

Masterarbeit

Konzeption und prototypische Entwicklung eines kanalbasierten Synthesizers für eine 3D-Audio- Umgebung

vorgelegt von

Hannes Kieselbach

im Studiengang

Audiovisual Media Creation And Technology,

3D Audio And Music Technology

an der

Hochschule der Medien, Stuttgart

am 18. Juli 2025

zur Erlangung des akademischen Grades

Master Of Science

Erstprüfer:

Prof. Oliver Curdt

Zweitprüfer:

Prof. Dr. Andreas Koch

Eidesstattliche Erklärung

„Hiermit versichere ich, Hannes Kieselbach, ehrenwörtlich, dass ich die vorliegende Masterarbeit mit dem Titel: „Konzeption und prototypische Entwicklung eines kanalbasierten Synthesizers für eine 3D-Audio-Umgebung“ 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. Ebenso sind alle Stellen, die mit Hilfe eines KI-basierten Schreibwerkzeugs erstellt oder überarbeitet wurden, 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 (§ 24 Abs. 2 Bachelor-SPO, § 23 Abs. 2 Master-SPO (Vollzeit)) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.



Stuttgart, 18. Juli 2025 – Hannes Kieselbach

Zusammenfassung

Der Einsatz von Synthesizern in der kommerziellen Musikproduktion beschränkt sich hauptsächlich auf zweikanalige Instrumente. Möchten Musiker:innen diese in einem immersiven Kontext verwenden, müssen mehrere Instanzen dieser Instrumente auf unterschiedliche Positionen geroutet werden und zeitgleich über komplexe Steuersignale automatisiert werden, um ein umhüllendes Instrument zu generieren. Dieser Workflow wird als umständlich aufgefasst und bremst den kreativen Umgang mit dem Instrument. Diese Arbeit beschäftigt sich mit der Konzeption und prototypischen Entwicklung eines kanalbasierten Synthesizers in der grafischen Programmierumgebung MAX/MSP. Es wird der Prototyp einer synthetischen Klangerzeugung in einer immersiven Audioumgebung konzipiert und ein Vorschlag einer nativ verwendbaren Software präsentiert ohne zusätzliche Programmiererweiterungen oder Hosts. Zusätzlich wird auf die Konzeption einer grafischen Benutzeroberfläche eingegangen, welche abstrahierte Parameter besitzt und dem Musiker ein kreatives und übersichtliches Arbeiten ermöglicht.

Abstract

The use of synthesizers in commercial music production is mainly limited to stereo based instruments. If musicians want to use them in an immersive context, several instances of these instruments need to be routed to different positions simultaneously via complex automated control signals to generate an enveloping sound scape. This workflow is perceived as cumbersome and interrupts the creative handling of the instrument. This thesis deals with the conception and prototypical development of a channel-based synthesizer in the graphical programming environment MAX/MSP. The prototype is designed to be used natively on your MAX/MSP system without additional program extensions or hosts. In addition, the conception of a graphical user interface is discussed, which has abstracted parameters and enables the musician to work creatively and clearly.

Inhaltsverzeichnis

Abbildungsverzeichnis	VI
Tabellenverzeichnis	VIII
Formelverzeichnis	IX
1 Einleitung	1
1.1 Problemstellung	1
1.2 Zielsetzung	2
1.3 Historische Entwicklung der Klangsynthese	4
2 Grundlagen	8
2.1 Synthetische Klangerzeugung	8
2.1.1 Additive Synthese	8
2.1.2 Subtraktive Synthese	10
2.1.3 Wavetable Synthese	13
2.1.4 Steuersignale	15
2.2 Psychoakustik im dreidimensionalen Raum	18
2.2.1 Ortung	18
2.2.2 Umhüllung	22
2.2.3 Beschreibung von Hörempfinden	24
2.3 Formate für 3D-Audio	30
2.3.1 Dolby Atmos	30
2.3.2 Ambisonics	32

3	Konzeption	35
3.1	Parameter der Nutzeroberfläche	36
3.2	Steuer- und Audiosignalfluss	41
3.3	Konzeption der Benutzeroberfläche	46
4	Implementierung	48
4.1	Hierarchie und verwendete Objekte	49
4.2	Visualisierungen	52
5	Analyse	55
6	Diskussion und Ausblick	58
A	Literaturverzeichnis	63
B	Digitaler Anhang	68

Abbildungsverzeichnis

Abbildung 1: Schematische Darstellung der additiven Klangsynthese(Stange-Elbe, 2015)	9
Abbildung 2: Schematische Darstellung des Spektrums einer Triangle und Square Welle(AvaTekh Technology Tutorials & demos NDLs vs. linear filters: an illustration, no date)	11
Abbildung 3: Screenshot zwei Wavetables in der DAW Ableton Live	14
Abbildung 4: Kopfbezogenes Koordinatensystem (Weinzierl, 2008)	19
Abbildung 5: Kopfbezogene Übertragungsfunktion des Außenohrs für dem Schalleinfall von vorne, oben, hinten (Weinzierl, 2008)	19
Abbildung 6: Blauertsche Bänder (Weinzierl, 2008)	21
Abbildung 7: Korreliertes Signal (blau) und dekorreliertes Signal (gelb) im Vektorskop (Weinzierl, 2008)	23
Abbildung 8: Sound Wheel (Force Technology, 2025)	28
Abbildung 9: Encoding für Dolby Atmos Objekte und Beds ('Dolby ® Atmos TM Next-Generation Audio for Cinema Overview', 2025)	31
Abbildung 10: Decoding für Dolby Atmos Objekte und Beds ('Dolby ® Atmos TM Next-Generation Audio for Cinema Overview', 2025)	31
Abbildung 11: Die Spärlichen Harmonischen (Zotter and Frank, 2019)	33
Abbildung 12: Steuer- und Audiosignalfluss der polyphonen Klangerzeugung	42
Abbildung 13: Steuer- und Audiosignalfluss des Pannings der Voices im Format Ambisonics	43
Abbildung 14: Steuer- und Audiosignalfluss der verwendeten Dekorrelationsmethoden	44

Abbildungsverzeichnis

Abbildung 15: Steuer- und Audiosignalfluss für die klangformenden Instanzen Fatness und Clarity	45
Abbildung 16: Schematische Darstellung der Nutzeroberfläche	47
Abbildung 17: Grafische Nutzeroberfläche – Vektorskop und Modulationskurve	52
Abbildung 18: Grafische Nutzeroberfläche - Spektrum der Klangsynthese	52
Abbildung 19: Die grafische Nutzeroberfläche	53
Abbildung 20: Nutzeroberfläche des Popup-Fensters spat5.viewer	54

Tabellenverzeichnis

Tabelle 1: MIDI - Note On (Manufacturers Assoc, 2020a)	16
Tabelle 2: MIDI - Note Off (Manufacturers Assoc, 2020a)	16
Tabelle 3: Vergleich der Standards MIDI 2.0 und OSC (Wright, 2002; Manufacturers Assoc, 2020b)	17
Tabelle 4: Richtungsabhängige spektrale Charakteristika unterschiedlicher Raumrichtungen basierend auf kopfbezogener Übertragungsfunktion(Blauert and Lindemann, 1986; Butler and Humanski, 1992; Jens Blauert, 1997)	22
Tabelle 5: Dimensionsanalyse nach Zwicker(Fastl and Zwicker, 2007)	26
Tabelle 6: Das Sound Wheel	27
Tabelle 7: Multidimensionale Skalierung(Gabrielsson and Sjögren, 1979)	29
Tabelle 8: Vergleich von Dolby Atmos zu Ambisonics	34
Tabelle 9: Parameter der grafischen Nutzeroberfläche	37
Tabelle 10: Verwendete Subpatches	51

Formelverzeichnis

Formel 1: Additive Synthese	9
Formel 2: Subtraktive Synthese	12
Formel 3: Vektor eines Wavetables	13
Formel 4: Interpolation zwischen zwei Speichern	14

1 Einleitung

Im Bereich der synthetischen Klangerzeugung haben sich über die letzten Jahrzehnte vielfältige Systeme etabliert, die es Musiker:innen ermöglichen, Klangformen unabhängig von physikalischen Klangquellen zu gestalten. Die Entwicklung immersiver Audioformate hat die Art und Weise, wie Musik erzeugt, produziert und erlebt wird, nachhaltig verändert. Dazu zählen auch immersive Formate wie Ambisonics oder Dolby Atmos. Vor allem das binaurale Rendering mit Headtracking kann diese Klangwelten einer breiten Masse erlebbar machen. Diese Technologie eröffnen einen neuen Raum für die Komposition und das Sounddesign im dreidimensionalen Raum. Diese parallelen Entwicklungen – auf der einen Seite die synthetische Klanggestaltung, auf der anderen Seite die dreidimensionale Raumabbildung – finden in der heutigen Musikpraxis jedoch nur begrenzt zusammen. Die meisten verfügbaren Software-Synthesizer sind weiterhin für eine klassische Stereo-Ausgabe ausgelegt. In der Folge müssen Produzent:innen bei der Gestaltung immersiver Audioinhalte auf umständliche Workarounds zurückgreifen.

Die vorliegende Arbeit setzt an dieser Schnittstelle an: Ziel ist es, ein kreatives Werkzeug zur Verfügung zu stellen, welches eine kreative Klanggestaltung im dreidimensionalen Raum ermöglicht und gleichzeitig intuitiv bedienbar ist. Dazu wird in der grafischen Entwicklungsumgebung MAX/MSP ein Prototyp eines kanalbasierten Synthesizers entwickelt, der den künstlerischen Zugriff auf dreidimensionale Klangräume erleichtert und neue Perspektiven für Produzierende, Musiker:innen, Klangkünstler:innen und Sounddesigner:innen eröffnet.

1.1 Problemstellung

Eine historische Perspektive ([Kapitel 1.3](#)) macht deutlich, dass sich die Synthetisierung von Musik kontinuierlich weiterentwickelt, von physikalisch-manipulativen Verfahren über analoge Schaltungen bis hin zu frei programmierbaren Softwareumgebungen. Dennoch verläuft der Übergang in eine native immersive Klangwelt bis heute schleppend. Neben weitreichenden technischen Möglichkeiten in der heutigen digitalen Audioproduktion ist eine Integration synthetischer Klangerzeugung in solche Audioumgebungen nur bedingt gelöst. Kommerziell verfügbare Synthesizer-Plug-ins

geben ihren Klang meist auf zwei Kanälen aus und sind für Stereo-Anwendungen entwickelt und optimiert. Sie bieten keine native Anbindung an höherkanalige Formate. Die meisten verfügbaren Software- oder Hardware-Synthesizer sind weiterhin für eine klassische Stereo-Ausgabe ausgelegt. Folglich müssen Produzierende bei der Arbeit mit immersivem Audio auf umständliche Workflows zurückgreifen. Hierzu zählen etwa parallele Instanzen, welche durch duplizierte Steuersignale getriggert werden und durch Routing im dreidimensionalen Raum angeordnet werden. Eine Automation dieser unterschiedlichen Instanzen ist sehr umständlich und fehleranfällig. Alternativ kann ein Monosignal erzeugt werden und anschließend durch eine Signalverarbeitung dekoreliert werden. Dieser Prozess ist jedoch wenig praxistauglich, erfordert fundamentales technisches Wissen und viel Erfahrung. Das steht einem anwendungsorientierten, kreativen Workflow entgegen.

1.2 Zielsetzung

Diese Arbeit verfolgt das Ziel, ein solches Werkzeug prototypisch zu entwickeln und intuitiv und übersichtlich zu gestalten. In der grafischen Entwicklungsumgebung MAX/MSP wird ein kanalbasierter Synthesizer entworfen, der es ermöglicht, synthetische Klänge direkt für ein fest definiertes Lautsprecher-Setup zu erzeugen. Die Besonderheit liegt in der Kombination von klanglicher Vielseitigkeit, räumlicher Manipulation und benutzerfreundlicher Oberfläche. Konkret umfasst die Zielsetzung:

1. Die Konzeption eines kanalbasierten Synthesizer-Designs mit Fokus auf räumlicher Ausdehnung und dessen Modulation
2. Die Umsetzung einer intuitiven Benutzeroberfläche. Sie verbindet komplexe Vorgänge der Steuersignale über abstrahierte Parameter und Visualisierungen
3. Die Analyse der Funktionalität, Bedienbarkeit und künstlerischer Anwendbarkeit

Kapitel 2 vermittelt die theoretischen Grundlagen der synthetischen Klangerzeugung, hierzu zählen die unterschiedlichen Syntheseverfahren der additiven Synthese, der subtraktiven Synthese und der Wavetable-Synthese. Zusätzlich werden die Steuersignale OSC und MIDI vorgestellt. Anschließend widmet sich dieses Kapitel der

der Psychoakustik im Raum, hierzu zählen die Aspekte der Ortung und Umhüllung. Es werden unterschiedliche Konzepte der Beschreibung von Klangempfinden vorgestellt. Am Ende dieses Kapitels werden die Formate Ambisonics und Dolby Atmos Music technisch vorgestellt.

Kapitel 3 befasst sich mit der Konzeption des Prototyps. Zunächst werden unterschiedliche Ideen zusammenfassend vorgestellt. Anschließend werden die finalen Ideen hierzu ausführlicher beschrieben. Hierzu zählt das Design des Steuersignal- und Audiosignalfusses. Zusätzlich wird ein Vorschlag einer Benutzeroberfläche vorgestellt, welche neben der einfachen Bedienbarkeit und Übersichtlichkeit auch die grafische Darstellung der Modulation ermöglichen soll. Dies fördert zusätzlich das Verständnis und den intuitiven Umgang mit dem Instrument.

Kapitel 4 beschreibt die konkrete Implementierung des Systems in MAX/MSP. Nach einer grundsätzlichen Vorstellung der Entwicklungsumgebung und dessen speziellen Anwendungsbereichen werden die implementierten Methoden, Objekte und Implementierungsebenen vorgestellt. Anhand einzelner Signalflüsse wird die Arbeitsweise erläutert.

Kapitel 5 analysiert die Funktionalität und Bedienbarkeit. **Kapitel 6** schließt mit einer Diskussion der vorangegangene Analyse sowie einem Ausblick auf zukünftige Schritte und Einsatzmöglichkeiten ab .

1.3 Historische Entwicklung der Klangsynthese

Die Geschichte der synthetischen Klangerzeugung reicht bis ins frühe 20. Jahrhundert zurück und ist eng mit der Entwicklung der elektroakustischen Musik verbunden. Fünf zentrale Meilensteine skizzieren folgend den Weg erster Experimente hin zur heutigen Vielfalt an Synthesizern.

Musique Concrète

Die *Musique Concrète* entsteht ab 1948 unter der Leitung des französischen Komponisten und Rundfunkingenieurs Pierre Schaeffer. Sie bildet eine der frühesten Formen elektroakustischer Musik und steht im Gegensatz zur traditionellen Komposition mit klassischen Instrumenten. Anstelle notierter Partituren und Instrumentalklänge verwendet Schaeffer aufgezeichnete Geräusche aus dem Alltag. Dazu zählen Maschinenklänge, Naturlaute oder menschliche Aktivitäten. Diese nutzt Schaeffer als Ausgangsmaterial für seine musikalische Arbeit. Er bearbeitet diese mithilfe von Schnitt, Wiederholung, Tonhöhenveränderung und Rückwärtsspiel. Pierre Schaeffer verfolgt mit seinem Ansatz das Ziel, den Klang aus seinem ursprünglichen Kontext zu lösen und ihn als *akustisches Objekt* zu betrachten. Die Klangquelle verliert dabei ihren bisherigen Konzert: entscheidend ist allein ihre klangliche Qualität. Damit etabliert Schaeffer ein neuartiges kompositorisches Denken, das sich auf die Bearbeitung klanglicher Rohmaterialien konzentriert, unabhängig von der bisherigen Herkunft. Die *Musique Concrète* markiert einen fundamentalen Wendepunkt in der Musikgeschichte: Zum ersten Mal ist es möglich, mit konkreten Klangereignissen zu komponieren, ohne auf traditionelle Notation oder Instrumentation angewiesen zu sein. Das Tonband entwickelt sich nun zu einem eigenständigen Instrument, das neue ästhetische Ausdrucksmöglichkeiten ermöglicht. Werke wie *Étude aux chemins de fer* (1948) inspirieren zahlreiche Komponist:innen und beeinflussen sowohl die elektroakustische als auch die elektronische Musik nachhaltig. Klang wird als formbares Material etabliert. (Taylor, 2001)

Karlheinz Stockhausen

In den 1950er Jahren etabliert sich in Deutschland eine technisch orientierte Form elektronischer Musik, die sich von der französischen *Musique Concrète* deutlich unterscheidet. Im Studio für elektronische Musik des WDR in Köln entwickelt

Stockhausen einen kompositorischen Zugang, der auf der systematischen Erzeugung und Bearbeitung synthetischer Klänge basiert. Anstatt aufgezeichnete Geräusche zu manipulieren, arbeitet Stockhausen mit rein elektronisch generierten Tönen, insbesondere Sinustönen, Impulsen und Rauschen. Er nutzt technische Geräte wie Generatoren, Filter und Tonbandmaschinen zur klanglichen Gestaltung. Stockhausen begreift die technologischen Mittel nicht lediglich als Werkzeuge zur Klangproduktion, sondern als grundlegenden Bestandteil seines kreativen Schaffensprozesses. In Werken wie *Studie II* (1954) komponiert er ausschließlich mit elektronisch erzeugten Sinustönen, die er in einer seriellen Struktur anordnet. Diese Methodik der additiven Klangsynthese beruht auf der Theorie der Fourier-Reihen und wird in [Kapitel 2.1.1](#) genauer vorgestellt. Dadurch entstehen musikalische Abläufe, die keinem traditionellen Formschema folgen, sondern aus einem kontrollierten Zusammenspiel von Tonhöhen, Dauern, Amplituden und räumlicher Anordnung hervorgehen. Diese Arbeitsweise dokumentiert den Übergang von der Bearbeitung gefundener Klänge zur gezielten Erzeugung klanglicher Elemente. Stockhausens Kompositionen verbinden technische Innovation und ästhetisches Denken, um neue Klangräume zu designen. Seine Forschungen am WDR-Studio beeinflussen nicht nur die Entwicklung der elektronischen Musik, sondern bereiten auch den theoretischen Boden für spätere Formen digitaler Klangsynthese. Stockhausen gilt als zentrale Figur bei der Etablierung einer neuen Klangästhetik, in der Technologie nicht als nachträgliches Mittel, sondern als eigentliches Instrument fungiert. Stockhausen legt somit einen wesentlichen Grundstein für die weitere Entwicklung elektronischer und experimenteller Musik. (Kurtz, 1988; Frisius, 1996)

Entwicklung des Moog-Synthesizers

In den 1960ern entwickelt der amerikanische Ingenieur Robert Moog den ersten kommerziell erfolgreichen modularen Synthesizer. Im Gegensatz zu den bisher technisch umständlichen Systemen stellt der Moog-Synthesizer ein modulares, benutzerfreundliches System dar, das speziell Musiker:innen anspricht. Das Instrument integriert elektronische Bauteile wie spannungsgesteuerte Oszillatoren, Filter, Verstärker sowie Hüllkurvengeneratoren in einem System, das mittels Patchkabel flexibel Steuersignale und Audiosignale verschalten kann. Diese Modularität ermöglicht Musiker:innen einen spielerischen Umgang mit elektronisch

erzeugtem Klang. Die Alltagstauglichkeit des Instruments trägt wesentlich zur Popularisierung elektronischer Musik bei. Elektronische Klanggestaltung wird nicht mehr nur technisch, sondern auch kompositorisch zugänglich. Durch eine intuitive Spielbarkeit erreicht das Instrument schnell Verbreitung in der populären Musik und beeinflusst Genres wie Jazz, House oder Techno. Dazu zählen Künstler wie Jean Michel Jarre oder Kraftwerk. Das von Moog eingeführte Prinzip der modularen Architektur bildet bis heute die Grundlage zahlreicher analoger und digitaler Syntheseumgebungen. (*Moog Synthesizer 1c/2c/3c | Vintage Synth Explorer, 2025; Deutsches Museum: Moog, 2025*)

Digitale Wavetable- und FM-Synthese

In den 1980er-Jahren setzt sich die digitale Klangsynthese zunehmend durch und erweitert das klangliche Spektrum elektronischer Instrumente. Einen großen Einfluss hat der Yamaha DX7, der 1983 auf den Markt kommt und die Frequenzmodulation (FM) als neuartige Synthesemethode in der Musikproduktion etabliert. Diese kommerziell genutzte Technologie basiert auf der Forschung von John Chowning. Er entdeckt, dass sich durch die Modulation der Frequenz eines Trägersignals mit einer zweiten Frequenz komplexe Spektren erzeugen lassen. Der Yamaha DX7 integriert diese Methode in einem kompakten, digital gesteuerten Instrument und macht sie erstmals für ein breites Publikum von Musiker:innen und Produzent:innen zugänglich. Parallel dazu entwickelt Wolfgang Palm das Konzept der Wavetable-Synthese. Diese basiert auf der sequenziellen Abtastung vorab gespeicherter Wellenformen, die im Gegensatz zu traditionellen Oszillatorformen eine größere Vielfalt an spektralen Eigenschaften aufweisen. Durch das gezielte Überblenden der Wavetables entstehen dynamische Klangverläufe. In [Kapitel 2.1.3](#) wird die Funktionsweise dieser Klangsynthese genauer erläutert. Beide Technologien tragen maßgeblich zur klanglichen Ästhetik der 1980er-Jahre bei. Sie ermöglichen dynamisch modulierte Klänge, die insbesondere in der Popmusik und im Sounddesign prägend werden. Darüber hinaus markieren sie den Übergang von analogen zu digital gesteuerten Synthesemethoden und bilden die Grundlage für zahlreiche Entwicklungen in der digitalen Klangsynthese, die bis heute relevant bleiben. (Bristow-Johnson, 1999)

Software-basierte Synthese

Die vom IRCAM in Paris in den 1980ern entwickelte Softwareplattformen *MAX/MSP* (*What is Max? | Cycling '74, 2025*) ermöglicht es, eigene Klangsynthese und Signalverarbeitung in Bild und Ton frei zu entwerfen. Die grafische Programmierumgebung erlaubt eine modulare Strukturierung von Signalflüssen und vereint grafisch Programmierung mit Echtzeit-Klangbearbeitung. Durch die Kombination von Audio-, Steuer- und Interfacedaten innerhalb einer gemeinsamen Umgebung ist die Klangsynthese frei programmierbar.

Im Zentrum von *MAX/MSP* steht die Idee einer offenen, kreativen Entwicklungsumgebung. Musiker:innen können die Umsetzung individueller klanglicher Vorstellungen unabhängig von auf dem Markt verfügbarer Software selbst realisieren. Besonders hervorzuheben ist das Potenzial solcher Plattformen im Kontext der räumlichen Klanggestaltung. Durch die Möglichkeit, mehrkanalige Signalpfade und Steuerdaten zu integrieren, lässt sich dreidimensionaler Klang generieren und modellieren. Damit eröffnet *MAX/MSP* nicht nur neue Wege im Bereich der Musikproduktion, sondern auch in Feldern wie der Klangkunst oder der interaktiven Medieninstallation. *MAX/MSP* bietet darüber hinaus viele frei verfügbare Programmiererweiterungen über eine offene Plattform an. Diese umfassen neben der Klangerzeugung und klassischer Signalverarbeitung auch Erweiterungen speziell für die Spezialisierung, beispielsweise die Software *spat5*. Die Verfügbarkeit solcher Werkzeuge demokratisiert den Markt durch den Austausch auf wissenschaftlicher Ebene sowie auf der semi professionellen Anwenderebene durch eine weltweit durch das Internet vernetzte Szene. (Cipriani, Giri and Dudas Richard, 2020)

2 Grundlagen

Um den Entwurf und die prototypische Entwicklung eines mehrkanaligen Synthesizers für eine immersive Klangumgebung zu konzipieren, ist es notwendig, die relevanten theoretischen und technischen Grundlagen zu erläutern. Dieses Kapitel liefert eine Einführung in die verschiedenen Bereiche der synthetischen Klangerzeugung, der psychoakustischen Wahrnehmung im Raum sowie der derzeit etablierten 3D-Audioformate Dolby Atmos Music und Ambisonics.

2.1 Synthetische Klangerzeugung

Die synthetische Klangerzeugung bildet das technische und gestalterische Fundament elektronischer Musikproduktion. Durch sie lassen sich Klänge unabhängig von akustischen Instrumenten erzeugen, modellieren und manipulieren. Dabei kommen verschiedene Syntheseverfahren zum Einsatz, die sich hinsichtlich Klangcharakteristik, Flexibilität und technischer Umsetzung deutlich unterscheiden. Im Folgenden werden die Verfahren der additiven, subtraktiven und Wavetable-Synthese beschrieben.

2.1.1 Additive Synthese

Die additive Klangsynthese zählt zu den ältesten und grundlegendsten Verfahren der elektronischen Klanggestaltung. Sie basiert auf der Annahme, dass sich jeder periodische Klang durch die Überlagerung einer Vielzahl sinusförmiger Teilschwingungen rekonstruieren lässt. Dieses Prinzip beruht auf der mathematischen Theorie der Fourier-Analyse. Fourier zeigt, dass sich beliebige periodische Funktionen, somit auch periodische akustische Signale, als Summe harmonischer Vielfache von Sinustönen unterschiedlicher Frequenz, Amplitude und Phase darstellen lassen. In der Klangsynthese werden diese Teilschwingungen als Partialtöne bezeichnet. (Strawn *et al.*, 1996) Die mathematische Beschreibung eines periodischen Signals $x(t)$ durch additive Synthese lässt sich als Summe dieser Partialschwingungen beschreiben.

$$x(t) = \sum_{n=1}^N A_n * \sin(2\pi * f_n * t + \phi_n)$$

Formel 1: Additive Synthese

- A_n Amplitude der n-ten Teilschwingung,
 f_n Frequenz [Hz] der n-ten Teilschwingung,
 ϕ_n für die Phasenverschiebung ϕ der n-ten Teilschwingung
 N für die Anzahl der eingesetzten Teiltöne,
 t Zeit [ms]
-

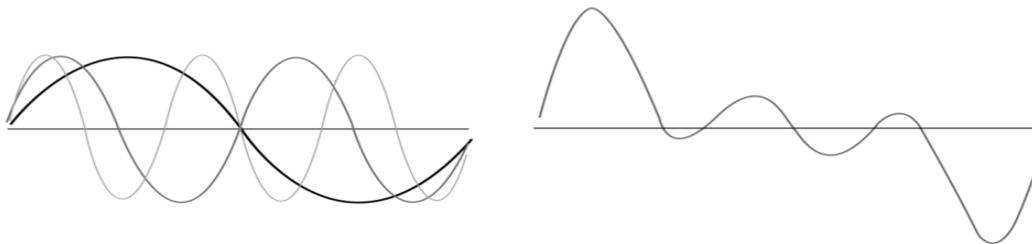


Abbildung 1: Schematische Darstellung der additiven Klangsynthese(Stange-Elbe, 2015)

Jeder Klang wird als Summe diskreter Sinusschwingungen erzeugt, deren Parameter individuell gesteuert werden können. Je höher die Anzahl N der Teilschwingungen, desto vielschichtiger lässt sich das resultierende Klangspektrum gestalten. (Stockhausen and Barkin, 1962) Insbesondere für spektral langsam modulierte oder glockenähnliche Klänge eignet sich dieses Verfahren besonders, da hier die Feinabstimmung einzelner Komponenten besonders deutlich wird. In der praktischen Anwendung erfolgt die additive Synthese durch die simultane Erzeugung einer Vielzahl von Oszillatoren. Jeder Oszillator generiert eine Sinusschwingung mit einer spezifischen Frequenz, Amplitude und Phasenlage. Durch die Addition dieser Einzelsignale entsteht das komplexe Klangereignis. Daran anknüpfend gibt folgende Möglichkeiten, die erzeugten Signale zu modulieren. (Strawn *et al.*, 1996)

- Attack, Duration, Sustain, Release für jede Teilschwingung,
 - Frequenzmodulation durch LFOs einzelner Teilschwingungen
-

Die klanglichen Möglichkeiten der additiven Synthese zeichnen sich durch ihre spektral feine Kontrolle über den Zeitverlauf des Klangs aus. Anders als bei subtraktiven Verfahren, bei denen ein obertonreiches Signal durch Filter reduziert wird, wird hier der Klang auf Basis der Fourier-Theorie von Grund auf geformt. (Stockhausen and Barkin, 1962) Ein wesentlicher Nachteil der additiven Synthese ist der hohe Ressourcenverbrauch. Da jeder einzelne Partialton separat berechnet und gesteuert werden muss, steigt die Komplexität exponentiell mit der Anzahl der gewünschten Spektralkomponenten. Für realistisch klingende Nachbildungen werden in der Praxis hunderte Oszillatoren benötigt. Bei dynamischer Steuerung durch zeitabhängige Amplitudenmodulation oder spektrale Transformationen erhöht sich die Rechenleistung zusätzlich. (Puckette, 2007)

2.1.2 Subtraktive Synthese

Die subtraktive Synthese zählt zu den bis heute am weitesten verbreiteten Klangsyntheseverfahren. Insbesondere in der analogen Klangästhetik gilt sie als prägendes Verfahren und formt Genres wie Synthpop, Techno oder Ambient. Aber auch für Filmmusik und Sounddesign werden diese Syntheseverfahren genutzt. (Strawn *et al.*, 1996) Der Klang subtraktiver Synthesizer wird oft als warm, organisch und musikalisch anschlussfähig beschrieben. Im Unterschied zur additiven Synthese, bei der ein Klang aus einzelnen Teilschwingungen konstruiert wird, beginnt die subtraktive Synthese mit einem komplexen, obertonreichen Klangmaterial. Im Anschluss wird dieses Spektrum gezielt durch Filterung reduziert. Der kreativ modulare Teil besteht also nicht in der Konstruktion, sondern in der Formung und Reduktion des akustischen Rohmaterials. Die Grundstruktur subtraktiver Synthesizer folgt einem seriellen Signalfluss, der typischerweise aus den folgenden Komponenten besteht.(Puckette, 2007)

Ein Oszillator erzeugt periodische Wellenformen mit reichlichem Obertonspektrum. Beispielsweise ein Sägezahn- oder Viereck-Oszillator. Ein Sägezahnsignal enthält alle ganzzahligen Harmonischen mit abnehmender Amplitude, besitzt daher ein besonders helles Klangbild. Das Rechtecksignal enthält nur ungerade Harmonische, ebenfalls mit abnehmender Amplitude, ergibt einen hohlen Klangcharakter. Das Dreiecksignal

enthält ebenfalls nur ungerade Harmonische, jedoch mit schnellerem Abfall der Amplitude, dadurch wird ein weicherer Klang generiert. (Uncini, 2022)

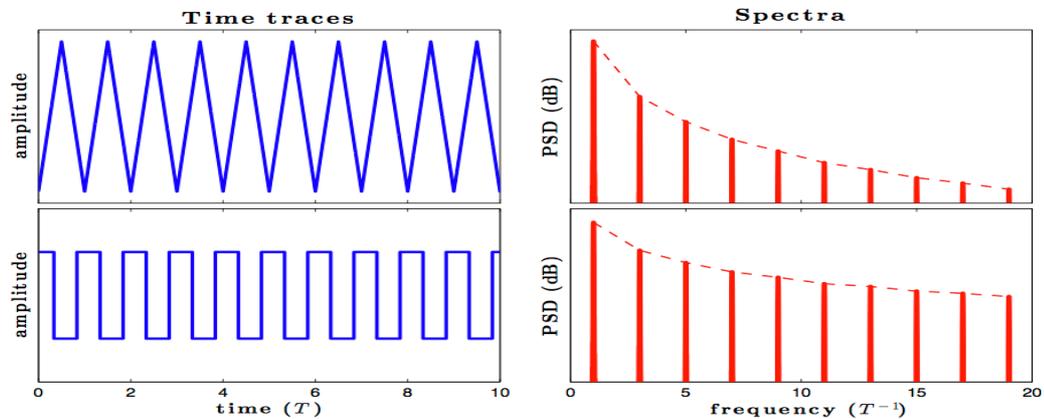


Abbildung 2: Schematische Darstellung des Spektrums einer Triangle und Square Welle(AvaTekh | Technology | Tutorials & demos | NDLs vs. linear filters: an illustration, no date)

Rauschgeneratoren erzeugen spektral breitbandiges Material– ideal für perkussive Klänge, Windgeräusche oder modulierende Layer. Die Auswahl der Oszillatorwellenform beeinflusst maßgeblich die Ausgangsbasis für die weitere klangliche Gestaltung. Die klangliche Vielfalt der subtraktiven Synthese entsteht maßgeblich durch die Auswahl und Kombination verschiedener Filtertypen. (Strawn *et al.*, 1996)

Die klanggestaltende Kernkomponente in der subtraktiven Synthese ist das Filter. Ein Filter formt das Spektrum durch selektive Abschwächung oder Betonung bestimmter Frequenzbereiche. Hierzu zählen Hoch-, Tiefpass-, Bandpass- oder Notch-Filter. Das Filter verleiht dem Klang seine charakteristische Klangfarbe. Das Tiefpassfilter lässt Frequenzen unterhalb einer Grenzfrequenz f_{cut} passieren und dämpft höherfrequente Anteile. Das Hochpassfilter lässt Frequenzen oberhalb einer Grenzfrequenz f_{cut} passieren und dämpft niederfrequente Anteile. Das Bandpassfilter lässt einen definierten Frequenzbereich mit Mittenfrequenz f_{mid} passieren und dämpft höhere und tiefere Frequenzen ab. Attack, Duration, Sustain und Release bestimmen das Ein- und Ausschwingverhalten des Klangs. (Strawn *et al.*, 1996)

Ein zentrales Merkmal subtraktiver Synthesizer ist die Möglichkeit, klangformende Parameter über die Zeit oder durch Sidechainsignale zu steuern. Durch die

Kombination mehrerer Modulationsquellen entsteht eine anpassbare Ausdrucksvielfalt. Gerade durch komplexe Envelopes oder LFO-Routings lassen sich organische, zeitlich variierende Klangverläufe erzeugen. Filter haben einen Sidechain-Eingang, an den das LFO-Signal geroutet werden kann. Dieses Signal moduliert bestimmte Parameter des Filters. Hierzu zählen beispielsweise Frequenz, Güte oder Gain. Das Prinzip der subtraktiven Synthese mit dem Ausgangssignal $x(t)$ lässt sich mathematisch durch die Faltung eines Signals mit der Impulsantwort eines Filters beschreiben. (Puckette, 2007)

$$x(t) = y(t) * adsr(t) * hpf(t) * lfo(t)$$

Formel 2: Subtraktive Synthese

$y(t)$	Obertonreicher Oszillator, z.B. Rechteck
$adsr(t)$	Hüllkurvenfilter, welcher den Einschwingvorgang bestimmt
$lfo(t)$	Niedrigfrequenter Sinus moduliert die Amplitude des Signals
$hpf(t)$	Filter, welcher ab einer Trenn- oder Mittenfrequenz Frequenzen absenkt

Die subtraktive Synthese eignet sich besonders gut zur Erzeugung klassischer synthetischer Klangtypen. Dazu zählen Bässe mit resonantem Tiefpassfilter, Leadsounds mit aggressiven Rechteckwellen, Pads mit sanften Filterverläufen und langen Hüllkurven oder perkussive Elementen, z. B. Hi-Hats oder Snare-Klänge auf Rauschbasis. (Friesecke, 2014) Moderne softwarebasierte Systeme wie MAX/MSP (*What is Max? | Cycling '74*, 2025), Pure Data (*documentation — Pd Community Site*, 2025) oder auch kommerzielle Plugin-Lösungen wie Reaktor (*Reaktor 6 – digital modular synthesizer | Komplete*, 2025) ermöglichen die komplexe Implementation subtraktiver Synthesestrukturen. Besonders hervorzuheben ist bei MAX/MSP die flexible Konfiguration des Signalflusses, etwa durch die Kombination mehrerer Oszillatoren, paralleler oder serieller Filterstrukturen, automatisierter Parameterverläufe und vor allem mehrkanaliger Ausgaben und komplexem Routing auf den DA-Wandler des Systems.

2.1.3 Wavetable Synthese

Die Wavetable-Synthese stellt ein weiteres Verfahren der Klangsynthese dar, das sich durch die Nutzung vorab gespeicherter, periodischer Wellenformen von den anderen Verfahren abgrenzt. Diese Wavetables sind als diskrete Samplings einzelner Schwingungsverläufe. Im Gegensatz zu additiven oder subtraktiven Verfahren, bei denen Klangspektren in Echtzeit berechnet werden, basiert die Wavetable-Synthese auf der gezielten Modifikation und zeitlichen Überblendung gespeicherter diskreter Audiosignale. Das Verfahren wurde ursprünglich in den 1970er-Jahren von Wolfgang Palm entwickelt und stellt heute eine zentrale Grundlage für zahlreiche moderne Software- und Hardware-Synthesizer dar. Ein Wavetable besteht typischerweise aus einer Anzahl von sogenannten Frames oder Wellenformen. Jeder Frame enthält eine vollständige Periode eines periodischen Signals und repräsentiert damit eine bestimmte spektrale Charakteristik. Diese Wellenformen sind im Speicher des Instruments abgelegt und werden durch einen Trigger ausgelesen. Sie können nacheinander oder parallel überblendend ausgelesen werden. Zur Steuerung der Tonhöhe wird die Leseposition innerhalb der Wellenform mit variabler Geschwindigkeit durchlaufen. Ein Phasor-Oszillator, typischerweise ein Dreiecksignal, zeigt als Vektor auf diskrete Werte dieser Liste, welcher das Signal zyklisch ausliest und ihn als eine fortlaufende Schwingung interpretiert. Mathematisch lässt sich ein einzelner Frame eines Wavetables $w[n]$, mit N Samples pro Periode als Vektor diskreter Werte darstellen. (Uncini, 2022)

$$w[n] = \{x_0, x_1, x_2, \dots, x_{N-1}\} \quad 0 \leq n \leq N$$

Formel 3: Vektor eines Wavetables

$w[n]$	Abtastwert an der Stelle n
N	Anzahl der Samples pro Periode

Eine der wesentlichen Erweiterungen klassischer Wavetable-Konzepte besteht in der Möglichkeit, kontinuierlich zwischen verschiedenen dieser Speicher zu überblenden. Dies wird durch eine gewichtete Interpolation zwischen zwei benachbarten Wellenformen realisiert.

$$s[n] = (1 - \alpha) * w_k[n] + \alpha * w_{k+1}[n] \text{ mit } \alpha \in [0,1]$$

Formel 4: Interpolation zwischen zwei Speichern

$w_k[n]$	Aktuelle Wellenform im Speicher
$w_{k+1}[n]$	Nachfolgende Wellenform im Speicher
α	Faktor

Der Faktor α empfängt ein Steuersignal, welches in der Praxis auch gerne durch ein Low Frequency Oszillator moduliert werden kann. Diese kontinuierliche Interpolation erzeugt sich überblendende spektrale Veränderungen, die charakteristisch für Wavetable-Klänge sind. Auf diese Weise lassen sich spektral animierte Klänge gestalten. Diese beruhen nicht direkt auf klassischen Filterprozessen, sondern zeichnen sich vielmehr durch die Änderung der spektralen Zusammensetzung der zugrundeliegenden Wellenform aus. Anders als bei der subtraktiven Synthese wird das Klangmaterial nicht durch externe Filter geformt, sondern durch die Wahl und Überlagerung intern gespeicherter Spektren. Dadurch lässt sich die spektrale Charakteristik eines Klangs deutlich präziser definieren und steuern. Da die Wavetable-Synthese auf die Wiederverwendung gespeicherter Wellenformen setzt, ist sie rechentechnisch effizienter als die additive Synthese, bei der jede Teilkomponente des Spektrums einzeln berechnet werden muss. (Verron *et al.*, 2009; Uncini, 2022)

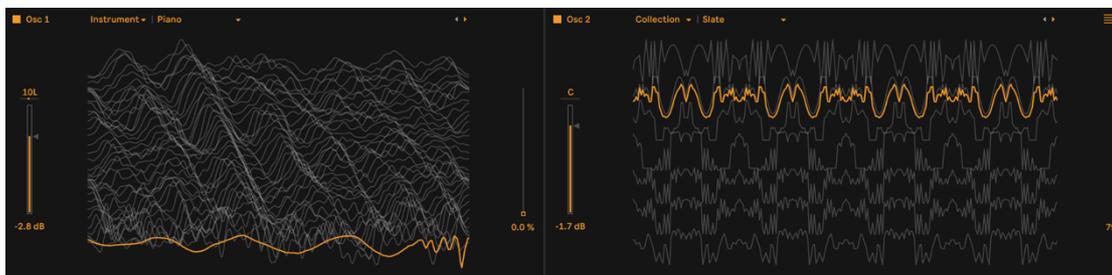


Abbildung 3: Screenshot zwei Wavetables in der DAW Ableton Live

2.1.4 Steuersignale

Die Steuerung klanglicher Parameter stellt einen essenziellen Bestandteil jeder Syntheseumgebung dar. Im Kontext elektronischer Musikproduktion ermöglichen Steuersignale die gezielte Kontrolle über Oszillatoren, Filter, Modulationseinheiten und Effekte. Diese Signale werden unabhängig vom Audiosignalfad übertragen und können statisch sein oder die klangliche Entwicklung über die Zeit strukturieren. (Bartetzki, 2022) Viele Programmierumgebungen wie MAX/MSP erlauben eine differenzierte Trennung von Audio- und Steuersignalen und eröffnen flexible programmierbare Möglichkeiten der Echtzeitsteuerung. Besondere Relevanz erhalten Steuerparameter im Zusammenhang mit dieser Arbeit bei der räumlichen Klanggestaltung. In einem spazialen Kontext lassen sich Steuerparameter nicht nur auf Tonhöhe, Lautstärke, und Anschlagsdynamik, sondern auch auf Raumparameter wie Ausdehnung oder Panning erweitern. (Verron, Aramaki and Kronland-Martinet, 2010) Zwei etablierte Standards – MIDI (Musical Instrument Digital Interface) und OSC (Open Sound Control) werden im folgenden Abschnitt vorgestellt.

Musical Instrument Digital Interface (MIDI)

Seit seiner Standardisierung im Jahr 1983 fungiert MIDI als digitales Protokoll zur Kommunikation zwischen elektronischen Musikinstrumenten, Computern und Peripheriegeräten. Anders als Audiodaten überträgt MIDI keine Audiostreams, sondern Steuerbefehle in Form von Datenpaketen. Diese beinhalten Angaben zu Tonhöhe, Anschlagstärke, Dauer, Programmwahl, Tonhöhenveränderung und weiteren Steuersignalkanäle. Dazu zählen Parameter wie Modulationsrad oder Filtersteuerung. Ein typisches MIDI-Datenpaket besteht aus einem Statusbyte und zwei Datenbytes. Das Statusbyte kodiert die Art des Befehls - beispielsweise *Note On* - sowie den MIDI-Kanal. Die Datenbytes enthalten numerische Parameter wie *Tonhöhenwert* und *Velocity*. Die Auflösung des ursprünglichen MIDI-Protokolls beträgt 7 Bit pro Wert. Im Jahr 2020 wurde das überarbeitete Protokoll MIDI 2.0 eingeführt und erweitert diese Auflösung auf 32 Bit. Die beiden folgenden Schaubilder zeigen die schematische Verteilung der Bits pro MIDI-2.0-Nachricht. (Manufacturers Assoc, 2020a)

mt=4	group	1 0 0 1	channel	r	note number	attribute type
velocity				attribute data		

Tabelle 1: MIDI - Note On (Manufacturers Assoc, 2020a)

mt=4	group	1 0 0 0	channel	r	note number	attribute type
velocity				attribute data		

Tabelle 2: MIDI - Note Off (Manufacturers Assoc, 2020a)

In der Klangsyntaxe steuert das MIDI-Protokoll typischerweise die Tonhöhe von Oszillatoren, die Öffnung von Filtern, die Modulationsintensität von Low Frequency Oszillators oder Hüllkurvengeneratoren. MIDI-Controllerdaten können einzelnen Klangparametern zugewiesen werden um Parameter wie *Tremolo*, *Vibrato* oder zu modulieren.

OSC – Open Sound Control

Open Sound Control stellt einen weiteren Standard zur Übertragung von Steuerdaten über IP-basierte Netzwerke dar und wird auch intern in der grafischen Programmierumgebung MAX/MSP verwendet. Im Gegensatz zu MIDI basiert OSC auf einer hierarchischen Adressstruktur, vergleichbar mit einem skriptbasierten Protokoll. Diese Struktur ermöglicht es, eine nahezu unbegrenzte Anzahl an Parametern, Adresspfaden und Datenaufösungen in einer Baumstruktur zu verschachteln. Entwickelt wurde OSC am Center for New Music and Audio Technologies der University of California, Berkeley. Ziel der Entwickler ist, eine flexible Alternative zu MIDI zu schaffen. Ein OSC-fähiges Gerät nutzt die Infrastruktur des Netzwerkprotokolls UDP. Hierzu wird jedes Gerät im Netzwerk mit einer IP-Adresse angesprochen. Folgend besteht jede OSC-Nachricht aus einem Adresspfad, beispielsweise `/synth/filter/cutoff/` und einem Datenpaket bestehend aus einem oder mehreren Werten. Insbesondere in komplexen Klangsyntaxestrukturen kann eine strukturierte Zuordnung von Steuerzielen für eine bessere Übersicht von Vorteil sein. Das Open Sound Protocol verwendet UDP oder TCP/IP-Protokolle aus der Netzwerktechnik zur Kommunikation und ist somit ideal für verteilte Systeme und netzwerkbasierte

Performances geeignet. Folgend sind zwei Beispiele für eine OSC-Nachricht dargestellt. (Wright, 2002)

```
192.168.4.20 maxpatch/textFeld/go! s "go_pressed"
```

```
192.198.22.1 synthMadeByHannes/filter/cutoff f 110
```

Ein signifikanter Vorteil von OSC liegt in seiner Datenauflösung: Es unterstützt die Datenformate *Float*, *Integer*, *Boolean* und *Strings*. Darüber hinaus erlaubt es die zeitliche Synchronisation über *Time Tags*. In modernen Syntheseumgebungen fungiert OSC häufig als Bindeglied zwischen mehreren Softwareinstanzen, externen Controllern oder Rendering Engines. Die freie Definition von Adressräumen erlaubt eine individuelle Gestaltung der Steuerstruktur. (Wright, 2002)

Obwohl MIDI und OSC unterschiedliche Paradigmen verfolgen, lassen sie sich in vielen Systemen parallel nutzen. Während MIDI durch seine breite Hardwarekompatibilität, Echtzeitfähigkeit und standardisierte Struktur überzeugt, liegen die Vorteile des Open Sound Protocol in Flexibilität, Datenkomplexität und Netzwerkfähigkeit.

Merkmal	MIDI 2.0	OSC
Protokolltyp	Seriell (DIN/USB)	Netzwerkbasiert (UDP/TCP/IP)
Datenstruktur	Fix (32Bit), MIDI-Kanäle	Frei definierbar, Baumstruktur
Datenformate	Integer	Float, Integer, String, Boolean
Zeitsteuerung	Keine Synchronisation	Synchronisation
Kompatibilität	Hardwareunterstützung	Software- und Netzwerkintegration

Tabelle 3: Vergleich der Standards MIDI 2.0 und OSC (Wright, 2002; Manufacturers Assoc, 2020b)

Die grafische Programmierumgebung MAX/MSP trennt konsequent zwischen Audio- und Steuersignalen. Während Audiosignale in der Regel mit einer Abtastrate von 44,1kHz oder höher verarbeitet werden, operieren Steuersignale in einem ereignisbasierten Takt mit einer niedrigeren Auflösung. In MAX/MSP sind Steuersignale typischerweise numerische Daten, die über Messages, Trigger-Objekte

oder Parameterveränderungen übertragen werden. Sie steuern beispielsweise Frequenzwerte von Oszillatoren, Amplitudenhüllkurven, Filterfrequenzen oder Positionierungsdaten im Raum. (Cipriani and Giri, 2010) Ein zentrales Objekt zur Steuerung in MAX ist `line~`, welches kontinuierliche Signalverläufe erzeugt und Sprünge im Signal glättet. Es ermöglicht die lineare Interpolation zwischen zwei Werten über eine definierte Zeitspanne, etwa um einen Filter gleichmäßig zu öffnen. Ein weiteres relevantes Objekt ist `gen~`, ein Low-Level-Modul, das samplegenaue Signalverarbeitung auf DSP-Ebene erlaubt. In `gen~` können sowohl Audio- als auch Steuerdaten gleichbehandelt werden. Das ermöglicht eine präzise Implementierung komplexer Steuerverläufe. MIDI-Daten lassen sich über Objekte wie `notein`, `ctlin`, `bendin` oder `midiparse` einlesen und gezielt weiterverarbeiten. OSC-Daten werden mittels `udpreceive` oder `oscroute` empfangen und können individuell adressiert werden. Dies erlaubt etwa das Mapping von externen Controllern auf Parameter des Programms. (Cipriani, Giri and Dudas Richard, 2020)

2.2 Psychoakustik im dreidimensionalen Raum

Die menschliche Wahrnehmung von Schall ist weit mehr als ein physikalischer Reiz. Sie ist ein komplexer Prozess aus neuronaler Verarbeitung und kognitiver Interpretation. Im Kontext dreidimensionaler Audioanwendungen ist das Verständnis psychoakustischer Grundlagen essenziell, um Klang gezielt zu positionieren, einzuhüllen und somit emotional erfahrbar zu machen. Die folgenden Abschnitte beleuchten drei zentrale Aspekte: die Ortung von Schallquellen, die Wahrnehmung von Einhüllung sowie die subjektive Beschreibung von Klangqualitäten.

2.2.1 Ortung

Die Ortung von Schallereignissen in der Horizontal- und Medianebene ist ein Zusammenspiel aus zeitlichen, pegelbezogenen und spektralen Informationen. Während in der Horizontalebene hauptsächlich Zeit- und Pegelunterschiede zur Lokalisation beitragen, sind in der Medianebene spektrale Veränderungen durch die Ohrmuschel ausschlaggebend (Weinzierl, 2008). Die räumliche Lokalisierung auditiver Reize ist ein komplexer psychoakustischer Vorgang, der auf einer Vielzahl

physikalischer Parameter basiert. Für die Beschreibung der Position einer Schallquelle wird ein kopfbezogenes Koordinatensystem etabliert, dessen Ursprung im Zentrum der gedachten Verbindungslinie zwischen den beiden Gehörgängen liegt. Dieses System unterteilt den Raum in drei Ebenen: die Horizontale, die Vertikale und die Frontalebene. Die Lokalisierung erfolgt auf Basis von drei Dimensionen: dem Azimutwinkel (φ), der Elevation (δ) und der Entfernung (r) zur Schallquelle.

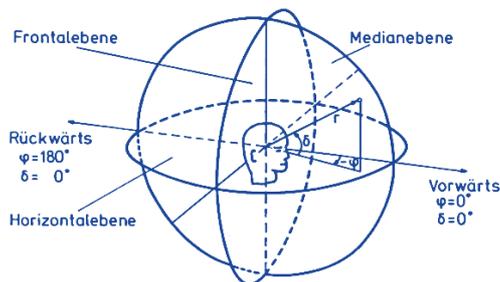


Abbildung 4: Kopfbezogenes Koordinatensystem (Weinzierl, 2008)

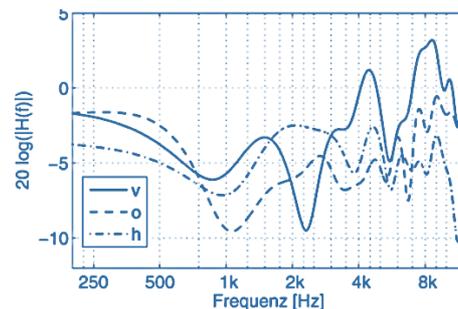


Abbildung 5: Kopfbezogene Übertragungsfunktion des Außenohrs für den Schalleinfall von vorne, oben, hinten (Weinzierl, 2008)

In der Horizontalebene sind die Interaural Time Difference (ITD) und die Interaural Level Difference (ILD) entscheidend. Diese interauralen Differenzen entstehen durch den Versatz der beiden Ohren auf der horizontalen Achse. Sie ermöglichen es unserem Gehirn, Unterschiede im Eintreffen des Schallsignals zwischen linkem und rechtem Ohr zu analysieren und daraus aus der Erfahrung heraus Rückschlüsse auf die Position der Quelle zu ziehen. Im Frequenzbereich $f \geq 2$ kHz treten Interaurale Pegelunterschiede (ILD) durch den Schallschatteneffekt des Kopfes auf, da dieser als akustisches Hindernis wirkt und der Schall sich nicht um dieses Hindernis herumbeugen kann. Diese Differenzen erreichen Werte von bis zu 20 dB SPL. Die ITD beschreibt ein Delay, welches durch den Laufzeitunterschied des Schalls zwischen dem einen und dem anderen Ohr entsteht. Diese Differenz kann in bis zu $\pm 0,7$ Millisekunden betragen. Besonders bei Frequenzen $f \leq 1,6$ kHz ist die ITD ein wirksames Lokalisationsmerkmal, da hier die Wellenlänge des Signals größer ist als der Kopfdurchmesser, was eine Phasenauswertung auf neuronaler Ebene erlaubt. (Weinzierl, 2008)

Neben den beschriebenen interauralen Differenzen spielen spektrale Unterschiede des durch das Außenohr gefilterten Schalls eine zentrale Rolle bei der Lokalisation. Diese

entstehen durch Reflexionen und Beugungen an der Pinna. Die resultierenden spektralen Veränderungen sind richtungsabhängig und liefern zusätzliche Informationen über den Ort des Schallereignisses, insbesondere in der Medianebene. Die Ohrmuschel fungiert als ein frequenzselektives Filter, das für bestimmte Richtungen durch Reflektionen charakteristische spektrale Muster erzeugt (siehe [Abbildung 5](#)). Diese Muster sind individuell geprägt und müssen durch Hörerfahrung erlernt werden. Diese richtungsabhängige Filterwirkung verändert die Energieverteilung im Frequenzspektrum und ermöglicht so auch bei identischen Zeit- und Pegelunterschieden der ILD und ITD eine Richtungsunterscheidung – etwa bei der Differenzierung zwischen vorne und hinten oder oben und unten. Mathematisch wird diese Funktion auch als Head-Related Transfer Function (HRTF) bezeichnet und beschreibt einen Filter in Abhängigkeit der Raumkoordinaten dem Azimut φ und Elevation δ . (Butler and Humanski, 1992; Weinzierl, 2008; Howard and Angus, 2009) Die Fähigkeit zur präzisen Lokalisierung ist in der Horizontalebene am höchsten. Studien zeigen, dass die mittlere Lokalisationsschärfe frontal ($\varphi = 0^\circ$) etwa $\pm 3,6^\circ$ beträgt. Diese Genauigkeit nimmt mit steigendem Azimutwinkel ab und verschlechtert sich auf $\pm 9,2^\circ$ bei seitlicher Schallquelle $\varphi = \pm 90^\circ$. Die Präzision hängt jedoch stark von den Frequenzanteilen des Signals und des Reflektionsmuster der Abhörumgebung ab. (Jens Blauert, 1997)

In der Medianebene sind ITD und ILD identisch. Die Ortung erfolgt hier hauptsächlich durch spektrale Unterschiede, womit die Lokalisation in der Medianebene abhängig von der spektralen Zusammensetzung des Schallereignisses ist. Da diese spektralen Muster stark frequenzabhängig sind, ist die Ortung in dieser Ebene ungenauer als in der Horizontalen. Lokalisationsfehler von $\pm 30^\circ$ sind bei einer Elevation von $\delta = 36^\circ$ sind möglich. (Weinzierl, 2008)

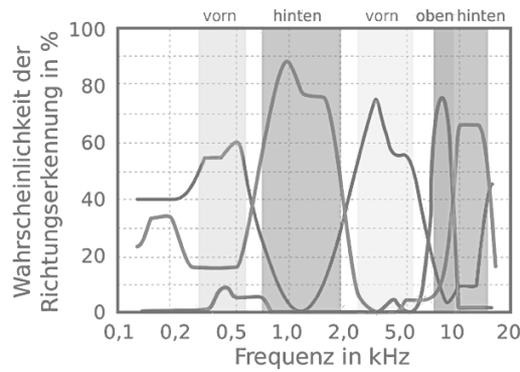


Abbildung 6: Blauertsche Bänder (Weinzierl, 2008)

Ein bekanntes Problem in der Medianebene ist die sogenannte Front-Back-Confusion. Unser auditiver Orientierungssinn ist in solchen Fällen auf subtile spektrale Hinweise angewiesen, die durch die Form der Ohrmuscheln hervorgerufen werden. Blauert beschreibt in seiner Untersuchung, dass bestimmte Frequenzbänder als besonders richtungsrelevant empfunden werden. So werden Frequenzen zwischen 250 Hz und 500 Hz sowie 3 kHz bis 5 kHz als Hinweis auf eine Position vor dem Hörer interpretiert, während Frequenzen zwischen 700 Hz und 1,5 kHz auf eine Position hinter dem Hörer hindeuten. Der Pitch-Height-Effekt beschreibt die psychoakustische Tendenz, höhere Frequenzen als höher im Raum liegend wahrzunehmen, auch wenn keine realen Höheninformationen im Signal enthalten. Diese Ortung kann sowohl die wahrgenommene Höhe als auch die Tiefe einer Quelle beeinflussen und wird visuelle Hinweise entscheidend beeinflusst. (Jens Blauert, 1997)

Raumrichtung	Spektrale Merkmale	Hörempfinden
Frontal (0°, 0°)	Geringe spektrale Veränderung	Neutrale Klangfarbe, direkte Ortung möglich
Seitlich (±90°, 0°)	Höhenabsenkung über 6–8 kHz, ITD/ILD sehr stark	Weite horizontale Positionierung, starke Ortung
Hinterkopf (180°, 0°)	Höhenabsenkung > 5 kHz, geringe ILD und ITD	Unschärfe Lokalisierung, Front-Back-Confusion

Oben (0°, +90°)	Betonung bei 8–10 kHz, Absenkung bei 2–4 kHz geringe ILD und ITD	Eindruck von Höhe, luftiger Klang
Unten (0°, -90°)	Absenkung um 1–2 kHz geringe ILD und ITD	Tiefer, dumpfer Eindruck, schlechter lokalisierbar
Oben hinten	Höhenbetonung bei 9–12 kHz, geringe ILD und ITD	Schwer lokalisierbar
Unten vorne	Geringe Höhen, Betonung bei 700–1200 Hz	Dumpfer Charakter, wenig Richtungssicherheit

Tabelle 4: Richtungsabhängige spektrale Charakteristika unterschiedlicher Raumrichtungen basierend auf kopfbezogener Übertragungsfunktion (Blauert and Lindemann, 1986; Butler and Humanski, 1992; Jens Blauert, 1997)

2.2.2 Umhüllung

Die Wahrnehmung einer räumlich ausgedehnten Schallquelle stellt ein zentrales Merkmal der umhüllenden Hörerfahrung dar. Im Unterschied zur punktuellen Ortung eines akustischen Ereignisses beschreibt Umhüllung das Gefühl, in eine klangliche Umgebung eingebettet zu sein. Dabei handelt es sich um ein Hörempfinden, das weniger durch präzise Lokalisationshinweise als vielmehr durch diffuse und richtungsunscharfe Signale hervorgerufen wird. Solche Signale erscheinen dem Hörenden nicht an einem spezifischen Punkt im Raum, sondern werden als „Fläche“, „Wolke“ oder „räumliche Atmosphäre“ wahrgenommen. (Weinzierl, 2008)

Allgemein formuliert lassen sich ausgedehnte Schallquellen erzeugen, indem Signale auf den beiden Kanälen gezielt voneinander abweichen, jedoch eine spektrale Ähnlichkeit besitzen. Technisch beschreibt dies eine Abnahme der Korrelation zwischen dem linken und rechten Audiokanal. Hierfür wird der Korrelationsgrad k eingeführt. Eine Korrelation von $k = 1$ beschreibt zwei identische Signale, was im Kontext eines Stereobilds als zentrierte Monoschallquelle resultiert. Demgegenüber führt ein Korrelationswert von $0 < k < 1$ zur Wahrnehmung einer breiten, nicht eindeutig lokalisierbaren Klangfläche. Zur visuellen Darstellung des Korrelationskoeffizienten k kann ein Vektorskop verwendet werden. Dieses Instrument

ist ein um 45 Grad gedrehtes Oszilloskop, welches die beiden Signale addiert. Eine positive Amplitude des ersten Signals zeigt nach links oben, eine positive Amplitude des zweiten Signals zeigt nach rechts oben. Wechsellspannungen werden als geschlossene Linien dargestellt. (Weinzierl, 2008)

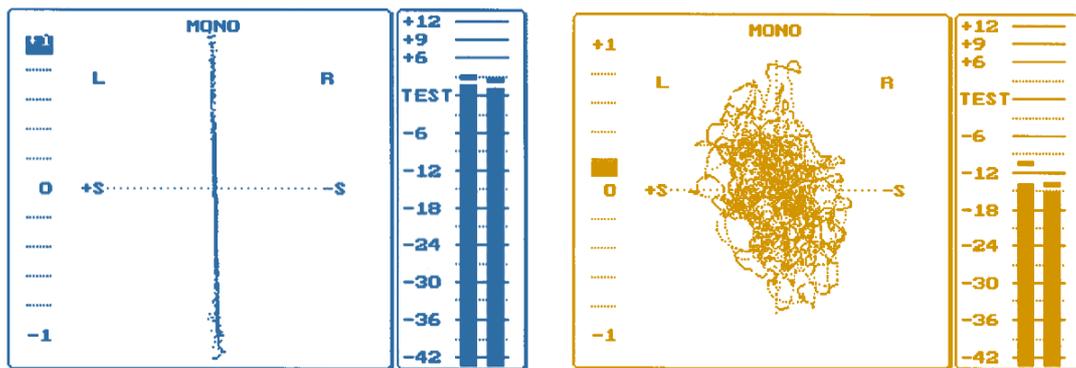


Abbildung 7: Korreliertes Signal (blau) und dekorreliertes Signal (gelb) im Vektorskop (Weinzierl, 2008)

Während in einem stereophonen Lautsprechersystem bereits durch die gezielte Dekorrelation zweier Signale ein subjektiver Eindruck von Breite entstehen kann, genügt das nicht um Umhüllung im dreidimensionalen Lautsprecherfeld zu generieren. Zwar können diese dekorrelierten Signale paarweise im Stereobild betrachtet als mehrfache, voneinander unabhängige Klangereignisse interpretiert werden. (Barbour, 2003) Zusätzlich ist es notwendig, dass die auf die verschiedenen Lautsprecher im 3D-Setup gerouteten Signale richtungsabhängige spektrale Charakteristika aufweisen, wie in [Tabelle 4](#) beschrieben. Diese positionsbezogene Filterung sorgt dafür, dass die Schalleinfallrichtung vom Hörenden differenziert wahrgenommen werden kann. Ein Höhenlautsprecher kann durch eine gezielte Betonung im Bereich zwischen 8 und 10 kHz zur wahrgenommen Umhüllung erheblich beitragen. Gleichzeitig kann ein Signal auf einem rückseitigen Lautsprecher durch eine Absenkung oberhalb von 5 kHz sowie eine subtile Änderung in der Phasenstruktur als von hinten kommend wahrgenommen werden. Dies fördert das Entstehen einer allumfassenden Klangfülle auch über und hinter dem Hörenden. (Butler and Humanski, 1992)

Zusammenfassend lässt sich festhalten: Die Generierung glaubhafter Umhüllung in einem dreidimensionalen Lautsprechersystem erfordert ein bewusstes Zusammenspiel von Signaldekorrelation und richtungsabhängiger Filterung.

2.2.3 Beschreibung von Hörempfinden

In den vorangehenden Kapiteln wurden technische und psychologische Grundsätze beschrieben. Dieser Abschnitt beschäftigt sich mit der empirischen Hörwahrnehmung. Die vorgestellte Forschung versucht objektive Raumparameter mit subjektiver Wahrnehmung zu verknüpfen – und ein Vokabular zu etablieren, das sowohl für die technische Konzeption als auch für das musikalische Design von Klangreglern des Instrumentenprototyps nützlich ist ([Kapitel 3.1](#)). Hierbei muss erwähnt werden, dass die Studien sich auf westlich kultivierte Personen beschränkt der kulturelle Unterschied nicht ausreichend berücksichtigt wird.

Auditory Scene Analysis

Die Theorie der Auditory Scene Analysis (ASA) wurde von dem Kognitionspsychologen Albert Bregman in den 1990er entwickelt und stellt ein grundlegendes Modell der auditiven Wahrnehmung dar. Ziel der ASA ist es zu beschreiben, wie das menschliche Gehör komplexe Klangumgebungen in sinnvolle Einheiten strukturiert. Bregmans (1990) zentrales Anliegen ist die Frage, ob der Hörsinn in der Lage ist gleichzeitig auftretende Schallereignisse zu segmentieren, zu gruppieren und so in distinkte auditive Objekte zu verwandeln. Ausgehend von einem kontinuierlichen Strom neuronaler Reize werden durch Gruppierungsprozesse kohärente Einheiten gebildet. Im auditiven Bereich erfolgt diese Gruppierung anhand von Ähnlichkeiten oder Unterschieden bestimmter Merkmale, wie Tonhöhe, Hüllkurve, Spektrum, Zeitverlauf, Ort und Bewegung. Diese Gruppierungsprozesse werden von Bregman die zwei Hauptkategorien *simultane* und *serielle Gruppierung* unterteilt. (Bregman Alb, 1990)

Ein zentrales Konzept der ASA ist das sogenannte Streaming, also die Aufspaltung des auditiven Inputs in mehrere parallele Wahrnehmungsströme. Dieses Phänomen lässt sich experimentell sehr gut demonstrieren, etwa durch abwechselnde Tonfolgen zweier unterschiedlicher Frequenzen. Werden die Töne schnell genug gespielt und ist der Frequenzabstand groß genug, so nimmt der Hörer sie nicht mehr als abwechselnde Töne, sondern als zwei getrennte Tonströme wahr. Dieser Effekt zeigt, dass die Wahrnehmung nicht direkt den physikalischen Gegebenheiten folgt, sondern durch neuronale Prozesse beeinflusst ist, die auf statistischer Wahrscheinlichkeitsverteilung beruhen. Bei der simultanen Gruppierung werden Bestandteile eines Klages, die zur gleichen Zeit auftreten, als Teil eines einzigen auditiven Ereignisses wahrgenommen.

Etwa die Obertöne eines Instruments, die gemeinsam eine Klangfarbe ergeben. (Yüksel, Önen and Seeber, 2024) Dagegen bezieht sich die *serielle Gruppierung* auf zeitlich aufeinanderfolgende Klangereignisse, die bei entsprechender Ähnlichkeit als ein kohärenter Strom wahrgenommen werden. Diese Mechanismen spielen insbesondere bei der Erkennung von Sprachmustern, Melodien oder sich bewegenden Klangquellen eine entscheidende Rolle. (Bregman, 1987) Ein weiterer relevanter Aspekt der ASA ist die Rolle von Gestaltprinzipien, wie Nähe, Ähnlichkeit, gute Fortsetzung oder Geschlossenheit. So werden beispielsweise räumlich benachbarte Klangquellen eher als zusammengehörig wahrgenommen, was direkt mit den zuvor beschriebenen Phänomenen der Ortung ([Kapitel 2.2.1](#)) und Umhüllung ([Kapitel 2.2.2](#)) in Zusammenhang steht.

Die auditive Szenenanalyse liefert damit nicht nur eine theoretische Grundlage zur Beschreibung der auditiven Wahrnehmung komplexer Szenen, sondern ermöglicht auch Rückschlüsse auf Gestaltungsprinzipien für Gestaltung von immersiven Audioinhalten. Beispielsweise kann durch gezielte Modulation, etwa der Klangfarbe, der räumlichen Position oder der Bewegung ein Eindruck getrennter oder zusammengehöriger auditiver Objekte erzeugt werden. Werden Schallereignisse mit ähnlicher spektraler Struktur, aber unterschiedlicher räumlicher Position erzeugt, so führt dies zu einer Trennung in der Wahrnehmung. Diese Effekte können gezielt genutzt werden, um dichte Klangtexturen aufzubrechen oder um einzelne Klangquellen trotz dichter Umhüllung hörbar zu machen. (Verron, Aramaki and Kronland-Martinet, 2010)

Dimensionsanalyse auditiver Wahrnehmung

Eberhard Zwicker prägt in den 1960er- und 1970er-Jahren die Grundlagen der Psychoakustik. Ziel seines Ansatzes ist es, subjektive Hörempfindungen mathematisch und messtechnisch zu fassen. Zwickers zentrale Methode zur Erhebung dieser Dimensionen basiert auf psychophysischen Urteilen von Testpersonen, die systematisch Paare von Klängen bewerten. Mithilfe der *multidimensionalen Skalierung* können aus diesen Daten Koordinatensysteme abstrahiert werden. Dadurch wird es möglich, sowohl Unterschiede zwischen Klang entlang bestimmter Wahrnehmungsachsen zu quantifizieren. Zwicker identifiziert insgesamt sechs

relevante Wahrnehmungsdimensionen. (Fastl and Zwicker, 2007) Die *Lautheit* bezeichnet dabei nicht die objektive Schalldruckstärke, sondern die subjektive Wahrnehmung von Lautstärke, welche logarithmisch zur physikalischen Intensität verläuft und stark frequenzabhängig ist. Die *Tonhaltigkeit* differenziert zwischen tonalen und rauschhaften Klanganteilen und gibt damit Auskunft über die Spektralstruktur eines Signals. Die *Schärfe* beschreibt die spektrale Gewichtung in Richtung höherer Frequenzen – ein hoher Anteil an oberen Teiltönen lässt den Klang als *scharf* oder *stechend* erscheinen. (Fastl and Zwicker, 2007) *Rauhigkeit* und *Schwankungsstärke* hingegen sind zeitliche Eigenschaften: sie messen die Schnelligkeit und Regelmäßigkeit von Amplitudenmodulationen innerhalb des Signals im Bereich von etwa $15 \text{ Hz} < f < 300 \text{ Hz}$ (*Rauhigkeit*) beziehungsweise $f < 20 \text{ Hz}$ (*Schwankungsstärke*). Diese Parameter beeinflussen die Lebendigkeit und Textur eines Klanges maßgeblich. Die *Klangfarbe* stellt als sechste Dimension einen komplexeren, oft metaphorisch verwendeten Begriff dar, der durch die Kombination der vorgenannten Merkmale entsteht und in der Alltagssprache mit Begriffen wie *warm*, *dumpf*, *hell* oder *brillant* beschrieben wird. (Fastl and Zwicker, 2007)

Lautheit	Subjektive Lautstärkeempfindung, nicht linear zur physikalischen Intensität, frequenzabhängig
Tonhaltigkeit	Anteil an tonalen vs. rauschhaften Bestandteilen des Klangs
Schärfe	Gewichtung hoher Frequenzen, bestimmt den „stechenden“ oder „hellen“ Charakter
Rauhigkeit	Wahrnehmung schneller Modulationen (ca. 15–300 Hz), beeinflusst die Textur und Lebendigkeit
Schwankungsstärke	Wahrnehmung langsamer Modulationen (<20 Hz), trägt zur Bewegung oder Pulsation bei
Klangfarbe	Zusammenspiel aller genannten Dimensionen, beeinflusst metaphorische Beschreibung wie „warm“, „dumpf“, „glänzend“

Tabelle 5: Dimensionsanalyse nach Zwicker (Fastl and Zwicker, 2007)

Sound Wheel

Das Sound Wheel (*Force Technology, 2025*) dient als Grundlage für den ITU-Standard ITU-R BS.2399-0 (2017), bietet ein hierarchisches Schema zur Beschreibung auditiver Wahrnehmung. Es beruht auf dem statistischen Verfahren der multidimensionalen Skalierung. Dieses Verfahren ermöglicht eine empirisch gestützte Strukturierung eines Klangvokabulars. Als Ergebnis stellt das Sound Wheel ein standardisiertes Vokabular zur Beschreibung subjektiver Klangwahrnehmung bereit und bietet sowohl für Laien, Musiker:innen aber auch für Ingenieur:innen und Wissenschaftler:innen eine übersetzbare Kommunikationsbasis. Im inneren Ring des Sound Wheels finden sich abstrahierte Hauptkategorien wie *Räumlichkeit*, *Artefakte*, oder *Timbre*. Diese Begriffe strukturieren das Erleben auditiver Reize in psychologisch nachvollziehbare Wahrnehmungsgruppen. Der mittlere Ring unterteilt die Hauptkategorien in spezifischere Eigenschaften, etwa *Färbung*, *Direktheit* oder *Lokalisation*. Die Stärke des Sound Wheel liegt in klar strukturierten Abstraktionen subjektiver Höreindrücke. (ITU-R BS.2399-0, 2017)

Innerer Ring Hauptgruppen	Bsp. Räumlichkeit, Präsenz, Klarheit, Artefakte
Mittlerer Ring Kategorien	Bsp. Ausdehnung, Diffusität, Intimität
Äußerer Ring Adjektive	Bsp. .nasal, brilliant, hallig

Tabelle 6: Das Sound Wheel

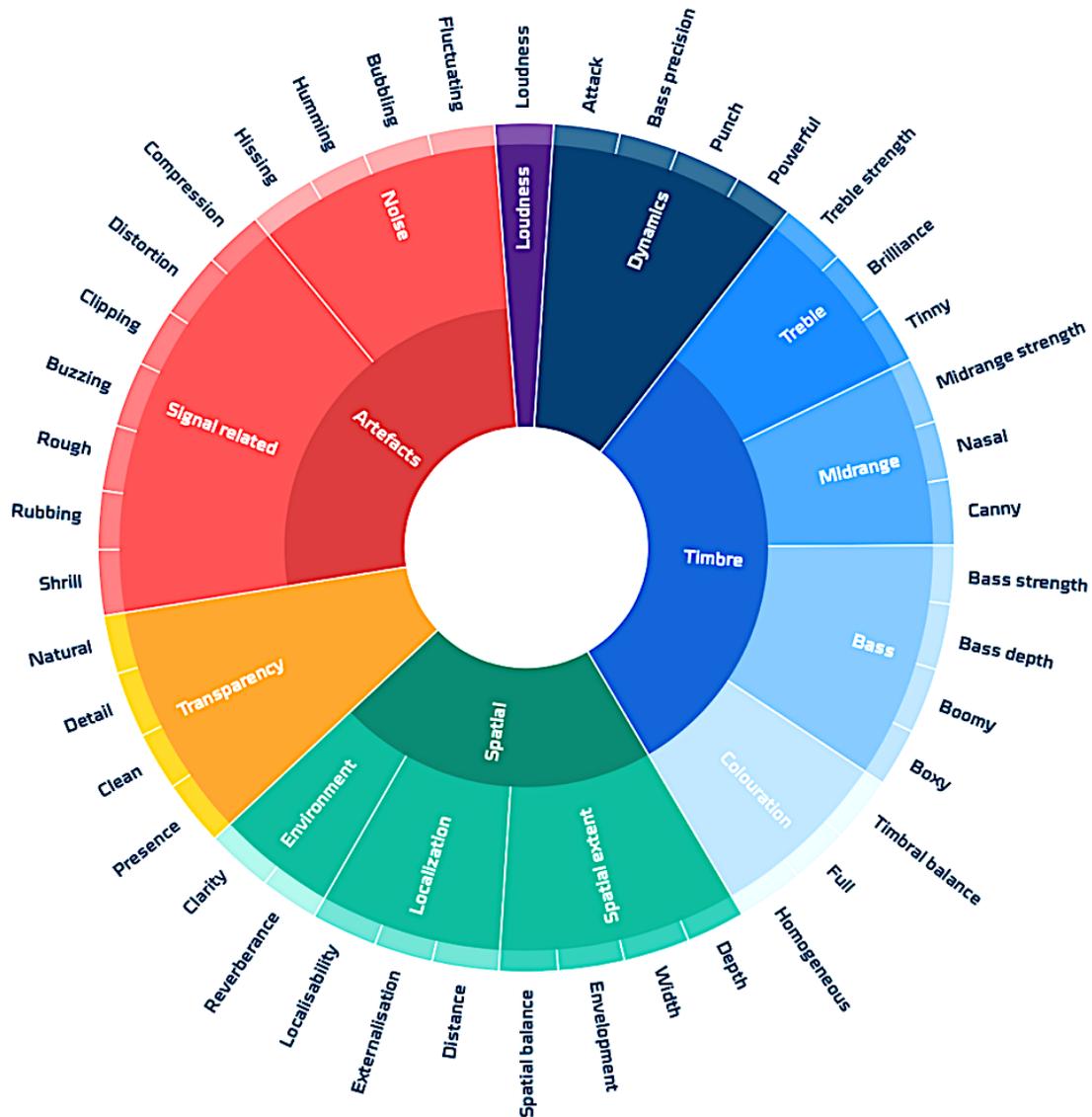


Abbildung 8: Sound Wheel (Force Technology, 2025)

Multidimensionale Skalierung

Ein weiteres sozialwissenschaftliches Prinzip der multidimensionalen Skalierung (MDS). Es dient dazu, die subjektive Wahrnehmung von Begriffen, Objekten oder systematisch zu erfassen. Die Methode basiert auf der Einschätzung eines Stimulus entlang gegensätzlicher Adjektivpaare. Sie bilden Skalen mit diskreten Abstufungen, auf denen Proband:innen ihre Wahrnehmung verorten. Ziel ist es, die Bedeutung und Bewertung eines Reizes zu quantifizieren. Ein klassisches Beispiel für diesen Ansatz ist

die Arbeit von Gabrielsson und Sjögren (1979), die auf Basis eines umfangreichen Datensatzes eine Vielzahl musikalischer Klänge hinsichtlich ihrer expressiven Qualität analysierte. Sie konnten mehrere Hauptdimensionen ableiten – darunter *Intensität*, *Komplexität*, *Helligkeit* und *Textur*. Durch die Gegensätzlichkeit können etwa verschiedene Schalter in der Klangsynthese, oder Mischverhältnisse nach semantischen Kriterien designt und in Bezug zu technischen Parametern parametrisiert werden. (Gabrielsson and Sjögren, 1979)

Dimension	Beispielhafte Gegensatzpaare	Psychoakustische Korrelate
Energie / Dynamik	laut – leise kräftig – sanft	Lautheit, Schallintensität, Transientenenergie
Klangfarbe / Helligkeit	hell – dunkel metallisch – holzig	Spektrale Verteilung, Formantstruktur
Textur / Temporalität	rau – statisch	Amplitudenmodulation Hüllkurvencharakteristik
Räumlichkeit / Körperlichkeit	nah – fern breit – eng	ITD/ILD, HRTF-Modulation, Interkanaldekoration
Komplexität / Struktur	einfach – komplex monoton – bewegt	Dichte spektraler und temporaler Ereignisse, rhythmische Variabilität

Tabelle 7: Multidimensionale Skalierung (Gabrielsson and Sjögren, 1979)

Im Kontext dieser Arbeit eröffnen die vorgestellten psychologischen Modelle die Möglichkeit, den gestalterischen Umgang mit Klang nicht nur auf Basis technischer Parameter wie Frequenz, Amplitude oder Hüllkurve zu strukturieren, sondern vielmehr auf psychoakustischer Ebene zu operieren. Die Nutzung des Instruments wird so nicht nur intuitiv nachvollziehbar, sondern auch nachweislich belegbar.

2.3 Formate für 3D-Audio

Um ein immersives Instrument zu entwickeln, gilt es das Format als Interface zwischen Instrument und Lautsprecher zu begreifen. Somit können Kanäle gezielt Lautsprechern zugeordnet werden und richtungsabhängig gefiltert werden. Im Folgenden werden zwei gängige Formate vorgestellt und verglichen.

2.3.1 Dolby Atmos

Dolby Atmos ist ein objektbasiertes 3D-Audioformat, das sich maßgeblich von traditionellen kanalbasierten Ansätzen unterscheidet. Während klassische Formate wie Dolby 5.1 oder Dolby 7.1 mit fest definierten Kanälen arbeiten, auch *bed* genannt, erlaubt der aktuelle Standard der Firma Dolby neben den bisher üblichen *beds* zusätzlich die freie Platzierung von Audioobjekten im dreidimensionalen Raum. Das *bed* bestehen üblicherweise aus einem festen Kanal-Layout, etwa 7.1.2, während bis zu 128 zusätzliche Objekte mit individuellen Metadaten für Position und Bewegung versehen werden können. Deckenlautsprecher oder nach oben abstrahlende Lautsprecher einer Soundbar ergänzen das horizontale Setup, um die dritte Dimension zu erschließen. Besonders in Kinosälen oder professionellen Studios kommt die Konfiguration 9.1.6 zum Einsatz, bei der sechs Höhenlautsprecher – je drei vorne und hinten – verbaut sind. Bei kommerziellen Musikproduktionen steht jedoch nur das Streamingformat 7.1.2 zur Verfügung. Im Decoding-Prozess analysiert das Wiedergabesystem die räumlichen Informationen und passt das Routing der Objekte und das angepasste *bed* an das vorhandene Lautsprecher-Layout an, welches im Renderer hinterlegt wird. Die Aufgabe des Renderers besteht darin, eine möglichst präzise räumliche Abbildung mit den vorhandenen Lautsprechern zu realisieren. Die Steuerdaten für die Position der Objekte werden gesondert in Form von Metadaten übermittelt. Die Steuerdaten für die Position der Objekte werden in Form von Metadaten übermittelt. (‘Dolby ® Atmos TM Next-Generation Audio for Cinema Overview’, 2025)

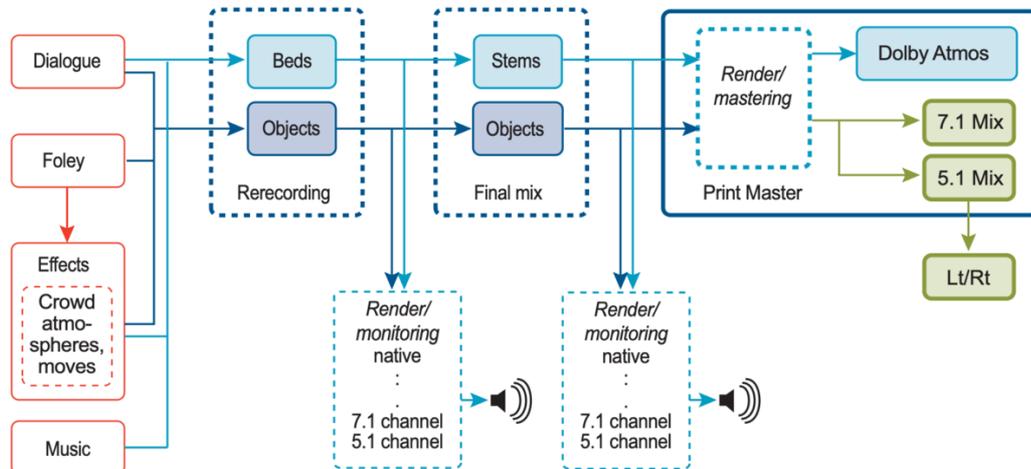


Abbildung 9: Encoding für Dolby Atmos Objekte und Beds
 ('Dolby ® Atmos TM Next-Generation Audio for Cinema Overview', 2025)

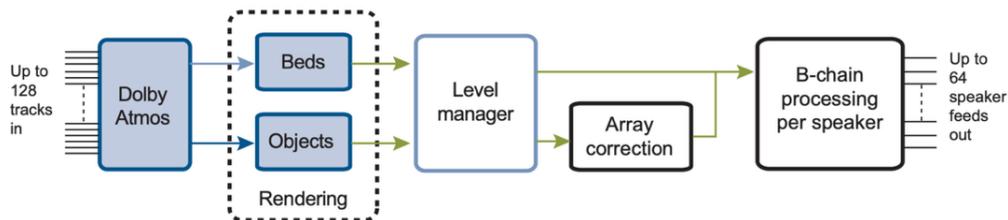


Abbildung 10: Decoding für Dolby Atmos Objekte und Beds
 ('Dolby ® Atmos TM Next-Generation Audio for Cinema Overview', 2025)

Die Empfehlung ITU-R BS.2051 (Radiocommunication Bureau, 2022) beschreibt Standardkonfigurationen für immersive Audio-Wiedergabe. Für ein 7.1.2-Setup, das Dolby Atmos-kompatibel ist, gelten die folgenden Positionierungen: Die horizontale Ebene besteht aus neun Lautsprechern: ein Center-Lautsprecher (0° Azimut), zwei Frontlautsprecher ($\pm 30^\circ$), zwei Side-Surrounds ($\pm 90^\circ$), zwei Rear-Surrounds ($\pm 135^\circ$). Die Höhe der Lautsprecher sollte sich idealerweise auf Ohrhöhe des Hörers befinden. Die Höhenebene umfasst die beiden Lautsprecher front-top-left und front-top-right. Die vertikale Anordnung erfolgt möglichst symmetrisch über der horizontalen Ebene.

2.3.2 Ambisonics

Ambisonics ist ein feldbasiertes Aufnahme- und Wiedergabesystem. Es basiert auf der Repräsentation eines Schallfelds über sphärische Harmonische. Diese mathematische Beschreibung erlaubt die Erfassung und Rekonstruktion eines vollständigen akustischen Raums in drei Dimensionen. Anders als bei klassischen kanalbasierten Formaten, in denen jedes Signal einem festen Lautsprecher zugeordnet ist, wird bei Ambisonics das Klangfeld selbst kodiert und kann unabhängig von der Lautsprecheranordnung encodiert werden. Die erste Ordnung von Ambisonics umfasst vier Kanäle: W (Kugel), X (vorn–hinten), Y (links–rechts) und Z (oben–unten). (Wittek, Haut and Keinath, 2006) Höhere Ordnungen erweitern diese Basis um weitere Richtungsinformationen durch zusätzliche sphärische Harmonische. Für die vierte Ordnung sind insgesamt 25 Kanäle notwendig. Eine Anpassung auf unterschiedliche Lautsprecher-Setups ist durch Dekodierung möglich, sodass beispielsweise auch für binaurale Kopfhörerwiedergabe oder Lautsprecherinstallationen mit nahezu beliebiger Anzahl und Anordnung eine konsistente räumliche Darstellung gewährleistet werden kann. Die Ambisonics-Technologie zeichnet sich gegenüber klassischen kanalbasierten Formaten durch die besondere Eigenschaft aus, das Klangfeld unabhängig von spezifischen Lautsprecher-konfigurationen zu kodieren. (Zotter and Frank, 2019)

Statt feste Kanäle für bestimmte Lautsprecherpositionen zu definieren, speichert Ambisonics die Schallinformation in einer mathematisch abstrahierten Form mithilfe sphärischer Harmonien. Diese Repräsentation des Schallfeldes erlaubt es, das auditiv erfahrbare Klanggeschehen im dreidimensionalen Raum beliebig zu rotieren, zu skalieren oder mit akustischen Effekten zu versehen, ohne dabei die Klangquelle selbst manipulieren zu müssen. Für Ambisonics sind die Anforderungen an das Lautsprecher-Array flexibler, erfordern aber nach dem generelle Prinzip eine symmetrische, möglichst sphärisch verteilte Anordnung, um die mathematisch berechnete Sphäre des Schallfeldes präzise abzubilden. (Zotter and Frank, 2019)

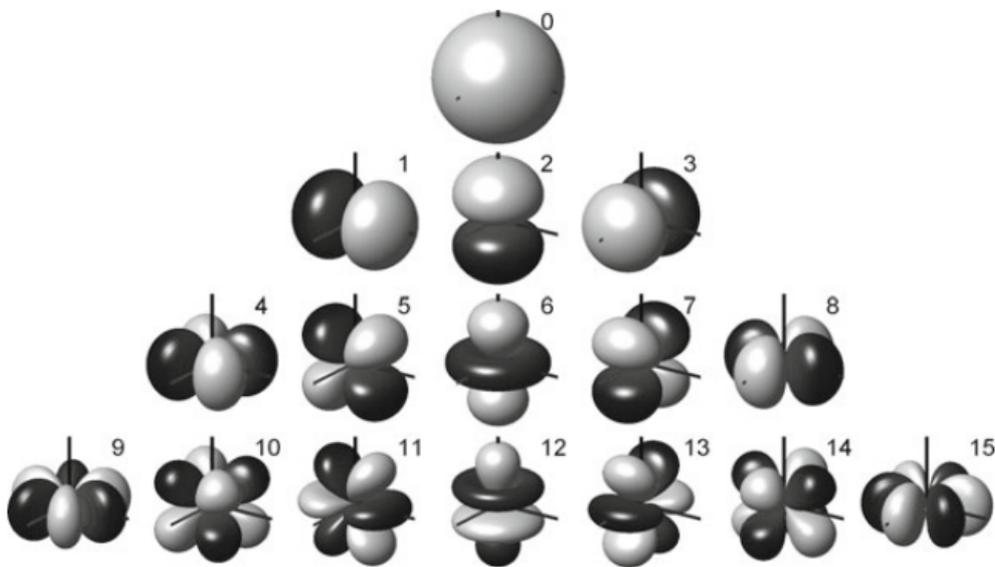


Abbildung 11: Die Sphärischen Harmonischen (Zotter and Frank, 2019)

Theoretisch werden für die vierte Ordnung Lautsprecherkonfigurationen mit 25 Kanälen in Kuppelform gestellt. Die Lautsprecher sollten gleichmäßig über die Horizontal- und Medianebene verteilt sein, um eine gleichwertige räumliche Auflösung in allen Richtungen zu gewährleisten. In der Praxis wird jedoch eine Kuppel realisiert. Hierbei kommt es weniger auf die Anzahl an, sondern auf die Gleichverteilung im Raum an. Decoder, beispielsweise durch Software wie *spat5* (*ForumMaxApps2-Spat5* | *Ircam Forum*, 2025) oder VST-basierte Plug-Ins, beispielsweise IEM-Plug-Ins (*IEM Plug-in Suite*, 2023) rechnen die virtuelle Szene in konkrete Lautsprecher signale um. Insbesondere im Kontext kreativer Audioprogrammierung bietet Ambisonics eine kreative Grundlage. Die experimentelle Verwendung von Ambisonics als Zwischenformat hat sich beispielsweise in modularen Audioframeworks wie *Spat5* in MAX/MSP etabliert. Diese Umgebung ermöglicht nicht nur die Erstellung, Rotation und Dekodierung auditiver Szenen, sondern auch die Integration räumlicher Effekte wie Faltungshalle, Allpass-Dekorrelation, Raumsimulation oder Positionstracking für binaurales Rendering. (Pulkki, 2000; Cipriani, Giri and Dudas Richard, 2020)

Dolby Atmos 7.1.2 vs. Ambisonics 4. Ordnung

Merkmal	Dolby Atmos 7.1.2	Ambisonics 4. Ordnung
Prinzip	Kanal- und objektbasiert Rendering auf genormte Lautsprecherstups	Feldbasiert über sphärische Harmonische, Rendering für beliebige Lautsprecherstups
Anzahl Kanäle	10	25
Höhenauflösung	Begrenzt (nur zwei Höhenkanäle)	Kontinuierlich, abhängig von Ordnung und Dekodierung
Lautsprecher- abhängigkeit	Hoch (spezifische Konfiguration erforderlich)	Gering, sehr flexibel durch Dekodierung
Anwendungsgebiete	Kino, Gaming, kommerzielle Musik	VR, AR, Installationen, Kunst

Tabelle 8: Vergleich von Dolby Atmos zu Ambisonics

3 Konzeption

Die vorangegangenen Kapitel wurden die erforderlichen Grundlagen für die Klangsynthese und das Verständnis der auditiven Wahrnehmung im dreidimensionalen Raum dargelegt. Insbesondere die Abschnitte zur Ortung von Schallquellen ([Kapitel 2.2.1](#)) und zur Wahrnehmung von Umhüllung ([Kapitel 2.2.2](#)) verdeutlichen, dass der Gehörsinn auf eine Vielzahl akustischer Hinweise reagiert und so Position und Ausdehnung wahrnimmt. [Kapitel 2.2.3](#) knüpft daran an, indem es die subjektive Beschreibung auditiver Eindrücke nach Zwicker (2007), Gabrielsson (1979) und der ITU-R BS.2399-0 (2017) beleuchtet und konkrete Werkzeuge zur Kategorisierung und Kommunikation von Klangerfahrungen anbietet. Diese theoretischen Erkenntnisse bilden das Fundament für das folgende Kapitel, in dem die konzeptionellen Entscheidungen zu dem Steuer- und Audiosignalfluss, zum Design der Parameter und zum Entwurf der grafischen Benutzeroberfläche des entworfenen Synthesizer-Prototyps dargestellt werden. Ziel ist es, eine intuitive, psychoakustisch inspirierte Steueroberfläche zu entwickeln, die Musiker:innen erlaubt, komplexe Steuersignale mit zusammengefassten und semantisch sinnvollen Bedienelementen zu steuern. Dabei geht es nicht allein um die technische Realisierung eines mehrkanaligen Instruments, sondern vielmehr um die Abbildung psychoakustischer Wirkprinzipien in eine kreative Steuerlogik, die technische Parameter abstrahiert. Diese Parameter greifen psychoakustisch relevante Parameter wie Filterargumente, Panning, Modulation und Routing auf. Klassische Syntheseparameter werden erweitert durch *Panning*, *Vividness*, *Size* und *Direct*. Musiker:innen erhalten damit Werkzeuge an die Hand, die nicht rein technische Begriffe wie Hüllkurvenzeit oder Filterfrequenz abbilden, sondern vielmehr klangliche Gesamtkonzepte. Die grafische Nutzeroberfläche wird so zu einem Vermittler zwischen der technisch unübersichtlichen Struktur des Synthesizers und der musikalischen Praxis der Musiker:innen.

Das folgende [Kapitel 3.1](#) beschreibt die Ideenfindung und finale Auswahl der konkreten Steuerparameter der Benutzeroberfläche. Aufbauend auf den psychoakustischen Wirkmechanismen wird erläutert, wie sich diese Parameter funktional abbilden lassen und welche semantischen Kategorien ihnen zugrunde liegen. Diese skizzierten Parameter dienen im weiteren Verlauf als Grundlage für die Entwicklung des grafischen

Interface und des zugrundeliegenden Audio- und Steuersignalfusses ([Kapitel 3.2](#) und [Kapitel 3.3](#)).

3.1 Parameter der Nutzeroberfläche

Folgend wird ein konzeptioneller Entwurf möglicher Steuerparameter vorgestellt. Ziel dieser Sammlung ist es, ein möglichst breites Spektrum an klanglichen und räumlichen Steuergrößen zusammenzutragen, die das kreative Potenzial eines mehrkanaligen Synthesizers in einer 3D-Audio-Umgebung abbilden. Die Parameter sind so konzipiert, dass sie sowohl technische Kontrollinstanzen wie Hüllkurven, Filter oder Modulatoren abbilden als auch intuitiv mit wahrnehmungspsychologischen Kategorien wie Präsenz, Räumlichkeit oder Bewegung verknüpft werden.

Es sei an dieser Stelle betont, dass nicht alle der nachfolgend aufgeführten Parameter im finalen Prototyp des Synthesizers implementiert werden. Vielmehr dient dieser Abschnitt als Grundlage des kreativen Ansatzes dieser Arbeit.

Parameter-name	Psychologische Wirkachsen	Psychoakustische Beschreibung	Technische Umsetzung
Timbre	warm – hell, weich – scharf	Klangfarbe, Obertonstruktur	Auswahl unterschiedlicher Wellenformen, Filtertypen
Charakter	weich – hart, fließend – kantig, lang – kurz	Zeit- und Dynamik- wahrnehmung	Hüllkurvenmodul (A, D, S, R)
Orientation	frontal – seitlich, unten – oben, nah – entfernt	Raumlage der Klangquelle	Steuerung der Parameter Azimut, Elevation und Entfernung in der auditiven Szene
Vividness	starr – lebendig, starr – vibrierend	Bewegung der Position	LFO-basierte Modulation der Orientation-Parameter

Parameter-name	Psychologische Wirkachsen	Psychoakustische Beschreibung	Technische Umsetzung
Width	kompakt – weit	Räumliche Ausdehnung und Umhüllung	Dekorrelation durch Allpassfilter , Mikro-Delays, Granularität
Fatness / Brilliance	dünn – fett, dumpf – brillant	Grundtonanhebung Höhenanhebung	EQ-Steuerung in Bässen und Höhen
Distance	nah – entfernt, direkt – diffus	Entfernungs-empfinden im Raum	Lautstärke, Hochfrequenzdämpfung, Hall Dry-Wet, Pre-Delay
Movement	statisch – wandernd, ruhig – beweglich	Dynamik durch Raumbewegung	Automatisierte Modulation von Orientation und Distance
Texture	glatt – rau, schlicht – komplex	Spektrale Feinstruktur des Klangs	Granulare Parameter, Spektralfilter
Sharpness	weich – scharf, diffus – definiert	Prägnanz und Lokalisation	EQ-Anhebung oberhalb 5 kHz, nichtlineare Verzerrungen
Spread	punktuell – flächig, eng – breit	Räumliche Verteilung	Abbildung auf mehrere benachbarte virtuelle Quellen
Presence	intim – präsent, mild – aggressiv	Vordergründigkeit im Abhörraum	Lautheit, Transientenshaping

Tabelle 9: Parameter der grafischen Nutzeroberfläche

Folglich sind einige dieser Parameter ausführlicher beschrieben in ihrem Steuer- und Audiosignalfluss. Dese Parameter werden im Verlauf dieser Arbeit weiterverfolgt und fließen in die Konzeption des gesamten Steuer- und Audiosignalflusses ein und werden Teil des Prototyps.

Timbre

Eine Liste stellt eine Auswahl von Presets bereit. Eine MIDI-Nachricht mit *Note-On*, *Note-Off*, *Pitch* und *Velocity* wird an die jeweilige Instanz der jeweiligen Voice übergeben. Mehrere parallel geschaltete Oszillatoren wie Sägezahn oder Rechteck, generieren jeweils unterschiedliche Wellenformen. Anschließend steuert ein Steuersignal die Pegelverhältnisse und die Cutoff-Frequenz eines resonierenden Tiefpassfilters. Die Audioausgabe der parallel geschalteten Oszillatoren wird mit einem individuellen Gain multipliziert und in einen Filterblock geleitet, in dem dynamisch die Obertonstruktur der einzelnen Audioströme reduziert wird. Zusätzlich zu obertonreichen Syntheseverfahren wird ein Sinusgenerator mit $Pitch = -12$ angesteuert. Er dient als Low-End und wird nicht spatial verarbeitet. Grundsätzlich sind alle Oszillatoren in einer Polyumgebung implementiert, das heißt für unterschiedliche Pitches können zusätzliche Instanzen erzeugt werden. Dies bedeutet in der musikalischen Praxis, das an einem Instrument mehrere Tasten gleichzeitig gespielt werden können. Das ermöglicht das Spielen von Harmonien auf der Klaviatur.

Charakter

Eine grafische Einheit steuert die Amplitudenhüllkurve durch visuelle Manipulation von vier Slidern für Attack, Duration, Sustain, Release. Die Interaktion mit der ADSR-Grafik erzeugt vier Parameterwerte, die über einen Multislidern an die jeweilige Instanz der Voice einen Hüllkurvengenerator weitergegeben werden. Diese Werte definieren die zeitliche Abfolge der Amplitudenverläufe der Hüllkurve. Die erzeugte Steuerkurve wird als Modulationsquelle auf die Amplitudensteuerung der jeweiligen Voice angewendet. Der Audioausgang des Oszillators wird durch eine vom Hüllkurvengenerator erzeugte Kurve multipliziert, wodurch der Lautstärkeverlauf des Klangs über die Zeit verändert wird und in den vier Phasen *Attack*, *Duration*, *Sustain*, *Release* erzeugt wird.

Fatness

Ein Regler steuert die wahrgenommene Präsenz des Low-Ends. Ein Steuersignal beeinflusst den Gain-Regler eines Verstärkers. Dem Audiosignal werden hierbei durch Übersteuerung nichtlineare Verzerrungen hinzugefügt. Das erhöht die Saturierung und die wahrgenommene Präsenz des Low-Ends.

Clarity

Ein Regler steuert die wahrgenommene Klarheit. Der Regler interpoliert zwischen zwei Filterkurven: Eine betont den Grundtonbereich (Low-Shelf-Anhebung), die andere die obere Partialtöne (High-Shelf-Anhebung). Je nach Position des Reglers werden Gain-Werte der Shelving-Filter angepasst. Die Audiostreams werden durch ein Filter geroutet. Möglich ist auch eine Modulation der Grenzfrequenz der Filter.

Voice Panning

Eine Nutzeroberfläche ermöglicht die räumliche Positionierung der einzelnen Voices des Synthesizers in der auditiven Szene über zwei interaktive Koordinatensysteme in horizontaler und vertikaler Ebene. Die Steuerdaten Azimut und Elevation werden für jeden Audiostream an einen Panner übertragen. Die einzelnen Audiostreams der Voices werden in einem zusammenfassenden Ambisonics Bus codiert. In diesem Zwischenformat können die einzelnen Stimmen oder die auditive Szene hinsichtlich Azimut und Elevation moduliert werden. Anschließend erfolgt die Decodierung auf das Lautsprecher-Layout über einen Decoder.

Parallele Audiosignalverarbeitung

Die Audiostreams werden nach der Klangerzeugung und -bearbeitung nun parallel zum Panning der Voices in der Ambisonics Domäne prozessiert. Dazu werden die einzelnen Stimmen mono summiert und in die spaziale Signalverarbeitung geführt. Diese besteht aus den drei Instanzen Allpassfilterung, Reverberation und Vividness. Diese drei Parameter der Nutzeroberfläche greifen in diese Signalverarbeitungskette ein. Das summierte Monosignal der Oszillatoren wird als Eingangssignal für die Reverb- und Allpassinstanzen geroutet. Anschließend entstehen je zwei 25-kanalige Audiostreams der beiden Dekorrelationsinstanzen und werden kanalweise summiert und in das Vividness-Objekt geroutet.

Size

Ein Regler bestimmt die empfundene Raumgröße. Das Steuersignal beeinflusst das Pegelverhältnis zwischen Direktsignalen und erzeugter Dekorrelation. Zusätzlich beeinflusst das Steuersignal die Duration und das Pre-Delay eines Reverbs. Die nun zusammengefassten Oszillatoren-Audiostreams werden als Monosignal in eine

Reverbengine mit 25 Instanzen prozessiert. Parallel dazu wird das übergebene Monosignal in eine Allpass-Filterbank mit 25 Instanzen prozessiert und dadurch in der Phase manipuliert. Reverbs-, Allpass- und Ambisonics-Domäne sind nun parallel zueinander.

Directness

Ein Regler steuert das Hörereignis der Entfernung. Das Steuersignal *direct* multipliziert das Steuersignal für die Duration und das Pre-Delay des Reglers *Size*. Hierdurch wird ein kurzes Pre-Delay für entfernte Wahrnehmung vorgeschlagen und ein langes Pre-Delay für direktere Wahrnehmung der Schallquelle. Der Vorteil von Trennung der Steuersignale *size* und *direct* liegt in der semantisch sinnvollen Verbindung von Allpass und Duration, um die *Size* zu modulieren. Bei einem hohen *size*-Wert kann es bei bestimmten Spielweisen anfangen zu schwimmen. Durch den *Directness* Regler kann in die Reverb Duration und das Reverb Pre-Delay eingegriffen werden.

Vividness

Ein Regler steuert zusammenfassend die wahrgenommene Ausdehnung im Raum. Hierzu werden in einer Matrix 25 Instanzen einer Modulation implementiert. Das erzeugte Modulationssignal jeder Instanz besteht aus einem Zufallsgenerator welches geglättet wird, um in der weiteren Verarbeitung Sprünge und hochfrequente Artefakte wie „Knackser“ zu vermeiden. Das Steuersignal *vividness* beeinflusst das randomisierte Modulationssignal und ändert dieses hinsichtlich seiner Amplitude und Frequenz. Das Modulationssignal jeder Instanz wird anschließend mit jeweils einem Kanal des Audiostreams multipliziert. Dadurch entstehen auf den 25 Kanälen zufällige und voneinander unabhängige Pegelsprünge.

Spatial Attack

Ein Schalter aktiviert oder deaktiviert die Modulation. Ein Regler steuert die Attack-Zeit. Aus der MIDI-Nachricht wird der Wert *velocity* extrahiert und anschließend als Steuersignal für einen Voltage Controlled Amplifier verwendet. Als gestalterisches Mittel wird das Velocity Signal in Abhängigkeit des Reglers geglättet. Hierdurch bestimmt die Anschlagsdynamik des MIDI-Instruments den Send-Level aller Ausdehnung. Das Audiosignal wird nach der Summierung der einzelnen Audiostreams der Oszillatoren über einen Voltage Controlled Amplifier verstärkt, ähnlich wie der

Send-Kanal an einem Mischpult, welcher die Intensität eines parallel prozessierten Effekts steuert. Zur Verdeutlichung des Effekts wird eine Pegelanzeige des Send-Levels schematisch abgebildet.

3.2 Steuer- und Audiosignalfluss

Damit Steuerlogik innerhalb der Software praktisch anwendbar wird, bedarf es eines klar definierten Audio- und Steuersignalflusses, der die technische Umsetzung der einzelnen Parameter in kohärente Verarbeitungsketten übersetzt. Die Signalflüsse verbinden die grafische Benutzeroberfläche mit der eigentlichen Klang- und Raumgestaltung im System. Neben der Organisation paralleler Audio- und Steuersignale müssen hierbei auch zentrale Instanzen wie Filter, Hüllkurven, Modulatoren, Panning-Einheiten sowie räumliche Effekte wie Reverb, Dekorrelation und randomisierte Modulation berücksichtigt werden. Die Parameterorganisation orientiert sich an der semantischen Gliederung der Klanggestaltung. Im folgenden Abschnitt werden schematische Signalablaufpläne vorgestellt.

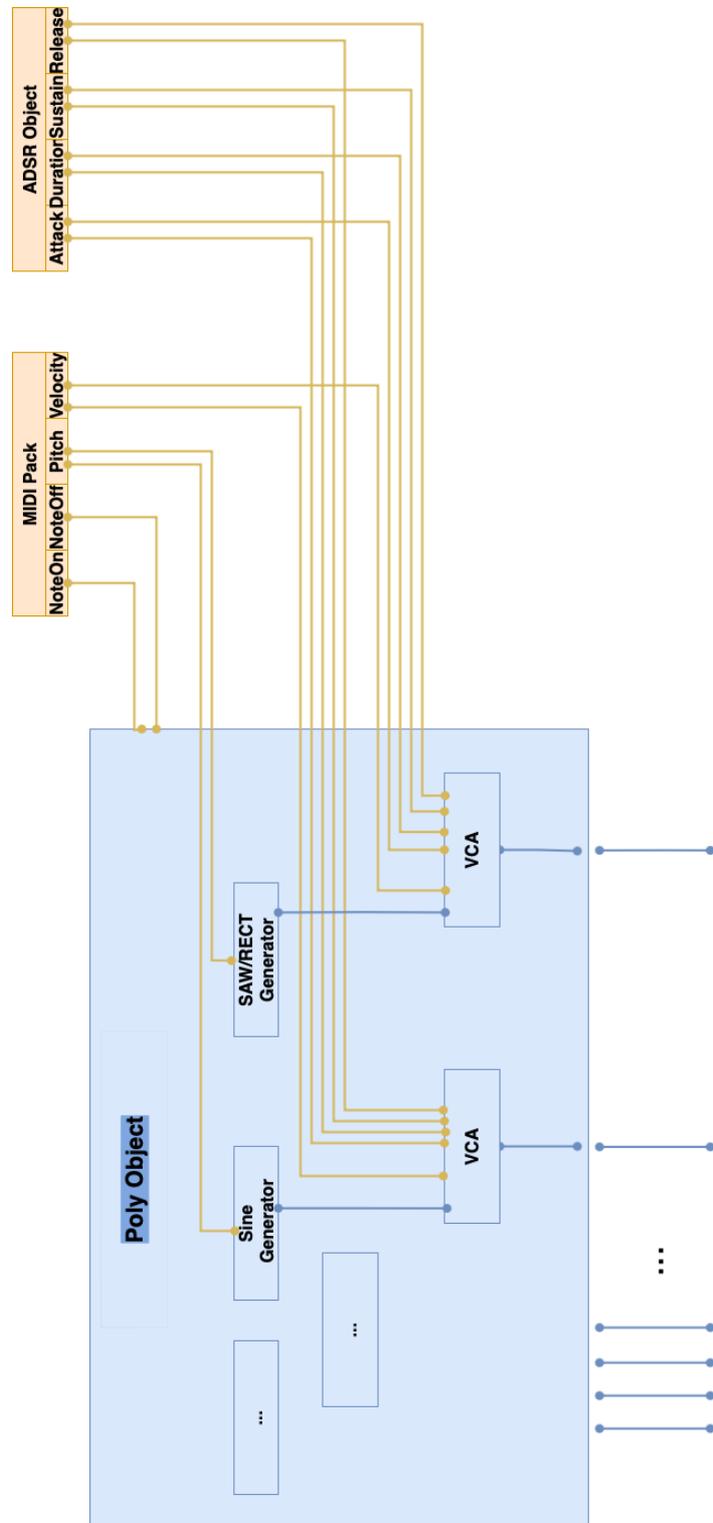


Abbildung 12: Steuer- und Audiosignalfluss der polyphonen Klangerzeugung

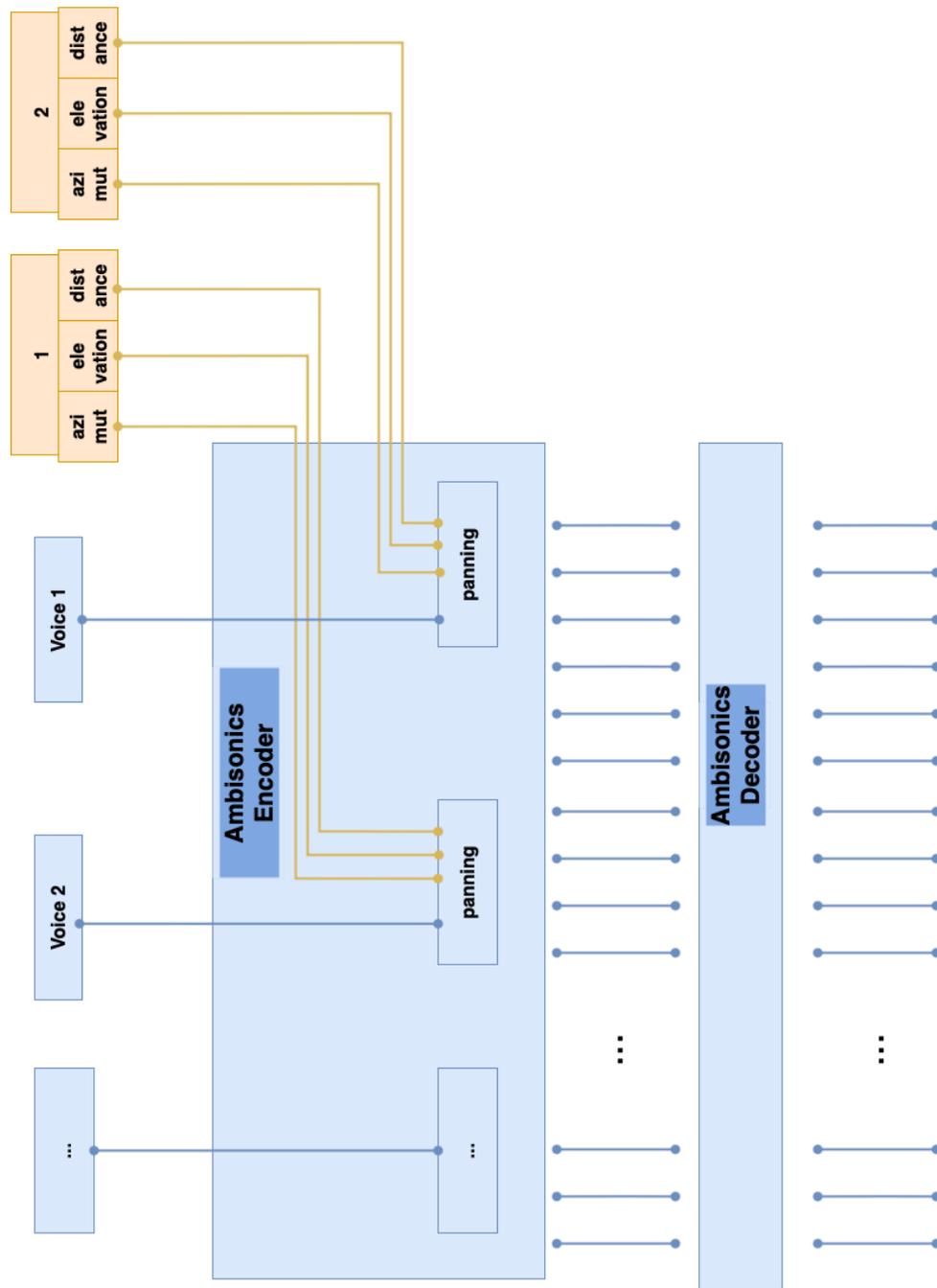


Abbildung 13: Steuer- und Audiosignalfluss des Pannings der Voices im Format Ambisonics

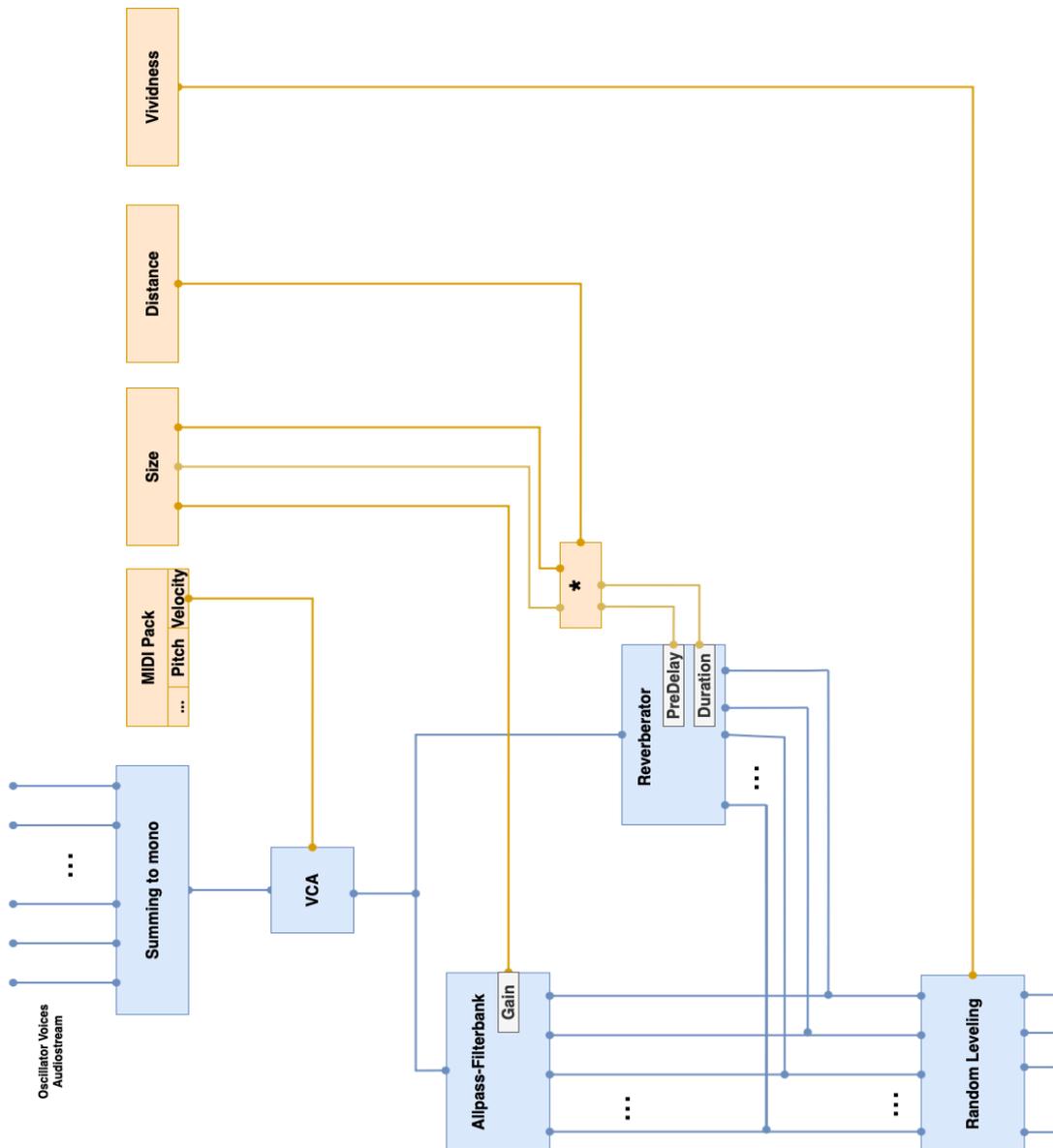


Abbildung 14: Steuer- und Audiosignalfluss der verwendeten Dekorrelationsmethoden

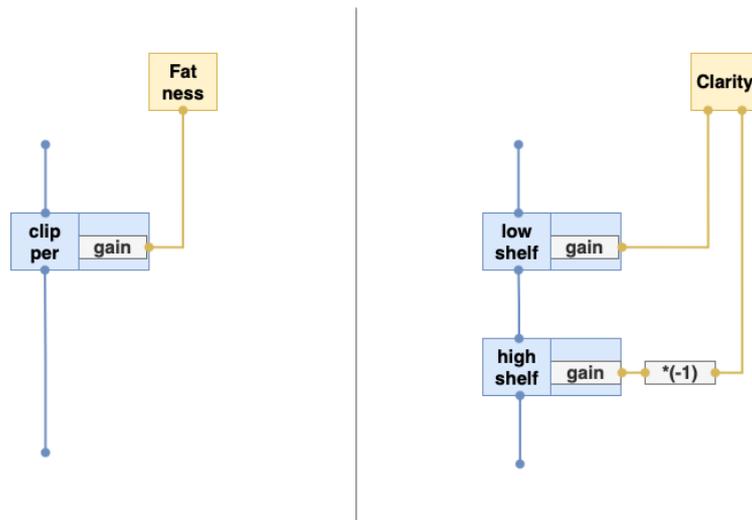
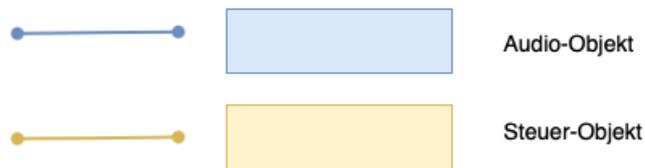


Abbildung 15: Steuer- und Audiosignalfluss für die klangformenden Instanzen Fatness und Clarity



3.3 Konzeption der Nutzeroberfläche

Der gestalterische Fokus liegt auf Übersichtlichkeit, durch die eine direkte Manipulation klangformender Prozesse einfach zu erreichen ist. Die grafische Gestaltung erfolgt in sich voneinander unterscheidenden Blöcken, welche die zuvor beschriebenen Parametergruppen aus [Kapitel 3.1](#) visuell und funktional gliedern. Diese Struktur folgt dem signaltechnischen Aufbau des Synthesizers und ermöglicht dem Nutzer ein assoziatives Arbeiten entlang der klanglichen Verarbeitungskette. Die grafische Nutzeroberfläche ist in drei funktionale Hauptbereiche unterteilt.

Im oberen Bereich der Oberfläche befindet sich die Steuerungseinheit für die Klangerzeugung. Die Auswahl des Grundklangs erfolgt über ein Drop-Down-Menü, das Presets für verschiedene Oszillator-Konfigurationen bereitstellt. Nachfolgend kann in der visuellen Nähe der Charakter über vier Regler verändert werden. Hier wird grafisch subtil ein Spektrum der summierten Voices dargestellt, um Unterschiede schnell wahrzunehmen. Fatness und Clarity bilden örtlich nah aneinander die nächste Einstellungseinheit. Die Steuerung der ADSR-Hüllkurve wird hierbei über dem Spektrum eingeblendet und ist über vier Slider einstellbar. Die beiden Klangcharakter-Regler Fatness und Clarity sind in der Nähe des Spektrums angeordnet und ermöglichen eine schnelle und gezielte Anpassung des Grundklangs. Als nächste visuelle Einheit werden zwei interaktive Koordinatensysteme dargestellt. Sie dienen der Kontrolle über Azimut und Elevation der einzelnen Stimmen. Durch Drag-and-Drop können einzelne Stimmen im Raum platziert werden.

Die räumliche Modulation wird über das Spatial Attack-Modul geregelt, das sich im Übergang zwischen klanglicher und immersiver Bereiche findet. Ein Schalter aktiviert die Modulation, während ein zusätzlicher Regler die Glättung des Steuersignals *velocity* für die Modulationsdaten beeinflusst. Eine Pegelanzeige bietet ein visuelles Feedback des Send-Level, wodurch sich die Verbindung von räumlicher Dynamik und der Spielweise des Musizierenden auch visuell nachvollziehen lässt. Der untere und präsenteste Bereich der grafischen Nutzeroberfläche umfasst die Parameter der räumlichen Klangumgebung: Size, Directness und Vividness. Die grafische Umsetzung folgt hier einem priorisierten Layout. Vividness ist am größten dargestellt und umfasst neben dem Regler auch ein Spektroskop, welches dekorrelierte Signale schematisch

darstellt und somit ein einfaches und intuitives visuelles Feedback gibt. Die Regler Size und Directness sind zwar im selben Bereich angebracht, sind jedoch visuell im Hintergrund. Sie dienen zur Feinjustierung.

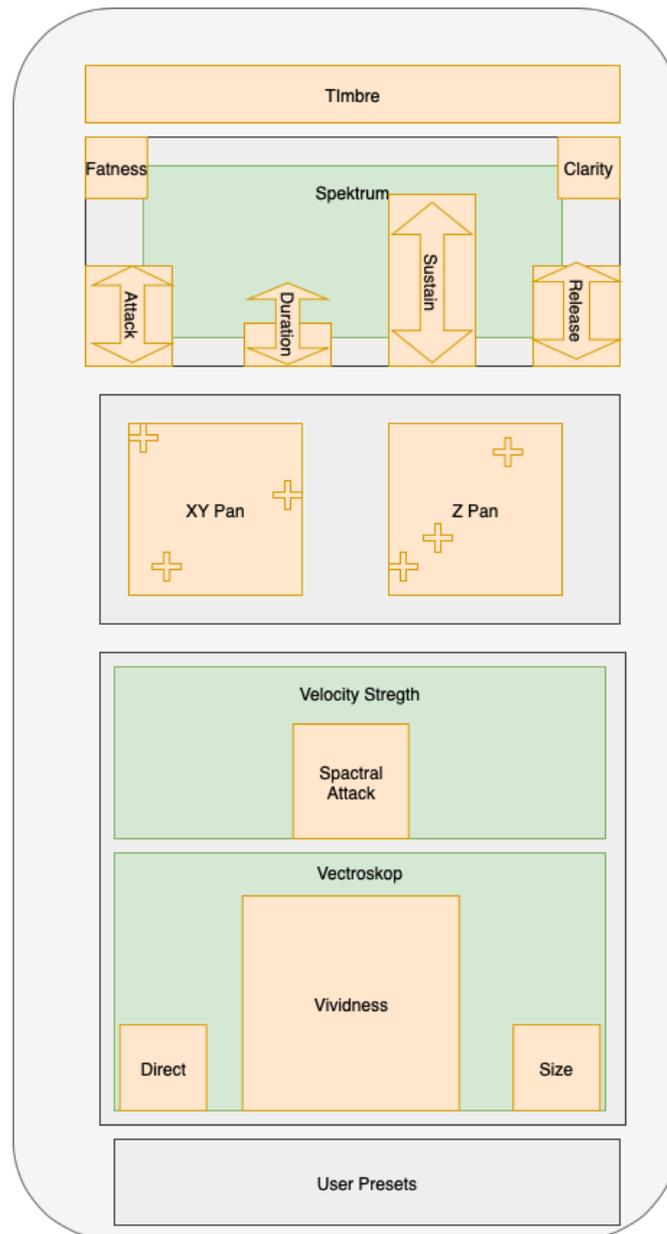


Abbildung 16: Schematische Darstellung der Benutzeroberfläche

4 Implementierung

Nach dem grundlegenden Steuer- und Audiosignalfluss und die Entwicklung der grafischen Nutzeroberfläche in Kapitel 3 abgeschlossen sind folgt in diesem Kapitel die praktische Umsetzung des Synthesizer-Prototyps in einer 3D-Audio-Umgebung. Ziel ist die Realisierung eines funktionalen Prototyps, der die in [Kapitel 3](#) beschriebenen Anforderungen erfüllt und das entworfene Interface mit einer konkreten Signalarchitektur verbindet. Die Implementierung erfolgt in der grafischen Programmierumgebung Max/MSP (*What is Max? | Cycling '74, 2025*), ergänzt um die Bibliothek *spat5* (*ForumMaxApps2-Spat5 | Ircam Forum, 2025*) zur Handhabung räumlicher Audioprojektionen. Max/MSP ist eine visuelle Programmierumgebung für die interaktive Audio- und Videoverarbeitung die von dem Unternehmen *Cycling '74* entwickelt wurde und sich in der Computermusik sowie der Medienkunst etabliert hat. Max/MSP unterscheidet sich fundamental von klassischen, textbasierten Programmiersprachen. Der Programmcode wird in Form von grafischen Objekten und Patchkabeln dargestellt. Diese visuelle Darstellung ermöglicht es komplexe Signalflüsse, Interaktionen und Parametersteuerungen intuitiv zu gestalten und im laufenden Betrieb zu manipulieren. Die Eignung von Max/MSP für die prototypische Entwicklung eines Synthesizers ergibt sich aus mehreren Aspekten: Max/MSP ermöglicht eine objektbasierte Strukturierung, bei der funktionale Einheiten wie Oszillatoren, Filter, Modulatoren oder Routinglogiken als Objekte implementiert und beliebig verschaltet werden können. Dies fördert sowohl die kreative Arbeit als auch die systematische Umsetzung komplexer Klangarchitekturen. Die grafische Oberfläche erlaubt es, Bedienelemente wie Slider oder Regler direkt in den Signalfluss zu integrieren. (Cipriani and Giri, 2010)

Um komplexere Setups zu realisieren, bedarf es daher einer zusätzlichen Infrastruktur zur Handhabung von virtuellen Schallquellen, Raumakustikmodellen und Codierungsalgorithmen. Hierzu wird in der Implementierung auf die frei verfügbare Erweiterung *spat5* gesetzt. Sie bietet eine umfassende Sammlung von Objekten zur räumlichen Audiowiedergabe. Im Rahmen dieser Arbeit kommt diese Erweiterung in mehreren Bereichen zum Einsatz:

1. Positionierung der Stimmen: Für jede Voice des Synthesizers kann eine eigene räumliche Position im Zwischenformat Ambisonics encodiert werden mithilfe des Objekts `spat5.hoa.encoder~`.
2. Räumliche Modulation: Die Steuersignale `direct` und `size` modifizieren die Objekte `spat5.early~`, `spat5.reverb~`, und `spat5.decorrelate~`.
3. Ausgabe auf Lautsprecher: Über `spat5.decoder~` wird das Zwischenformat Ambisonics 4. Ordnung auf ein fest definiertes Kanal-Layout decodiert.

Die Signalverarbeitung ist modular aufgebaut und hierarchisch gegliedert. Dies wahrt die Übersicht und bietet die Möglichkeit bestimmte Subpatches zu ersetzen oder im Verlauf der Entwicklung zu modifizieren.

4.1 Hierarchie und verwendete Objekte

Presentation

Das oberste Layer ist die grafische Nutzeroberfläche und ist dem Endnutzer zugänglich. Es beinhaltet zur Strukturierung der Oberfläche die skalierbaren Oberflächen `panel` und die Textfelder `text`. Zur visuellen Darstellung von Signalflüssen werden die Objekte `scope~` und `slider` verwendet. Diese stellen Modulationspegel, Korrelation und spektrale Verteilung der Klangsynthese visuell dar. Die Implementierung der visuellen Darstellungen wird in [Kapitel 4.2](#) dargestellt. Nutzereingaben der Oberfläche werden durch die Objekte `slider`, `umenu` und `dial` entgegengenommen. Mit dem Objekt `preset` können Einstellungen im Patch gespeichert werden, um personalisierte Zustände des Instruments zu speichern.

Subpatches

Im Hauptpatch sind alle Objekte aus dem Presentation Mode hinterlegt. Zusätzlich finden sich in dieser Ebene alle Subpatches `p midi`, `poly~ cycle`, `poly~ rect`, `poly~ saw`, `p timbreLeveling+Filter`, `p shelvingFilter`, `p ambiPanning`, `p allpassGenerator`, `p reverbGenerator`, `p randomAmplitude`, `p velocityMod`.

Die folgende Tabelle beschreibt diese Subpatches hinsichtlich ihrer Ein- und Ausgabewerte sowie der essenziell verwendeten Objekte.

Subpatch	Aufgabe
p midi	Nimmt ein externes MIDI-Signal eines Controllers oder einer Software an und rechnet es in die Steuersignale <i>Pitch -12</i> , <i>Pitch 0</i> , <i>Pitch +3</i> , <i>Pitch+12</i> um und leitet es weiter an einen Verteilerbus <i>midi</i> .
poly~ cycle	Eine Instanz von <i>cycle~</i> erstellt einen Sinus-Signal der entsprechenden Tonhöhe. Anschließend wird das Signal durch <i>adsr~</i> mit einer entsprechenden Hüllkurve versehen. Die Steuersignale werden von Verteilerbus von <i>midi</i> und dem Verteilerbus <i>adsr</i> empfangen. Ausgegeben wird ein Audiosignal in Form eines Sinus in Abhängigkeit von Frequenz und Hüllkurve.
poly~ rect	Eine Instanz von <i>rect~</i> erstellt einen Rechteck-Signal der entsprechenden Tonhöhe. (siehe <i>poly~ cycle</i>)
poly~ saw	Eine Instanz von <i>saw~</i> erstellt einen Sägezahn-Signal der entsprechenden Tonhöhe. (siehe <i>poly~ cycle</i>)
p timbre Leveling+ Filter	Dieser Patch nimmt die einzelnen Voices der <i>poly~</i> Objekte entgegen. Es gibt die Möglichkeit mithilfe von resonierenden Hoch- und Tiefpassfiltern und VCAs die einzelnen Voices klanglich zu formen. Es nimmt vier Audiostreams entgegen und gibt vier Audiostreams aus.
p shelving Filter	Nach der allgemeinen Formung des Klangs können hier mithilfe von <i>biquad~</i> die Obertonstruktur angehoben oder abgesenkt werden. Zusätzlich kann mit dem Objekt <i>overdrive~</i> ein Signal angezerrt werden und eine größere Saturation erstellt werden. Neben den Audiostreams nimmt dieses Objekt die Steuerdaten <i>fatness</i> und <i>clarity</i> entgegen. Ausgegeben werden wieder die vier Audiostreams.

<p>p ambi Panning</p>	<p>Das Objekt nimmt vier Audiostreams entgegen und encodiert diese mit dem Objekt <code>spat5.hoa.decoder~</code> in das Ambisonics Format 4. Ordnung. Hierzu nimmt das Objekt OSC-Steuerdaten der Steueroberfläche <code>spat5.viewer</code> entgegen.</p>
<p>p allpass Generator</p>	<p>Das Objekt nimmt einen Audiokanal <i>summedVoices</i> und ein Steuersignal <i>size</i> entgegen. Das Steuersignal steuert die Amplitude des Audiostreams. Mithilfe des Objekts <code>spat5.decorrelate~ @channels 25</code> wird die Phasenlage des Eingangssignals manipuliert und ein 25-kanaliger Audiostream <i>mc.allpassed</i> mit dekorrelierten Versionen des Eingangssignals ausgegeben.</p>
<p>p reverb Generator</p>	<p>Das Objekt nimmt einen Audiokanal <i>summedVoices</i> entgegen und die zwei Steuersignale <i>size</i> und <i>direct</i> entgegen. <i>size</i> hat Einfluss auf die Nachhallzeit und <i>direct</i> hat Einfluss auf die Pre-Delay Zeit der Reverbengine <code>spat5.reverb~ @channels 25</code>. Ausgegeben wird ein 25-kanaliger Audiostream <i>mc.reverbed</i>.</p>
<p>p random Amplitude</p>	<p>Das Objekt nimmt die beiden Audiostreams <i>mc.reverbed</i> und <i>mc.allpassed</i> entgegen. Diese werden kanalweise summiert und ein neuer 25-kanaliger Audiostream entsteht. Dieser Stream wird kanalweise durch <code>randomMultiply</code> moduliert. Das Steuersignal <i>vividness</i> nimmt Einfluss auf dieses Objekt. Es werden die Häufigkeit der Zufallswerte und deren Dämpfung auf die Audiostreams beeinflusst. Ausgegeben wird ein 25-kanaliger Audiostream <i>mc.decorrelated</i>.</p>
<p>p velocityMod</p>	<p>Der Subpatch nimmt <i>midi</i> entgegen und extrahiert den Parameter <i>velocity</i>. Dieses Steuersignal wird mithilfe des Objects <code>line 20. 200.</code> geglättet und als Modulationssignal für einen VCA verwendet. Dieser steuert kanalweise den Pegel von <i>summedVoices</i>.</p>

Tabelle 10: Verwendete Subpatches

4.2 Visualisierungen

Die Berechnung der Visualisierungen findet im Hauptpatch statt. Zur Darstellung der Korrelation wird das Objekt `scope~` verwendet. An dessen beiden Eingängen werden zwei Kanäle des Audiostreams `mc.decorrelated` übergeben. Somit lässt sich eine Differenz der beiden Signale grafisch darstellen. Um die Visualisierung noch intuitiver zu gestalten, werden beide Eingangssignale mit dem Wert `velocity` multipliziert. Dadurch werden die Visualisierungen bei steigender Anschlagdynamik intensiver. Um die Darstellung weniger sprunghaft zu gestalten, werden zusätzlich die Werte von `velocity` mit dem Objekt `rampsmooth~ 700 5000` geglättet. Dadurch lässt sich eine kontinuierlichere Darstellung erzeugen.

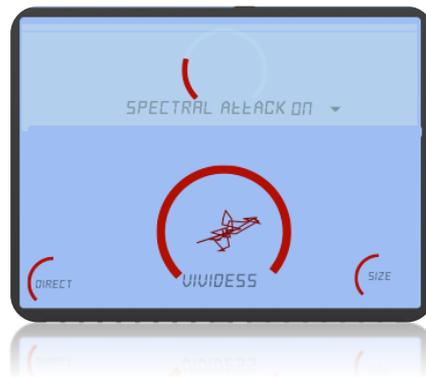


Abbildung 17: Grafische Nutzeroberfläche – Vektorskop und Modulationskurve



Abbildung 18: Grafische Nutzeroberfläche - Spektrum der Klangsynthese

Das Vektorskop wird mit dem Objekt `scope~` realisiert. Als Eingangssignal dient die nicht-manipulierte Version des Signals. Auch dieses wird zur angenehmeren Darstellung durch das Objekt `rampsmooth~ 700 5000` geglättet.

Die gesamte Nutzeroberfläche wurde nach dem Konzept in [Kapitel 3.3](#) konzipiert. Das interaktive Koordinatensystem `spat5.viewer` konnte nicht in die Nutzeroberfläche integriert werden und muss durch einen Doppelklick in einem Popup-Fenster geöffnet werden.

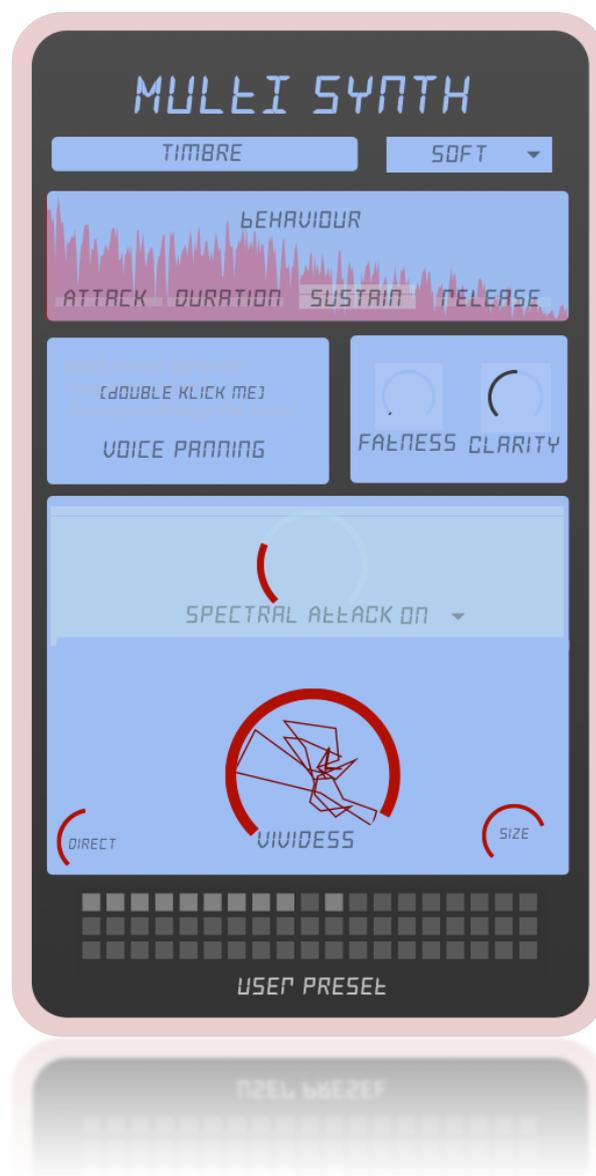


Abbildung 19: Die grafische Nutzeroberfläche

4 Implementierung

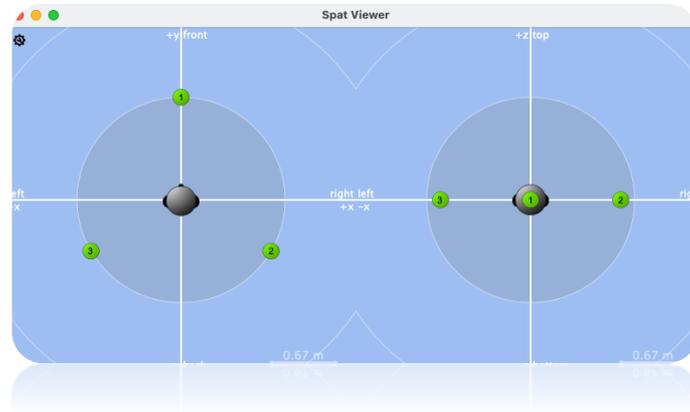


Abbildung 20: Nutzeroberfläche des Popup-Fensters spat5.viewer

5 Analyse

Der entwickelte Prototyp eines kanalbasierten Synthesizers wurde hinsichtlich seiner funktionalen Eigenschaften überprüft. Ziel ist es, die technische Umsetzung der im [Kapitel 3](#) formulierten Konzepte zu überprüfen. Das System ist in der Lage, die angestrebten Funktionen konsistent und zuverlässig auszuführen. Zustände sind reproduzierbar. Es zeigt sich, dass alle relevanten Steuersignale korrekt übertragen und verarbeitet werden. Auch die in spat5 integrierten Objekte zur Dekorrelation und Hallgenerierung wurden erfolgreich mit individuellen Steuersignalen angesteuert. Besondere Herausforderungen ergaben sich bei der Modulation einzelner Audiokanäle über das Vividness-Modul. Die Zufallsmodulation pro Kanal musste mithilfe von mathematischen Funktionen bearbeitet werden, um ungewollte Artefakte zu vermeiden. Der Einsatz von Ambisonics als Zwischenformat ermöglicht eine flexible Steuerung der Raumposition einzelner Klangquellen. Hauptaspekte der Signalverarbeitung sind visuell dargestellt, damit sie unmittelbar und erfahrbar sind. Auch die Darstellung räumlicher Bewegung durch dynamische Vektorskope dient dem besseren Klangverständnis. Die Steuerung über ein virtuelles Koordinatensystem ermöglicht eine präzise Platzierung von Klangereignissen in der auditiven Szene.

Ein wesentlicher Aspekt der Konzeption ist das Design der Nutzerparameter. Ziel ist es, eine Steuerlogik zu implementieren, die sich weniger an technischen Parametern als an wahrnehmungsbezogenen Kategorien orientiert. Durch das Konzept der semantischen Differenziale werden diese Nutzerparameter einem technisch unversierten Publikum begreifbar. Ein zentrales Anliegen dieser Arbeit ist die Übertragbarkeit der entwickelten Steuerkonzepte in die künstlerische Praxis. Die Parameter Size, Directness und Vividness bieten direkten Zugang zu rumorientierter Signalverarbeitung, die in klassischen Synthesizer-Systemen nur mit erheblichem technischem Aufwand zu realisieren sind. Die gezielte Dekorrelation einzelner Audiosignale durch Allpassfilter, Reverberation und randomisierten Pegelsprüngen ermöglicht eine kontrollierbare Umhüllung. Musikalisch ergeben sich dadurch neuartige Ausdrucksformen, etwa durch die Kombination perkussiver Spatial-Attack-Zeiten mit weiträumigen Reverb-Feldern oder die durch anschlagsdynamisch abhängige Ausdehnung. Die Möglichkeit,

alle Prozesse live zu steuern, eröffnet neue Wege für performative Musikdarbietungen im immersiven Kontext.

Die Konzeption eines kanalbasierten Synthesizers für eine dreidimensionale Audio-Umgebung stellt besondere Herausforderungen an die Gestaltung und Organisation von Steuer- und Audiosignalflüssen. Während klassische Synthesizer auf eine begrenzte Anzahl an Parametern zurückgreifen, die in der Regel klanglich-technisch kategorisiert sind, verfolgt die vorliegende Arbeit einen anderen Ansatz: Die Steuerparameter sollen nicht in erster Linie technische, sondern psychoakustisch wirksame Kategorien manipulieren. Dies bedeutet, dass semantische Begriffe wie *Directness*, *Vividness*, *Size* oder *Clarity* konkrete technische Signalverarbeitungsketten steuern. Hierbei besteht eine zentrale Herausforderung in der Abbildung dieser semantischen Konzepte in technische Prozesse. Während beispielsweise eine Änderung des Parameters *Timbre* durch den Wechsel von Wellenformen oder Filtern relativ eindeutig technisch übersetzt werden kann, erfordern Steuersignale wie *vividness* eine Kombination mehrerer Signalverarbeitungsschritte: es müssen Zufallssignale generiert werden, geglättet und kanalweise auf Gain-Werte angewendet werden, um eine lebendige, räumlich diffuse Bewegung zu erzeugen, zusätzlich werden durch dieses Steuersignal die Häufigkeit der Änderung der Zufallszahlen und deren Gewichtung gesteuert. Die Steuerung dieser Prozesse muss dabei stets intuitiv innerhalb der GUI dargestellt werden. Abstrahierte Parameter können zwar den kreativen Umgang fördern, da sie schnell und einfach zu erreichen sind und eine große Wirkung entfalten. Tiefe Eingriffe in klangliche Strukturen werden dadurch jedoch verhindert. Hierbei gilt es einen guten Mittelweg zwischen technischer Abbildungstiefe und gestalterischer Abstraktion zu beschreiten.

Die Realisierung der des Prototyos erfolgt in der grafischen Entwicklungsumgebung Max/MSP. Diese bietet einerseits die notwendige Modularität, um die vielen parallelen Signalflüsse zu strukturieren, andererseits aber auch Herausforderungen bei der Performance-Optimierung. Der modulare Aufbau führt bei steigender Stimmenzahl rasch zu einer hohen Komplexität: Jede Voice ist als separate Instanz eines umschließenden Objekts implementiert, was zu einem exponentiellen Anstieg der Komplexität führt wenn mehrere Stimmen und Kanäle verarbeitet werden. Durch die begrenzte Rechenleistung eines Hosts kann diese Komplexität begrenzt sein.

Die Implementierung spatialer Signalverarbeitung basierend auf der Erweiterung *spat5* erweist sich als geeignetes Werkzeug. Das Format Ambisonics erweist sich als geeignetes Zwischenformat, um die Voices der Oszillatoren im Raum zu positionieren.

Da jede Voice einzeln encodiert und moduliert werden kann, müssen Steuerdaten präzise adressiert und synchronisiert werden. Hierbei hat sich MAX/MSP als intuitive Programmierumgebung erwiesen. Des Weiteren ist zu beachten, dass bestimmte räumliche Effekte stark vom verwendeten Lautsprecher-setup abhängig sind. Eine genaue Kalibrierung und Einmessung des Abhör-raums sind essenziell, um die Höreindrücke tatsächlich wahrnehmbar zu machen. Vor allem bei ungleichen Lautsprecherabständen oder asymmetrischen Positionen entstehen Verzerrungen in der Wahrnehmung, die dem kreativen Ziel entgegenwirken können. Dieser Aspekt wurde in dieser Arbeit nicht berücksichtigt und ist zusätzlich zu beachten.

Zusammenfassend lässt sich feststellen, dass die Gestaltung eines mehrkanaligen, psychoakustisch inspirierten Synthesizers nicht nur eine kreative Aufgabe, sondern auch eine technische Herausforderung darstellt. Insbesondere das Spannungsfeld zwischen gestalterischer Semantik und technischer Implementierung erfordert eine präzise Konzeption der Signalflüsse. Die grafische Programmierweise von Max/MSP erweist sich dabei als hilfreich.

6 Diskussion und Ausblick

Die Entwicklung eines kanalbasierten Synthesizers für eine 3D-Audio-Umgebung stellt nicht nur eine technische und gestalterische Herausforderung dar, sondern wirft zugleich grundlegende Fragen zur zukünftigen Praxis elektroakustischer Musikproduktion auf. Die Einordnung des vorliegenden Prototyps erfolgt daher im Spannungsfeld zwischen etablierten Stereoproduktionsmethoden und dem musikalischen Wunsch, Werkzeuge für immersive Klanggestaltung zu schaffen, die intuitiv bedienbar sind. In der gegenwärtigen Produktionspraxis dominieren nach wie vor zweikanalige Synthesizer-Instrumente, die innerhalb digitaler Audio-Workstations wie Ableton Live, Logic Pro oder Reaper genutzt werden. Die Erweiterung dieser Instrumente auf mehrkanalige Umgebungen erfolgt im produzierenden Alltag durch aufwendiges Routing: eine einzelne Klangquelle wird dupliziert, auf verschiedene Spuren verteilt und anschließend über externe Panner-Plug-ins räumlich positioniert. Die Kontrolle dieser Manipulation geschieht dabei nicht über das Synthesizer-Interface selbst, sondern über externe Parameter außerhalb des eigentlichen Instruments in einer Host-Umgebung. Dieses Verfahren ist technisch möglich, jedoch wenig kreativitätsfördernd und vor allem bei Automationen aufwendig und fehleranfällig. Spontane musikalische Entscheidungen werden dadurch ausgebremst. Der hier entwickelte Prototyp verfolgt einen grundlegend anderen Ansatz. Er integriert die räumliche Steuerung von Klangquellen direkt in die Klangerzeugung. Dadurch wird die Raumdimension nicht als nachträglicher Effekt oder als separate Mischinstanz behandelt, sondern als integraler Bestandteil des Klangdesigns verstanden. Die musikalische Steuerung räumlicher Parameter erfolgt dabei über semantisch abstrahierte Bedienelemente, die in Abhängigkeit der Anschlagsdynamik räumlich wirksame Prozesse beeinflussen. Dadurch wird eine neue Schnittstelle zwischen Komposition und Raumgestaltung geschaffen. Der kreative Zugang zu Umhüllung und Raumgröße als musikalisch beeinflussbare Parameter eröffnet neue gestalterische Perspektiven, etwa für elektroakustische Kompositionen oder immersive Installationen. Ein zentrales Anliegen der vorliegenden Arbeit besteht darin, psychoakustische Erkenntnisse nicht nur zu beschreiben, sondern sie aktiv in die Gestaltung des Prototyps zu überführen. Die im zweiten Kapitel dargestellten Modelle der auditiven Wahrnehmung – insbesondere hinsichtlich der räumlichen Ortung

(Kapitel 2.2.1), der Umhüllung durch dekorrelierte Signale (Kapitel 2.2.2) sowie der subjektiven Beschreibung von Höreindrücken (Kapitel 2.2.3) geben dem Instrument eine wahrnehmungsorientierte Färbung. Besonders deutlich wird dies bei den Steuergrößen *Orientation*, *Vividness* und *Directness*, die direkte Anknüpfungspunkte an auditive Lokalisation und räumliche Tiefe darstellen. Während herkömmliche Synthesizer diese Kategorien entweder gar nicht oder nur über externe Routinglösungen abbilden können, ermöglicht der entwickelte Prototyp deren direkte Gestaltung innerhalb der Klangquelle. Räumlichkeit ist nicht länger ein nachgelagerter Effekt, sondern ein integrativer Bestandteil der Klangbildung selbst. Dieses Konzept findet sich in Theorien der immersiven Mediengestaltung wieder. Ein Raum ist nicht nur das erforderliche Medium für Klang, sondern kann auch als aktives Ausdrucksmittel genutzt werden. (Rumsey, 2013) Eine besondere Stärke des entwickelten Synthesizers liegt darin, dass er psychoakustische Effekte nicht als fixe Algorithmen implementiert, sondern als interaktive Steuerachsen zugänglich macht. Der Parameter *Vividness* etwa generiert keine vordefinierte Bewegungskurve, sondern moduliert die Gewichtung diffuser Signalpfade in Abhängigkeit von zufallsgesteuerten VCAs – ein Verfahren, das weniger eine technisch präzise Raumbewegung erzeugt, als vielmehr eine lebendige, sich ständig verändernde Klangszene erzeugt. Damit wird dem Anspruch der Arbeit Rechnung getragen, psychoakustisches Wissen in der Interaktion mit dem Instrument erfahrbar zu machen. Zusammenfassend lässt sich feststellen, dass das psychoakustische Fundament eine musikalisch sinnvolle Abstraktion technischer Steuerprozesse abbildet.

Die Entwicklung des vorliegenden Prototyps zielt nicht auf eine technische Machbarkeitsstudie ab, sondern verfolgt ausdrücklich eine künstlerische Perspektive. Der Synthesizer soll als gestalterisches Werkzeug dienen, das Musiker:innen eine kreative, psychoakustisch fundierte Interaktion mit Klang im Raum ermöglicht. Diese Zielsetzung wirft die Frage nach dem musikalischen Potenzial und den ästhetischen Grenzen des entwickelten Instruments auf.

Zunächst ist festzuhalten, dass der Synthesizer nicht als ein universelles Allweckinstrument konzipiert wurde, sondern als spezialisiertes Werkzeug zur Erzeugung räumlich modulierter Klangtexturen. Bereits in der Auswahl und Verknüpfung der Steuerparameter wird diese ästhetische Ausrichtung sichtbar. Unterstützt durch eine polyphone Oszillatorarchitektur und Filterung können dabei

dichte Klangfelder erzeugt werden, deren Spektrum sich im Raum ausbreitet und über Zeit moduliert wird. Diese Eigenschaften prädestinieren das Instrument für Einsatzgebiete, in denen klangliche Flächen, Texturen, spektrale Komplexität und räumliche Tiefe zentrale Gestaltungsmittel darstellen.

Während *Size* für eine kanalweise Dekorrelation der Signalanteile sorgt und damit die Quelle vom Punkt- zum Flächenklang transformiert, bewirkt *sividness* eine kontinuierliche, zufallsmodulierte Veränderung der Pegelverhältnisse über die Lautsprecherkanäle und schafft Lebendigkeit. Diese Kombination erzeugt ein Hörempfinden, das nicht mehr auf statische Ortung abzielt, sondern auf Bewegung, Lebendigkeit und räumliches Chaos. In Verbindung mit allpassgefilterten Signalkomponenten und Reverbs entsteht eine Wolke, die sich der präzisen Lokalisierbarkeit entzieht und vielmehr das Raumerleben selbst in den Vordergrund stellt. Die klangliche Qualität der von den Oszillatoren erzeugten Voices lässt sich nicht mit allgemeinen Synthesizern oder kommerziellen Plugins vergleichen und ist nicht der Schwerpunkt dieser Arbeit. Hierzu können in Zukunft feinere und spektral spannendere Syntheseverfahren wie eine Wavetable-Synthese implementiert werden.

Der Prototyp entfaltet sein volles Potenzial primär in dichten, langsamen, ambientartigen Kontexten. Rasche rhythmische Prozesse oder perkussive Impulse lassen sich durch die Hüllkurvensteuerung zwar umsetzen, verlieren jedoch im Zusammenspiel mit der räumlichen Modulation schnell an Klarheit und Durchsetzungskraft. Dies liegt an Modulationsprozessen wie Allpass-Filterung oder Amplitudenmodulation durch Zufallswerte. Auch das Zusammenspiel mehrerer simultaner Stimmen kann – insbesondere bei hohem *size*-Wert – zu einer auditiven Überlagerung führen, die die Differenzierung einzelner Voices erschwert. Ein limitierender Faktor ist die Klangdichte in Echtzeitperformance. Bei bis zu zwanzig Instanzen pro Oszillator und vielfacher Modulation können je nach verwendetem Computersystem Audioaussetzer entstehen. Diese können zwar durch die Buffergröße der Digital-Analog-Wandlung kompensiert werden. Dadurch entstehen jedoch Latenzen, welche einer Live-Performance im Weg stehen.

Trotz dieser Einschränkungen eröffnet der Prototyp ein neues und spannendes Feld klanglicher Möglichkeiten, das durch konventionelle Produktionsmittel nur schwer erschlossen werden kann. Er bietet insbesondere für experimentelle Komponist:innen und Klangkünstler:innen ein kreatives Werkzeug. Die Möglichkeit, räumliche

Bewegung, Umhüllung und Tiefe nicht nur als statisches Panning, sondern als dynamisch gestaltbare Klangdimension zu nutzen, eröffnet neue Ausdrucksformen für Live-Performer:innen. Die visuelle Rückmeldung in der Nutzeroberfläche sowie die strukturierte Zuweisung semantischer Steuerachsen erleichtern den kreativ performativen Zugriff auch in komplexen und unübersichtlichen Live-Performance Szenarien. Die Möglichkeit, den Synthesizer durch standardisierte Steuersignale zu kontrollieren, erlaubt eine einfache Integration in bestehende Systeme – etwa in Kombination mit Bewegungstracking, Kamerasystemen, Touch-Interfaces oder generativen Steuerdaten über das Open Sound Protocol. So kann etwa ein Ausstellungsbesucher durch Position im Raum die Ausdehnung eines Klangereignisses modulieren. Die vorliegende Arbeit stellt einen kleinen Schritt zur Entwicklung eines Synthesizers für die 3D-Audio-Produktion dar. Mit dem entworfenen Prototyp wurde gezeigt, dass es möglich ist, komplexe räumliche Klangereignisse direkt im Prozess der Klangsynthese zu gestalten und dabei Parameter der auditiven Wahrnehmung in eine intuitive Parametrisierung zu übertragen.

Eine mögliche Weiterentwicklung besteht in der Implementierung des Prototyps als VST/AU-Plug-ins oder als Standalone-Anwendung in einer verbreiteten Programmiersprache wie Python oder C++. Denkbar wäre auch die Integration zusätzlicher psychoakustischer Phänomene, wie etwa weiteren Aspekten der Auditiven Szenenanalyse, welche auf der kognitiven Trennung und Gruppierung auditiver Ereignisse beruht. Der Einsatz dynamischer Maskierungseffekte oder die Beeinflussung von Raumgefühl durch gezielte Manipulation von Modulationsverläufen könnte stärker in den Fokus rücken. Darüber hinaus wären Nutzerstudien denkbar, die untersuchen, inwieweit Musiker:innen mit unterschiedlichem Erfahrungsstand die parametrisierte Steuerung als intuitiv, kreativ oder kontrollierbar empfinden. Besonders im Vergleich zu klassischen Synthesizern könnten qualitative Interviews und empirische Tests neue Erkenntnisse darüber liefern, wie sinnvoll eine semantische Steuerungslogik in kreativen Produktionsprozessen tatsächlich ist. Auch bei der Interaktivität bestehen Erweiterungspotenziale: Eine berührungsbasierte Steuerung über Touchscreens, Bewegungssensoren könnte in einer erweiterten Version den künstlerischen Ausdruck und die performative Kontrolle des Instruments weiter verbessern. Besonders im Kontext von Live-Aufführungen oder interaktiven Installationen eröffnet dies spannende Möglichkeiten.

Abschließend lässt sich festhalten, dass der Prototyp eine eigenständige Position innerhalb aktueller Produktionspraktiken einnimmt. Er verknüpft etablierte Synthesetechniken mit aktueller 3D-Audiotechnologie und überträgt zentrale Erkenntnisse der Psychoakustik in semantisch orientierte Steuerlogik. Damit bietet er eine funktionale Alternative zu bestehenden Workflows und setzt zugleich einen Impuls für die Weiterentwicklung musikalischer Werkzeuge im Bereich immersiver Medienproduktion. Ein nächster Schritt könnten Nutzerstudien im Bereich der klanglich ästhetischen und psychoakustisch Wahrnehmung sein.

A Literaturverzeichnis

2025 (no date) *VCV Rack Manual*. Available at: <https://vcvrack.com/manual/> (Accessed: 16 July 2025).

AvaTekh | Technology | Tutorials & demos | NDLs vs. linear filters: an illustration (no date). Available at: https://avatekh.com/tutorials/simple_illustration.html (Accessed: 15 July 2025).

Barbour, J.L. (2003) 'ELEVATION PERCEPTION: PHANTOM IMAGES IN THE VERTICAL HEMI-SPHERE', *AES: A Journal of the Audio Engineering Society* [Preprint].

Bartetzki, A. (2022) 'Klangsynthese', in *Handbuch der Audiotechnik*. Springer Berlin Heidelberg, pp. 1–30. Available at: https://doi.org/10.1007/978-3-662-60357-4_9-1.

Blauert, J. and Lindemann, W. (1986) 'Spatial mapping of intracranial auditory events for various degrees of interaural coherence', *The Journal of the Acoustical Society of America*, 79(3), pp. 806–813. Available at: <https://doi.org/10.1121/1.393471>.

Bregman Albert S. (1990) 'Chapter 8: Summary and Conclusions: What We Do and Do Not Know about Auditory Scene Analysis', *Audit*, pp. 641–705. Available at: <http://books.google.ca/books?id=jI8muSpAC5AC&printsec=frontcover&dq=Bregman+%22Auditory+scene+analysis%22#PPR9,M1> (Accessed: 16 July 2025).

Bregman, A.S. (1987) 'Auditory Stream Segregation and the Control of Dissonance in Polyphonic Music.', *Contemporary Music Review*, 2(1), pp. 63–92. Available at: <https://doi.org/10.1080/07494468708567054>.

Bristow-Johnson, R. (1999) *Wavetable Synthesis 101, A Fundamental Perspective*. Available at: <https://www.researchgate.net/publication/228992574>.

Butler, R.A. and Humanski, R.A. (1992) 'Localization of sound in the vertical plane with and without high-frequency spectral cues', *Perception & Psychophysics*, pp. 182–186.

Cipriani, A. and Giri, M. (2010) *This is a demo copy of ELECTRONIC MUSIC AND SOUND DESIGN Theory and Practice with Max 8-volume 1*. Available at: www.contemponet.comwww.virtual-sound.com.

Cipriani, A., Giri, M. and Dudas Richard (2020) *Electronic music and sound design : theory and practice with Max 8. Volume 2*.

Deutsches Museum: Moog (2025). Available at: <https://web.archive.org/web/20191230182608/https://www.deutsches-museum.de/presse/presse-2019/moog/> (Accessed: 16 July 2025).

documentation — Pd Community Site (2025). Available at: <https://puredata.info/docs> (Accessed: 16 July 2025).

‘Dolby ® Atmos TM Next-Generation Audio for Cinema Overview’ (2025).

Fastl, H. and Zwicker, E. (2007) ‘Psychoacoustics: Facts and models’, *Psychoacoustics: Facts and Models* [Preprint]. Available at: <https://doi.org/10.1007/978-3-540-68888-4/COVER>.

Force Technology (2025). Available at: <https://forcetechnology.com/en/services/acoustics-noise-sound-quality/senselab-sound-wheel/download-version-sound-wheel-high-resolution> (Accessed: 9 July 2025).

ForumMaxApps2-Spat5 | Ircam Forum (2025). Available at: <https://forum.ircam.fr/projects/detail/forummaxapps2-spat5/> (Accessed: 16 July 2025).

Friesecke, A. (2014) *Die Audio-Enzyklopädie*.

Frisius, R.. (1996) ‘Karlheinz Stockhausen : Einführung in das Gesamtwerk’.

Gabrielsson, A. and Sjögren, H. (1979) 'Perceived sound quality of sound-reproducing systems', *The Journal of the Acoustical Society of America*, 65(4), pp. 1019–1033.

Available at: <https://doi.org/10.1121/1.382579>.

Howard, D.M. (David M. and Angus, Jamie. (2009) *Acoustics and psychoacoustics*. Focal.

IEM Plug-in Suite (2023). Available at: <https://plugins.iem.at/> (Accessed: 22 August 2023).

ITU-R BS.2399-0 (2017) *Methods for selecting and describing attributes and terms*.

Available at: <http://www.itu.int/ITU-R/go/patents/en>.

Jens Blauert (1997) *Räumliches Hören*.

Kurtz, M. (1988) *Stockhausen - eine Biographie*. Kassel/Basel: Bärenreiter.

Manufacturers Assoc, M. (2020a) *Universal MIDI Packet (UMP) Format and MIDI 2.0 Protocol Association of Musical Electronics Industry AMEI and MIDI Manufacturers Association MMA M2-104-UM*. Available at: <http://www.midi.org>.

Manufacturers Assoc, M. (2020b) *Universal MIDI Packet (UMP) Format and MIDI 2.0 Protocol Association of Musical Electronics Industry AMEI and MIDI Manufacturers Association MMA M2-104-UM*. Available at: <http://www.midi.org>.

Moog Synthesizer 1c/2c/3c | Vintage Synth Explorer (2025). Available at: <https://www.vintagesynth.com/moog/synthesizer-1c2c3c> (Accessed: 16 July 2025).

Puckette, M. (2007) *The theory and techniques of electronic music, The Theory and Techniques of Electronic Music*. World Scientific Publishing Co. Available at: <https://doi.org/10.1142/6277>.

Pulkki, V. (2000) *Generic panning tools for MAXXMSP*. Available at: www.acoustics.hut.fi/villee.

Radiocommunication Bureau, I. (2022) *Advanced sound system for programme production BS Series Broadcasting service (sound)*. Available at: <http://www.itu.int/ITU-R/go/patents/en>.

Reaktor 6 – digital modular synthesizer | Komplete (2025). Available at: <https://www.native-instruments.com/de/products/komplete/synths/reaktor-6/?srsltid=AfmBOoqhJErIsmNBhkup6RtTrejo9X73-e9-ZKMhvfFbhoulJ7BEO-2f2> (Accessed: 16 July 2025).

Stange-Elbe, J. (2015) *Computer und Musik*. DE GRUYTER. Available at: <https://doi.org/10.1515/9783486720198>.

Stockhausen, K. and Barkin, E. (1962) *The Concept of Unity in Electronic Music*, *Source: Perspectives of New Music*.

Strawn, J. *et al.* (1996) 'The Computer Music Tutorial', *The MIT Press* [Preprint].

Taylor, T.D. (2001) *Strange Sounds: Music, Technology and Culture*. New York, London: Routledge.

Uncini, A. (2022) *Springer Topics in Signal Processing Digital Audio Processing Fundamentals, Volume 21*.

Verron, C. *et al.* (2009) 'Controlling a spatialized environmental sound synthesizer', *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 321–324. Available at: <https://doi.org/10.1109/ASPAA.2009.5346504>.

Verron, C., Aramaki, M. and Kronland-Martinet, R. (2010) *A spatialized additive synthesizer*. Available at: <http://marcs.uws.edu.au/links/ICoMusic>.

Weinzierl, S. (2008) *Handbuch der Audiotechnik, Handbuch der Audiotechnik*.

Springer Berlin Heidelberg. Available at: <https://doi.org/10.1007/978-3-540-34301-1>.

What is Max? | Cycling '74 (2025). Available at: <https://cycling74.com/products/max> (Accessed: 29 August 2023).

Wittek, H., Haut, C. and Keinath, D. (2006) 'Doppel-MS-eine Surround-Aufnahmetechnik unter der Lupe Double M/S-a Surround recording technique put to test'. Available at: www.schoeps.de. (Accessed: 28 August 2023).

Wright, M. (2002) *OSC spec 1_o*. Available at: https://opensoundcontrol.stanford.edu/spec-1_o.html (Accessed: 16 July 2025).

Yüksel, M., Önen, O. and Seeber, B.U. (2024) 'Acoustics, Psychoacoustics, and Properties of Sound', pp. 113–124. Available at: https://doi.org/10.1007/978-3-031-76173-7_7.

Zotter, F. and Frank, M. (2019) *Springer Topics in Signal Processing Ambisonics A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality, Volume 19*. Available at: <http://www.springer.com/series/8109>.

B Digitaler Anhang

Dieser Arbeit liegt ein digitaler Anhang in folgender Ordnerstruktur bei:

./PDF Die hier vorliegende Arbeit als PDF-Dokument

./MAX MAX-MSP Projektdatei