finde die Idee prinzipiell gut, weil offensichtlich gibt's ja fast noch gar keine solche Tools. Von daher ist es gut sowas zu haben und zu gucken ah ok wird das auch angenommen und wenn nicht ist das halt nicht so, aber dann weiß man das wenigstens.“ (Rieger, 2021b, ab 39:25).

Zorn hält den Vorschlag für sinnig. "Man kann schöne Sachen damit machen, es ist eine weitere Differenzierungsebene, mit der sich musikalische Dinge darstellen lassen" (Zorn, 2021, ab 19:25). Er fragt sich aber allgemein, wie komplexe Spatialisierung beim Hörer abgebildet, sprich adäquat wiedergegeben werden soll. Das Hauptproblem sieht er im Markt und einer entsprechenden Verwertungskette.

Scuda findet den Vorschlag spannend, da die komplexe Raumgestaltung ‚content driven‘ ist, also abhängig vom Signal selbst. Er sieht in dem Vorschlag aber auch eine Zeitersparnis und erweiterte gestalterische Möglichkeiten. Auch wenn der Vorschlag des SWS nicht seiner bisherigen Arbeitsweise entspricht, erkennt er die grundlegende Idee an. „Was ich super finde an dem Tool, darin sehe ich auch einen großen Wert oder eine große Leistung von diesem Instrument, ist, dass der Raum sehr viel dichter an den Anfang der Produktionskette gebracht wird. Also das derjenige, der am Keyboard steht und Knöpfe drückt, dass der quasi sehr direkt vor Augen geführt bekommt oder gezwungen wird sich über den Raum am Ende Gedanken zu machen. Ich kann mir vorstellen,

dass viele in Ableton oder sonst wie, weil da nur Stereo rauskommt, alles nur Stereo produzieren und sagen später schiebe ich das im Raum zusammen oder ich geb's meinem Kumpel, der kann das und dann kommt der Raum als künstlerisches Element später dazu als am Anfang gleich mit dabei zu sein. Und dadurch, dass man dem am Anfang gleich sehr unmittelbar dabei hat und in der Hand auch hat, macht sich dann hoffentlich [...] nen Komponist gleich Gedanken darüber: was macht dieser Raum denn eigentlich, wie klingt der denn? Wie kann ich den denn in das was ich tue integrieren als weiteres Ausdrucksmittel. Dass er nicht nur ein Stückchen Effekt der dazu kommt ist, sondern, dass er gleich Bestandteil von dem ist was ich mache und das ist halt vielleicht die größte Leistung dieses Dinges, dass der Raum einfach sofort am Anfang da ist und sagt: hallo übrigens ich bin Raum, mich gibt es, ich will auch irgendwie bedient und bespielt werden" (Scuda, 2021, ab 1:02:35).

Günther beschreibt den Vorschlag als praktischen Ansatz, dessen Umsetzung auf jeden Fall Arbeit abnimmt.

5.2.2 Objektifizierung von Klangkomponenten und objektbasierter Output

Nachdem die Aussagen zum generellen Vorschlag der Integration von 3D-Audio in den Klangerzeuger dargelegt wurden, soll nun verstärkt Bezug auf den Prototyp genommen werden. Es wird auf den Vorschlag der Objektifizierung von Klangkomponenten eingegangen sowie den objektbasierten Output mit sich anschließendem ausgelagerten Rendering.

Koch hält die Objektifizierung von Klangkomponenten für unbedingt nötig: dies ermögliche neue künstlerische Möglichkeiten. Schall sei komplex und quasi nie nur eine Punktschallquelle. Die Wahrnehmung räumlicher Reflexion entspreche dem natürlichen Hören. Der objektbasierte Audiooutput stellt für ihn nur einen möglichen Ansatz dar.

Langhammer hält die Objektifizierung von Klangkomponenten für nützlich, objektbasiertes Audio sinnvoll für ein Zwischenformat im Produktionsprozess und zieht den Vergleich zu Dolby Atmos sowie den ADM²⁷ Export/ Import. Er weist darauf hin, dass sich bezüglich eines Endformats noch nichts final durchgesetzt habe und in dem Bereich viel Bewegung herrsche. Für ihn sind bezüglich des Outputs eines virtuellen Instruments wie dem SWS sowohl Audioobjekte als auch eine HOA Szene vorstellbar. Aktuell nutzt er keine Zwischenformate in seinem Workflow. Ist eine Produktion objektbasiert, ist sie das von Anfang bis Ende. Ein objektbasiertes Zwischenformat hält er aber für wünschenswert: „Ich würde mir schon wünschen, dass man sozusagen ein

²⁷ Audio Definition Model; hier mit Bezug auf das Broadcast Wave File (BWF) Format, das Metadaten für z.B. die Position der Audiosignale enthält (EBU, 2018).

objektbasiertes Zwischenformat nutzen kann, um ja, das sozusagen in andere Mixe als schon vorgemischter Teil einer immersiven Audio-Mischung einzubinden." (Langhammer, 2021, ab 34:32)

Die Möglichkeit zur Objektifizierung von entstehenden Klangkomponenten eines Synthesizers zu haben findet auch Regnier gut. Seiner Meinung nach sei der Zusammenhang auch für ungeübte Nutzer einfach zu verstehen. Für eine weitere Form der Umsetzung verweist er erneut auf den Zusammenhang zwischen Partikeln und Audioobjekten. Wichtig wäre für ihn eine flexible Formatierung des objektbasierten Outputs, sodass dieser von einer Vielzahl an 3D-Audio-Panning-Tools verarbeitet werden kann.

Für Romanov stellt die Objektifizierung einen logischen Ansatz dar. Für ihn sei das eine konsequente Weiterentwicklung von der Verteilung einzelnen Stimmen eines Stereo-Synthesizers im Stereobild.

Hoppe nimmt erneut Bezug zur Wahrnehmung und Wirkung und sagt zu dem Vorschlag: „Das finde ich prinzipiell richtig. Also man muss natürlich gucken gerade wenn's dann irgendwie zu kleinteilig wird verliert es immer seine Wirkung. Wenn sich zu viele Quellen bewegen, der Komplexitätsgrad für die Leute im Publikum ist immer schnell erreicht. Viel schneller als man im Studio das denkt. Daher weiß ich nicht, ob man so super kleinteilig alles gut wahrnimmt. Vom Grundprinzip finde ich das völlig richtig" (Hoppe, 2021, ab 29:38). Er sieht den SWS eher in einem ‚Expert Mode‘, was sich mit der Aussage von Koch über den SWS hinsichtlich eines Tools zum Lernen, Experimentieren und Forschen in gewisser Weise deckt.

Für Rieger ist der Vorschlag zur Objektifizierung von Klangkomponenten auf den ersten Blick intuitiv. Ein Kernproblem sieht er darin, dass es für Komponisten ohne Erfahrung mit 3D-Audio schwer zu beurteilen ist, was gut funktioniert. Daher schlägt er wie auch Hoppe vor, die Möglichkeiten in dem Tool auf das zu reduzieren, was gut funktioniert. Aber das müsse man erstmal ermitteln. Den Komplexitätsgrad der Umsetzung bewertet er als enorm hoch.

Zorn bewertet den Ansatz direkt auf ästhetischer Ebene: „Auch super sinnvoll, ist aber schon ästhetisch ein riesen Paradigma. Weil das hat man schon in den Demos [...] gehört. Je komplexer du geworden bist, je mehr du gelayert hast, desto unpräziser sind die Sachen teilweise geworden." (Zorn, 2021, ab 20:50). Er gibt zu bedenken, dass das Gehirn viel mehr Zeit brauche, um Räumlichkeit zu verstehen, als das beim Verstehen von Identitäten der Fall ist. Die Komponenten seien theoretisch dar, werden aber nicht zwangsläufig begriffen.

Scuda findet den ganzen Vorschlag spannend. Er betrachtet die Positionierung und Bewegung einzelner Klangkomponenten skeptisch hinsichtlich kognitiver Unterscheidbarkeit. Die Wirkung

müsse für ihn erst einmal überprüft werden. Er fragt sich, wie sich die Klänge des SWS mit anderen Elementen der Musik, beispielsweise perkussiven Beats, verbinden. Und wie man zwischen den musikalischen Elementen Abhängigkeiten hinsichtlich der Spatialisierung schaffen kann.

Günther hält den Ansatz für „total sinnvoll“ (Günther, 2021, ab 23:12). Speziell den formatagnostischen, offenen Ansatz, mit Möglichkeit zur Verwendung unterschiedlicher 3D-Audio Panning Tools, begrüßt er explizit (ebd., ab 1:05:03). Für ihn gilt es herauszufinden, welche anderen Arten von Klangkomponenten noch interessant sein können. Im gestalterischen Prozess findet er es wichtig, darüber möglichst viel Kontrolle zu haben.

5.2.3 ‚Layer-Mode‘ und ‚Voice-Mode‘

Nach der Einschätzung des Vorschlags zur Objektifizierung von Klangkomponenten und dem objektbasierten Ausgang des virtuellen Instruments, werden nun die Aussagen der zwei vorgeschlagenen Modi – ‚Layer-Mode‘ und anschließend ‚Voice-Mode‘ – gegenübergestellt.

Für Koch funktioniert der ‚Layer-Mode‘ Ansatz gut. Die Aufteilung in diese Art von Klangkomponenten sei wahrnehmbar und daher gerechtfertigt. Für ihn ist das ein interessantes Stilmittel. Er würde den Ansatz erweitern und mit Reflexionen aus einem Hallgerät kombinieren.

Langhammer begrüßt diesen Vorschlag, er entspräche dem bisherigen Aufbau von Synthesizern – die Layer ergeben eine musikalische Sinneinheit. Zudem entspricht der Ansatz auch der Anlieferung durch die Künstler in seinem zweigeteilten Workflow: „Also dadurch, dass das jetzt dem bisherigen Aufbau eines Synthesizers entspricht, mit multiplen Oszillatoren einen Sound zu formen, find ich das jetzt einfach nur straight forward. Ok man macht das halt auch in nem Spatial Synthesizer so. Ich kann jetzt halt noch nicht sagen, ob es in der Synthese oder in dem kreativen Prozess unbedingt nötig ist, also z.B. den Output von nem Noise Oszillator anders [räumlich zu positionieren] als nen Triangle oder Sawtooth oder was immer. Aber die Möglichkeit zu haben das zu tun, ist auf alle Fälle gut“ (Langhammer, 2021, ab 37:10). Er gibt aber auch zu bedenken, dass die zusätzliche Ebene für eine höhere Komplexität sorgt. „Es ist halt ne zusätzliche Ebene im Sounddesign die der Künstler dann beim Erstellen des Sounds in der Hand hat. Macht halt mehr Komplexität. Aber ich glaube es würden einige Leute kommen und sagen: hätte ich da jetzt nicht Zugriff drauf oder könnten Sie mir da Zugriff geben, dass ich Noise-Anteil anders positionieren kann als den tonalen Anteil“ (ebd., ab 38:00).

Regnier assoziiert mit der Aufteilung in Layer eine „full clouds of sounds“ (Regnier, 2021, ab 35:36) und befindet den Vorschlag als gut. Dabei stellt sich für ihn die Frage, ob diese Klangwolke voll genug für eine vollständige Umhüllung sei und welche Möglichkeiten bestehen, noch dichtere

Klänge zu erzeugen. Den ‚Spread‘-Parameter sieht er analog zu einer Unison-Funktion²⁸ und nennt es eine „good strategy and for users and easy to understand“ (ebd., 37:20).

Romanov findet den Vorschlag gut und würde sich für die einzelnen, einen Layer darstellenden, Oszillatorgruppen unabhängige Tiefpassfilter-Grenzfrequenzen wünschen.

Hoppe zieht den Vergleich zu einem Plug-In von *Blueripple Audio*, das ein beliebiges Audiosignal eines Live-Inputs in mehrere Frequenzbänder zerlegt und die spektralen Komponenten automatisiert im Raum verteilt – für ihn ein gelungenes Beispiel eines gebündelten Effekts, steuerbar über einen Knopf und direkt spielbar. Den darüber erzeugten Effekt bewertete er wahrnehmungstechnisch besser als das klangliche Ergebnis des ‚Layer-Mode‘. Auch Rieger zieht den Vergleich zu Tools, die ein Live-Input Audiosignal über Filterungen in Frequenzbänder zerlegen und diese räumlich verteilen. Rieger würde untersuchen, für welche Kategorie von Klängen der Layer Vorschlag Sinn ergibt. Generell sei das ein guter Gedanke, die Strategie der Spatialisierung stärker an die musikalische Funktion oder Beschaffenheit des Klangs zu koppeln. Bei ‚Pad-Sounds‘ kann er sich die unterschiedliche räumliche Verteilung verschiedener Layer gut vorstellen.

Zorn hält den Vorschlag prinzipiell für eine gute Idee und kann den Ansatz aus seiner kompositorischen Tätigkeit heraus bestätigen. Die räumliche Kopplung an Klangfarben sei auch wieder eine ästhetische Entscheidung. Er wünscht sich mehr Differenzierungsmöglichkeiten: welche Elemente stehen im Vordergrund, welche im Hintergrund?

Scuda hält sich bei Synthesizer spezifischen Fragen zurück. Er plädiert allgemein für transientenreiche Klänge, um die räumliche Aufteilung besser wahrzunehmen.

Günther bewertet die Entkopplung der einzelnen Oszillatoren als sinnvoll. Allerdings fehlen ihm in der Umsetzung mehrere Optionen: Diese könnten über der Zeit nicht präzise genug bearbeitet werden – unabhängige Hüllkurven pro Oszillatorgruppe wären wünschenswert (vgl. time-varying Wavetable Synthesizer, Kapitel 2.1). Die ‚MW-Zones‘ seien aber ein guter Schritt in die Richtung. Darüber hinaus bräuchte es separat einstellbare Tiefpassfilter-Grenzfrequenzen und feinere Tonhöhenvariationsmöglichkeiten pro Oszillatorgruppe in Cents und Halbtönen, nicht nur in Oktaven. Zudem wünscht sich Günther mehr Kontrolle über den ‚Spread‘-Parameter. Die Geschwindigkeit der entstehenden Schwebung solle auf das Tempo synchronisierbar sein.

Im Folgenden werden die Bewertungen zum vorgeschlagenen ‚Voice-Mode‘ dargelegt.

²⁸ Unison meint im Synthesizer Kontext mehrere, leicht gegeneinander verstimimte Suboszillatoren eines Klangsyntheseblocks, die zu einer räumlichen Ausdehnung und einem volleren Klangbild führen können.

Die räumliche Verteilung einzelner polyphoner Voices hat für Koch „super gut funktioniert“ (Koch, 2021, ab 29:13). Dies sei ein „Effekt, [...] [der] ganz klar den Raum aufmacht“ (ebd., 13:08). Auch Langhammer hat einen sehr guten Eindruck von dem Vorschlag. Er berichtet von einem Projekt mit Hardware Modular-Synthesizern, bei dem für ein gespieltes Arpeggio jede Note an eine andere räumliche Position verteilt wurde und bestätigt dabei, dass es sich bei diesem Ansatz auch aus seiner Erfahrung um einen „kreativen Wunsch“ der Musiker handele (Langhammer, 2021, ab 41:43).

Für Regnier ist die Verteilung einzelner Voices offensichtlich. Er gibt zu bedenken, dass das Voice Management die Spatialisierung beeinflusse und es mit beispielsweise ‚last note played‘, ‚first note played‘, ‚round robin‘ mehrere Optionen im Voice Management bestehen. Regnier wünscht sich mehr Kontrolle über die Verteilung der Voices auf die Audioobjekte. Im Dialog entwickelt sich die Idee eines ‚Spatial Step Sequencers‘: für jeden Step gibt es einen zusätzlichen Parameter, mit dem sich eine Note auf ein bestimmtes Audioobjekt im Raum routen lässt. "You can create a sort of narration, a story [in space]" (Regnier, 2021, ab 39:10).

Für Rieger ergibt der ‚Voice-Mode‘ Vorschlag Sinn, allerdings ist auch ihm die Verteilung der Voices auf Audioobjekte zu willkürlich. Es brauche musikalische Kontrolle: beispielsweise der Grundton immer auf das erste Audioobjekt. Das räumliche Verteilungsmuster der Objekte stehe in der Abhängigkeit zum musikalischen Material.

Romanov sieht die Aufteilung in Voices, wie auch schon den Ansatz der Objektifizierung, generell als logischen Schritt. Aber „wenn ständig Sound aus einer anderen Richtung kommt [findet er] das eigentlich eher störend“ (Romanov, 2021, ab 37:34). In Zusammenspiel mit einem Arpeggiator könne er es sich den Ansatz allerdings gut vorstellen, auch für Klangwolken ergäbe dieser Sinn. Wie Regnier weist er auf die verschiedenen Voice Management Algorithmen hin und regt an, diese, um individuelle Varianten zu erweitern. Beispielhaft schlägt er eine Gauß-Verteilung der Audioobjekte um einen Ankerpunkt vor. Für Romanov gibt es hier nichts Absolutes, da es sich immer auch eine künstlerische Entscheidung handele.

Hoppe begrüßt diese einfache Form der Aufteilung, bei der man nichts erklären müsse und sagt: „Das finde ich sehr nah an nem Musiker zum Beispiel, das finde ich sehr barrierearm. Also wenn man sowas ganz einfaches - wir hatten auch mal sowas gemacht, wenn man sagt, ok ein Pianist zum Beispiel der hat jetzt mal – wenn er nicht Nils Frahm ist – keine Ahnung von sowas. Ob man nicht einfach ne Klaviatur quasi aufteilt im Raum" (Hoppe, 2021, ab 32:48).

Für Zorn ergibt die räumliche Verteilung des harmonischen Gefüges Sinn. Wie die anderen Komponisten vermisst auch er eine bessere Kontrolle über die räumliche Zuteilung des musikalischen

Materials. „Das finde ich ist eins der guten Potentiale der Sache. [...]. Das Einzige was ich bisschen - was mir dabei Bauchschmerzen machen würde ist, wie man das sinnig aufteilt, dass sich diese Voice, dass die nicht irgendwie gestaffelt von 1 nach 5 gehen. Das man da Presets baut, die ein schnelles Ergebnis liefern, mit einer einfachen Oberfläche" (Zorn, 2021, ab 26:45). Das feste Routing von Voices auf Audioobjekte verschenke kreatives Potential. Auch Zorn dachte intuitiv an die Kopplung mit einem Arpeggiator oder Step Sequencer, außerdem an die Öffnung, beziehungsweise die Kopplung mit einem beliebigen monophonen virtuellen Instrument.

Günther sagt zu der räumlichen Verteilung polyphoner Voices: „Leute kennen das so nicht" (Günther, 2021, ab 35:05). Auf diese Weise räumlich arbeiten zu können habe großes Potential. Er zieht den Vergleich zu einem Projekt, bei dem die einzelnen Saiten einer elektrischen Gitarre auf mehrere Lautsprecher im Raum verteilt wurden. Allerdings muss auch für ihn die Verteilung der Noten auf Objekte nachvollziehbar und reproduzierbar sein. Die Kombination mit einem Step Sequencer könne er sich gut vorstellen.

5.2.4 Bewegung der Audioobjekte

Abschließend werden die Aussagen zu den Vorschlägen zur Bewegung der Audioobjekte dargestellt. Diese umfassen die Modulation der richtungsbestimmenden Winkel über Modulationssignale der Klangsynthese und MIDI-Befehle sowie die Makrobildung ‚Anchor-Mode‘.

Koch wünscht sich für eine produktnahe Umsetzung mehr Presets und die Verweigerung schlecht klingender Konfigurationen. Es steigere nicht zwingend die Qualität der Spatialisierung, wenn beispielsweise der gleiche LFO für die Modulation der Tiefpassfilter-Grenzfrequenz und eine Kreisbewegung zuständig ist. Der Hersteller eines Instruments trage eine gewissen Verantwortung, da er mit der Gestaltung Einfluss auf den Einsatz und die Nutzung des Instruments nimmt. Den ‚Anchor-Mode‘ hält Koch für eine gelungene Funktion, die eine einfache Gestaltung von räumlichen Kontrasten ermögliche.

Für Langhammer hat Bewegung viel mit der Dynamik des Stücks zu tun, in das sich der Sound einbettet. Für das Sounddesign findet er die Option zur Kopplung über Modulationsquellen wichtig und die Umsetzung gelungen. Den allgemeinen Ansatz hält er für richtig: „Parameter aus der Synthese zu verwenden, um damit Positionsdaten zu erzeugen ist durchaus, ja, wünschenswert oder zielführend für die meisten Musiker denke ich" (Langhammer, 2021, ab 49:21). Speziell die Kopplung von Änderung der Tiefpassfilter-Grenzfrequenz mit der Position des Audioobjekts bewertet er als gut. Den ‚Anchor‘ begreift er als eine Art Gruppenobjekt, über dieses in die Dynamik der Bewegung eingegriffen werden kann. Er wünscht sich Zugriff auf den ‚Anchor‘ auch in der

Mischung durch den Toningenieur und sieht diesen Ansatz als Teil einer Lösung eines Problems speziell bei dichten räumlichen Arrangements.

Regnier zeigt sich ebenfalls begeistert von dem ‚Anchor-Modus: „der ‚Anchor‘-Idee finde ich super eigentlich. Das ist wirklich etwas das ich würde das extrem nützlich finden. Das ist wirklich die Art von Sachen das für Spatial Audio sehr nutzbar wäre. Es gibt ein Ort im Raum und etwas passiert im Zusammenhang mit diesem Anchor. Das finde ich eine gute Idee und etwas - ohne das, ich würde finden es fehlt. Das finde ich sehr sehr gut zu haben [...] this is where you start connect things together" (Regnier, 2021, ab 42:35). Für die Modulation von Audioobjekt-Positionen wünscht sich Regnier mehr Feinjustierung. Beispielsweise würde er gerne die Hüllkurve für den Tiefpassfilter-Grenzfrequenz logarithmisch und für die Positionsänderung linear gestalten können. Wichtig sei, was die Parameter wahrnehmungstechnisch verbindet. Die Kopplung über die Steuersignale ist Regnier tendenziell etwas technisch gedacht. Zudem stellt sich für ihn die Frage, welche Optionen der Modulationsmatrix dem Nutzer überhaupt zur Verfügung gestellt werden sollten.

Romanov bewertet den Vorschlag als gut. Er stelle eine einfache Möglichkeit der räumlichen Mikro-romodulation dar, die speziell bei elektronischen Klängen für Lebendigkeit und organische Sounds Sorge. Die Frequenz des LFO auf das Merum der Musik zu synchronisieren, ist für ihn ein wichtiger Punkt: dies stärke das musikalische Ineinandergreifen. Auch er gibt zu bedenken, dass nicht jede Bewegung wahrnehmungstechnisch sinnig ist und vor allem gelernte Klänge räumlich gut wahrnehmbar sind. Viele der Bewegungen mit fremden, elektronischen Quellen würden für das räumliche Hören nicht sonderlich gut funktionieren. Zudem stelle die Modulation von Azimut und Elevation sowie die drei Achsenrotationen noch nicht das Optimum da. Romanov schlägt vor, über Koordinatentransformationen auf weitere Bewegungsformen zu kommen und den gesamten Block zur Richtungsberechnung auf wahrnehmungstechnisch relevante Optionen begrenzen. Zudem stellt er die Frage nach räumlichen Wiedergabesystemen ohne elevierte Lautsprecher und dem einhergehenden Informationsverlust der Elevation von Klängen bei der Wiedergabe darüber. Den ‚Anchor-Mode‘ hält Romanov für eine sinnvolle Funktion.

Aus der Erfahrung von Hoppe seien Bewegungen über „Preset-Kurven nicht wirklich expressiv“ und deren künstlerische Wirkung „schnell durch“ (Hoppe, 2021, ab 34:15). Für ihn hat die Bewegung der Audioobjekte über Modulationssignale aber funktioniert. Auch er wünscht sich hier einen verstärkt wirkungsorientierteren Fokus. Den ‚Anchor-Mode‘ bezeichnet er als einen von ihm geforderten „One-Knob Spektakel Effekt“ und befindet ihn für gut (Hoppe, 2021, ab 40:12).

Rieger würde bei der initialen räumlichen Verteilung und Bewegung der Objekte verstärkt auf wahrnehmungstechnisch gut funktionierende Presets zurückgreifen und den Distanz-Parameter dazu nehmen. Darüber hinaus gibt er zu bedenken: „Dann ist das noch nur ein Sound. Du hast ja mehrere Sounds in einer Komposition. Willst du wirklich fünfmal mit diesem Synthesizer arbeiten? Du hast so viele Sounds die so kreuz und quer durch die Gegend bewegen weiß ich nicht, ob das dann am Ende auch gut klingt - müsste man mal ausprobieren" (Rieger, 2021b, ab 44:21). Den ‚Anchor-Mode‘ sieht er sehr positiv – diese Form der gruppierten Bewegung erscheint ihm sinniger als die Bewegung einzelner Objekte. Aus seiner praktischen Erfahrung bestätigt er die Wichtigkeit akustischer Anker, die dem Hörer zur Orientierung dient und auch ohne Center-Lautsprecher einen räumlichen Bezugspunkt etablieren. Er empfiehlt weitere wahrnehmungstechnisch sinnvolle Makrobildungen zu entwickeln.

Für Zorn ergibt der Vorschlag Sinn. Die räumliche Positionierung und Bewegung käme aus dem Klang heraus und sei keine zusätzliche Ebene, die am Ende dazukommt. Der „Anchor-Modus [ist die] beste Möglichkeit, um zum Beispiel Pads zu rhythmisieren" (Zorn, 2021, ab 35:12). Allgemein erscheint ihm die Benutzeroberfläche für die Positionierung für zu voll. Parameter sollten reduziert oder zumindest versteckt werden.

Scuda erkenne bei den Demo-Videos vor allem visuell viel von dem was an Bewegung von Statuen geht. Schließt er die Augen, nehme er nicht mehr ganz so viel davon wahr. Es gälte zu überprüfen, inwiefern sich ein wahrnehmbarer Effekt auch über ein simples „Auto-Panning Ping-Pong Delay erreichen“ ließe (Scuda, 2021, ab 34:28). Ein wichtiger Punkt stellt für Scuda die Interaktion mit anderen musikalischen Elementen dar. Mit einem Instrument ließe sich schließlich keine zufriedenstellende Komposition realisieren. Daher schlägt Scuda vor, die Form der halbautomatisierten Panoramisierung über verschiedene Instrumente und Spuren hinweg zugänglich zu machen. Dahinter steckt der Vorschlag, sich bei den Signalen für die Positionsmodulation von Audioobjekten nicht auf das eine virtuelle Instrument zu beschränken, sondern auf Signale verschiedener virtueller Instrumente oder Audioeffekte innerhalb der DAW zugreifen zu können. Der ‚Anchor-Modus‘ ist für ihn der Ausgangspunkt spannender Sounddesign Experimente. In Kombination mit einem objektbasierten Hall könne er sich ein „Atmen des Raums“ über das ‚Zusammenziehen‘ und ‚Loslassen‘ der Hallkomponenten vorstellen (Scuda, 2021, ab 43:37).

Günther bewertet die Vorschläge zur Bewegung für grundlegend gut – räumliche Geometrie und Klang lägen eng beieinander. Die Art und Weise wie Klangparameter und räumliche Parameter zusammenhängen, müsse aber noch besser erforscht werden. Die Modulationssignale der Klangsynthese seien dafür ein guter Anfang. Wahrnehmungstechnisch funktionierende Kopplungen müssen gefunden werden. Den ‚Anchor-Mode‘ findet SG theoretisch ebenfalls gut und kann ihn

sich in der kompositorischen Praxis vor allem für langsame räumliche Entwicklung vorstellen. Wie auch Regnier nennt Günther das Schwarmverhalten einer Partikel-Simulation als eine gute Organisationsform für viele Audioobjekte und eine Erweiterung zu LFOs und Hüllkurven.

5.3 Diskussion

In der folgenden Diskussion werden für jeden Themenbereich die Bewertungen der Experten zu den Vorschlägen gegenübergestellt und so ein Stimmungsbild hinsichtlich der einzelnen Themenbereiche gezeichnet. Wie in der Auswertung ersichtlich wurde, gingen die Ausführungen der Experten meist über die Bewertung der konkreten Vorschläge hinaus. Dieser zusätzliche Erkenntnisgewinn, der sich aus der Expertise der Experten ergibt, wird zusammenfassend dargelegt und eingeordnet. Wie bereits erwähnt ist der Experte Repräsentant seiner ‚Zunft‘. Ziel der folgenden Diskussion ist im Vergleich zur deskriptiven Auswertung das ‚Überindividuell-Gemeinsame‘ der Deutungen herauszuarbeiten (Meuser & Nagel, 2009, S. 82). Allerdings ist dies bei den nicht trennscharf kategorisierbaren beruflichen Hintergründen der Experten nicht immer möglich.

Konzept

Der Vorschlag zur Integration von 3D-Audio in das virtuelle Instrument und der damit einhergehenden Verlagerung der Spatialisierung zum Komponisten wird von allen Experten als ein richtiger und wichtiger Schritt angesehen. Speziell für die Komponisten steht diese Arbeitsweise nicht zur Disposition und ist ein uneingeschränkt sinnvoller Ansatz – die Gestaltung der räumlichen Attribute ist fester Bestandteil des Kompositionsprozesses. Regnier und Hoppe betonen, dass es zu anderen kompositorischen Ergebnissen führt, wenn die Spatialisierung auf einer Ebene mit der Klangerzeugung und nicht in einem nachgelagerten Schritt stattfindet. Koch und Langhammer gewinnen dem Vorschlag ebenfalls viel Positives ab, da er Orientierung für ihre Arbeit in der Mischung geben kann. Sie fordern aber den Erhalt zur Möglichkeit der Einflussnahme auf die räumlichen Attribute in der Postproduktion.

Nach Zorn und Langhammer darf die erhöhte Komplexität eines virtuellen Instruments durch die Integration von 3D-Audio nicht unterschätzt werden. Wie es Zorn formuliert, stellt diese eine zusätzliche Differenzierungsebene dar, die vom Komponisten bedacht und bewusst gestaltet werden muss. Romanov, Scuda und Günther machen zudem auf die potenzielle Zeitersparnis aufmerksam, die eine Integration von 3D-Audio in das virtuelle Instrument mit sich brächte. Komplexe Spatialisierung kann in der Postproduktion einen erheblichen Mehraufwand darstellen.

Für Scuda und auch Rieger ist es essenziell, das virtuelle 3D-Audio Instrument in einen Kontext mit anderen musikalischen Elementen zu bringen, da für sie die Wirkung des 3D-Audio Instruments erst im Arrangement beurteilt werden kann.

Auch wenn Künstlern ein Abstraktionsvermögen hinsichtlich der räumlichen Gestaltung haben wenn sie Klänge mit der Intention zur Spatialisierung in 3D-Audio erzeugen, ist speziell bei Regnier deutlich geworden, dass eine Klangvorstellung hörbar gemacht werden muss, um sie in ihrer Wirkung zu überprüfen. Speziell bei der Spatialisierung von 3D-Audio bestehen noch wenig etablierten Hörgewohnheiten und Konventionen, die zur Orientierung dienen können. Daher ist ein wirkungsorientiertes Experimentieren mit den Gestaltungsmöglichkeiten eines 3D-Audiosystems hinsichtlich der Spatialisierung eine Aufgabe für die künstlerische Forschung. Wie Koch und Scuda bestätigen, können virtuelle Instrumente, welche Möglichkeiten zur Spatialisierung für 3D-Audio integrieren, wichtige Werkzeuge für diesen Prozess sein.

Workflow

Der Vorschlag zur Objektivierung von Klangkomponenten des Syntheseprozess und deren Panoramisierung im virtuellen Instrument wird von allen Experten begrüßt oder für einen interessanten Ansatz befunden. Die die kleinteilige Spatialisierung von Klangkomponente eröffne gestalterisch neue Möglichkeiten - Hoppe, Zorn und Scuda äußern sich allerdings skeptisch hinsichtlich der Wahrnehmbarkeit. Hoppe und Rieger sehen zudem die durch die Möglichkeiten zur Objektivierung und Panoramisierung entstehende Komplexität der Benutzeroberfläche als hoch an und fordern eine Reduktion auf wirkungstechnisch sinnvolle Optionen. Für sie stellt der aktuelle Stand einen ‚Expert-Mode‘ dar.

Dem Vorschlag zum objektbasierten, formatagnostischen Output des Instruments und dem sich anschließenden, ausgelagerten 3D-Audio Rendering wird von Langhammer, Regnier und Günther viel Potential zugesprochen. Die Komponisten Regnier und Günther wünschen zudem eine flexible Formatierung des Outputs, um das virtuelle Instrument mit möglichst vielen 3D-Audio-Panning-Tools verwenden zu können. Für Langhammer stellt auch ein in das virtuelle Instrument integriertes 3D-Audio Rendering eine sinnvolle Option dar. Auch Koch hält den objektbasierten Output nicht für die einzig mögliche Lösung.

Klangkomponenten

Der vorgeschlagene ‚Voice-Mode‘ wird von acht Experten entweder als gut oder logische Konsequenz beurteilt, Scuda äußerte sich hierzu nicht. Laut Koch öffnet die Verteilung der Voices den Raum was laut Günther einen Effekt ermöglicht, den man so nicht kennt. Romanov schränkt ein,

dass sich diese Art von räumlicher Verteilung einzelner Noten vorrangig für Stilmittel wie beispielsweise Arpeggios eignet. Regnier, Rieger, Romanov und Zorn fordern zudem eine Form der musikalischen Kontrolle zur Verteilung der Voices auf die Audioobjekte. Wie Regnier und Romanov beschreiben, hat die Umsetzung des Voice Management Auswirkungen auf die Verteilung. Die Komponisten & Sounddesigner Regnier, Zorn und Romanov können sich die Kopplung an einen Step Sequencer vorstellen.

Der vorgeschlagene ‚Layer-Mode‘ wird von den Experten ebenfalls weitestgehend als gut bewertet. Koch hält ihn für wahrnehmungstechnisch gerechtfertigt und Langhammer erachtet den Zugriff auf die einzelnen Oszillatoren für sinnvoll. Hoppe zeigt sich nicht überzeugt. Er zieht den Vergleich zu einem Audio-Effekt Plug-In, das den Live-Input über Filter spektral zerlegt und anschließend spatialisiert. Dieser Ansatz sei flexibler und in seiner Wirkung überlegen. Auch Rieger stellt den Vergleich zu einem solchen Audio-Effekt her.

Regnier ist nicht sicher, ob die Dichte des Klangbilds, das sich über die räumliche Verteilung mehrerer Layer und deren Dekorrelation über den ‚Spread‘-Parameter erzeugen lässt, für eine wahrgenommene Umhüllung ausreicht. Er schlägt vor, jeden Layer um weitere Unisono ‚Suboszillatoren‘ zu erweitern, die ein dichteres Klangbild ermöglichen. Eine Form der gebündelten Kontrolle über die räumliche Verteilung der ‚Suboszillatoren‘ eines Layers, könnte in Anlehnung an den ‚Anchor-Mode‘ erfolgen.

Die Toningenieure Koch und Rieger wünschen sich zudem die Kombination der Layer mit einem 3D-Hallgerät – Koch um komplexe Reflexionen simulieren zu können, Rieger für die Entfernungsdarstellung.

Mit Romanov, Zorn, Regnier und Günther äußern sich vier Experten mit Hintergrund in der Synthesizer-Programmierung zur detaillierteren Einflussnahme auf die Oszillatoren der Layer. Romanov fordert unabhängige Tiefpassfilter-Grenzfrequenzen für jeden Layer. Günther wünscht die Erweiterung der Transponierung um Halbtöne und Cents sowie mehr Kontrolle über den ‚Spread‘-Parameter. Regnier würde unabhängige Amplitudenhüllkurven für jeden Layer begrüßen.

Aus den Aussagen der Experten kann abgeleitet werden, dass die beiden Modi für die weitere Entwicklung besser auf zwei verschiedene virtuelle Instrumente verteilt werden sollten. Für einen wirkungsorientierten Aufbau des Instruments scheinen in Abhängigkeit der objektifizierten Klangkomponenten spezifische Optionen erforderlich (wie beispielsweise ein Step Sequencer im ‚Voice-Mode‘ und weitere Suboszillatoren im ‚Layer-Mode‘).

Bewegung

Den Vorschlag zur Bewegung der Audioobjekte über Modulationssignale des Klangsinteseprozess und Eingabe des Musikers beschreiben die Experten überwiegend als guten Ansatz. Die Aussagen von Koch, Günther, Scuda, Rieger, Hoppe, Romanov und Regnier decken sich aber dahingehend, dass sie einen wirkungsorientierten Ausbau der Modulationsoptionen fordern. Den auf das Metrum der DAW synchronisierten LFO hält Romanov für ein gutes Modulationsquellsignal, da dies für eine musikalische Verzahnung Sorge. Regnier fehlen zudem Optionen zur Feinjustierung, wie beispielsweise die Wahl zwischen linearen und logarithmischen Hüllkurvenverläufen.

Laut Langhammer sind die Tiefpassfilter-Grenzfrequenz und laut Regnier die Wavetable-Position prinzipiell richtige klangformende Parameter für die Kopplung mit einer Bewegung. Hoppe, Zorn und Rieger fordern für die Bewegungen Preset-Bildungen sowie die Beschränkung auf in der Wirkung überzeugende Optionen.

Hinsichtlich der Bewegung der Audioobjekte sollte auf den Vorschlag des ‚Anchor-Mode‘ aufgebaut werden. Sowohl die Experten mit Kompositions- als auch die Experten mit Toningenieur-Perspektive sehen in dieser Form der räumlichen Makrobildung einen Mehrwert. Hoppe sieht die Stärke des ‚Anchor-Mode‘ vor allem in der einfachen Bedienbarkeit und in dem klar wahrnehmbaren Effekt. Neben der einfachen und intuitiven Handhabung stellt dieser für Rieger eine einfache Möglichkeit dar, einen räumlichen Bezugspunkt bzw. eine Orientierung zu geben. Der Postproduktionstool-Entwickler Langhammer kann sich einen ‚Anchor‘ auch in der Postproduktion als eine Art Gruppenobjekt gut vorstellen. Für die weitere Entwicklung wünschen die Sounddesigner Rieger, Hoppe und Romanov mehr solcher makroskopischen Bewegungssteuerungen. Für weitere Organisationsprinzipien für Gruppen von Audioobjekten schlagen die Komponisten Regnier und Günther das Schwarmverhalten von Partikelsimulationen vor. Der ‚Anchor-Mode‘ könne auch Teil eines objektbasierten Audio-Effekts mit Live-Input sein, um arbiträre Signale in ihrer räumlichen Ausdehnung einfach zu kontrastieren.

Es gilt festzuhalten: Bewegungen müssen wahrnehmungstechnisch nicht zwangsläufig bewusst identifizierbar sein. Wie Romanov beschreibt, können sie dennoch als eine Form der Mikromodulation dem Klang zu Lebendigkeit verhelfen und elektronische Klänge organischer wirken lassen. Diese Mikromodulationen sind beispielsweise auch bei leichten Tonhöhenmodulationen oder Veränderungen der Wavetable Position üblich. Speziell für diese Art von Klängen – einer Art Klangwolke mit modulierender räumlicher Ausdehnung, die den Hintergrund einer Komposition darstellen können – kann das von Langhammer beschriebene integrierte Rendering und der kanal- oder szenenbasierte Output eine sinnvolle Option darstellen.

Zudem wird resümiert, dass die Modulation von Koordinaten und Achsenrotationen nur einen Ausgangspunkt darstellt. Andere Formen der Positionsmodulation oder Kombinationen der bestehenden Optionen sollten explorativ untersucht werden. An dieser Stelle bräuchte es ausführlichere Studien, um die genauen wirkungstechnischen Zusammenhänge besser zu verstehen. Dies war im Rahmen der fünf ein-minütigen Demo-Videos im Rahmen dieser Arbeit nicht zu leisten.

Bezüglich der dynamischen Panoramisierung kam es zu einem interessanten Vorschlag von Scuda. Möglicherweise kann es sinnvoll sein, eine halbautomatisierte Verknüpfung von Parametern der Panoramisierung eines virtuellen Instruments zu anderen Elementen der musikalischen Produktion zu schaffen. Es könnte die Möglichkeit eröffnet werden, Signale anderer Instrumente und Audioeffekte für die Steuerung zu verwenden. In eingeschränktem Maße wäre dies in *Ableton Live 10* über Mapping-Konstruktionen möglich. Der Vorschlag weist eine gewisse Parallele zur ‚Sidechain‘-Kompression²⁹ auf, die eine stilprägende Produktionstechnik in der populären Musik darstellt (Ableton, 2018). So wäre es denkbar, die Audioobjekte in der Ausdehnung ihrer räumlichen Verteilung zu variieren (vgl. ‚Anchor-Mode‘) und diese Ausdehnung beispielsweise in eine Abhängigkeit des Signals eines Envelope Followers auf einem separaten Track zu setzen.

Das Maß an Kontrolle in Nutzerhand stellt eine wiederkehrende Frage dar. In den Aussagen der Experten lassen sich zwei Pole erkennen, deren Vertreter alle einen Komposition & Sounddesign Hintergrund aufweisen: Regnier, Günther und Romanov sind viele und kleinteilige Optionen wichtig, um einen facettenreichen, hochwertig klingenden Sound zu designen, Hoppe, Rieger und Zorn fordern mehr Makro-Bedienelemente und Presets, mit denen sich schnell überzeugende Sounds abrufen lassen. Diese Forderungen schließen sich prinzipiell nicht aus und entsprechen im Kern dem Aufbau kommerzieller virtueller Instrumente. Meist lassen sich in einem ‚Expert-Mode‘ viele kleinteilige Einstellungen treffen, die in Makros und Presets gebündelt und in einem ‚Easy-Mode‘ einfach zugänglich gemacht werden.

Abschließend betrachtet wurden die vorgestellten Vorschläge für gute Ansätze befunden und der Prototyp positiv bewertet. Die Komponisten baten um eine Demo-Version des ‚Spatial Wavetable Synthesizers‘ für ihre eigene Arbeit. Deutlich wurde aber auch, dass an vielen Stellen weiterer Entwicklungsbedarf besteht.

²⁹ Bei der ‚Sidechain‘-Kompression eines Eingangssignals ist das Steuersignal der Kompression ein externes Signal und nicht das Eingangssignal (Ableton, 2018).

6 Schlussbetrachtung

In der folgenden Schlussbetrachtung wird zunächst in Kapitel 6.1 das methodische Vorgehen des Praxisteils kritisch beleuchtet. In Kapitel 6.2 werden die Ergebnisse dieser Arbeit zusammengefasst und die Forschungsfragen abschließend beantwortet. Sich anschließende Forschungsbereiche werden in Kapitel 6.3 ausblickend beschrieben und eine kurze kontextuale Einordnung dieser Arbeit vollzogen.

6.1 Kritische Beleuchtung

Im Folgenden wird zunächst auf Herausforderungen und Limitationen der Implementierung des Prototyps und anschließend auf die der Experten-Evaluation eingegangen.

Prototypische Implementierung

Max 8 war für die Implementierung des SWS eine sinnvolle Wahl, da die Flexibilität der graphischen Programmierumgebung für das explorative Vorgehen Vorteile hatte und auf fertige Blöcke zur Signalverarbeitung (wie beispielsweise einen Tiefpassfilter und einen 2D-Wavetable-Oszillator) zurückgegriffen werden konnte, um ein komplexes Instrument innerhalb der begrenzten Zeit aufzubauen. Allerdings hat der SWS in seiner Umsetzung als M4L-Device im Vergleich zu einem virtuellen Instrument mit beispielsweise einer VSTi-Schnittstelle den Nachteil, dass es ausschließlich in die DAW *Ableton Live 10* integrierbar ist, wobei sich ein virtuelles Instrument mit VSTi Schnittstelle in jede als VST-Host fungierende DAW integrieren lässt. Da *Ableton Live 10* ausschließlich über Stereo-Kanäle verfügt, wurde das System für das 3D-Audio Rendering auf *Reaper* erweitert. In der Praxis hat sich der Workflow mit zwei DAWs als fehleranfällig herausgestellt. Die zeitliche Synchronisierung über MTC erwies sich als instabil, der Workaround über die MacOS spezifische Funktion des *Aggregate Device* lässt sich auf einem Windows System nicht exakt reproduzieren und die Audiotreiber bedingte Latenz zwischen den DAWs sowie die auf 16 Kanäle beschränkte Kanalanzahl der virtuellen Soundkarte *Blackhole* machen diesen Workaround unständig für einen Produktions-Workflow.

Darüber hinaus ist die CPU-Auslastung der Implementierung hoch. Das hat zum einen mit der Entwicklungsumgebung *Max 8* zu tun, da die in der Handhabung komfortablen fertigen Blöcke einen gewissen ‚Overhead‘ produzieren, der sich in der beanspruchten Rechenkapazität niederschlägt. Zudem trägt der notwendige Aufbau des objektbasierten virtuellen 3D-Audio Instruments zu einer erhöhten CPU-Auslastung bei. Für jede objektifizierte Klangkomponente ist eine parallele Signalverarbeitung notwendig, da diese nicht wie in einem Stereo-Synthesizer zusammengemischt werden können.

Hinsichtlich der 2D-Wavetable Synthese wurde die Chance verpasst, ein zentrales klangliches Charakteristikum dieses Synthese-Verfahrens, das Timbre-Morphing plakativ mit der dynamischen Panoramisierung zu verknüpfen. Dies ist zu einem gewissen Grad dem verwendeten 2D-Wavetable Objekt in *Max 8* geschuldet, dessen Implementierung artefaktbehaftet ist. Im SWS ist es theoretisch möglich, beispielsweise mit Hilfe eines LFOs sowohl die Wavetable-Position als auch die Elevation der Audioobjekte zu modulieren. Dieser Effekt wurde in den Demo-Videos nicht herausgestellt, da die Modulation der Wavetable-Position mit störenden klanglichen Artefakten einhergeht. Mit einer sauberen 2D-Wavetable Implementierung ließen sich interessante Studien zur Wechselwirkung des Timbre-Morphings und einer dynamischen Panoramisierung durchführen.

In der Modulationsmatrix der Panoramisierung ist neben der Filter-Hüllkurve und zwei LFOs auch ‚Key‘ als Modulationsquellsignal verfügbar. Hier wurde auf den Wert der entsprechenden MIDI Note Number n für die Berechnung zurückgegriffen. Wahrnehmungstechnisch wäre es sinnvoller gewesen, auf die im Klangsyntheseblock modulierte Tonhöhe $n_{\text{mod}}(t)$ für die weitere Berechnung aufzubauen, da diese Tonhöhenmodulationen enthält. Somit hätte sich ein Tonhöhenvibrato direkt auf die dynamische Panoramisierung eines Audioobjekts übertragen lassen. Dies wurde erst nach Beendigung der Implementierungsphase bemerkt.

Das verwendete Tiefpassfilter *Max 8* Objekt, das ‚lowpass resonant filter‘, ist zwar einfach in der Handhabung, weist aber klanglich Defizite im Vergleich zu anderen Filter-Designs auf. Der Experte Johannes Regnier hat auf diesen Punkt hingewiesen und aus seiner praktischen Erfahrung den Einsatz des ‚state-variable filter‘ empfohlen, welcher als Objekt in *Max 8* verfügbar ist.

Die Abmessung der Benutzeroberfläche des SWS ist verglichen mit anderen M4L-Devices ausladend. Dies ist zum einen der Komplexität des Instruments, zum anderen dem frühen Entwicklungsstand und prototypischen Charakter der Implementierung geschuldet. Über Nutzertests, iterative Feedbackschleifen sowie weiteren Makrobildungen ließe sich die Benutzeroberfläche simplifizieren und auf die für die klangliche Gestaltung relevanten Bedienelemente reduzieren.

Darüber hinaus gibt es mehrere Detail-Schwachstellen, die das Vorgehen nicht weiter beeinträchtigt haben. Z.B. bleibt auf der Benutzeroberfläche die numerische Eingabe der Audioobjekt-Koordinaten aller Audioobjekte aktiv, unabhängig von der ausgewählten Anzahl an Audioobjekten.

Experten-Evaluation

Hinsichtlich der Experten-Evaluation kommt es zu methodenbedingten Schwächen qualitativer Experteninterviews. Trotz der Relevanz des Deutungswissens der Experten, kommt es auch

aufgrund der begrenzten Expertenauswahl (auch Beschränkung auf deutschsprachige Experten) zu einer gewissen Verzerrung des gezeichneten Stimmungsbildes.

Auch stellt die statische Binauralisierung der Hörbeispiele in den Demo-Videos keine ideale Wiedergabeform für die Beurteilung von 3D-Audio Inhalten dar. Da aber primär die konzeptuellen Vorschläge des SWS und Workflow-Designs evaluiert wurden und den Experten ein Abstraktionsvermögen hinsichtlich des gewählten Renderings zuzutrauen ist, wird davon ausgegangen, dass diese Einschränkung keinen maßgeblichen Einfluss auf die Aussagen der Experten hatte.

Im Rahmen der Auswertung wurde nicht der für theoriegenerierende Experteninterviews gängige Weg über ein vollständiges Transkript aller Interviews gewählt, an das sich klassischerweise eine Codierung des Materials auf Basis deduktiver Kategorien des Leitfadens, zur Reduktion auf die Kernaussagen, anschließt (Kaiser, 2014, S. 91f.). Es wurde auf Basis der Tonaufnahmen und Gedächtnisprotokolle eine um direkte Zitate ergänzte Paraphrasierung angefertigt, die Grundlage für die Auswertung war. Trotz gewissenhaftem Vorgehen und fachlicher Vertrautheit ist die unmittelbare Paraphrasierung ein interpretativer Akt und entspricht nicht dem Goldstandard der qualitativen Inhaltsanalyse. Das gewählte Vorgehen wird aber im Rahmen der Machbarkeit und dem frühen prototypischen Entwicklungsstand der Implementierung als angemessen betrachtet.

6.2 Fazit

Im folgenden Fazit steht die Beantwortung der beiden Forschungsfragen im Vordergrund. Ausgangspunkt dieser Arbeit ist die erkennbare Trennung in den Produktions-Workflows populärer Musik in 3D-Audio zwischen der beim Komponisten liegenden Klangerzeugung und der Spatialisierung, die meist einen Arbeitsschritt des Toningenieurs in der Postproduktion darstellt. Da die Gestaltung räumlicher Attribute seit jeher Teil des Kompositionsprozesses ist, ist es der Ansatz dieser Arbeit, diese verstärkt beim Komponisten anzusiedeln – speziell bei der Produktion in 3D-Audio. Die Integration von 3D-Audio in ein virtuelles Instrument kann dafür eine Basis darstellen. Daraus folgt Forschungsfrage 1:

Wie kann 3D-Audio in ein virtuelles Instrument der Wavetable Synthese integriert werden?

Zur Beantwortung wurde mit dem ‚Spatial Wavetable Synthesizer‘ ein Prototyp implementiert, welcher mehrere konkrete Umsetzungsvorschläge zur Integration von 3D-Audio in das virtuelle Instrument beinhaltet. Die räumlichen Klangattribute, welche ein 3D-Audiosystem in besonderer Weise darzustellen vermag, wurden dabei auf Richtung und Ausdehnung begrenzt.

Der SWS ist in die DAW *Ableton Live 10* integriert und somit zugänglich für Produktions-Workflows populärerer Musik. Der SWS basiert auf dem Ansatz, Klangkomponenten des Synthese-

prozesses zu objektifizieren und diese innerhalb des virtuellen Instruments zu panoramisieren. Dieser Vorschlag hat zum Ziel, eine variable Ausdehnung des Klangs des Instruments zu ermöglichen und den Klang des Instruments in dem zu gestaltenden Raum, den ein 3D-Audiosystem eröffnet, um den Hörer herum zu verteilen.

Die Objektifizierung zwei verschiedener Arten von Klangkomponenten wurde vorgeschlagen: Mit der Verteilung einzelner polyphoner Voices („Voice-Mode“) wird ein räumliches Auffächern des harmonischen Gefüges eines polyphonen Klangs erreicht. Die Verteilung unabhängiger Oszillatorgruppen („Layer-Mode“) eines komplex gelayerten Sounds ist der Vorschlag zu einer räumlichen Verteilung der innerklanglichen Struktur.

Die Modulation von Koordinaten der Audioobjekte durch Hüllkurven und LFOs sowie MIDI-Befehle stellt den Vorschlag zur Bewegung der Audioobjekte über Signale dar, die ebenso für die Modulation klangformender Parameter der Wavetable Synthese zur Verfügung stehen und somit die Klangsynthese und 3D-Audio-Domäne verzahnen. Mit dem „Anchor-Mode“ wurde eine Form der Makroformbildung für die einfache Zugänglichmachung des Kontrastierens zwischen einem gering ausgedehnten Klang und einer breiten Ausdehnung dieses Klangs vorgeschlagen.

Es ist zudem Teil der Implementierung, dass der Output des Instruments in Anlehnung an das objektbasierte Paradigma geschieht und hinsichtlich der Vielzahl an 3D-Audioformaten einen formatagnostischen Ansatz verfolgt. Das 3D-Audio Rendering ist somit aus dem virtuellen Instrument ausgelagert. In dieser Arbeit wurde ein High Order Ambisonics Workflow mit anschließender Virtualisierung für die binaurale Wiedergabe verwendet.

Die implementierten Vorschläge sind Ergebnisse eines explorativen Entwicklungsprozesses. Forschungsfrage 2:

Wie bewerten Experten aus den Bereichen Klangerzeugung und Postproduktion die Vorschläge der prototypischen Implementierung?

sieht eine Evaluation dieser Vorschläge mit Hilfe von leitfadengestützten Experteninterviews vor. Den Experten wurden die Vorschläge des SWS anhand kurzer Demo-Videos präsentiert.

Hinsichtlich des generellen Konzepts der Integration von 3D-Audio in ein virtuelles Instrument und der damit einhergehenden Verlagerung der Spatialisierung zum Komponisten, zeigten sich die Experten angetan und sehen dieses Vorgehen als richtigen und wichtigen Schritt an.

Es wurde zudem generell als sinnvoll erachtet, die Option auf die Gestaltung der Ausdehnung eines einzelnen Instruments über die Panoramisierung von Klangkomponenten zu haben. Hier

wurde zudem deutlich, dass die angebotenen Optionen stärker hinsichtlich ihrer Wirkung abgestimmt werden müssen. Das Design des virtuellen Instruments suggeriert eine gewisse Nutzung und trägt damit auch bis zu einem gewissen Grad Verantwortung für die klangästhetische Arbeit.

Die Objektifizierung polyphoner Voices zählt zu den gewinnbringenden Vorschlägen der prototypischen Implementierung. Sie sorgt für eine interessante Form der Ausdehnung, öffnet den Raum und ist ein bisher wenig bekannter Effekt. Primär von den Komponisten und Sounddesign-Experten wurde eine musikalische Kontrolle über die Verteilung der gespielten Noten auf die Audioobjekte gefordert.

Die Objektifizierung einzelner Layer wurde ebenfalls begrüßt, aber hinsichtlich ihrer Wirkung tendenziell etwas geringer eingestuft. Die einzelnen Oszillatorgruppen sollten über mehr und detaillierte Parameter zur Klanggestaltung verfügen, um dichtere Klangbilder erzeugen zu können.

Der Vorschlag der Option zur dynamischen Panoramisierung der Audioobjekte über Modulationssignale der Klangsynthese stellt eine Verzahnung dar, muss aber weitergedacht werden – die aktuelle Umsetzung ist zu stark aus dem technischen Prozess heraus motiviert und zu wenig auf das räumliche Hören abgestimmt. Gute Aspekte sind die Realisierung komplexer Bewegungen, die sonst nicht möglich oder in ihrer Erstellung in der Postproduktion aufwändig wären.

Der ‚Anchor-Mode‘ erfuhr einstimmig positiven Zuspruch und funktioniert für das Kontrastieren zwischen geringer und großer räumlicher Ausdehnung auch in seiner Wirkung gut. Oberste Maxime für jede Art der dynamischen Panoramisierung sollte dessen Wirkung sein. Weitere Makro-bildungen müssen daher aus Erkenntnissen künstlerischer Forschung und der Psychoakustik resultieren beziehungsweise mit diesen überprüft werden.

Das vorgeschlagene objektbasierte Zwischenformat für den Produktionsprozess kann sinnvoll sein, um dem Toningenieur einen Anhaltspunkt zur Intention des Komponisten hinsichtlich der räumlichen Gestaltung zu geben und gleichzeitig die Möglichkeit zu eröffnen, in der Postproduktion in das räumliche Arrangement einzugreifen. Der objektbasierte Output sollte über mehrere Formatierungs-Optionen anschlussfähig an möglichst viele 3D-Audio-Panning-Tools sein. Zudem kann die Spatialisierung im virtuellen Instrument durch den Komponisten eine Zeitersparnis im Workflow darstellen, da dieser Schritt in der Postproduktion entfällt, beziehungsweise schneller vonstatten geht.

Abschließend lässt sich zusammenfassen, dass die Integration von 3D-Audio in das virtuelle Instrument der richtige Weg zu sein scheint, die Spatialisierung zum Komponisten zu verlagern und einer Trennung im Produktionsprozess entgegenzuwirken. Die Objektifizierung und

Panoramisierung polyphoner Voices ist ein vielversprechender Ansatz. Wichtig ist eine musikalisch verknüpfte Kontrolle der Spatialisierung zu ermöglichen³⁰ sowie die Bedienelemente auf wahrnehmungstechnisch sinnvolle Optionen zu begrenzen und diese über Preset- und Makrobildung in Anlehnung an den ‚Anchor-Mode‘ leicht zugänglich zu machen.

6.3 Ausblick

Für den weiteren Entwicklungsprozess kann abgeleitet werden, dass in einem nächsten Schritt eine stabile Version des SWS nötig ist, um Komponisten und Sounddesignern das Experimentieren mit dem Instrument zu ermöglichen. Dafür sollte die Implementierung des virtuellen Instruments mit VSTi Schnittstelle erfolgen, um die Integration in diverse, als VST-Hosts fungierende und 3D-Audio fähige DAWs zu ermöglichen. Dadurch könnte auch eine Effizienzsteigerung hinsichtlich der CPU-Auslastung erreicht werden können. Für die künstlerischen Experimente wird es wichtig sein, den SWS in ein musikalisches Arrangement mit weiteren musikalischen Elementen zu bringen. Nicht alle Elemente müssen dabei ‚3D‘ sein. Die Erkenntnisse dieser künstlerischen Experimente können in einem nächsten Iterationsschritt in die Weiterentwicklung des virtuellen 3D-Audio-Instruments einfließen.

Perspektivisch ist die Anbindung an 3D-Audio-Panning-Tools kommerzieller Formate wie *Dolby Atmos Music* und *Sony 360 Reality Audio* denkbar. Im Sinne eines barrierearmen Workflows sollte die zusätzliche Option auf ein internes Rendering, beispielsweise mit dritter Ordnung Ambisonics³¹ in das virtuelle Instrument integriert werden. Dies hätte zudem den Vorteil, die 3D-Audio-Szene in der DAW als WAV-File aufnehmen zu können. Das wiederum könnte einen Ausgangspunkt für das Sampling³² von 3D-Audio-Szenen darstellen, das bisher kaum untersucht ist.

Für das Design des virtuellen Instruments hinsichtlich des Funktionsumfangs stellt sich die Frage nach Generalität oder Spezialisierung. Für eine produktähnliche Entwicklung wäre ein spezialisiertes, stark wirkungsorientiertes Instrument ein sinnvoller Schritt. Ein Beispiel für die Spezialisierung ist die Kopplung des ‚Voice-Mode‘ mit einem Step Sequencer. Über die Integration des Step Sequencers in das virtuelle Instrument kann die musikalische Kontrolle über die räumliche Verteilung der Noten gelingen. Dabei könnte jedem Step neben gängigen Werten wie Tonhöhe und Velocity auch ein Audioobjekt für das Routing zugewiesen werden. Alternativ könnte die Anzahl der Audioobjekte an die vorhandenen Steps gekoppelt sein und für jeden Step über Kugelkoordinaten dessen Panoramisierung angegeben werden. Als Klangsyntheseverfahren könnte

³⁰ Beispielsweise über zum Metrum der Musik synchrone LFOs und Kontrolle über die Verteilung musikalischer Noten auf Audioobjekte.

³¹ Die DAWs *Pro Tools*, *Nuendo* und *Reaper* unterstützen die Arbeit mit bis zu dritter Ordnung Ambisonics.

³² Dem Sampling wird eine transformierende Auswirkung auf die Produktion populärer Musik zugesprochen (Ronson, 2014).

weiterhin Wavetable Synthese zum Einsatz kommen, mit sauber implementierter 2D-Wavetable Synthese für eine Kopplung von Timbre-Morphing mit der Bewegung der Audioobjekte. Ein solches Instrument mag hinsichtlich der klanglichen Bandbreite limitiert sein. Möglicherweise ist aber diese klangliche Limitierung die Basis für eine wirkungsvolle Spatialisierung.

Zudem sollte perspektivisch die Kopplung des virtuellen Instruments an einen 3D-Hall untersucht werden, beziehungsweise die in den SWS integrierte Panoramisierung um Möglichkeiten zur Einflussnahme auf Entfernung und akustische Umgebung ergänzt werden. Die erzeugten Metadaten könnten beispielsweise von 3D-Audio-Panning-Tools mit integriertem 3D-Hall umgesetzt werden.

Es hat sich gezeigt, dass die für die Objektifizierung in Frage kommenden Klangkomponenten aus dem Klangerzeugungsverfahren abgeleitet werden. Bei der Granularsynthese liegt die Objektifizierung einzelner Grains nahe, bei der additiven Synthese die Objektifizierung der einzelnen Teiltonoszillatoren und bei der Wavetable Synthese die Objektifizierung paralleler Oszillatoren sowie polyphoner Voices. Denkt man diesen Ansatz weiter, könnten beispielsweise beim Vocoder die Klangkomponenten einzelner Filterbänke objektifiziert werden (Cook, 2003, S. 80f.). Wie mehrfach beschrieben spielt die Beschaffenheit des Signals bei einer wirkungsorientierten Spatialisierung eine große Rolle. Vor allem transiente und breitbandige Signale eignen sich für die Spatialisierung, da sie gut lokalisiert werden können. Aus dieser Perspektive könnte die Entwicklung einer 3D-Audio-Drum-Machine interessant sein.

Offen bleibt auch die Frage, ob virtuelle Instrumente mit integrierten Möglichkeiten zur Spatialisierung für 3D-Audio in der breiten Masse von Komponisten und Produzenten angenommen werden. Vermutlich ist eine Akzeptanz solcher Tools abhängig von der zukünftigen Verbreitung von 3D-Audio-Formaten in der Musikproduktion. Mit der Ankündigung von Apple, die DAW *Logic Pro* mit dem nächsten Update um ein Spatial Audio Authoring zu erweitern, könnte sich ein Produktions-Workflow mit anschließender Verwertungsmöglichkeit über *Apple Music* im *Dolby Atmos Music* Format etablieren und einer breiten Masse ermöglicht werden, Musik in 3D-Audio zu produzieren (Rogerson, 2021).

Literaturverzeichnis

- Ableton. (2018, März 27). Sidechain Compression: Part 1—Concepts and History. *Tutorials*.
<https://www.ableton.com/en/blog/sidechain-compression-part-1/>
- Ableton AG. (2020). *Max for Live Production Guidelines*. Ableton AG.
- Ableton AG. (2021). *Wavetable*. Ableton AG. <https://www.ableton.com/en/packs/wavetable/>
- Ableton & Amazing Noises. (2020). *Creative Extensions*. Ableton AG. <https://www.ableton.com/en/packs/creative-extensions/>
- Agrawal, S., Simon, A., Bech, S., Bærentsen, K., & Forchhammer, S. (2020). Defining Immersion: Literature Review and Implications for Research on Immersive Audiovisual Experiences. *J. Audio Eng. Soc.*, 68(6), 404–417.
- Arthur, S., & Nazroo, J. (2003). Designing fieldwork strategies and materials. In *Qualitative research practice. A guide for social science students and researchers* (S. 109–137). Sage.
- Arturia. (2021). *Pigments—Polychrome Software Synthesizer*. Arturia. <https://www.arturia.com/products/analog-classics/pigments/sounds>
- Bates, E. (2009). *The Composition and Performance of Spatial Music* [PhD Thesis].
- Berg, J. (2009, Mai). The Contrasting and Conflicting Definitions of Envelopment. *126th AES Convention*. AES Convention, Munich, Germany.
- Bilbao, S. (2009). *Numerical Sound Synthesis*. John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9780470749012>
- Blauert, J. (Hrsg.). (2005). *Communication acoustics*. Springer-Verlag.
- Blauert, J., & Braasch, J. (2008). Räumliches Hören. In *Handbuch der Audiotechnik*. Springer Berlin Heidelberg.
- Bogner, A., Littig, B., & Menz, W. (2014). *Interviews mit Experten*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-19416-5>
- Bohn, J. (2002). Watershed IV (review). *Computer Music Journal*, 26(3), 102–103.

- Bristow-Johnson, R. (1996). Wavetable synthesis 101, a fundamental perspective. *Audio Engineering Society Convention 101*, 1–27.
- Buff, H.-M. (2020). *Musikproduktion in 3D-Audio für Kopfhörer*. Ebner Media Group GmbH & Co. KG.
- Chowning, J. M. (1971). The Simulation of Moving Sound Sources. *Journal of the Audio Engineering Society*, 19(1), 2–6.
- Coleman, P., Franck, A., Francombe, J., Liu, Q., de Campos, T., Hughes, R. J., Menzies, D., Galvez, M. F. S., Tang, Y., Woodcock, J., Jackson, P. J. B., Melchior, F., Pike, C., Fazi, F. M., Cox, T. J., & Hilton, A. (2018). An Audio-Visual System for Object-Based Audio: From Recording to Listening. *IEEE Transactions on Multimedia*, 20(8), 1919–1931. <https://doi.org/10.1109/TMM.2018.2794780>
- Cook, P. R. (2003). *Real Sound Synthesis for Interactive Applications*. CRC Press.
- Cycling '74. (2020). How MSP Works: Max Patches and the MSP Signal Network. *Max Tutorials*. https://docs.cycling74.com/max8/tutorials/03_msphowmsspworks
- Dean, R. T. (Hrsg.). (2009). *The Oxford handbook of computer music*. Oxford University Press.
- Dickreiter, M., & Goeres-Petri, J. (2013). Schallwahrnehmung. In *Handbuch der Tonstudioteknik* (Bd. 7). De Gruyter.
- Duda, S. (2021). *Serum—Advanced Wavetable Synthesizer*. xfer Records. <https://xfer-records.com/products/serum>
- EBU. (2018). Audio Definition Model (ADM). *Tech 3364*, 2(June).
- Encircled Audio. (2020). *ENCIRCLED Spatial Audio Production for immersive sound experiences*. https://encircled-audio.com/static/iosono_contentproduction_encircled-453891057e2031bec3c29b4066428672.pdf
- Essl, K. (2019). Wandlungen der elektroakustischen Musik. *Anklaenge 2007*, 47–84. <https://doi.org/10.2307/j.ctvdjrnq6.6>
- Fontana, A., & Frey, J. H. (1998). Interviewing—The art of science. In *Collecting and interpreting qualitative materials* (S. 47–78). Thousand Oaks.

- Frank, M., & Zotter, F. (2020). Ambisonics—Das szenenbasierte 3D-Audio Produktionsformat. *vdt magazin - die Fachzeitschrift für Tonmeister*, 3, 27–31.
- Frank, M., Zotter, F., & Sontacchi, A. (2015, März). *Producing 3D Audio in Ambisonics*. AES 57th International Conference, Hollywood, USA.
- Geier, M., Ahrens, J., & Spors, S. (2010). Object-based Audio Reproduction and the Audio Scene Description Format. *Organised Sound*, 15(03), 219–227. <https://doi.org/10.1017/S1355771810000324>
- Görne, T. (2015). *Tontechnik* (4. Aufl.). Hanser Verlag.
- Görne, T. (2017). *Sounddesign* (1. Aufl.). Hanser Verlag.
- Griesinger, D. (1997). General Overview of Spatial Impression, Envelopment, Localization, and Externalization. *AES 15th International Conference*, 163–149.
- Grote, F. (o. J.). Subtrakt 5.1. *KVR Audio*. Abgerufen 21. Juni 2021, von https://www.kvraudio.com/product/subtrakt_5_1_by_florian_grote
- Günther, S. (2021, Mai 26). *Experteninterview Steffen Günther* [Audio Recording].
- Harley, M. A. (1997). An American in Space: Henry Brant's „Spatial Music“. In *American Music* (1. Aufl., Bd. 15). University of Illinois Press.
- Herre, J., Hilpert, J., Kuntz, A., & Plogsties, J. (2015). MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding. *Journal of the Audio Engineering Society*, 62(12), 821–830. <https://doi.org/10.17743/jaes.2014.0049>
- Hitzler, R. (1994). Wissen und Wesen des Experten. Ein Annäherungsversuch – zur Einleitung. In *Expertenwissen. Die institutionalisierte Kompetenz zur Konstruktion von Wirklichkeit*. Westdeutscher.
- Hoppe, A. (2021, Mai 20). *Experteninterview Andy Hoppe* [Audio Recording].
- Howard, D. M., & Angus, J. (2009). *Acoustics and psychoacoustics* (4th ed). Focal.
- ITU-R. (2018). *Advanced sound system for programme production* (Rec. ITU-R BS.2051-2). ITU.
- Kaiser, R. (2014). *Qualitative Experteninterviews*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-02479-6>

- Kaufmann, J.-C. (1999). *Das verstehende Interview – Theorie und Praxis*. UVK.
- Kermit-Canfield, E., & Abel, J. (2016). Signal decorrelation using perceptually informed allpass filters. *DAFx 2016 - Proceedings of the 19th International Conference on Digital Audio Effects*, 225–231.
- Kim-Boyle, D. (2006). *Spectral and Granular Spatialization with Boids*. University of Maryland, Department of Music.
- Koch, T. (2021, Mai 28). *Experteninterview Thomas Koch* [Audio Recording].
- Kurtz, M. (1993). *Stockhausen: A Biography*. Faber & Faber.
- Langhammer, J. (2021, Mai 18). *Experteninterview Jan Langhammer* [Audio Recording].
- Lazzarini, V. (2019). *Computer Music Instruments II* (1. Aufl.). Springer, Cham.
- Lehrman, P. D., & Tully, T. (2017). What Is MIDI? *MIDI for the professional*, 1–22. <https://doi.org/10.4324/9781315670836-1>
- Li, S., Schlieper, R., & Peissig, J. (2019). Externalization enhancement for headphone-reproduced virtual frontal and rear sound images. *Proceedings of the AES International Conference*.
- Matt. (2021, Juni 4). What is OSC? [What is OSC?]. *OpenSoundControl.org*. <http://opensoundcontrol.org/>
- Melchior, F. (2020). *Vorlesungsunterlagen 3D-Audio*. EinführungGeschichte_3DAudio, Stuttgart.
- Meuser, M., & Nagel, U. (2009). Experteninterview und der Wandel der Wissensproduktion. In *Experteninterviews. Theorien, Methoden, Anwendungsfelder* (3. Aufl., S. 35–61). VS Verlag für Sozialwissenschaften.
- Miyama, C., Schacher, J. C., & Peters, N. (2013). *SPATDIF LIBRARY – Implementing the spatial sound descriptor interchange format* (Bd. 5, Nummer 3, S. 1–5).
- MMA. (1996). MIDI musical instrument digital interface specification 1.0. In *Los Angeles*. The MIDI Manufacturers Association.
- Olivieri, F., Peters, N., & Sen, D. (2019). *A technology overview and application to Next-Generation Audio, VR and 360° Video*. 28.

- Papula, L. (2016). *Mathematik für Ingenieure und Naturwissenschaftler* (7. Aufl., Bd. 3). Springer Fachmedien Wiesbaden.
- Potard, G., & Burnett, I. (2004). *Decorrelation Techniques for the Rendering of Apparent Sound Source Width in 3D Audio Displays*. 5.
- Puckette, M. (2007). The theory and techniques of electronic music. *The Theory and Techniques of Electronic Music*, 1–331. <https://doi.org/10.1142/6277>
- Raffaseder, H. (2010). *Audiodesign* (2. Aufl.). Carl Hanser Verlag München.
- Recording Academy. (2017). *Final Nominations List 60th Annual GRAMMY® Awards*. THE NATIONAL ACADEMY OF RECORDING ARTS & SCIENCES, INC. <https://www.grammy.com/grammys/awards/60th-annual-grammy-awards-2017>
- Regnier, J. (2021, Mai 25). *Experteninterview Johannes Regnier* [Audio Recording].
- Reuter, C., Auhagen, W., & Altenburg, D. (Hrsg.). (2014). *Musikalische Akustik*. Laaber-Verl.
- Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21(4), 25–34.
- Rieger, M. (2021a, Mai). Spatial Audio on demand: Music Streaming Services at a Glance. *VRTonung*. <https://www.vrtonung.de/en/music-streaming-spatial-audio/>
- Rieger, M. (2021b, Mai 21). *Experteninterview Martin Rieger* [Audio Recording].
- Rill, G., & Schaeffer, T. (2014). *Grundlagen und Methodik der Mehrkörpersimulation* (Bd. 2). Springer Vieweg.
- Roads, C. (1996). *The computer music tutorial*. The MIT Press.
- Roads, C. (2001). *Microsound*. MIT Press.
- Rogerson, B. (2021, Juni 9). Apple confirms Logic Pro update with Spatial Audio authoring tools. *Music Radar*. <https://www.musicradar.com/news/apple-spatial-audio-logic-pro>
- Roginska, A., & Geluso, P. (2018). Immersive sound: The art and science of binaural and multi-channel audio. In *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Focal Press. <https://doi.org/10.4324/9781315707525>
- Romanov, M. (2017). *MrHeadTracker*. <https://git.iem.at/DIY/MrHeadTracker>

- Romanov, M. (2021, Mai 24). *Experteninterview Michael Romanov* [Audio Recording].
- Ronson, M. (2014, März). *How sampling transformed music*. TED2014. https://www.ted.com/talks/mark_ronson_how_sampling_transformed_music?language=en
- Rossetti, D., & Manzolli, J. (2019). *Creating Timbre in Space: Granular Synthesis and Ambisonics Spatialization Study and Composition*. 4.
- Roth, D. (2021). *Blackhole* [MacOS]. Existential Audio. <https://github.com/ExistentialAudio/BlackHole>
- Rudrich, D. (2021). *IEM Plug-In Suite (v1.12.0)* [Computer software]. IEM. <https://git.iem.at/audioplugins/IEMPluginSuite>
- Ruschkowski, A. (2019). *Elektronische Klänge und musikalische Entdeckungen*. Reclam Philipp Jun. https://books.google.de/books?id=Z_RkxQEACAAJ
- Schacher, J. C., Miyama, C., & Lossius, T. (2014). The SpatDIF library—Concepts and practical applications in audio software. *Proceedings - 40th International Computer Music Conference, ICMC 2014 and 11th Sound and Music Computing Conference, SMC 2014 - Music Technology Meets Philosophy: From Digital Echos to Virtual Ethos, April, 861–868*.
- Scherliess, V. (2020). *„Werkstatt Musikgeschichte“—Annäherungen an Musik und Musiker: Reden—Vorträge—Einführungen* (1. Aufl.). Schmidt-Römhild.
- Schmeder, A., Freed, A., & Wessel, D. (2010). *Best Practices for Open Sound Control* (S. 10). CNMAT.
- Schoerhuber, C., Zaunschirm, M., & Hoeldrich, R. (2018). Binaural Rendering of Ambisonic Signals via Magnitude Least Squares. *Fortschritte der Akustik*.
- Scuda, U. (2021, Mai 27). *Experteninterview Ulli Scuda* [Audio Recording].
- Sengpiel, E. (2008). *Frequenzverhältnisse von Tonabständen*. UdK. <http://www.sengpielaudio.com/FrequenzverhaeltnisseAntworten.pdf>
- Sidi, N. (2009). *Absynth 5—Benutzerhandbuch*. Native Instruments GmbH.
- Snyder, J. P. (1993). *Two Thousand Years of Map Projections*. The University of Chicago Press.

- Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M., & Zotter, F. (2013). Spatial Sound With Loudspeakers and Its Perception: A Review of the Current State. *Proceedings of the IEEE*, 101(9), 1920–1938. <https://doi.org/10.1109/JPROC.2013.2264784>
- Stockhausen, K. (1978). *Vier Kriterien der Elektronischen Musik*. http://www.elektropolis.de/ssb_story_stockhausen.htm
- Topper, D., Burtner, M., & Serafin, S. (2002). SPATIO-OPERATIONAL SPECTRAL (S.O.S.) SYNTHESIS. *Proc. of the 5th Int. Conference on Digital Audio Effects, DAFX-02*, 5.
- Verron, C., Aramaki, M., Kronland-Martinet, R., & Pallone, G. (2008). Spatialized additive synthesis of environmental sounds. *125th Convention of the Audio Engineering Society*, 10.
- von Coler, H. (2019). A Jack-based application for spectro-spatial additive synthesis. *Proceedings of the 17th Linux Audio Conference (LAC-19)*, 5.
- von Coler, H. (2020). *A System for Expressive Spectro-Spatial Sound Synthesis*. TU Berlin.
- Weinzierl, S. (Hrsg.). (2008). *Handbuch der Audiotechnik*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-34301-1>
- Welch, C. (2021, Juni 9). Apple Music's spatial audio is sometimes amazing but mostly inconsistent. *The Verge*. <https://www.theverge.com/2021/6/9/22525028/apple-music-spatial-audio-dolby-atmos-hands-on>
- Wendt, T. (2018, Mai 16). A Grammy for Kraftwerk's Blu-ray Album—3-D The Catalogue—Dolby Atmos Surround Sound with Pro Tools. *Avid Blogs*. <http://www.avidblogs.com/kraftwerk-3d-the-catalogue/>
- Zacharov, N., Pedersen, T., & Pike, C. (2016). A common lexicon for spatial sound quality assessment—Latest developments. *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, 1–6. <https://doi.org/10.1109/QoMEX.2016.7498967>
- Zorn, D. (2021, Mai 21). *Experteninterview Dustin Zorn [Audio Recording]*.
- Zotter, F., & Frank, M. (2019). *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (Bd. 19). Springer International Publishing. <https://doi.org/10.1007/978-3-030-17207-7>

Zvonar, R. (2005). *A History of Spatial Music*. eContact! https://econtact.ca/7_4/zvonar_spatial-music.html

Anhang

A1 Leitfaden

Leitfaden	Slide	Zeit
1 Vorstellung Interviewer und Ablauf		1'
Mitschnitt Erlaubnis Interview? Kopfhörer? Videos? Namentliche Nennung?		
Student HdM, Audiovisuelle Medien mit Schwerpunkt Ton; Masterarbeit zum Thema <i>Integration von 3D-Audio in einen virtuellen Wavetable Synthesizer</i>		
2 Soziodemographie Experte		4'
Verortung in der Matrix zum Expertensampling. Beruflicher Hintergrund? Erfahrung mit 3D-Audio?	S1	1'
Aktueller Workflow bei der Produktion von Inhalten in 3D-Audio?		3'
3 Einleitende Worte		2'
Weit verbreiteter Workflow bei 3D-Audio Studioproduktion mit elektronischen Klängen: Komponist erzeugt Mono- oder Stereoquellen mit Klangerzeuger. Toningenieur positioniert im Mixing mit einem 3D-Panner den Klang im Raum. Zur Einordnung: bezieht sich nicht auf Experimentelle Musik, Computermusik o.ä. bei der die Komponisten häufig ihre eigenen Instrumente entwickelt (OpenMusic, Pure Data, Max/MSP, ...), sondern auf „standard“ Studioproduktionen.		
Alternativer Workflowvorschlag: Die Panoramisierung des Klangs wird in den Klangsyntheseprozess und damit in den Synthesizer integriert. In die Hände des Komponisten legen.		
Dazu wurde ein prototypischer objekt-basierten Wavetable-Synthesizer entwickelt. Das virtuelle Instrument ist in einem frühen Stadium (Zwischenstand) und soll als Diskussionsgrundlage dienen. Es geht im Wesentlichen um die Evaluierung der zugrundeliegenden Konzepte. (Keine Produktevaluation)		
4 Vorstellung des Spatial Wavetable Synthesizers		6'
4.1 Signalkette auf Makroebene schematisch	S2	1'
4.2 Wesentliche Prinzipien des Spatial Wavetable Synthesizers	S3	2'
<ul style="list-style-type: none">• Teiltöne der Synthese (Layer oder Voices) werden objektifiziert (initiale Position auf Kugelfläche).		

	<ul style="list-style-type: none"> • Objekte werden über Steuerdaten der Klangsynthese auf der Kugelfläche bewegt (Änderung der Position über die Zeit). 		
4.3	Verortung auf der Bedienoberfläche	S4&5	0.5'
4.4	Bedienoberfläche Wavetable-Synthese	S6&7	1'
4.5	Bedienoberfläche Positionsberechnung	S8	1'
4.6	Bedienoberfläche Routing	S9	0.5'
5	Fragen		30'
5.1	Hauptfragen:		20'
	<ul style="list-style-type: none"> • Wie bewerten Sie allgemein die Idee der Verzahnung von Klangsynthese und Spatialisierung? (Allg. als Workflow, mehr zum Komponisten verlagern und in die Klangerzeugung zu integrieren) • Wie bewerten Sie das vorgeschlagene Konzept der Objektifizierung einzelner Signalanteile? (Zum einen das Objektifizieren, ginge auch kanalbasiert. Zum anderen das Panning in externem Plugin mit Daten aus der Synthese. Was gilt es zu beachten?) • Was ist ihr Eindruck von der Aufteilung in mehrere Layer? <ul style="list-style-type: none"> ○ Auffächerung des Klangs, mehr auf der Sounddesign-/Timbre-Ebene. ○ Kontrolle über einzelne Layer? Wenn mehr, was genau? Grad Komplexität? • Was ist ihr Eindruck von der Aufteilung in polyphone Voices? <ul style="list-style-type: none"> ○ Mehr auf einer kompositorischen Ebene. Räumliches Auffächern des harmonischen Gefüges. • Was ist ihre Meinung zur Bewegung der Objekte über Modulationssignale des Klangsyntheseprozesses ? <ul style="list-style-type: none"> ○ Was ist ihr Eindruck vom vorgeschlagenen Anchor Modus? 		
5.2	Detail Bonusfragen:		10'
	<ul style="list-style-type: none"> • Ansatz der Dekorrelation (Spread) soll ausgebaut werden. Auch als Audio FX in Kombination mit vorhandenen Klangerzeugern denkbar? Eine Art Wrapper Plug-In 		

um vorhandene virtuelle Synthesizer. Braucht es überhaupt eine eigenständige Klangerzeugung in einem solchen Tool wie dem SWS?

- Mit welchen Problemen sind sie in ihrer täglichen Arbeit mit 3D-Audio konfrontiert?
- Kognitive Komplexität der Bedienoberfläche? Intuitiv verständlich?
- Darstellung Objektpositionen (Kugelflächenprojektion, Polarkoordinaten) gut? Wenn nein, Alternativen?
- Welche Steuersignale, bzw. Modulationsquellen sind für Bewegung der Objekte unnötig? Welche fehlen?
- Layer Signalanteile lieber verschiedene Syntheseverfahren? Wenn ja, welche?

6 Abschluss

3'

Danksagung

Gibt es noch etwas, was Sie mir aus ihrer Sicht zu dem Projekt mitteilen möchten? Gerne individuell, subjektiv?

Ausblick auf weiteren Verlauf der Arbeit

Total: 46'

A2 Lautsprecherlayout ‚Produktionsstudio‘

Kanal	Bezeichnung	Azimut	Elevation
1	Left	30°	0°
2	Right	-30°	0°
3	Center	0°	0°
4	Subwoofer		
5	Left Surround	135°	0°
6	Right Surround	-135°	0°
7	Left Side Surround	90°	0°
8	Right Side Surround	-90°	0°
9	Left Top Front	45°	40°
10	Right Top Front	-45°	40°
11	Left Top Back	135°	40°
12	Right Top Back	-135°	40°
13	Voice of God	0°	90°

Tabelle 3: Lautsprecherlayout 'Produktionsstudio'.

A3 Digitaler Anhang

/00_MA_JonasKieser

Entwicklung und Evaluation eines Prototyps zur Integration von 3D-Audio in einen virtuellen Wavetable Synthesizer.pdf

/01_Spatial Wavetable Synthesizer

SpatialWavetableSynthesizer_v0.1.amxd.mxf

/02_Demo-Videos

Demo1.mov

Demo2.mov

Demo3.mov

Demo4.mov

Demo5.mov

/03_Interviews

Andy Hoppe.m4a

Dustin Zorn.m4a

Jan Langhammer.m4a

Johannes Regnier.m4a

Martin Rieger.m4a

Michael Romanov.m4a

Steffen Günther.m4a

ThomasKoch.m4a

Ulli Scuda.m4a

Präsentation SWS Experteninterviews.pdf