

Bachelorarbeit

im Studiengang Audiovisuelle Medien

CHANCEN UND GRENZEN DER BINAURALTECHNIK FÜR DEN HÖRFUNK

Eine gestalterische Perspektive

vorgelegt von

Christoph Oberlin

36786

an der Hochschule der Medien Stuttgart

am 06. 09. 2022

zur Erlangung des akademischen Grades eines Bachelor of Engineering

Erstprüfer: Prof. Dr. Frank Melchior

Zweitprüfer: Prof. Oliver Curdt

Eidesstattliche Erklärung

Hiermit versichere ich, Christoph Oberlin, ehrenwortlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Chancen und Grenzen der Binauraltechnik für den Hörfunk“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden. Ich habe die Bedeutung der ehrenwortlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwortlichen Versicherung zur Kenntnis genommen.

Datum: _____ Unterschrift: _____

Danksagungen

Vielen Dank an meine Familie und Freunde, welche mich über den gesamten Bearbeitungszeitraum der Bachelorarbeit emotional unterstützt haben und immer wieder bereit waren die gleichen Hörbeispiele anzuhören und mir ihren subjektiven Eindruck zu schildern.

Außerdem möchte ich mich explizit bei Linda Quast, Marc Hindrichsmeyer, Daniel Knüttel, Sebastian Oberlin und der gesamten Studioproduktion Ton des Sommersemesters 2022 bedanken, welche mir das nötige Ausgangsmaterial für die Hörbeispiele bereitgestellt haben.

Ein ganz besonderer Dank gilt außerdem meinem Erstprüfer Prof. Dr. Frank Melchior, sowie meinem Zweitprüfer Prof. Oliver Curdt. Beide haben meinen Werdegang durch ihre Lehrveranstaltungen und die Gespräche abseits der Hochschulveranstaltungen ganz entscheidend geprägt und standen mir jederzeit bei Fragen mit ihrem fundierten Fachwissen zur Verfügung.

Zusammenfassung / Abstract

Die Binauralsynthese ermöglicht, dreidimensionale Mischungen über Kopfhörer wiederzugeben und diese so einer Vielzahl von Konsumentinnen und Konsumenten zugänglich zu machen. Die vorliegende Arbeit verdeutlicht die Chancen und Grenzen dieser Technologie für den Hörfunk, mit Schwerpunkt auf die gestalterische Ebene.

Im ersten Teil der Arbeit wird die menschliche Psychoakustik mit besonderem Fokus auf das räumliche Hören behandelt. Zudem wird eine exemplarische Produktionsumgebung für die Erstellung von Binauralmischungen vorgestellt.

Im zweiten Teil wird anhand von Hörbeispielen demonstriert, wie gängige Hörfunkformate, bestehend aus Sprache und Musik, in 3D Audio umzusetzen wären und inwiefern sich ein jeweiliger Mehrwert oder auch Nachteile daraus ergeben. Dabei wird aufgezeigt, dass sich in vielen Fällen eine Mischung von binauralen- und Stereoelementen am besten eignet.

Abstract

The binaural synthesis enables a huge number of clients to play back threedimensional mixes over headphones. Focusing on the creative abilities, this thesis shows the chances and limits of this technologie for the radiomarket.

The first part addresses the human psychoacoustics and threedimensional hearing and also shows an exemplary way of mixing for binaural.

The second half opens a discussion how 3D audio could be implemented in different radio formats and where the advanteges and disadvantage lie. Most of times the best way is a mix between binaural and stereo elements.

Inhaltsverzeichnis

| | |
|---|-----------|
| Eidesstattliche Erklärung | 2 |
| Danksagungen | 3 |
| Zusammenfassung / Abstract | 4 |
| 1 Einleitung | 9 |
| 2 Psychoakustik | 10 |
| 2.1 Anatomie und Physiologie des menschlichen Ohres | 10 |
| 2.2 räumliches Hören | 12 |
| 2.2.1 Räumliches hören bei einer Schallquelle | 14 |
| 2.2.2 Entfernungshören | 18 |
| 2.2.3 Räumliches Hören bei mehreren Schallquellen | 19 |
| 2.2.4 Interaurale Kohärenz | 20 |
| 2.3 auditive Szenenanalyse | 22 |
| 3 Binauraltechnik | 23 |
| 3.1 Kunstkopfstereophonie | 23 |
| 3.2 Binauralsynthese und individualisierte HRTF | 25 |
| 3.3 Dynamische Binauralsynthese | 27 |
| 4 Exemplarische Binaurale Produktionumgebung | 31 |
| 4.1 Ambisonics | 31 |
| 4.1.1 Mathematische Grundlagen | 31 |
| 4.1.2 Ambisonic erster Ordnung | 32 |
| 4.1.3 Kodierung | 33 |
| 4.1.4 Higher Order Ambisonics | 35 |
| 4.2 Tools | 35 |
| 4.2.1 Reaper | 35 |
| 4.2.2 Ambisonic Plugins | 35 |

| | | |
|----------|--------------------------------------|-----------|
| 4.2.3 | Sound Particles | 37 |
| 4.3 | Busstruktur | 37 |
| 4.4 | Effekte | 39 |
| 4.4.1 | Räumlichkeit | 39 |
| 4.4.2 | Delay | 41 |
| 5 | Binaurale Inhalte | 42 |
| 5.1 | Sprache | 42 |
| 5.1.1 | Informative Sprachbeiträge | 42 |
| 5.1.2 | Fiktive Hörfunkbeiträge | 45 |
| 5.2 | Musik | 50 |
| 5.2.1 | E-Musik | 50 |
| 5.2.2 | U Musik | 51 |
| 5.3 | Jingles | 57 |
| 6 | Fazit | 58 |
| 6.1 | Ausblick | 59 |
| 7 | Hörbeispielverzeichnis | 60 |

Abbildungsverzeichnis

| | | |
|------|--|----|
| 2.1 | Außenohr, Mittelohr und Innenohr [Weinzierl, 2008, S. 42] | 10 |
| 2.2 | Kopfbezogenes Koordinatensystem [Weinzierl, 2008, S. 88] | 12 |
| 2.3 | Frequenzgänge des Betrages der Außenohr-Übertragungsfunktionen für Schalleinfall von links (90° Azimut, 0° Elevation) [Weinzierl, 2008, S. 91] | 13 |
| 2.4 | Richtungsbestimmende Frequenzbänder in der Medianebene [Dickreiter et al., 2014, S. 131] | 15 |
| 2.5 | Lokalisationsgenauigkeit in der Medianebene [Melchior, 2021] | 15 |
| 2.6 | Lateralisationskurve für Töne und Rauschen bei interauralen Pegeldifferenzen [Weinzierl, 2008, S. 97] | 16 |
| 2.7 | Lateralisationskurve interauraler Phasenlaufzeitdifferenzen [Weinzierl, 2008, S. 98] | 17 |
| 2.8 | Lokalisationsgenauigkeit in der Horizontalebene [Melchior, 2021] | 17 |
| 2.9 | Summenlokalisationskurven für Pegel- bzw. Laufzeitdifferenzen der beiden Lautsprecher-signale. Breitbandsignale, Zuhörerkopf fixiert (nach Wendt 1963) [Weinzierl, 2008, S. 101] | 20 |
| 2.10 | Hörereignislage und -ausdehnung in Abhängigkeit vom Kohärenzgrad bei unentzerrter Kopfhörerarbeit. Die Prozentzahlen kennzeichnen die relativen Häufigkeiten, mit denen die Hörereignisse in den entsprechenden Bereichen auftraten (nach Cherniak u. Dubrovsky 1968). [Weinzierl, 2008, S. 106] | 21 |
| 3.1 | System zur Binauralsynthese [Melchior, 2021] | 26 |
| 3.2 | angepasstes System zur dynamischen Binauralsynthese [Melchior, 2021] | 28 |
| 4.1 | Sphärische Harmonische 0. bis 2. Ordnung [Weinzierl, 2008, S. 661] | 32 |
| 4.2 | Natives 3D FOA Aufnahmesetup [Zotter and Frank, 2019, S. 10] | 33 |
| 4.3 | Tetraedrisch angeordnete Nierenmikrofone (links) und deren Encodermatrix (rechts) [Zotter and Frank, 2019, S. 11] | 33 |

| | | |
|-----|---|----|
| 4.4 | 2D FOA Binauraldecoder mit Head-Tracking [Zotter and Frank, 2019, S. 9] | 34 |
| 4.5 | Exemplarisches Blockschaltbild der Busstruktur | 38 |
| 4.6 | Stereo-Reverb mit unabhängiger Regelung von frühen Reflexionen und Nachhall | 40 |
| 5.1 | Directivity Shaper | 46 |
| 5.2 | Positionierung der Einzelkanäle der Steroaufnahme eines vorbeifahrenden Autos | 48 |
| 5.3 | Positionierung der Einzelkanäle des ORTF 3D | 50 |
| 5.4 | Visualisierung möglicher Hörereignisorte bei der Kopfhörerwiedergabe (rot = Mono; blau = Stereo; grau = 3D Audio) | 52 |
| 5.5 | Wahrnehmung einer binauralisierten dreidimensionalen Soundszene [Buff, 2020, S. 30] | 53 |
| 7.1 | | 61 |

1 Einleitung

Bei der Wiedergabe von Stereoinhalten über Kopfhörer befinden sich die darstellbaren Hörereignisorte innerhalb des Kopfes. Dies steht im Widerspruch zum natürlichen Hören, bei dem sich alle Hörereignisorte um den Kopf herum befinden. Durch Kunstkopfaufnahmen oder die Anwendung von Binauraltechnik können bei Kopfhörerwiedergabe Hörereignisse dargestellt werden, deren Position außerhalb des Kopfes verortet wird.

In den 80er- und 90er-Jahren gab es bereits Radiosendungen, welche binaurale Hörspiele und Konzertübertragungen ausstrahlten. Jedoch setzten sich diese aufgrund von klangfarblichen Mängeln, großen Lokalisationsfehlern, sowie der Inkompatibilität zu Lautsprechersystemen nicht durch und wurden wieder eingestellt.

Durch technische Innovationen konnte ein Teil der Probleme relativiert werden und durch die Markteinführung von objektbasierten Formaten, wie "Dolby Atmos Home", sowie die damit verbundenen Werbekampagnen wurde das Interesse der Öffentlichkeit und der Industrie wieder geweckt. Inzwischen stehen viele der neu erscheinenden Singles und Alben auch in einer 3D-Mischung zur Verfügung und der Konsum von binauralen Inhalten steigt.

Aus diesem Grund stellt sich die Frage wie der Hörfunk mit der Binauraltechnik umgeht und welchen Mehrwert diese bieten kann.

2 Psychoakustik

Die Psychoakustik ist eine interdisziplinäre Wissenschaft, welche die Wahrnehmung von akustischen Reizen untersucht. Dabei kann man die untersuchten Hörphänomene unterscheiden zwischen *Schwellen* und *überschwelligen Empfindungsgrößen*. Für *Schwellen* lässt sich als Beispiel die Hörschwelle anführen. Sie beschreibt denjenigen Schalldruckpegel, ab welchem der Proband einen akustischen Reiz gerade so wahrnimmt. Unter *überschwelligen Empfindungsgrößen* versteht man Größen wie die empfundene Lautheit oder das Maß an Umhüllung. [Weinzierl, 2008, S. 52 ff.]

Um die menschliche Psychoakustik zu verstehen, ist es zunächst wichtig den Aufbau des menschlichen Ohres zu betrachten.

2.1 Anatomie und Physiologie des menschlichen Ohres

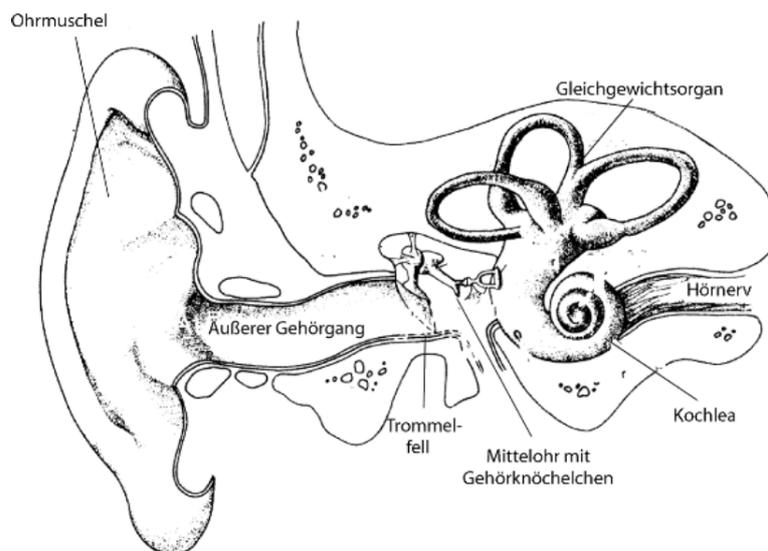


Abbildung 2.1: Außenohr, Mittelohr und Innenohr [Weinzierl, 2008, S. 42]

Das menschliche Ohr kann in drei Sektionen untergliedert werden: Das Außenohr mit Ohrmuschel, äußerem Gehörgang und Trommelfell. Das Mittelohr mit den Gehörknöchelchen Hammer, Amboss und Steigbügel, sowie dem ovalen Fenster, auf welchem letzterer aufliegt. Und das Innenohr mit Kochlea, Hörnerv und dem Gleichgewichtsorgan (siehe Abb 2.1).

Der Weg von akustischen Ereignissen aus unserer Umwelt führt über Reflexionen an Corso und Ohrmuschel in den äußeren Gehörgang. Durch Interferenzen von Direkt-schall und Reflexionen verändert sich der Frequenz- und Phasengang des jeweiligen Schallereignisses in Abhängigkeit zur Schalleinfallrichtung.

Das Außenohr jedes Menschen hat dabei seine individuelle Form. Dementsprechend ist die richtungsabhängige Frequenz- und Phasenveränderung auch individuell und nicht pauschalisierbar.

Anders als andere Säugetiere, wie Hunde oder Katzen, haben sich die Muskeln, welche der Bewegung der Ohren dienen, sehr weit zurückgebildet. Deshalb muss der Mensch seinen Kopf drehen um ein Ohr in Richtung eines Schallereignisses zu wenden.

Der äußere Gehörgang leitet den Schall weiter zum Trommelfell. Unter anderem wird durch seine Eigenresonanz der Frequenzbereich zwischen 2000 und 5000 kHz angehoben. Diese Anhebung liegt genau in dem Frequenzbereich, welcher für die Verständlichkeit der menschlichen Sprache am wichtigsten ist.

Das Trommelfell trennt das Außen- und Mittelohr räumlich ab und hat einen Durchmesser von ca. 10 mm [Weinzierl, 2008, S. 42 und S. 46].

Im Mittelohr sind die drei Gehörknöchelchen über Knorpel, ähnlich einem Gelenk, miteinander verbunden. Die eine Seite des Hammers, der Hammergriff, ist mit dem Trommelfell verwachsen. Der Steigbügelfuß liegt auf dem ovalen Fenster. Die Gehörknöchelchen passen die Impedanz zwischen Luft (geringe Dichte) und der mit einer Flüssigkeit gefüllten Kochlea (hohe Dichte) an. Durch die Hebelwirkung der Gehörknöchelchen, sowie dem Flächenunterschied von Trommelfell zum Steigbügelfuß wird der akustische Reiz um mehr als das 20 fache verstärkt [Weinzierl, 2008, S. 46].

Das ovale Fenster markiert den Übergang von Mittel- zu Innenohr und ist Teil der mit Flüssigkeit gefüllten Kochlea. Die Flüssigkeit der Kochlea wird von der Basilarmembran eingeschlossen, auf welcher sich die Haarzellen, die eigentlichen Hörnerven befinden. Durch Druckbewegungen des Steigbügelfußes am ovalen Fenster werden innerhalb der Kochlea Wellen angeregt. "Aufgrund der Elastizitätsbedingungen der Basilarmembran und der damit verbundenen Dämpfungseigenschaften steilen sich diese Wellen an unterschiedlichen Orten auf, und zwar abhängig von der frequenzabhängigen Wellen-

länge.” [Weinzierl, 2008, S.47f.] Das Innenohr dient also unter anderem der Zerlegung einer Welle aus dem Zeitbereich in ihre Frequenzanteile.

An den Haarzellen liegen afferente Nervenfasern an, welche über den inneren Gehörgang zum Kortex führen. Auf dem Weg zum Kortex passieren die Reize immer wieder Neurone, welche auf bestimmte Muster der Reizsignale reagieren. Umso näher die Neuronen dem Kortex sind, desto komplexer werden diese Reizmuster.

Unter anderem werden auch die Reizsignale beider Ohren miteinander verglichen. Dabei sind Unterschiede im Frequenzgang, der Phasenlage und der Amplitude ausschlaggebend für unsere räumliche Wahrnehmung von Schall, wobei minimale Differenzen von unter 30 μs eine Rolle spielen. [Weinzierl, 2008, S. 51]

”Die räumliche Ortung von Schallquellen muss als eine Leistung des gesamten Hörsystems angesehen werden.”[Weinzierl, 2008, S. 51]

2.2 räumliches Hören

Schallereignisse aus unserer Umwelt haben stets eine relative Position zum Kopf des jeweiligen Hörers. Räumliches Hören beschreibt den Zusammenhang zwischen den Orten und der räumlichen Ausdehnung der verschiedenen Hörereignisse zueinander. Zur örtlichen Definition wird ein kopfbezogenes Koordinatensystem herangezogen (siehe Abb. 2.2). Dabei befindet sich der Ursprung auf der interauralen Achse, genau zwischen den beiden Gehörkanaleingängen. Die Position einer Schallquelle kann durch drei Polarkoordinaten ausgedrückt werden. Durch *Azimat* - φ , welcher die Schalleinfallrichtung in der Horizontalebene definiert, der *Elevation* - δ , welche die Höhe in der Medianebene beschreibt und der *Entfernung* - r . [Weinzierl, 2008, S. 88]

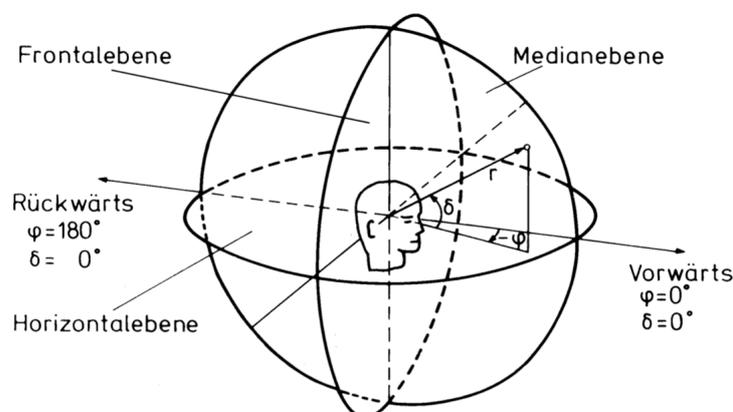


Abbildung 2.2: Kopfbezogenes Koordinatensystem [Weinzierl, 2008, S. 88]

Wie bereits in 2.1 beschrieben, werden räumliche Informationen unter anderem durch den Vergleich der beiden vorliegenden Ohrsignale gewonnen. Es wird unter *Schallereignisort*, dem tatsächlichen Ort der Schallquelle und *Hörereignisort*, dem empfundenen Ort der Schallquelle unterschieden.

Dabei gibt es sowohl *monaurale Ohrsignalmerkmale* (ein Ohr ist für den Empfang hinreichend) als auch *interaurale Ohrsignalmerkmale* (beide Ohren sind für den Empfang notwendig). Außerdem wird unter den Beschallungsarten *monotisch* (nur ein Ohr wird beschallt), *diotisch* (beide Ohren werden identisch beschallt) und *dichotisch* (beide Ohren werden unterschiedlich beschallt) unterschieden.

Unser Körper "[...] kann als eine bewegliche Antenne zum Empfang von Schallsignalen betrachtet werden. Diese Antenne hat eine richtungs-, entfernungs- und frequenzabhängige Richtcharakteristik." [Weinzierl, 2008, S. 89 f.]

Abhängig von der Schalleinfallrichtung werden akustischen Reize bezüglich ihres Frequenzgangs und ihrer Phasenlage linear verzerrt. Je nach Schallereignisort sind diese bei beiden Ohren unterschiedlich. Sie lassen sich durch messtechnische Methoden erfassen und können in einer Außenohr-Übertragungsfunktion oder auch Head-Related Transfer Function (HRTF) mathematisch erfasst und wie folgt beschrieben werden: $H_l(f, r, \varphi, \delta)$ und $H_r(f, r, \varphi, \delta)$. Übertragen in den Zeitbereich spricht man von einer kopfbezogenen Impulantwort beziehungsweise Head-Related Impulse Response (HRIR).

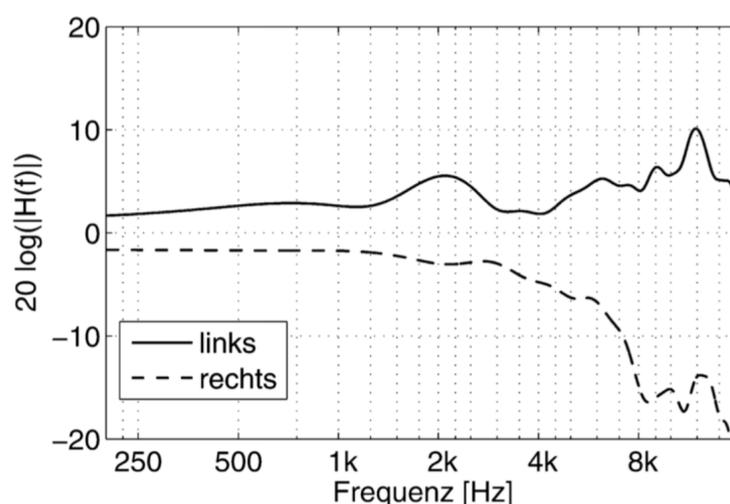


Abbildung 2.3: Frequenzgänge des Betrages der Außenohr-Übertragungsfunktionen für Schalleinfall von links (90° Azimut, 0° Elevation) [Weinzierl, 2008, S. 91]

Die *interaurale Außenohr-Übertragungsfunktion*, welche die Unterschiede der beiden Ohrsignale beschreibt und folglich für das räumliche Hören essentiell ist wird über die Division der linken und rechten HRTF bestimmt:

$$H_i(f, r, \varphi, \delta) = \frac{H_l(f, r, \varphi, \delta)}{H_r(f, r, \varphi, \delta)} \quad (2.1)$$

Für die interauralen Phasen- und Gruppenlaufzeiten ergibt sich dadurch

$$\tau_{\text{ph}}(f, r, \varphi, \delta) = \frac{b_i(f, r, \varphi, \delta)}{f} \quad (2.2)$$

sowie

$$\tau_{\text{gr}}(f, r, \varphi, \delta) = \frac{db_i(f, r, \varphi, \delta)}{df} \quad (2.3)$$

[Weinzierl, 2008, S. 91]

Beim Richtungshören wird unter anderem auch die Kopf- beziehungsweise Körperhaltung relativ zur Schallquelle registriert. Durch eine Änderung dieser ändern sich proportional dazu auch die Ohrsignale. Durch sogenannte "Peilbewegungen" kann das auditive System zusätzliche Informationen sammeln und bereits gesammelte Informationen validieren. Man unterscheidet dabei zwischen unbewussten Peilbewegungen und bewussten, dem Umhören.

Informationen, welche durch bewusstes Umhören gewonnen wurden, dominieren dabei in aller Regel solche, welche statisch wahrgenommen wurden. Außerdem spielen auch weitere Sinneseindrücke, wie zum Beispiel der Seh-, oder Tastsinn, eine wichtige Rolle bei der Hörereignisbildung. [Weinzierl, 2008, S. 88]

2.2.1 Räumliches hören bei einer Schallquelle

Lokalisation in der Medianebene

Wenn Schall aus Richtung der Medianebene auf den Kopf trifft, werden die Ohren nahezu diotisch beschallt. Dementsprechend bleiben dem auditiven System nur monoaurale Ohrsignalmerkmale, um das Hörereignis zu lokalisieren.

Bei schmalbandigen Signalen hängt die Hörereignisrichtung weniger von der Schalleinfallrichtung sondern überwiegend von der spektralen Verteilung des Schmalbandsignals ab. Man spricht von Richtungsbestimmenden Bändern (Abb 2.4). Diese können

jedoch von Person zu Person leicht variieren. [Weinzierl, 2008, S. 94]

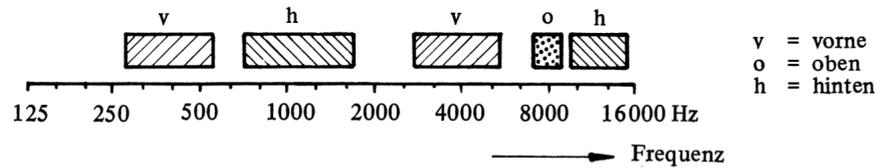
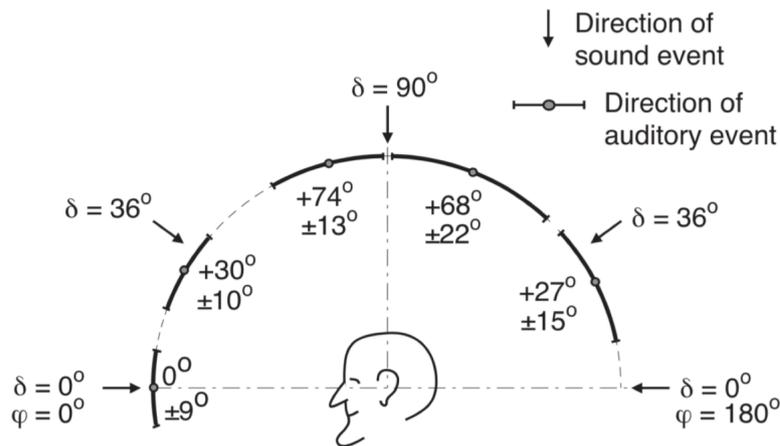


Abbildung 2.4: Richtungsbestimmende Frequenzbänder in der Medianebene [Dickreiter et al., 2014, S. 131]

Bei Breitbandsignalen, mit welchen wir in unserem Alltag hauptsächlich konfrontiert sind, entspricht meist der Hörereignisort der Schalleinfallrichtung. Durch Reflexionen an der Ohrmuschel bilden sich Kammfilter, über die unser Gehirn auf die Hörereignisrichtung schließen kann. Diese Frequenzanhebungen entsprechen zumeist den Richtungsbestimmenden Bändern. Dabei ist in der Regel das Frequenzband ausschlaggebend, bei welchem die Ohrsignale die größten Anteile aufweisen. [Weinzierl, 2008, S. 95]

Dadurch lässt sich auch der *Elevationseffekt* erklären, bei dem der Hörereignisort eines Instruments mit besonders starken Frequenzanteilen um 8kHz in einer Stereomischung räumlich über den Hörereignisorten der anderen Instrumente verortet wird.



Adapted from Damaske and Wagener (1969)

Abbildung 2.5: Lokalisationsgenauigkeit in der Medianebene [Melchior, 2021]

Dadurch dass dem auditiven System nur monoaurale Ohrsignalmerkmale zur Verfügung stehen, ist der Lokalisationsfehler in der Medianebene relativ groß (siehe Abb 2.5).

Lokalisation in der Horizontalebene

Bei der Hörereignisbildung in der Horizontalebene stehen neben den monoauralen Ohrsignalmerkmalen zusätzliche interaurale Ohrsignalmerkmale zur Verfügung. Diese ergeben sich aus der interauralen Außenohr-Übertragungsfunktion H_i . Die interauralen Ohrsignalmerkmale können in *interaurale Pegeldifferenzen* ΔP und *interaurale Zeitdifferenzen* $\Delta\tau$ untergliedert werden.

Interaurale Pegeldifferenzen führen frequenzunabhängig zu einer seitlichen Auslenkung des Hörereignisses. Die Lateralisationskurve verläuft äußerst linear und ist sowohl für einzelne Sinustöne als auch für breitbandige Signale sehr ähnlich (vgl. Abb. 2.6).

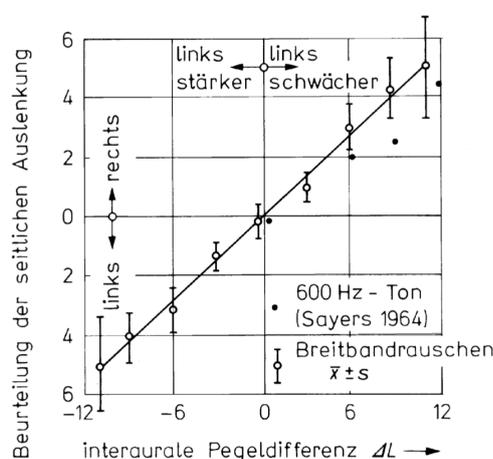


Abbildung 2.6: Lateralisationskurve für Töne und Rauschen bei interauralen Pegeldifferenzen [Weinzierl, 2008, S. 97]

Interaurale Zeitdifferenzen müssen spezifischer betrachtet werden. Das Gehirn wertet sowohl die interaurale Gruppenlaufzeit τ_{gr} , als auch die interaurale Phasenlaufzeit τ_{ph} aus. Für die Lateralisation spielen bei τ_{ph} nur Frequenzen unterhalb von 1,6 kHz eine Rolle. Oberhalb von 1,6 kHz analysiert das Gehirn die frequenzspezifische Hüllkurve des Schallereignisses (τ_{gr}).

Im Vergleich zur interauralen Pegeldifferenz verläuft die Lateralisationkurve im Bereich von $\tau_{ph} = \pm 500 \mu s$ fast linear und flacht dann ab (vgl. Abb. 2.7).

Bei der Bildung des Hörereignisortes wertet das auditive System gleichermaßen interaurale Pegeldifferenzen und interaurale Phasen- und Gruppenlaufzeitdifferenzen aus. Dabei können sich die interauralen Merkmale gegenseitig bedingen und kompensieren. Beispiel hierfür sind Schlagzeug-Overheadaufnahmen mit einer AB-Stereomikrofonierung bei welchen die Snare aufgrund von Phasenlaufzeitdifferenzen leicht nach links rückt.

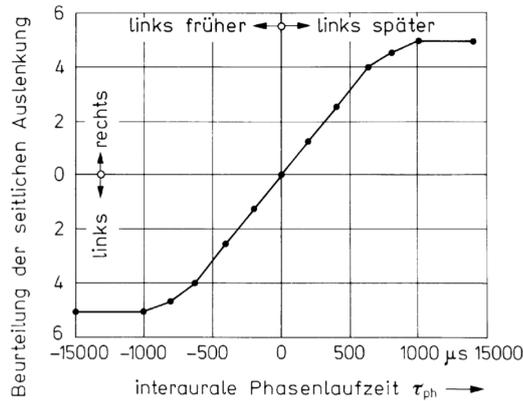


Abbildung 2.7: Lateralisationskurve interauraler Phasenlaufzeitdifferenzen [Weinzierl, 2008, S. 98]

Senkt man den linken Kanal leicht ab beziehungsweise hebt den rechten Kanal an, so wird der Hörereignisort wieder in der Phantommittle wahrgenommen.

Werden die Widersprüche der interauralen Merkmale zu groß, zerfällt das Hörereignis in mehrere Anteile aus verschiedenen Richtungen. [Weinzierl, 2008, S. 98]

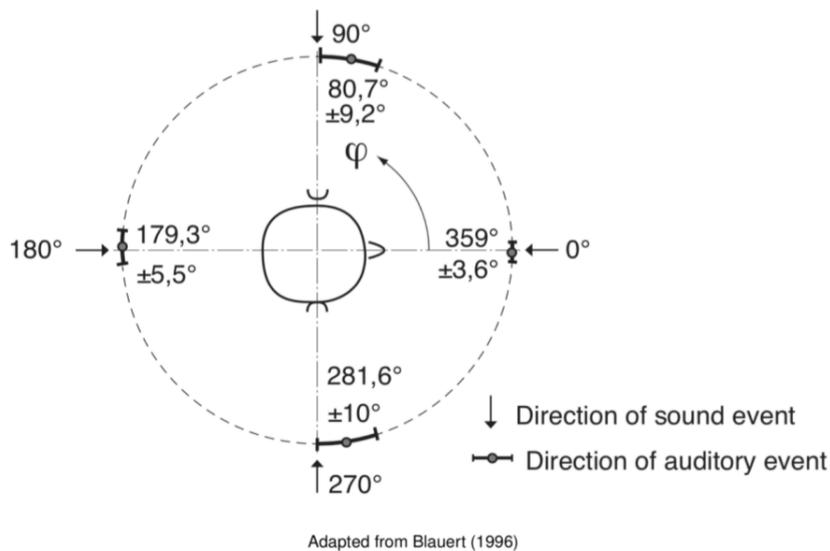


Abbildung 2.8: Lokalisationsgenauigkeit in der Horizontalebene [Melchior, 2021]

Die Lokalisationsgenauigkeit in der Horizontalebene liegt für die Schalleinfallrichtung aus frontaler Richtung bei circa drei Grad, sodass sehr geringe Auslenkungen registriert werden. Aus anderen Schalleinfallrichtungen ist der Lokalisationsfehler deutlich größer

(vgl. Abb. 2.8).

Dadurch dass bei einem Schalleinfallswinkel von 0 Grad und 180 Grad nur monoaurale Ohrsignalmerkmale vorliegen, ist der einzig wahrnehmbare Unterschied der beiden Schalleinfallsrichtungen im Frequenzspektrum zu finden. Daher ist in diesem Bereich bei fixiertem Kopf der Lokalisationsfehler einer nicht bekannten Schallquelle sehr hoch. Man spricht von "Front-Back-Confusion". [Frank and Zotter, 2018]

2.2.2 Entfernungshören

Sowohl die monoaurale als auch die interaurale Übertragungsfunktion verändern sich mit der Schallquellenentfernung. Man kann diese Änderungen in vier Entfernungsbereiche untergliedern:

1.) Nahfeld - $r \leq 25\text{cm}$

Im Nahfeld entstehen starke spektrale Änderungen der Schallereignisse und die interauralen Pegeldifferenzen steigen aufgrund der Schirmwirkung des Kopfes. Außerdem steigt der Signalpegel mit abnehmender Entfernung signifikant an.

2.) mittlerer Entfernungsbereich - $r = 25\text{cm} - 15\text{m}$

Im Freifeld ändert sich der Signalpegel durch das $1/r$ -Gesetz mit 6dB pro Entfernungsverdopplung. Dabei kann sich auch aufgrund der Isophonen (Kurven gleicher Lautstärke) der Klangeindruck ändern. Dem Gehirn bekannte Signale können dadurch auf ihre Entfernung hin ausgewertet werden.

Ein weiteres Merkmal ergibt sich durch das Verhältnis von Direktschall zu reflektiertem Schall. "Die Gewichtung zwischen den beiden Hauptmerkmalen (Schallquellenpegel und Intensitätsverhältnis von Direktschall zu reflektiertem Schall) hängt bei der Entfernungsbeurteilung stark von der Schalleinfallsrichtung und anderen Eigenschaften des Schalls ab (Zahorik 2002)."[Weinzierl, 2008, S. 99]

3.) große Entfernungen - $r > 15\text{m}$

Bei Entfernungen über 15m entsteht eine wahrnehmbare frequenzabhängige Zusatzdämpfung, welche mit steigender Frequenz steigt. Der akustische Horizont unseres auditiven Systems liegt bei etwa 15m. Jedoch kann unser Gehirn aufgrund von Er-

fahrungen und anderen Sinneseindrücken auf größere Entfernungen schließen.

4.) Im-Kopf-Lokalisation

Wenn zwei Schallereignisse identische beziehungsweise sehr ähnliche Signale abstrahlen und sich beide sehr nah an linken und rechten Ohr befinden (zum Beispiel Kopfhörerwiedergabe), so entsteht das Hörereignis im Kopf.

[Weinzierl, 2008, S. 98 f.]

2.2.3 Räumliches Hören bei mehreren Schallquellen

Eine einzelne Schallquelle existiert nur im Freifeld und liegt in unserem Alltag nur sehr selten vor. Sobald ein Schallereignis in einem Raum stattfindet, entstehen mehrere Schallquellen, bestehend aus Direktschall und Reflexionen.

Es können zwar Schallfelder von Schallquellen überlagert werden, die Hörereignisse, welche den einzelnen Schallquellen zugeordnet werden, aber nicht. In einer exemplarischen Situation, in der zwei Schallquellen identische oder sehr ähnliche Signale abstrahlen, wobei das eine Signal gegenüber dem anderen verzögert oder mit variiertem Pegel abgestrahlt wird, lassen sich drei Fälle definieren, welche auch auf mehr als zwei Schallquellen übertragbar sind. [Weinzierl, 2008, S. 100]

Summenlokalisierung

Als Beispiel für die Summenlokalisierung kann man eine klassische Lautsprecheraufstellung mit einem Basiswinkel von $\alpha = 60$ Grad anführen. Spielt man über beide Lautsprecher das selbe Signal ab, so entsteht in der Mitte der beiden Lautsprecher eine Phantomschallquelle. Diese ist jedoch weniger scharf lokalisierbar als eine tatsächliche Einzelschallquelle.

Verändert man die Signalpegel oder verzögert die beiden Schallquellen zueinander, verschiebt sich die Phantomschallquelle auf der Stereobasis der beiden Schallquellen in Richtung des zuerst eintreffenden beziehungsweise lautereren Schallereignisses (vgl. Abb 2.9).

Bei seitlicher Lautsprecheraufstellung (Basismitte befindet sich bei 90 Grad) wird die Summenlokalisationskurve um einiges steiler und es entstehen sehr starke individualspezifische Abweichungen.

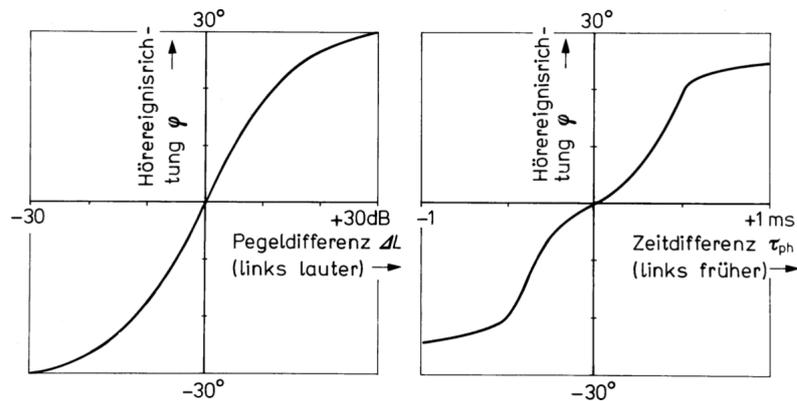


Abbildung 2.9: Summenlokalisationskurven für Pegel- bzw. Laufzeitdifferenzen der beiden Lautsprecher signale. Breitbandsignale, Zuhörer kopf fixiert (nach Wendt 1963) [Weinzierl, 2008, S. 101]

Präzedenzeffekt und Echo

Verzögert man die beiden Schallquellen mehr als eine Millisekunde zueinander, wird der Hörereignisort in Richtung des zuerst eintreffenden Schallereignisses wahrgenommen. Dieses Phänomen nennt man *Präzedenzeffekt* oder auch das *Gesetz der ersten Wellenfront*.

Durch diesen Effekt bildet das auditive System in einem Raum den Hörereignisort in Richtung der eigentlichen Schallquelle und nicht in Richtung der schallreflektierenden Wände. Überschreitet die Verzögerung der später eintreffenden Wellenfronten jedoch einen gewissen signal- und pegelabhängigen Grenzwert, so werden die verzögerten Schallereignisse als separate, räumlich unabhängige Hörereignisse (Echos) wahrgenommen. Diese Grenze wird *Echoschwelle* genannt. Sie variiert je nach vorliegendem Schallsignal zwischen einer und mehreren hundert Millisekunden. [Weinzierl, 2008, S. 103]

2.2.4 Interaurale Kohärenz

Die interaurale Kohärenz beschreibt die Ähnlichkeit der beiden Ohrsignale. Um die interaurale Kohärenz mathematisch zu bestimmen wird in der Regel die normierte Kreuzkorrelationsfunktion in der Schreibweise für Leistungssignale verwendet. Dabei steht $\mathbf{x}(t)$ für das rechte, $\mathbf{y}(t)$ für das linke Ohrsignal.

$$R_{\text{norm, } x, y}(\tau) = \frac{\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} x(t) * y(t + \tau) dt}{x(t)_{\text{eff}} * y(t)_{\text{eff}}} \quad (2.4)$$

Dabei wird das Maximum ihres Betrags im physiologischen Bereich von τ in der Psychoakustik als *interaauraler Kohärenzgrad* bezeichnet. Dieser kann Werte zwischen 1 (Ohrsignale sind identisch) und 0 (Ohrsignale sind komplett verschieden) annehmen. Dabei haben sowohl frequenzunabhängige Pegeldifferenzen als auch verzerrungsfreie Verschiebungen eines der beiden Signale keinen Einfluss auf den interauralen Kohärenzgrad. [Weinzierl, 2008, S. 105]

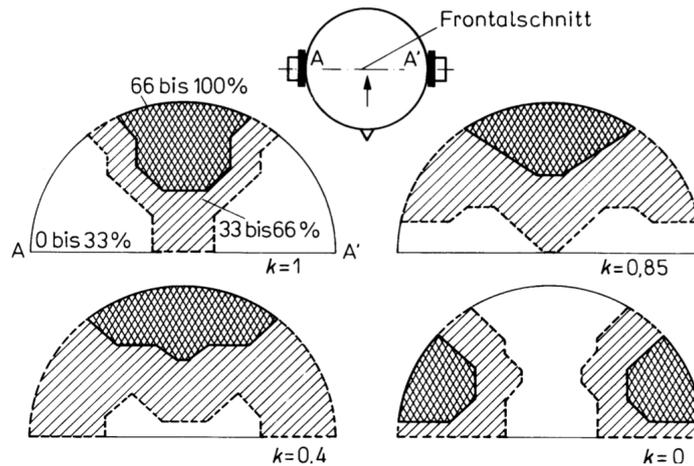


Abbildung 2.10: Hörereignislage und -ausdehnung in Abhängigkeit vom Kohärenzgrad bei unentzerrter Kopfhörerdarbietung. Die Prozentzahlen kennzeichnen die relativen Häufigkeiten, mit denen die Hörereignisse in den entsprechenden Bereichen auftraten (nach Cherniak u. Dubrovsky 1968). [Weinzierl, 2008, S. 106]

In Laborumgebung lassen sich mithilfe von drei Rauschgeneratoren beliebige Kohärenzgrade generieren. Dabei werden zwei unabhängige Rauschsignale pro Ohr mit einem gemeinsamen für beide Ohren überlagert. Besteht das Hörereignis nur aus dem gemeinsamen Rauschsignal, so ergibt sich der Kohärenzgrad $k = 1$. Das Hörereignis wird recht Scharf lokalisiert in der Kopfmitte wahrgenommen.

Werden die unabhängigen Rauschsignale dazu gemischt, sinkt der Kohärenzgrad und das Hörereignis dehnt sich aus. Eine scharfe Lokalisation ist nicht mehr möglich. Unter einem bestimmten Schwellenwert zerfällt das Hörereignis in zwei Anteile, welche der jeweiligen Schallquelle zugeordnet werden können (vgl. Abb 2.10). [Weinzierl, 2008, S. 106]

In einem freien Schallfeld sind die beiden Ohrsignale weder völlig kohärent noch komplett inkohärent. Durch Übersprechen bleibt - außer beim Hören über Kopfhörer - ein

gewisser Kohärenzgrad immer vorhanden und durch Reflexionen an Boden, Decke und Wänden dekorrelieren die beiden Ohrsignale.

In einem diffusen Schallfeld ist der gesamte Raum vom Hörereignis ausgefüllt und es entsteht der Effekt von *Einhüllung*. Dieser Effekt wird gerade bei Musik als positives Qualitätsmerkmal wahrgenommen.

Bei teilkohärenten Ohrsignalen kann das auditive System stark kohärente Anteile herausfiltern und einem schärfer lokalisierten Hörereignis zuordnen. [Weinzierl, 2008, S. 106]

2.3 auditive Szenenanalyse

Bei der auditiven Szenenanalyse geht es darum, wie das auditive System akustische Umwelteinflüsse analysiert und diese in sinnvolle Einheiten gliedert. Dabei wird zwischen der *simultanen Gruppierung* und der *sequentiellen Gruppierung* unterschieden. Die *simultanen Gruppierung* beschäftigt sich mit Reizparametern, welche das auditive System zur Unterscheidung gleichzeitig stattfindender akustischer Reize nutzen kann. Um Schallanteile einer oder mehrerer Schallquellen zuzuordnen werden folgende Attribute genutzt: “(1) Gemeinsames Ein- und Aussetzen, (2) gemeinsame Grundfrequenz (Harmonizität), (3) gemeinsame räumliche Herkunft, (4) kohärente Amplituden- oder Frequenzmodulationen und (5) spektrale Form (Klangfarbe) [...]” [Weinzierl, 2008, S. 72]

Bei der *sequentiellen Gruppierung* geht es darum, nacheinander eintreffende Schallergebnisse zusammenzufassen beziehungsweise auf verschiedene Ursprungsquellen zurückzuführen. Dabei kann zum Beispiel eine Tonfolge, bei der sich wiederholend Töne aus einer hohen Lage mit Tönen einer tieferen Lage abwechseln, abhängig von der Wiedergabegeschwindigkeit, als eine gemeinsame Melodie oder als zwei unabhängige Melodien wahrgenommen werden. *Bregman und Campbell* (1971) sprechen dabei von *auditory stream segregation*. [Weinzierl, 2008, S. 72]

Neben der Wiedergabegeschwindigkeit der einzelnen Töne spielen auch Faktoren wie die Differenz der Lage (umso näher diese beieinander liegen, desto eher wird die Tonfolge als eine Einheit wahrgenommen), die spektrale Struktur und die räumliche Lokalisation eine Rolle. [Albert S. Bregman, 1990]

3 Binauraltechnik

Die Binauraltechnik ermöglicht die Wiedergabe von dreidimensionalen Audioinhalten über Kopfhörer.

Bei der Wiedergabe von klassischen Stereomischungen über Kopfhörer entstehen Hörereignisorte innerhalb des Kopfes. Analog zur stereophonen Lautsprecherwiedergabe, bei der Hörereignisorte zwischen linkem und rechtem Lautsprecher darstellbar sind, lassen sich bei Kopfhörerwiedergabe Hörereignisse zwischen linkem und rechtem Ohr darstellen.

Durch die Anwendung von Binauraltechnik entsteht eine *Externalisierung*. Der Hörereignisort befindet sich also außerhalb des Kopfes, was dem natürlichen Hören entspricht. Im Folgenden werden verschiedene Techniken aufgezeigt, welche zur Externalisierung verwendet werden können.

3.1 Kunstkopfstereophonie

Die Idee der Kunstkopfstereophonie ist die Aufnahme beziehungsweise Übertragung der beiden Ohrsignale am Aufnahmeort. Dabei werden die Ohrsignale an den Eingängen der äußeren Hörkanäle abgegriffen, damit die akustischen Eigenschaften des Gehörgangs nicht mehrfach berücksichtigt werden. Durch die Reproduktion über Kopfhörer wird die Außenohr-Übertragungsfunktion des Hörers außer acht gelassen und es ergibt sich im Idealfall der gleiche Höreindruck wie am Aufnahmeort. [Dickreiter et al., 2014, S. 345]

Geschichte der Kunstkopfstereophonie

Bereits um 1930 wurde die binaurale Tonübertragung erforscht und es entstanden die ersten Kunstköpfe, welche jedoch erst einmal für experimentelle Zwecke beziehungsweise spezielle Anwendungsgebiete bestimmt waren.

Der erste kommerzielle Kunstkopf, der "KU 80" der Firma "Neumann", wurde 1973 auf der IFA mit dem Hörspiel "Demolition" vorgestellt. Dabei beeindruckten vor allem die

Darstellung von Räumlichkeit und Entfernung, gleichzeitig gab es erhebliche Mängel der Klangfarbe und der Lokalisation in der Horizontalebene.

Der Nachfolger "KU 81" und andere vergleichbare Kunstköpfe dieser Zeit nahmen sich der Mängeln an, indem sie die Kopf- und Ohrform genauer evaluierten und nachbildeten. Dadurch stieg die Lokalisationsgenauigkeit, die Problematik aber, dass einige Hörer*innen Schallereignisse, welche bei der Aufnahme frontal vor dem Kunstkopf verortet waren, hinter den Hörern*innen lokalisierten blieb erhalten. Die klangfarblichen Mängel wurden unter anderem durch eine Diffusfeldentzerrung der eingebauten Druckempfänger vermindert. [Dickreiter et al., 2014, S. 346]

Vor- und Nachteile der Kunstkopfstereophonie

Kunstkopfaufnahmen bieten ein hohes Maß an räumlicher Detailtreue, an Einhüllung und der Darstellung verschiedener Entfernungen. Deshalb eignen sie sich sehr gut für Atmosphären in Hörspielen, als Hauptmikrofon bei binauralen Musikaufnahmen oder auch für Effekte wie das Flüstern in ein Ohr.

Zudem ist eine nachträgliche binaurale Aufnahme einer raumbezogenen Stereo oder Mehrkanalmischung möglich. Dabei wird der Kunstkopf in den Sweet-Spot gestellt und die Mischung in Echtzeit aufgezeichnet. Die Qualität der Aufnahme steht dabei in voller Abhängigkeit zum Wiedergabesystem und dem Abhörraum.

Nachteil bleibt, dass die Außenohr-Übertragungsfunktion schon innerhalb des übertragenden beziehungsweise aufgenommenen Signals vorhanden ist, wodurch die Aufnahme für die Wiedergabe über Lautsprechersysteme inkompatibel wird. Die Kopf- und Ohrmaße des Kunstkopfes entsprechen zwar Durchschnittswerten, doch kleine individuelle Abweichungen können schon zu großen Unterschieden in der räumlichen Wahrnehmung führen.

Eine weitere Problematik bildet das statische Aufnahmeprinzip. Wie in Kapitel 2.2 beschrieben fundiert die korrekte räumliche Lokalisation auf unbewussten Peilbewegungen. Durch die statische Aufnahme rotiert die Soundszene bei Kopfbewegungen mit, was dem natürlichen Hören widerspricht und zu Lokalisationsfehlern führt. [Dickreiter et al., 2014, S. 218]

Außerdem ist eine nachträgliche Klanggestaltung nur bedingt möglich. Bei einer Stereoaufnahme kann über das Stützmikrofonverfahren die Balance zwischen Raum- und Direktsignal manipuliert werden. Zudem können die Lautstärkeverhältnisse und die

räumliche Balance der Instrumente untereinander angepasst werden. Bei einer Kunstkopfaufnahme sind diese Verhältnisse nach der Aufnahme nicht mehr manipulierbar und können nur durch die Positionierung des Kunstkopfes im Raum und die Anordnung und Dynamik der einzelnen Instrumente beeinflusst werden.

3.2 Binauralsynthese und individualisierte HRTF

Um die Problematik der mangelnden Manipulationsmöglichkeiten der Kunstkopfste-reophonie anzugehen, ist es notwendig die Einzelkomponenten, welche zum Hörereignis selbst und dessen Verortung beitragen, zu trennen. Wie in Kapitel 2.2 erläutert, wird ein Schallsignal mit den jeweilig ortsabhängigen Außenohr-Übertragungsfunktionen (HRTF) gefaltet, woraus das auditive System auf den Hörereignisort schließen kann. [Weinzierl, 2008, S. 671] Liegen für jeden beliebigen Ort die jeweiligen HRTFs vor, so lässt sich durch digitale Faltung der Hörereignisort eines Direktsignals in der Theorie frei manipulieren.

System zur Binauralsynthese

Um ein vorliegendes Quellsignal zu binauralisieren, muss das Signal selbst wie auch dessen gewünschte Schallquellenposition vorliegen. Über die Schallquellenposition können aus einer vorliegenden Datenbank sowohl die passenden Filterkoeffizienten, als auch die dazugehörige interaurale Laufzeitdifferenz ermittelt werden. Das Quellsignal wird mit den jeweiligen Filtern für linkes und rechtes Ohr gefaltet und anschließend abhängig der - im Optimalfall skalierten - Laufzeitdifferenz verzögert (vgl. Abbildung 3.1).

Messung von HRTFs

Um die persönliche HRTF eines Probanden zu messen, ist ein reflexionsarmer Raum notwendig, damit Raumeinflüsse minimiert werden. Der Proband bekommt an den beiden Gehörgangeingängen Druckempfänger platziert und wird auf einen Drehteller gestellt. Um den Probanden herum sind in festen Abständen im Halbkreis Lautsprecher positioniert. Aus jedem Lautsprecher wird nacheinander ein Breitbandimpuls abgespielt, während die beiden Ohrsignale gemessen werden. Es liegt eine *Head-Related Impulsresponse (HRIR)* im Zeitbereich vor, aus welcher die HRTF im Frequenzbereich abgeleitet werden kann. Nachdem die Ohrsignale für alle Lautsprecher des Halbkreises

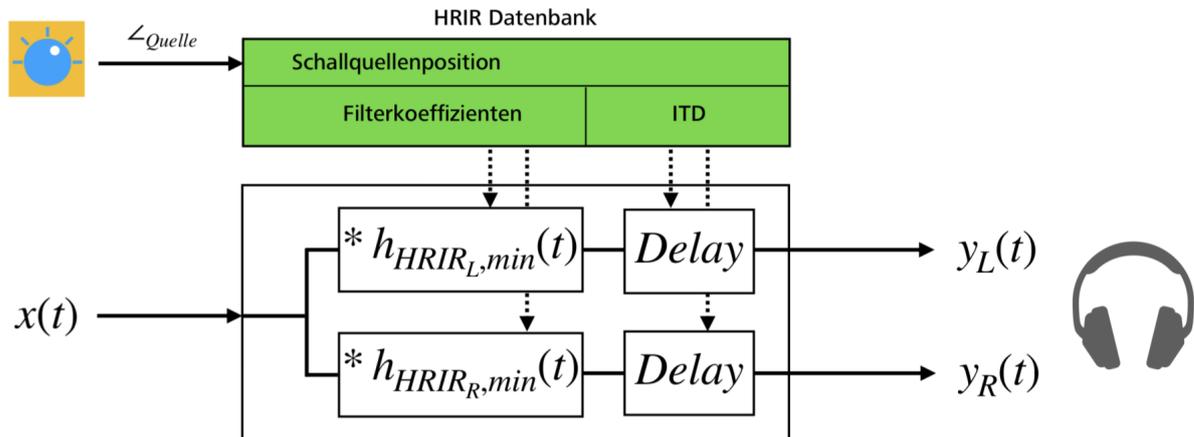


Abbildung 3.1: System zur Binauralsynthese [Melchior, 2021]

gemessen wurden, wird der Proband durch den Drehteller gedreht und der Vorgang so lange wiederholt, bis eine Kuppel aus Messpunkten vorhanden ist.

Jeder Messpunkt entspricht einer richtungsabhängigen Übertragungsfunktion, welche aus zwei *FIR-Filtern*, und einer *interauralen Laufzeitdifferenz (ITD)* besteht. Wenn die gemessene HRTF für eine Datenbank bestimmt ist, muss die ITD extrahiert und entfernt werden, da sie abhängig von der individuellen Kopfbreite ist und im Nachhinein skalierbar sein sollte.

Um eine verzerrungsfreie Messung zu erhalten, müssen die akustischen Eigenschaften der Lautsprecher und Mikrofone berücksichtigt und aus der Übertragungsfunktion herausgerechnet werden. Zudem sollte im Mittel über alle Richtungen ein neutraler Klangeindruck entstehen, weshalb die Messpunkte diffusfeldentzerrt werden. Da die für die Messung verwendeten Lautsprecher meist sehr kompakt sind und dementsprechend nur bis zu einer gewissen unteren Grundfrequenz beschallen, muss ebenfalls der Bassbereich angepasst und erweitert werden.

Eine weitere Möglichkeit, eine personalisierte HRTF zu ermitteln ist eine optische, häufig KI-gestützte Methode. Dabei werden Fotos und Videos des Kopfes, der Ohren und der Schultern - teilweise mit Maßband - gemacht und ausgehend von diesen optischen Informationen eine individuelle HRTF berechnet. Ein Beispiel dafür ist "Aural ID" von Genelec, bei der im Anschluss an die Berechnung der HRTF, diese auch als SOFA-Datei (allgemeingültige Austauschformat für HRTFs) nutzbar ist. [Genelec,]

Binaurale Raumimpulsantworten

Während bei der Messung einer HRTF der Raumeinfluss möglichst weit minimiert wird, geht es bei der Messung von *binauralen Raumimpulsantworten* (engl.: *Binaural Room Impulsresponse (BRIR)*) genau darum, den räumlichen Einfluss der zu virtualisierenden Räumlichkeit aufzunehmen. Daher muss die Länge der BRIR mindestens die Dauer der Nachhallzeit des entsprechenden Raumes betragen, was einen vielfachen Rechenaufwand im Vergleich zu einer HRIR bedeutet, welche nur wenige Millisekunden lang ist. [Weinzierl, 2008, S.673]

Da die Bildung des Hörereignisortes durch Reflexionen beeinflusst wird, funktioniert die Ortung und Externalisierung von Hörereignissen dadurch meist besser. Nachteil ist jedoch, dass die Hörereignisse räumlich ausgedehnter und weniger scharf lokalisierbar sind.

BRIRs werden zum Beispiel für die Virtualisierung von Regieräumen verwendet, wie zum Beispiel bei "Waves NX" [WavesAudioLtd,] oder "dearVR Monitor".

[DearRealityGmbH, 2022a] Es gibt auch Binauralisierer wie den "dearVR Pro", bei welchem frühe Reflexionen und Nachhall angepasst und in beliebigem Verhältnis dazu gemischt werden können [DearRealityGmbH, 2022b].

3.3 Dynamische Binauralsynthese

Wie schon in Kapitel 2.2 erwähnt, spielen Peilbewegungen des Kopfes beim räumlichen Hören und der Bildung von Hörereignisorten eine erhebliche Rolle. Um eine auditive Szene natürlich hören zu können, ist es deshalb notwendig, dass Kopfbewegungen registriert werden und die auditive Szene sich dementsprechend anpasst, da dies vom auditiven System erwartet wird. [Dickreiter et al., 2014, S. 347]

Die Technik, um solche Kopfbewegungen zu registrieren, wird Headtracking genannt. Dabei wird die Kopfhaltung des Hörers zum Beispiel über einen Gyrosensor in bis zu drei Winkelgrößen erfasst und an die Binauralisierungsanwendung weitergegeben. Die Winkelgrößen beschreiben die Kopfneigung um die z-Achse (engl.: *Yaw*), die vertikale Kopfneigung um die y-Achse (engl.: *Pitch*) und das laterale Rollen des Kopfes um die x-Achse (engl.: *Roll*).

Um Headtracking zu realisieren, muss das System zur Binauralsynthese (Abbildung 3.1) angepasst und erweitert werden (vgl. Abbildung 3.2).

Zunächst muss die Verarbeitung im Binauralisierer blockweise mit einem vordefinierten Zeitintervall geschehen. Zwischen diesen Blöcken muss interpoliert werden, um Sprünge im Signal zu vermeiden. Zusätzlich zur Schallquellenposition liegen dem Binauralisierer die Winkelgrößen der Kopfdrotation des Hörers vor. Diese werden miteinander verrechnet und anschließend die entsprechenden FIR-Filter und die ITD gewählt. Da die gemessene HRTF eine endliche Auflösung an Messpunkten hat, muss der Binauralisierer zwischen den verschiedenen Messpunkten interpolieren. Wenn Schallquellenpositionen zwischen zwei Messpunkten dargestellt werden sollen, wird je nach Binauralisierer der räumlich nächste Messpunkt gewählt oder die umliegenden Messpunkte werden mit einer Gewichtung interpoliert.

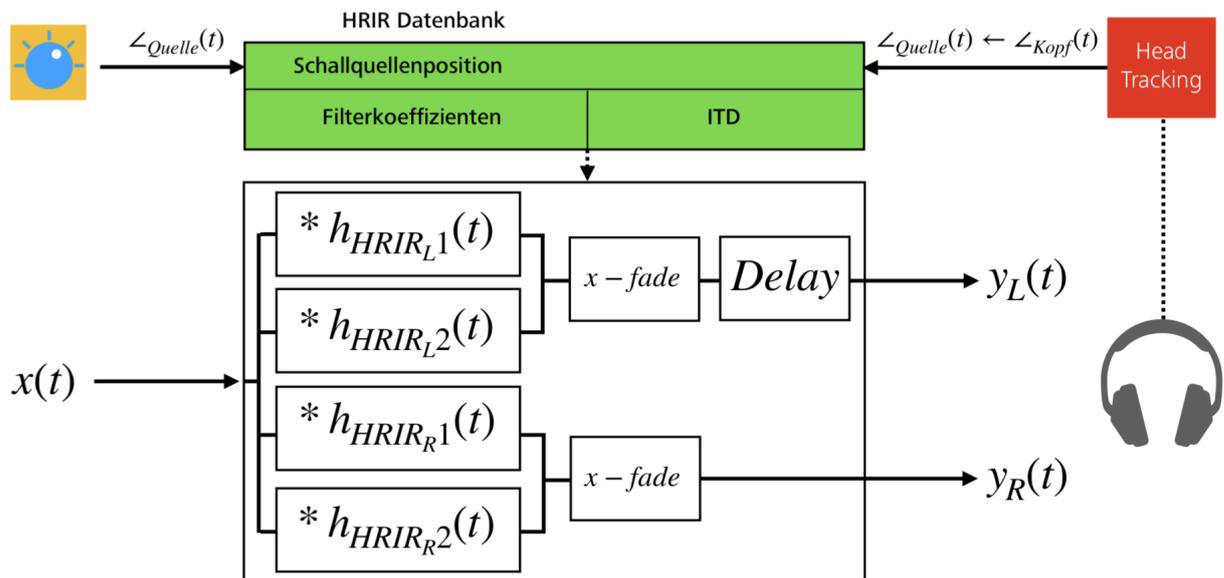


Abbildung 3.2: angepasstes System zur dynamischen Binauralsynthese [Melchior, 2021]

Durch Headtracking wird unter anderem das Problem der "Front-Back-Confusion" gelöst, da der Hörer mit wenigen Kopfbewegungen die Hörereignisorte validieren kann. [Weinzierl, 2008, S. 677]

Individualisierung der ITD

Durch dynamische Binauralsynthese ist der Einfluss einer nicht ganz passenden HRTF geringer und eine genaue Lokalisation ist weiterhin möglich. Jedoch ist eine passende interaurale Laufzeitdifferenz essentiell für einen stabilen Hörereignisort.

Ist diese nicht passend gewählt, bleibt das Hörereignis bei Kopfdrehungen nicht ortstabil, was die Lokalisation erschwert. Die Individualisierung der ITD kann über zwei

Verfahren erreicht werden: entweder messtechnisch, indem der Abstand der Ohren bestimmt und auf ein Kopfmodell übertragen wird oder interaktiv in einem dynamischen Binauralsystem. Dabei kann der Hörer die ITD skalieren, bis der Hörereignisort bei Kopfdrehungen stabil bleibt. [Lindau et al., 2010]

Relatives Headtracking

Ein Problem, welches sich durch Headtracking ergibt, ist die im Alltag ständig wandelnde Perspektive des Konsumenten, wenn dieser in Bewegung ist. Gerade bei Musikmischungen ist es essenziell, dass die grundsätzliche Hörrichtung frontal orientiert ist.

Um das zu erreichen muss sich die Normalausrichtung (Blickrichtung 0 Grad) immer wieder auf die gegebenen Umstände rekalisieren. Die Implementation eines solchen Algorithmus hat außerdem den Vorteil, dass es keine absolute Neutralausrichtung mehr gibt und deshalb einfache Gyrosensoren zum Erfassen der relativen Kopfbewegungen ausreichend sind.

Zusätzlich kann man die möglichen Rotationwinkel, die sich durch Kopfbewegungen ergeben, auf einige Grad beschränken. So sind Peilbewegungen noch möglich, welche die räumliche Auflösung erheblich verbessern. Die Soundszene jedoch dreht sich nicht um 90 Grad, wenn man zum Beispiel um eine Ecke läuft. Dadurch ist aber auch ein bewusstes Umhören in einer solchen Soundszene nicht mehr ohne weiteres möglich. [Faller et al., 2015]

Pseudo Headtracking

Ein Grund, weshalb sich die dynamische Binauralsynthese bisher nicht etabliert hat, sind die vielen Anforderungen an ein solches System.

Die Audiomischung muss als Soundszene (zum Beispiel "Higher Order Ambisonics") beziehungsweise als mehrkanaliges Audiofile vorliegen, was im Vergleich zu Stereo- beziehungsweise statischen Binauralmischungen einen vielfachen Datenaufwand bedeutet. Das Wiedergabegerät braucht deshalb genug Rechenleistung, um die dynamische Binauralsynthese in Echtzeit zu berechnen. Zudem fehlt es an Trackersystemen (zum Beispiel eingebaut in Kopfhörern) und einem etablierten Kommunikationsstandard zwischen Tracker und Renderer.

Ein Lösungsansatz ist Pseudo Headtracking. Dabei werden zufällige Kopfbewegungen

von wenigen Grad simuliert und zwischen diesen interpoliert. Auf diese Weise haben die Hörereignisse keine absolut feste Position, wodurch deren Verortung erleichtert wird. Gerade die Problematik der Differenzierung vom vorderen zum hinteren Bereich wird dadurch erleichtert und auch die Wahrnehmung von Räumlichkeit wird im Vergleich zu statischen Binauralsystemen gesteigert. [Faller et al., 2015]

4 Exemplarische Binaurale Produktionsumgebung

Da zu Demonstrationszwecken die Hörbeispiele individualisierbar und mit der Möglichkeit zur dynamischen Binauralsynthese vorliegen sollen, habe ich mich für das szenenbasierte Zwischenformat Ambisonics entschieden. Da mit Ambisonics nur Hörereignisse außerhalb des Kopfes darstellbar sind, ergibt sich die Kombination aus einem Higher Order Ambisonics- und einem Stereomaster.

4.1 Ambisonics

Ambisonics wurde in den 1970er Jahren im Wesentlichen von Michael Gerzon, einem britischen Mathematiker entwickelt. Dabei ist eine theoretisch beliebig genaue Übertragung von dreidimensionalen Schallfeldern möglich. Die Genauigkeit des Schallfeldes steigt dabei mit der Anzahl der übertragenen Kanäle und der Anzahl an Lautsprechern auf der Wiedergabeseite. [Weinzierl, 2008, S. 659] Im Gegensatz zu kanalbasierten Formaten ist weder eine feste Lautsprecheranzahl noch eine bestimmte Lautsprecheraufstellung vorgeschrieben.

4.1.1 Mathematische Grundlagen

Ambisonics basiert auf dem Prinzip der Zerlegung eines Schallfeldes in Kugelflächenfunktionen (*spherical harmonics*). Das Schallfeld $p(r)$ lässt sich durch eine Überlagerung der radial gewichteten Kugelflächenfunktionen der Ordnung m mit den sphärischen Besselfunktionen $j_m(kr)$ und einem Phasenfaktor i^m beschreiben. Die Richtung wird dabei aus der Perspektive des Hörers durch Kugelkoordinaten (Radius r , Azimuth φ und Elevation δ) angegeben. [Weinzierl, 2008, S. 660]

Es gilt folgender Zusammenhang:

$$p(r) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{0 \leq n \leq m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\varphi, \delta) \quad (4.1)$$

In Abbildung 4.1 wird die winkelabhängige Amplitude der sphärischen Harmonischen Y_{mn}^{σ} illustriert. Die Reproduktion des Schallfeldes geschieht durch die Übertragung der Komponenten B_{mn}^{σ} . Umso mehr Komponenten übertragen werden, desto hochauflösender ist die Resynthetisierung des Schallfeldes. [Weinzierl, 2008, S. 661]

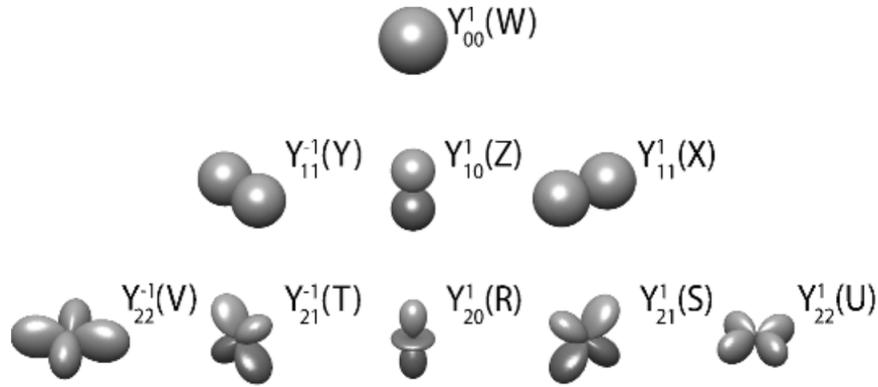


Abbildung 4.1: Sphärische Harmonische 0. bis 2. Ordnung [Weinzierl, 2008, S. 661]

Für die Übertragung eines Schallfeldes mit Ambisonics m -ter Ordnung werden $(m+1)^2$ Komponenten beziehungsweise Übertragungskanäle benötigt. [Weinzierl, 2008, S. 662]

4.1.2 Ambisonic erster Ordnung

Bei Ambisonics erster Ordnung (FOA) werden die ersten vier Komponenten der Fourier-Bessel-Reihe übertragen. Diese werden als W, X, Y und Z bezeichnet und bilden das vierkanalige FOA B-Format.

Die Komponente W stellt den Schalldruck im Ursprung dar und ist richtungsunabhängig. X, Y und Z stellen die Druckgradienten beziehungsweise die Schnellekomponenten in Richtung der drei Raumachsen dar.

Ambisonics erster Ordnung lässt sich nativ durch einen Druckempfänger im Ursprung und drei Druckgradientenempfänger mit Achtercharakteristik entlang der drei Raumachsen aufnehmen (vgl. Abb. 4.2).

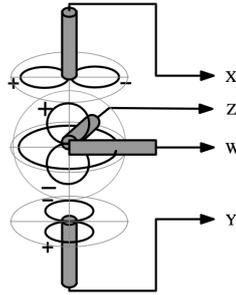


Abbildung 4.2: Natives 3D FOA Aufnahmesetup [Zotter and Frank, 2019, S. 10]

Alternativ lässt sich eine Ambisonics-Aufnahme erster Ordnung mit vier tetraedrisch angeordneten Druckgradientenempfängern mit Nierencharakteristik realisieren, was als FOA A-Format bezeichnet wird. Durch eine Matrizierung der Einzelkanäle kann zum B-Format encodiert werden (vgl. Abb. 4.3). Die Mikrofone sollten dabei so weit wie möglich koinzident platziert werden. Da dies in der Realität aufgrund der Bauform der Mikrofone nicht möglich ist, resultiert neben einem Lokalisationsfehler für hohe Frequenzen, im Diffusfeld eine Absenkung im Höhenbereich. [Zotter and Frank, 2019, S. 11]

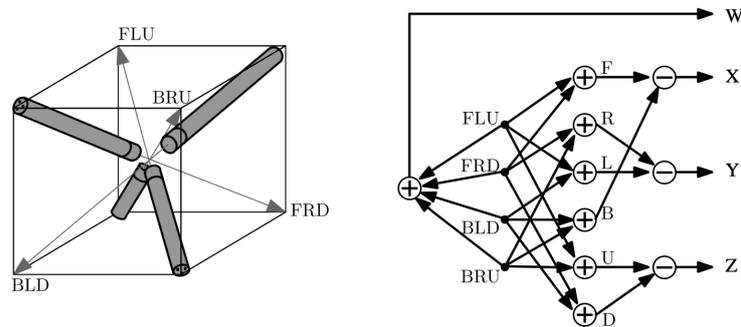


Abbildung 4.3: Tetraedrisch angeordnete Nierenmikrofone (links) und deren Encodermatrix (rechts) [Zotter and Frank, 2019, S. 11]

4.1.3 Kodierung

In Bezug auf Ambisonic bedeutet Encodierung, die Komponenten B_{mn}^σ eines realen oder virtuellen Schallfeldes zu gewinnen. Dies geschieht in realen Schallfeldern durch ein Mikrofonarray (vgl. Abb. 4.2 und Abb. 4.3) und in virtuellen Schallfeldern mithilfe eines Schallfeld-Modells.

Decodierung beschreibt die Resynthesierung des Schallfeldes für einen zentralen Hörerpunkt durch die Überlagerung von Lautsprechersignalen. Die Größe des korrekt resynthetisierten Bereichs hängt dabei von der Anzahl der übertragenen Komponenten ab.

Die Kodierung setzt voraus, dass das aufgenommene Schallfeld eine ebene Welle ist und dass die wiedergebenden Lautsprecher ebene Wellen abstrahlen. [Weinzierl, 2008, S. 662]

Eine ebene Welle wird in Abhängigkeit ihrer Einfallsrichtung durch reelle Gewichtungsfaktoren zu den Ambisonics-Komponenten encodiert, welche die Übertragungskanäle bilden. Diese müssen bei der Wiedergabe für eine konkrete Lautsprecherkonfiguration decodiert werden.

Eine Möglichkeit der binauralen Wiedergabe ist, die bereits dekodierten Lautsprecher-signale mit den HRIRs ihrer jeweiligen Position zu falten. In Abbildung 4.4 wird schematisch für den 2D Fall die dynamische Wiedergabe mit Head-Tracking beschrieben. Dabei wird das encodierte Schallfeld mithilfe einer Rotationsmatrix, entsprechend der Kopfhaltung des Hörers, rotiert, bevor es zu den einzelnen Lautsprechersignalen dekodiert wird, welche wiederum mit den HRIRs gefaltet werden.

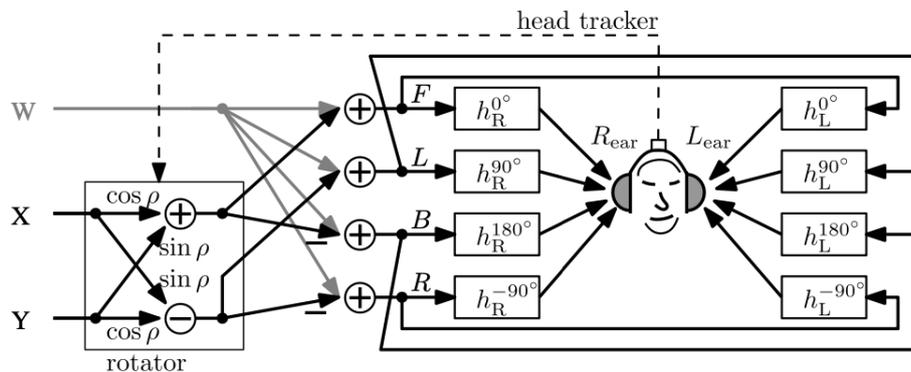


Abbildung 4.4: 2D FOA Binauraldecoder mit Head-Tracking [Zotter and Frank, 2019, S. 9]

Ein großer Vorteil eines solchen Systems ist, dass die Rotationen mathematisch leicht zu realisieren sind und somit in sehr kurzen Zeitabständen möglich werden. Außerdem bleiben die HRIRs des Binauralisierers konstant. [Zotter and Frank, 2019, S. 9]

4.1.4 Higher Order Ambisonics

Ordnungen größer als FOA werden als Higher Order Ambisonics (HOA) bezeichnet. Dadurch, dass die räumliche Auflösung von FOA sehr begrenzt ist, ähneln sich die Lautsprechersignale zwei beieinander liegender Lautsprecher stark. Mit Erhöhung der Ordnung wird die räumliche Auflösung vergrößert und damit die Korrelation benachbarter Lautsprecher herabgesetzt, was in einer verbesserten Wahrnehmung von Räumlichkeit und Tiefe resultiert. [Zotter and Frank, 2019, S. 20]

4.2 Tools

4.2.1 Reaper

Um mit HOA in größtmöglicher Auflösung als Zwischenformat arbeiten zu können, wird eine DAW mit möglichst großer Busbreite benötigt. Hierbei bietet Reaper viele Möglichkeiten, da eine Busbreite von bis zu 64 Kanälen zur Verfügung steht, womit sich HOA in der 7. Ordnung realisieren lässt.

Weitere Vorteile von Reaper sind die kostengünstige Preispolitik, die flexible Routingstruktur (die Busbreite eines Kanals kann jederzeit angepasst werden) und die Modularität. Jeder Plugin-Parameter kann mit wenig Aufwand durch einen LFO moduliert, oder über OSC ferngesteuert werden. Außerdem gibt es eine JavaScript-Einbindung, womit sich iterative Prozesse automatisieren lassen. [Francis, 2015]

4.2.2 Ambisonic Plugins

Für die Encodierung und die Bearbeitung innerhalb der Ambisonic Domäne habe ich fast ausschließlich mit der "IEM Plugin Suite" gearbeitet.

Für die Decodierung bevorzuge ich den "SPARTA AmbiBIN" der Aalto-Universität, da dieser sämtliche Decoder-Methoden unterstützt und die Möglichkeit bietet eine individuelle HRTF im SOFA-Format zu laden. [McCormack and Politis, 2019]

Im Folgenden will ich die von mir am meisten verwendeten Plugins kurz vorstellen:

IEM MultiEncoder

Der "IEM Multiencoder" bietet eine einfache Möglichkeit Mono-, Stereo- und Multichannelsignale (bis zu 64) zu encodieren. Dabei wird jeder Kanal wie ein Monokanal

behandelt und lässt sich unabhängig positionieren. Zur verbesserten Übersicht können die einzelnen Elemente unterschiedlich eingefärbt werden. Über einen Masterpunkt können alle Elemente relativ zu diesem rotiert werden.

Außerdem gibt es mit Solo, Mute und Gain einfache Mixingfunktionen, sowie die Möglichkeit der Visualisierung der gerade spielenden Elemente. [IEM, 2022]

IEM RoomEncoder

Der "IEM RoomEncoder" ermöglicht die Ambisonicencodierung in einem virtuellen quaderförmigen Raum mit über 200 Reflexionen. Die Raumdimensionen, sowie Quell- und Hörposition lassen sich dabei frei wählen. Außerdem lässt sich der Reflexionskoeffizient der einzelnen Wände und die Raumcharakteristik über Kuhschwanzfilter für Höhen und Tiefen anpassen. Mehrere "RoomEncoder"-Instanzen können miteinander synchronisiert werden, sodass Raumdimensionen, Klangcharakteristik und das Verhältnis von Direktsignal zu Reflexionen gleich bleiben.

Ferner ist es möglich den Direktschallanteil auszuschalten, womit der "RoomEncoder" als reiner Reflexionsgenerator genutzt werden kann. [IEM, 2022]

IEM MultiEQ

Der "IEM MultiEQ" ist ein einfacher Multichannel-EQ welcher die simultane Bearbeitung von bis zu 64 Kanälen ermöglicht. Dadurch lässt sich die Klangfarbe bereits encodierter und summierter HOA-Signale bearbeiten. [IEM, 2022]

IEM OmniCompressor

Der "IEM OmniCompressor" arbeitet von seiner Funktionsweise wie ein normaler Mono- oder Stereokompressor. Als Steuersignal wird der Omnidirektionale W-Kanal verwendet, wobei die resultierende Kompression auf alle Kanäle angewendet wird.

Durch die Möglichkeit einer "Lookahead"-Funktion und einer unendlichen Ratio, kann der "OmniCompressor" außerdem als Limiter eingesetzt werden. [IEM, 2022]

IEM SceneRotator

Mit dem "IEM SceneRotator" kann die Ambisonic-Soundszene entlang der drei Raumachsen rotiert werden. Dadurch lässt sich die Position einer Gruppe von Hörereignissen automatisieren. Über eine OSC-Schnittstelle können Informationen eines "Head-

Trackers" empfangen werden und die Soundszene entsprechend der Kopfposition rotiert werden. [IEM, 2022]

4.2.3 Sound Particles

Sound Particles ist eine Sounddesignsoftware, mit welcher immersive Atmos und Soundeffekte erstellen werden können. Im dreidimensionalen Raum können Partikel - welche Audiosamples repräsentieren - statisch positioniert, über einen gewählten Zeitraum freigesetzt und bewegt werden. Dabei können zahlreiche Zufallsfaktoren einbezogen werden um zum Beispiel die Position der Partikel zu variieren oder die Audiosamples zu verstimmen und zu verzögern.

Der Anwender definiert die Abhörposition, welche automatisiert werden kann und wählt den "Output-Typ". Dabei ist unter anderem die Berechnung in Ambisonics sechster Ordnung möglich. [Fonseca, 2020]

Ein Nachteil der Software ist, dass die Audiosamples vorbereitet werden müssen und innerhalb der Software nicht editierbar sind, wodurch die Erstellung und Veränderung einer Atmo sehr zeitintensiv sein kann.

4.3 Busstruktur

Der Aufbau einer Reapersession ist natürlich stark abhängig von der jeweiligen Projektanforderung. Es haben sich für mich jedoch Busstrukturen etabliert, welche ich immer wieder nutze.

Zunächst liegt das Ausgangsmaterial in Mono-, Stereo-, oder Mehrkanalspuren vor und kann, wie auch im Stereo- oder Surroundumfeld, mit beliebigen Plugins bearbeitet werden. Die Ausgänge dieser Kanäle führen zu logisch gegliederten Binauralisierungsbussen. Dabei ist zu beachten, dass die Quellkanäle auf jeweils eigenen Eingangskanälen des Binauralisierungsbusses anliegen. Diese führen in den auf dem Bus liegenden "IEM MultiEncoder" und werden in eine beliebige Ambisonicordnung encodiert.

Dadurch wird zum einen Rechenleistung gespart und zum anderen wird die räumliche Übersicht erleichtert. Von hier können die HOA-Signale in beliebig vielen weiteren HOA-Gruppen zusammengefasst und in ihrer Dynamik und Klangfarbe bearbeitet werden.

Diese führen letztendlich in einen der beiden HOA Premaster. Auf einem der beiden

liegt ein "IEM SceneRotator", welcher per OSC die Kopfhaltung der Hörerin oder des Hörers empfängt und die Soundszene entsprechend rotiert. Der zweite HOA Premaster ist für Elemente gedacht, welche trotz dynamischer Binauralsynthese statisch bleiben sollen. Die beiden HOA Premaster werden anschließend in einem HOA Master zusammengefasst und von dort an den Binauralisierer geschickt.

Parallel zum Ambisonic-Signalfad gibt es einen Stereo-Signalfad. Dieser kann einerseits für Kunstkopfaufnahmen oder Binauralisierer ohne Ambisonic-Output verwendet werden, andererseits für Hörereignisse, welche innerhalb des Kopfes lokalisiert werden sollen. Dabei sollte beachtet werden, dass die Summenbearbeitung der Stereoelemente nicht auf die binauralisierten Elemente angewendet wird, weshalb auch hier der Umweg über zwei Stereopremaster gegangen wird (vgl. Abbildung 4.5).

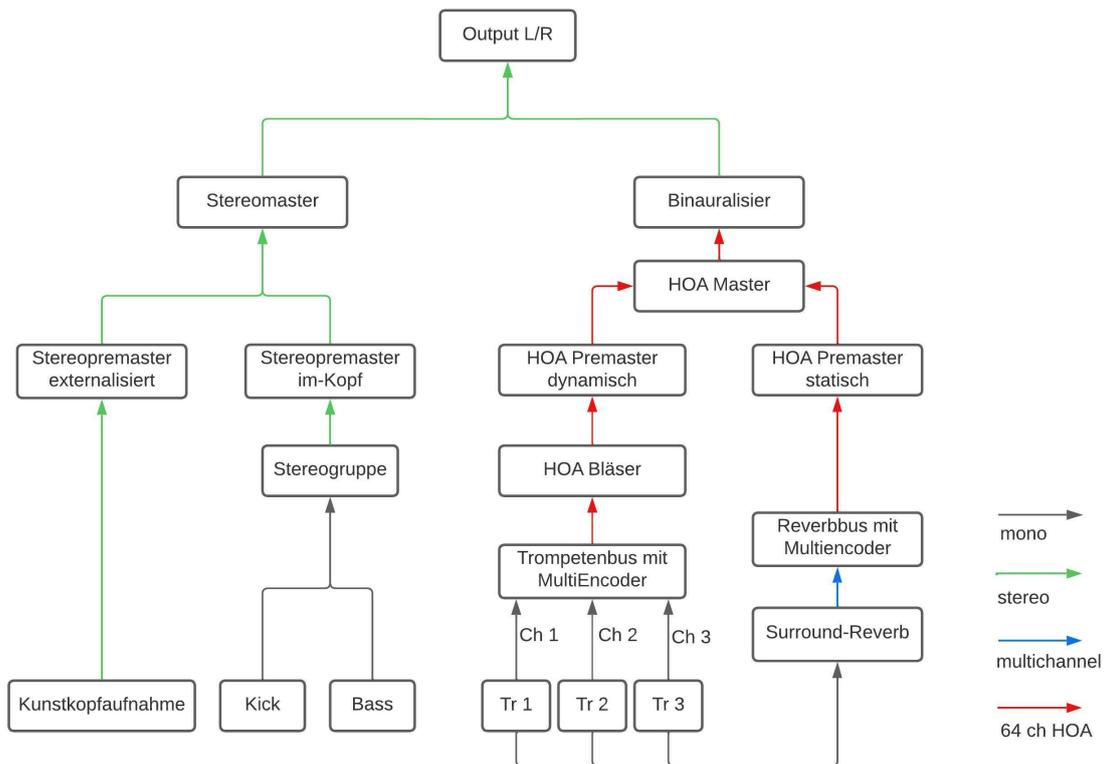


Abbildung 4.5: Exemplarisches Blockschnittbild der Busstruktur

Da mit dieser Busstruktur nur ein Binauralisierer mit Fernfeld-HRTF vorgesehen ist, lassen sich keine Hörereignisse im Nahfeld darstellen. Bei Bedarf lässt sich die Struktur durch einen weiteren "HOA Nahfeldmaster" ergänzen, welcher in einen zweiten

Binauralisierer mit Nahfeld-HRTF führt.

4.4 Effekte

Effekte, welche bei der Produktion von Stereoinhalten verwendet werden, können für die Produktion von binauralen Inhalten weiterhin verwendet werden. Die Effekte können entweder mit dem Ausgangssignal summiert und anschließend gemeinsam binauralisiert werden, oder sie werden unabhängig voneinander binauralisiert (vgl. Hörbeispiel Effekte 01). Außerdem ist eine Kombination von Ursprungssignal innerhalb des Kopfes und externalisierten Effekten möglich, wodurch sich diese besser in den räumlichen Gesamtkontext einordnen lassen (vgl. Hörbeispiel Effekte 02).

Dabei hat sich für mich aus Übersichtsgründen auch hier die Binauralisierung der Effekte eines Hörereignisses innerhalb eines "IEM MultiEncoders" bewährt.

4.4.1 Räumlichkeit

Die Räumlichkeit ist einer der wichtigsten Parameter um eine authentische dreidimensionale Soundszene zu erschaffen. Durch teilkohärente Signale aus verschiedenen Richtungen fällt die räumliche Zuordnung leichter und es entsteht der Effekt von Einhüllung. Ein räumlicher Ausklang kann in Frühe Reflexionen (Early Reflections) und Nachhall untergliedert werden.

Frühe Reflexionen

Frühe Reflexionen können dazu beitragen, die Hörereignisrichtung besser zu definieren. Dabei kann zum Beispiel der "IEM RoomEncoder" verwendet werden, welcher die Reflexionen aus allen Raumrichtungen simuliert und einen realistischen dreidimensionalen Raumeindruck vermittelt. Das ist von Vorteil, wenn nur wenige Hörereignisse aus einer Richtung zu hören sind, wie zum Beispiel bei Sprachbeiträgen (siehe Kapitel 5.1).

Bei einer komplexeren Soundszene ist der räumliche Eindruck meist schon durch die Anordnung der Hörereignisse und einen dreidimensionalen Nachhall gegeben. Hierbei können frühe Reflexionen dazu genutzt werden, die Richtung eines Hörereignisses besser zu definieren. Dabei reicht oft ein einfacher Stereohall aus. Viele Hallplugins bieten die Möglichkeit, die Lautstärke von Nachhall und frühen Reflexionen unabhängig voneinander zu regeln (vgl. Abbildung 4.6).



Abbildung 4.6: Stereo-Reverb mit unabhängiger Regelung von frühen Reflexionen und Nachhall

Die Reflexionen können zum Beispiel links und rechts vom eigentlichen Hörereignis positioniert werden. Frühe Reflexionen bieten außerdem eine Möglichkeit Hörereignisse, welche innerhalb des Kopfes verortet werden, eine externalisierte Richtung zu geben (vgl. Hörbeispiele Effekte 03 und Effekte 04).

Nachhall

Es gibt verschiedene Ansätze, einen dreidimensionalen Nachhall zu synthetisieren. Dies kann innerhalb der Ambisonic Domäne durch ein Feedback-Delay-Netzwerk (z.B. "IEM Fdn Reverb") oder durch die Faltung mit Raumimpulsantworten geschehen. Eine weitere Möglichkeit ist die Binauralisierung eines Mehrkanalreverbs beziehungsweise mehrerer Stereoreverbs (vgl. Hörbeispiel Effekte 05).

Ich habe mich in den meisten Produktionen für binauralisierte Stereo- beziehungsweise Mehrkanalreverbs entschieden, da diese viele und bekannte Einstellungsmöglichkeiten bieten und die Möglichkeit besteht, mir aus der Stereoumgebung bekannte Halltypen zu verwenden.

Als Hall-Plugin habe ich den "HOFA-IQ-Reverb" gewählt, wobei es sich um einen 5.1-fähigen Faltungshall handelt. Dabei habe ich je eine Instanz für unteres und oberes Layer verwendet. Im Vergleich von oberem zu unterem Halllayer habe ich für das obere eine längere Predelay-Zeit und eine etwas längere Nachhallzeit gewählt, wodurch der Hall nach oben hin ausklingt. Außerdem können beim oberen Layer die gesamten Tief-Mitten weggenommen werden, da diese für die obere Ebene keinen Mehrwert bringen und den Mix "vermatschen". Durch eine Anhebung der oberen Mitten (circa 2 kHz),

kann der Effekt des oberen Ausklangs noch verstärkt werden (vgl. Hörbeispiel Effekte 06).

4.4.2 Delay

Delays bieten eine gute Möglichkeit, den durch die Binauraltechnik entstandenen Raum zu nutzen und besser zu definieren. Dabei kann zum Beispiel ein Slap-Delay verwendet werden, um ein Hörereignis in eine gewünschte Richtung zu lenken und gleichzeitig auszudehnen (vgl. Hörbeispiel Effekte 07).

Außerdem ist die Platzierung mehrerer Stereo-Delays mit verschiedenen Einstellungen um den Hörer denkbar. Diese können zum Beispiel auf unterschiedliche Zählzeiten verzögert werden (vgl. Hörbeispiel Effekte 08) oder miteinander verschachtelt sein, wodurch ein komplexeres Delaynetzwerk entsteht (vgl. Hörbeispiel Effekte 09).

In ersterem Fall steigt der Grad an Umhüllung, da ein ähnliches Signal mehrfach aus verschiedenen Richtungen wahrgenommen wird, was dem auditiven System eine gewisse Lernmöglichkeit gibt.

Bei zweiterem handelt es sich nicht um ein Ping-Pong-Delay, wodurch dem auditiven System nur monoaurale Ohrsignalmerkmale zur Verfügung stehen.

5 Binaurale Inhalte

Die binauralen Gestaltungsmöglichkeiten stehen in starker Abhängigkeit zu den jeweiligen Hörfunkinhalten. Daher ist eine Unterteilung in Wort- und Musikbeiträge sinnvoll.

5.1 Sprache

Im Bereich der Wortbeiträge kann zwischen rein informativen und fiktionalen Inhalten unterschieden werden. Bei rein informativen Beiträgen steht die Sprachverständlichkeit an oberster Stelle. Außerdem ist eine mögliche Differenzierung verschiedener Sprecher wichtig, da gerade bei Gesprächsrunden auch einander ins Wort gefallen werden kann. Bei fiktionalen Inhalten steht die Vermittlung einer Geschichte im Vordergrund. Durch binaurale Elemente kann der Immersionsgrad einer Geschichte gesteigert werden, wodurch diese besser vermittelt werden kann.

5.1.1 Informative Sprachbeiträge

Zu rein informativen Sprachbeiträgen zähle ich unter anderem:

- Moderation
- Nachrichten, Wetter und Verkehr
- Interview
- Diskussionsrunden

Beim Konsum von Stereo-Hörfunkinhalten über Kopfhörer ist in der Regel die Hörereignisposition des Sprechers in der Mitte des Kopfes. Die Tonalität ist durch den Nahbesprechungseffekt geprägt, die Dynamik stark beschränkt und es ist wenig bis keine Räumlichkeit vorhanden.

Sprachverständlichkeit

Um die Sprachverständlichkeit zu ermitteln wird zwischen Silben-, Wort- und Satzverständlichkeit unterschieden. Die Satzverständlichkeit liegt grundsätzlich über der Wortverständlichkeit und diese wiederum über der Silbenverständlichkeit. Bei einer Silbenverständlichkeit von etwa 80% ist eine Satzverständlichkeit von 100% gegeben, was als eine gute Sprachverständlichkeit gilt. [Dickreiter et al., 2014, S. 72]

Die Silbenverständlichkeit ist abhängig vom Frequenzgang, dem Abstand von Nutz- zu Störsignal und den räumlichen Gegebenheiten beziehungsweise dem Verhältnis von Direkt- zu Diffusschall. Die Verständlichkeit steigt mit der Erhöhung der Grenzfrequenz, wobei bei etwa 4 kHz eine Wortverständlichkeit von 100% erreicht ist. [Dickreiter et al., 2014, S. 73] Daher begrenzen die meisten Sprachübertragungsanwendungen die Signale auf 4 kHz. Bei der Übertragung von binauralen Inhalten sind jedoch die Frequenzen oberhalb von 4 kHz wichtig für die Hörereignisbildung (siehe Abbildung 2.3), weshalb der Frequenzgang nicht beschränkt werden sollte.

eine Sprecherin/ein Sprecher

Bei einzelnen Sprecher*innen ist die erwartete Hörereignisposition, entsprechend einer Rednerin oder eines Redners, frontal vor den Hörer*innen, bei mittlerer Entfernung (> 2m).

Dadurch ergibt sich bei statischer Binauralsynthese die Problematik der "Front-Back-Confusion". In *Hörbeispiel Sprache 01* ist ein einzelner Sprecher zu hören, welcher ohne zusätzliche Raumanteile erst vorne, dann hinten binauralisiert ist. Aus subjektiver Sicht klingen die Textabschnitte, welche nach hinten binauralisiert wurden lediglich dumpfer, was im direkten Vergleich auf die richtige Hörereignisposition schließen lässt. Ohne diesen direkten Vergleich wäre es vermutlich schwierig, den richtigen Hörereignisort zu bestimmen.

Fügt man der trockenen Sprachaufnahme räumliche Reflexionen hinzu (Hörbeispiel Sprache 02), so wird der Unterschied deutlicher und die räumliche Einordnung fällt leichter. Allerdings ist dadurch eine gewisse Räumlichkeit gegeben, welche gerade bei Moderationen untypisch ist.

Durch die Binauralisierung verändert sich außerdem die Klangfarbe des Sprechers. Diese Veränderung lässt sich in gewissem Maße kompensieren, indem man den Grundtonbereich des Sprechers leicht anhebt. Durch die hörfunktypische starke Kompression und den ausgeprägten Grundton wirkt der Sprecher trotz Binauralisierung sehr na-

he und die Externalisierung ist nicht so stark ausgeprägt. Dieses Phänomen ließe sich durch mehr Räumlichkeit kompensieren, was meiner Meinung nach für Moderationen oder Ähnliches nicht notwendig ist.

Ich verfolge den Ansatz, die typischen Klangideale beizubehalten und nur eine gewisse Externalisierung zu erreichen, um nicht mit anderen binauralen Hörfunkinhalten zu brechen. Ein weiterer möglicher Ansatz wäre die Sprecher*innen nicht zu binauralisieren und nur binauralisierte Reflexionen hinzuzufügen (vgl. Hörbeispiel Sprache 03).

mehrere Sprecher*innen

Wenn bei einem Interview oder Diskussionsrunden mehrere Personen beteiligt sind, ist es wichtig, dass die Hörer*innen zwischen den Sprecher*innen differenzieren können, damit die jeweiligen Aussagen der richtigen Person zugeordnet werden können. Wie in Kapitel 2.3 erläutert, spielen hierbei unter anderem Tonalität und räumliche Herkunft eine wichtige Rolle. Die Tonalität ist durch die jeweiligen Sprecher*innen gegeben und lässt sich entsprechend nur mit der Auswahl dieser beeinflussen. Die räumliche Herkunft, beziehungsweise der Hörereignisort lassen sich in einem binauralen System und wenn die einzelnen Direktsignale vorliegen, weitestgehend frei manipulieren.

Bei der Wiedergabe von Stereo-Radio über Kopfhörer befinden sich die sprechenden Personen entweder in der Kopfmittle und lassen sich aufgrund ihrer Tonalität unterscheiden (zum Beispiel eine Frau interviewt einen Mann) oder die Sprechenden werden über Amplitudenpanning auf der interauralen Achse verteilt um eine bessere Differenzierung zu ermöglichen.

In einer Binauralmischung können die beteiligten Sprecher*innen, entsprechend einer Podiumsdiskussion vor der Hörerin oder dem Hörer positioniert werden. Dadurch ist eine dem Hörenden bekannte Szene gegeben, was eine Einordnung dieser leichter macht. Da frontal vor dem Hörenden eine Lokalisationsgenauigkeit von wenigen Grad gegeben ist, fallen bereits geringfügige Abweichungen auf und erleichtern die Zuordnung. Gleichzeitig lassen zu große Abstände zwischen den Hörereignisorten die auditive Szene zerfallen und die Sprecher*innen wirken nicht so, als ob sie in einem Raum sitzen würden. Durch das Hinzufügen von Reflexionen kann man diesem Effekt in gewissem Maße entgegenwirken, wodurch größere Abstände und damit eine bessere Differenzierung möglich ist (vgl. Hörbeispiel Sprache 04).

Bei statischer Binauralsynthese ergibt sich die Problematik, dass der Kopf automatisch

der sprechenden Person zugewendet wird, die Soundszene aber statisch bleibt.

In Gesprächssituationen, in welchen sich mehrere Sprechende überlagern, fällt ein deutlicher Unterschied der Sprachverständlichkeit zwischen diotischer und dichotischer Beschallung auf. Beim binauralen Ansatz habe ich wieder versucht eine reale Gesprächssituation nachzubilden (vgl. Hörbeispiel Sprache 05). Bei dynamischer Binauralsynthese ergibt sich der Mehrwert, dass der Hörende seine Aufmerksamkeit, wie in einer realen Hörsituation, besser auf einen Sprechenden lenken kann. Während für mich bei statischer Binauralsynthese die Positionierung der Sprecher recht breit und damit weit voneinander entfernt wirkt, fällt diese bei dynamischer Binauralsynthese weniger ins Gewicht.

5.1.2 Fiktive Hörfunkbeiträge

Bei fiktiven Hörfunkinhalten steht die Vermittlung einer Geschichte im Vordergrund. Durch die zusätzlichen Gestaltungsmöglichkeiten, welche die Binauralsynthese ermöglicht, lässt sich der Grad an Immersion steigern, wodurch diese besser vermittelt werden können.

Im Folgenden will ich anhand einer circa einminütigen Soundszene auf unterschiedliche Gestaltungsmöglichkeiten eingehen (vgl. Hörbeispiel Hörspiel 01).

Atmo und Einhüllung

Für die Basis einer akustischen Atmosphäre (Atmo) bieten Kunstkopfaufnahmen eine gute Grundlage. Da die Hörereignisorte einer Atmo in der Regel kein kritischer Faktor für die Handlung ist, können Grundproblematiken wie die "Front-Back-Confusion" hierbei vernachlässigt werden. Diese Grundatmo kann durch weitere, nachträglich binauralisierte Elemente ergänzt werden, um die akustische Umgebung besser zu definieren (vgl. Hörbeispiel Hörspiel 02). Um Hörereignisse (in meinem Beispiel das Motorengeräusch) ausgedehnt und nicht so scharf lokalisierbar darzustellen, kann zum Beispiel der "Directivity Shaper" verwendet werden (vgl. Hörbeispiel Hörspiel 03).

Mit dem "Directivity Shaper" können die verschiedenen Frequenzbereiche des Ausgangssignals an unterschiedlichen Positionen im Raum verteilt werden. Die Richtwirkung der Frequenzbänder kann über die Ambisonic-Ordnung und die Wahl der Gewichtungsfaktoren beeinflusst werden. So ist es beispielsweise möglich, das Dröhnen des Motors, also den Bassanteil, mit Ambisonic 0. Ordnung zu binauralisieren, wo-

durch dieser nur auf dem W-Kanal liegt und aus allen Richtungen wiedergegeben wird (vgl. Kapitel 4.1). Dadurch steigt die Lautstärke des Bassanteils, weshalb dieser zur Kompensation abgesenkt werden muss (siehe Abbildung 5.1).

Naheliegend ist die Positionierung entsprechend der "Richtungsbestimmenden Bänder", mit tieffrequenten Anteilen im unteren Bereich und hochfrequenten Anteilen weiter oben.[Buff, 2020, S. 87]

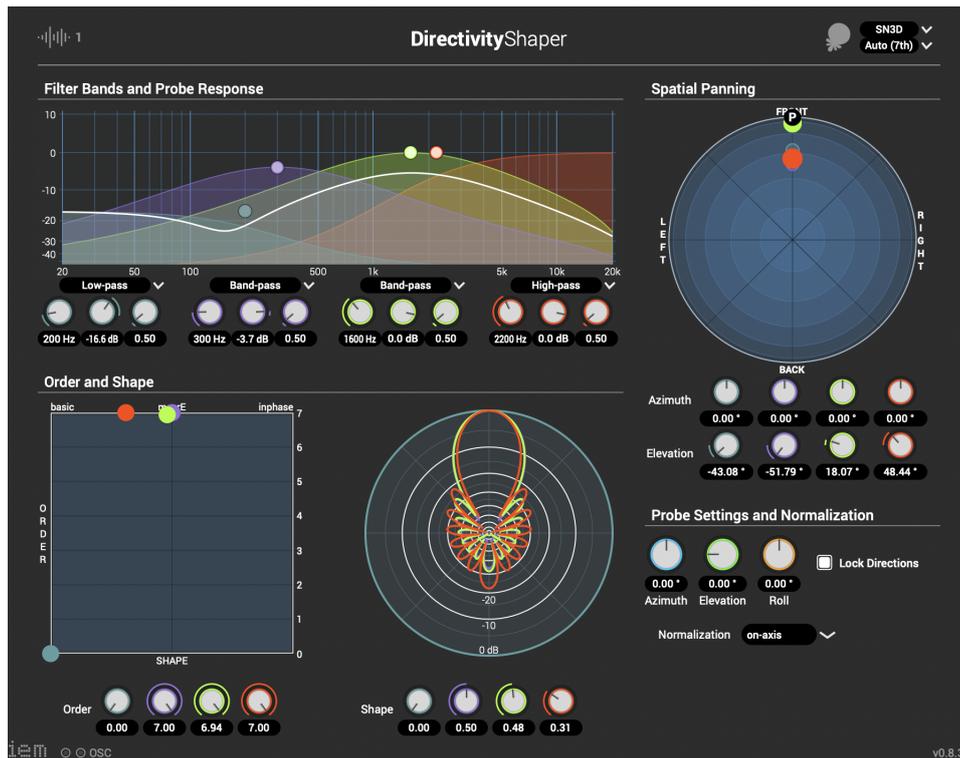


Abbildung 5.1: Directivity Shaper

Eine weitere Möglichkeit, eine Stereo-Atmoaufnahme einhüllender zu gestalten, ist diese zu duplizieren, an der x- beziehungsweise y-Achse zu spiegeln und zeitlich zu versetzten (vgl. Hörbeispiel Hörspiel 04). Gerade bei dynamischer Binauralsynthese ist es wichtig, dass Atmoelemente wie das Fahrgeräusch oder der Regen auf dem Autodach nicht zu konkret lokalisierbar sind, was durch die oben beschriebenen Methoden erreicht werden kann.

Durch weitere konkrete Hörereignisse aus verschiedenen Richtungen, kann das Maß an Einhüllung gesteigert werden. Hierbei eignen sich Stereoaufnahmen, in welchen sich bewegende Schallquellen aufgenommen wurden, oder welche, die schon im Stereo breit

und einhüllend wirken, besonders gut. Durch das Phänomen der Summenlokalisation (vgl. Kapitel 2.2) entsteht ein Hörereignis zwischen den beiden virtuell im Raum platzierten Schallquellen.

Im seitlichen Bereich weist die Summenlokalisation starke individualspezifische Abweichungen auf, was bei einer statischen Binauralsynthese zu sehr unterschiedlichen Hörereignisorten führen kann. Da diese Elemente aber nur dazu dienen, das Maß an Einhüllung zu steigern und keine wichtigen Elemente der Handlung enthalten, kann diese Problematik meines Erachtens vernachlässigt werden.

Bei konkreten Hörereignissen ist jedoch zu beachten, dass durch die "Front-Back-Confusion" bei statischer Binauralsynthese, Hörereignisse an Orten wahrgenommen werden, welche mit Erfahrungswerten brechen (z.B. ein Blinkergeräusch von hinten). Dadurch kann die auditive Szene unglaublich werden, was das Ziel einer Erhöhung des Immersionsgrades verfehlt.

Beispiel: vorbeifahrende Autos

Als Beispiel für bewegliche Atmoelemente lassen sich in meiner exemplarischen Soundscene die vorbeifahrenden Autos anführen. In der Realität nehmen wir von vorne nach hinten fahrende Elemente wahr, indem das Element sich mit seinem charakteristischen Klang aus der Frontalebene ankündigt, dann an der zugewandten Seite vorbeizieht (wobei je nach Geschwindigkeit ein mehr oder weniger ausgeprägter Dopplereffekt entsteht) und anschließend nach hinten hin verklingt. [Buff, 2020, S.29]

Realisiert werden kann der Effekt eines vorbeifahrenden Autos, indem bei einer Stereoaufnahme ein Kanal mit seitlicher Auslenkung vorne und ein Kanal mit gleicher Auslenkung nach hinten binauralisiert wird (vgl. Abbildung 5.2).

Dabei kündigt sich das Fahrzeug von links-vorne an. Der Dopplereffekt wird unmissverständlich mit einem vorbeifahrenden Auto assoziiert und durch die interaurale Pegeldifferenz ist es eindeutig auf welcher Seite das Fahrzeug vorbei fährt. Anschließend verklingt es in hinterer Richtung (vgl. Hörbeispiel Hörspiel 05).

Erzählperspektiven

Aufgrund der Tatsache dass die Hörerin oder der Hörer von der Soundscene umgeben ist, wird sie/er automatisch Teil der Szene. Dabei kann sie/er zum Beispiel als stiller, unabhängiger Betrachter auftreten und die Geschichte aus Sicht des Erzählers verfol-

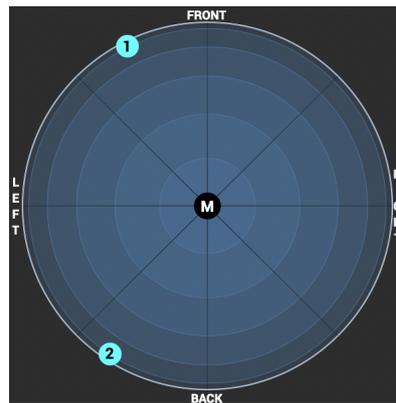


Abbildung 5.2: Positionierung der Einzelkanäle der Stereoaufnahme eines vorbeifahrenden Autos

gen. Durch die entsprechend distanzierte Perspektive würde sich, für eine realitätsnahe Darstellung, ein Großteil der Handlung im Frontalbereich abspielen, da dies der Blickrichtung des Erzählers entspricht. Handlungskritische Hörereignisse aus hinteren- und seitlichen Richtungen würden eher zu Verwirrung führen, deshalb wären die Richtungen hauptsächlich für Atmos nutzbar. Eine Erzählerstimme innerhalb des Kopfes wäre denkbar, um einen Kontrast zwischen dem Erzähler und der "realen Umgebung" zu schaffen.

Eine weitere Möglichkeit ist, die Geschichte aus der Sicht eines Protagonisten zu erzählen. Ich habe mich bei meiner Soundszene für diese "akustische Ego-Perspektive" entschieden, da meines Erachtens gerade bei dynamischer Binauralsynthese der erreichbare Immersionsgrad am größten ist. Die grundlegende Hörereignisrichtung sollte trotzdem nach vorne orientiert sein, damit sich der Hörer nicht ständig neu ausrichten muss, um der Handlung zu folgen.

Wenn aus Sicht des Protagonisten erzählt wird, können alle Schallereignisse der Umwelt entsprechend der Realität externalisiert dargestellt werden. Es bleibt jedoch die Frage, wie die Stimme des Protagonisten selbst dargestellt wird.

Der naheliegendste Ansatz wäre, diese in der Kopfmittle zu platzieren und die tonalen Veränderungen, welche sich durch den Körper- beziehungsweise Knochenschall ergeben, nachzuvollziehen. Dies würde in einer Anhebung des Grundton- und Bassbereichs des Sprechenden resultieren. Da die Hörerin oder der Hörer fremde Stimmen immer ohne diese Anhebung hört, könnten Irritationen auftreten, weshalb ich mich für eine neutrale Klangcharakteristik entschieden habe. Um den Protagonisten trotz der Stimme

innerhalb des Kopfes in der akustischen Szene zu etablieren, kann mit externalisierten Reflexionen gearbeitet werden (vgl. Hörbeispiel Hörspiel 06).

In einer "akustischen Ego-Perspektive" wären zusätzlich die Gedanken und Gefühle des Protagonisten darstellbar. Als Hörereignisort würde sich der Bereich innerhalb des Kopfes anbieten. Dabei könnte sich während der Gedankengänge des Protagonisten eine Art "akustischer Schleier" zum Beispiel in Form eines Tiefpassfilters oder Halls über die Schallereignisse der Umwelt legen, um den Kontrast zwischen der Realität und den Gedanken zu verstärken.

Beispiel: Telefonat

In der vorliegenden Soundszene wird ein Telefonat etabliert, welches eine weitere akustische Perspektive bietet. Dadurch dass der Ohrhörer an eines der Ohren gehalten wird, entsteht eine monotische Beschallung. Die Stimme ist außerdem aufgrund der Telefonübertragung sowie des kleinen Lautsprechers in der Hörmuschel in ihrem Frequenzgang beschränkt und hat einen relativ hohen Anteil an unharmonischen Verzerrungen.

Diese Phänomene können nachvollzogen werden, indem das Sprachsignal mit einem Bandpassfilter beschränkt und mit einem Verzerrungsplugin verzerrt wird. Anschließend wird es nur auf einen der beiden Stereokanäle geroutet. Um den Telefoneffekt zu verstärken habe ich außerdem ein Rauschen zum Ohrsignal dazugemischt.

Eine Alternative wäre die Aufnahme mit einem Kunstkopf, an dessen Ohr eine Hörmuschel gehalten wird (vgl. Hörbeispiel Hörspiel 07).

Um eine realistische Darstellung der Soundszene während des Telefonats zu gewährleisten, müssten eigentlich die externalisierten Hörereignisse für das rechte Ohrsignal abgedämpft werden. Dies wäre technisch realisierbar, indem hinter den Binauralisierer auf den jeweiligen Kanal die Höhen gedämpft würden. Da dies bei einer dynamischen Binauralsynthese schwer zu realisieren ist, habe ich mich dagegen entschieden.

Beim Anhören einer solchen Soundszene entsteht im Kopf des Hörenden ein Bild der akustischen Szene. Je bekannter die Szene ist, desto leichter ist es diese plausibel darzustellen. Außerdem können visuelle Hinweise in Form von Cover-Bildern oder Booklets helfen die auditive Szene richtig einzuordnen. Am Beispiel meiner Szene könnte man exemplarisch das Bild einer Frau zeigen, die mit einem Handy am Ohr am Steuer zeigen.

5.2 Musik

5.2.1 E-Musik

Bei ernster Musik spielt der Raum, in welchem die musikalische Darbietung statt findet, eine erhebliche Rolle und trägt zu einem großen Teil zur Gestaltung des Klangbildes bei. Aus diesem Grund wird in der Regel bei klassischen Aufnahmen ein Hauptmikrofon mit viel Raumanteil verwendet, zu dem Stützmikrofone mit wenig Raumanteil dazu gemischt werden können.

Bei Stereoaufnahmen wird als Hauptmikrofon meist eine AB-Stereofonie verwendet. Dabei werden normalerweise zwei Mikrofone mit Kugelcharakteristik parallel versetzt zueinander aufgestellt. Bei Schallquellen, welche sich nicht im Nahfeld befinden, ist die resultierende Pegeldifferenz minimal. Stattdessen entsteht eine räumliche Abbildung aufgrund von Laufzeitdifferenzen, deren Ausprägung mit der gewählten Basisbreite beeinflusst werden kann. [Weinzierl, 2008, S. 112]

Für eine binaurale Aufnahme kann anstatt einer AB-Stereofonie, ein Kunstkopf oder ein 3D-Mikrofonarray als Hauptmikrofon verwendet werden. Die Stützmikrofone müssten dann entsprechend des Hörereignisortes des jeweiligen Instrumentes binauralisiert werden. In *Hörbeispiel Musik 01* ist eine Orgelaufnahme zu hören, welche mit einer AB-Stereofonie, einem Neumann KU-100 Kunstkopf und einen Shoeps ORTF 3D aufgenommen wurde. Die Mikrofone des ORTF 3D wurden an den typischen Lautsprecherpositionen binauralisiert (siehe Abb. 5.3) [ITU, 2022].

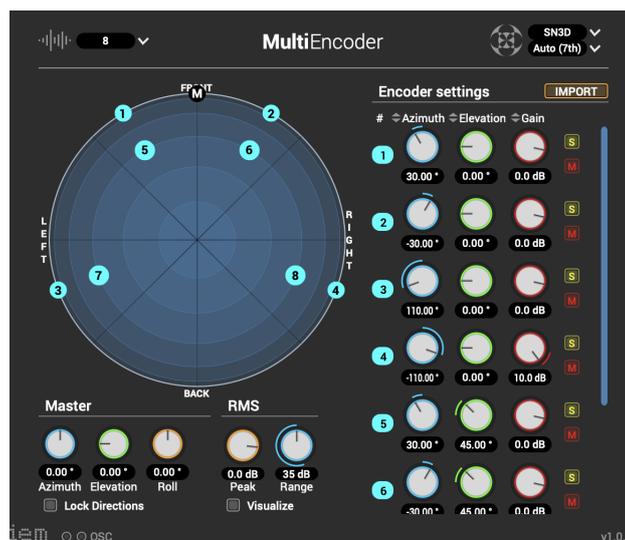


Abbildung 5.3: Positionierung der Einzelkanäle des ORTF 3D

Im direkten Vergleich fallen neben klangfarblichen Unterschieden, große Unterschiede in der Darstellung der Räumlichkeit auf.

Die AB-Stereofonie hat gerade für die hohen Frequenzen eine sehr hohe räumliche Auflösung, bietet aber im Vergleich zu den beiden anderen Aufnahmen durch die Im-Kopf-Lokalisation ein geringeres Maß an Umhüllung. Die Kunstkopfaufnahme wirkt ähnliche detailliert und gleichzeitig sehr einhüllend. Im Vergleich zur AB Variante, bei welcher der Tiefmitten- und Bassbereich recht eindeutig in der Kopfmittle lokalisiert ist, wirken die tiefen Frequenzen in der Kunstkopfaufnahme sehr diffus und drückend, was den Effekt der Umhüllung verstärkt aber im Vergleich zum Stereo weniger differenziert wirkt.

Beim ORTF 3D ist die Klangfarbe und das Maß an Umhüllung abhängig von der Binauralisierungsmethode. Bei der Faltung mit einer HRIR ist das Maß an Umhüllung im Vergleich zur Kunstkopfaufnahme recht gering. Durch die Binauralisierung mithilfe einer BRIR kann sowohl der räumliche Eindruck als auch die Klangfarbe beeinflusst werden (vgl. Hörbeispiel Musik 02).

Im direkten Vergleich wirken sowohl die Kunstkopfaufnahme als auch das binauralisierte ORTF 3D räumlicher und bilden daher die auditive Szene für meinen Geschmack besser ab. Jedoch entsteht eine Art "Leere" im Kopf, wo zuvor das Hörereignis der AB-Aufnahme befand. Daher habe ich den Ansatz verfolgt, die beiden Aufnahmen zu mischen. Dabei habe ich den Kunstkopf beziehungsweise das ORTF 3D als Hauptmikrofon verwendet und die AB-Stereofonie dazugemischt (vgl. Hörbeispiel Musik 04).

5.2.2 U Musik

U-Musik verfolgt nicht den Anspruch der Abbildung eines realistischen Raumes, in welchem sich alle Hörereignisse des Musikstückes abspielen. Vielmehr ist die Räumlichkeit und die Lokalisation der Hörereignisse eines von vielen Gestaltungsmerkmalen, welche eingesetzt werden können um die Intention einer Komposition zu unterstützen. Im folgenden Abschnitt will ich anhand von 3 Binauralmischungen auf mögliche Gestaltungsansätze eingehen (vgl. Hörbeispiele "replace me", "Feeling Good" und "Change"). Im Klangideal der Populärmusik spielt die Direktheit von Hörereignissen eine wichtige Rolle. Bei der Wiedergabe über Kopfhörer werden Lead-Vocal, Kick-Drum und Snare-Drum fast ausschließlich in der Kopfmittle, der Mono-Mittle wahrgenommen, denn "Mono ist Druck und Botschaft" und damit "[...] das Fundament von Popmusik [...]"

[Buff, 2020, S. 25]

Die vordergründigen Mono-Elemente können mit Stereo-Elementen, wie zum Beispiel Dopplungen ergänzt werden, wodurch deren Durchsetzungskraft noch verstärkt werden kann.

Mono, Stereo und 3D Audio



Abbildung 5.4: Visualisierung möglicher Hörereignisorte bei der Kopfhörerwiedergabe (rot = Mono; blau = Stereo; grau = 3D Audio)

In Bezug auf die binaurale Reproduktion von dreidimensionalen Audioinhalten kann die Mono-Mitte entweder wie bei einem Lautsprecher-Setup als Frontal vor der Hörerin oder dem Hörer, oder wie bei der Wiedergabe von Stereoinhalten über Kopfhörer als in der Kopfmitte betrachtet werden. Da eine Binauralisierung immer mit einer gewissen Verfärbung und Indirektheit einhergeht, habe ich mich für das Mono in der Kopfmitte entschieden. In diesem Ansatz kann Stereo als eine Erweiterung der Mono-Elemente betrachtet werden und die binauralisierten Elemente als Erweiterung des Mono-Stereo-Konstrukts (siehe Abb. 5.4).

Bei einem Big-Band-Arrangement kann man dieses Modell anwenden, indem Gesang, die Kick-Drum, Snare und E-Bass in der Kopfmitte platziert werden. Auf Stereoebene werden die Elemente durch die anderen Instrumente der Rhythmusgruppe erweitert.

Die binaurale Dimension stellen Bläser, Percussion, Schlagzeug-Overheads, sowie Effekte und Hall dar. (vgl. Hörbeispiel Musik 05) Dieser Ansatz verfolgt dabei weniger den Ansatz einer realistischen Darstellung, als vielmehr die Unterstützung des vertikalen Arrangements. (vgl. Hörbeispiel Musik 06)

Positionierung und Lokalisation

Bei der Positionierung von einzelnen Klangelementen sind der Kreativität prinzipiell keine Grenzen gesetzt. Dabei sollte aber bedacht werden, dass die erwartete Hörrichtung eines Musikstücks, entsprechend einer Bühne, in frontaler Richtung liegt, weshalb tragende Elemente im vorderen Bereich positioniert werden sollten.

Außerdem ist zu bedenken, dass die binaurale Darstellung einer 360 Grad Soundszene, bei statischer Binauralsynthese an ihre Grenzen kommt. Die Sphäre, welche in Abbildung 5.4 in grau dargestellt ist, entspricht nicht der Wahrnehmung einer binauralisierten dreidimensionalen Soundszene. Stattdessen ist der darstellbare Raum im vorderen und hinteren Bereich sehr beschränkt und zu den Seiten hin sehr weiter ausgedehnt (siehe Abb. 5.5).



Abbildung 5.5: Wahrnehmung einer binauralisierten dreidimensionalen Soundszene [Buff, 2020, S. 30]

Dieser Effekt der "bohnenförmigen Darstellung" relativiert sich bei der Wiedergabe mit dynamischer Binauralsynthese, da die subjektive Verzerrung des Raumes immer entsprechend der Kopfhaltung geschieht. Zudem kann man diesem Phänomen in der Mischung entgegenwirken, indem man zum Beispiel in die jeweiligen Bereiche keine vordergründigen Elemente platziert (vgl. Hörbeispiel Musik 07) oder einen zusätzlichen Stereohall oder Delays nach vorne positioniert, wodurch in diesem Bereich mehr Tiefe entsteht.

Wie oben beschrieben ist der für mich bevorzugte Hörereignisort der Gesangsstimme in der Kopfmittle. Dadurch wird die Stimme zwar nicht verfärbt und hat so das größtmögliche Maß an Direktheit, wird aber im räumlichen Kontext ohne konkrete Richtung dargestellt. Dies kann sie etwas separiert im Kopfinneren wirken lassen, was bei dynamischer Binauralsynthese zu Irritationen führen könnte.

Durch binauralisierte Reflexionen kann dem entgegengewirkt werden und auch Slap-Delays tragen dazu bei, die Gesangsstimme einer konkreten Richtung zuzuordnen (vgl. Hörbeispiel Musik 08).

Dadurch dass ich die Direktsignale eines Schlagzeugs, aufgrund der Direktheit von Mono, im Kopffinneren platziere, ergibt sich die Frage, was mit den Schlagzeug-Overheads geschieht. Ein Ansatz ist, diese ebenfalls nicht zu binauralisieren. Dies führt jedoch zu einer starken Trennung des Schlagzeugs gegenüber den externalisierten Hörereignissen, was mit einem externalisierten Hall in gewissem Maße kompensierbar wäre.

Eine weiterer Ansatz ist, die Overheads zu binauralisieren. Durch die Positionierung nach vorne und einen Basiswinkel von 60 Grad bekommt das Schlagzeug eine konkrete Hörereignisrichtung und ist damit besser lokalisierbar. Im Vergleich zur Version ohne binauralisierte Overheads wirkt das Schlagzeug aber weniger einhüllend. Durch eine Vergrößerung des Basiswinkels kann das Maß an Einhüllung gesteigert werden, aber gleichzeitig wird das Schlagzeug wieder schwerer lokalisierbar und es entstehen hörbare Verfärbungen (vgl. Hörbeispiel Musik 09).

Umhüllung

Das Gefühl von Umhüllung wird beim Konsum von Musik als Qualitätsmerkmal wahrgenommen. Eine naheliegende Möglichkeit Umhüllung zu erzeugen, ist Hall. Wie in Abschnitt 4.4 beschrieben bevorzuge ich einen nachträglich binauralisierten Stereo- beziehungsweise Mehrkanalhall. Ein entscheidender Aspekt bei der räumlichen Wahrnehmung des Halls ist die Positionierung der einzelnen Kanäle. Bei einem 5.1-Surroundhall können die Signale wie bei einem 5.1-Kreis verteilt werden (± 30 Grad und ± 110 Grad). Verteilt man die Signale weiter im Raum (± 45 Grad und ± 135 Grad), wirkt der Hall größer und einhüllender. Im Frontalbereich entsteht jedoch eine Art Lücke in der Räumlichkeit, welche durch einen binauralisierten Stereohall mit relativ kleiner Basisbreite (± 20) und längerem Pre-Delay gefüllt werden kann. Dadurch wirkt man gleichzeitig der bohnenförmigen binauralen Wahrnehmung (siehe Abb. 5.5) entgegen (vgl. Hörbeispiel Musik 10).

Eine weitere Möglichkeit das Maß an Umhüllung zu beeinflussen, ergibt sich durch die Positionierung der einzelnen Hörereignisse. Allgemein lässt sich feststellen: liegen die Hörereignisse nah beieinander und in der selben Richtung so ist das Maß an Umhüllung gering. Liegen sie hingegen weiter auseinander, steigt das Maß an Umhüllung. Dabei ist

jedoch zu beachten, dass ähnlich wie bei Stereo bei zu großem Abstand die Hörereignisse nicht mehr als eine Einheit wahrgenommen werden und zerfallen.

Bei der Wahl der Basisbreite einer Stimmgruppe, sprich der Winkelbereich in welchem die Instrumente der Stimmgruppe verteilt sind, ist dieses Phänomen entscheidend. Wählt man zum Beispiel einen Winkelbereich von 360 Grad, sind die einzelnen Hörereignisse nicht mehr konkret lokalisierbar, wodurch eine drückende Wirkung entsteht. Als Resultat wirkt die auditive Szene eher weniger einhüllend, da die Stimmgruppen keiner direkten Richtung zugeordnet werden können (vgl. Hörbeispiel Musik 11 und Hörbeispiel Musik 12).

Vocalproduction

In der aktuellen Pop-Musik spielt die Vocalproduction eine sehr wichtige Rolle. Neben der Hauptgesangslinie, besteht diese unter anderem aus Dopplungen, Harmonien und hintergründigen Chören.

Die Vielzahl an inkohärenten Spuren bietet eine gute Möglichkeit, um die Hörerin oder den Hörer mit der Stimme der Sängerin oder des Sängers einzuhüllen, sodass das Gefühl entsteht, von der Stimme umgeben zu sein. Dabei muss bei der Aufnahme und dem Editieren der Aufnahmen sehr sauber und genau gearbeitet werden, da kleine Phrasierungs- und Tonhöhendifferenzen, welche sich innerhalb einer Stereomischung verspielen, bei einer Binauralmischung, je nach Platzierung der einzelnen Hörereignisse, deutlicher wahrnehmbar sind. Im folgenden Absatz will ich darauf eingehen, welche Positionierungen welcher Elemente sich für mich bewährt hat.

Dopplungen dienen der Betonung einzelner Wörter und Phrasen. In Stereomischungen werden diese meist ganz links und ganz rechts positioniert, um die Durchsetzungskraft zu verstärken und einen Kontrast zur sonst recht scharf lokalisierten Hauptgesangslinie zu erzeugen. Bei einer Binauralmischung kann entweder der gleiche Ansatz wie bei Stereo gewählt werden, um die Stimme innerhalb des Kopfes zu verbreitern, oder die Dopplungen werden mit großem Basiswinkel breit binauralisiert.

Ich habe mich bei der Mischung von "replace me" für letzteren Ansatz entschieden, da die Stimme dadurch mehr Durchsetzungskraft gegenüber anderen binauralisierten Elementen gewinnt und außerdem die Hörereignisrichtung der Stimme in den Frontalbereich gelenkt wird (vgl. Hörbeispiel Musik 13).

Harmonien erweitern die Gesangsmelodie zu einem harmonischen Konstrukt und werden oft so dezent zur Hauptmelodie gemischt, dass sie eher unterbewusst wahrgenom-

men werden. Bei dem mir zu Demonstrationszwecken vorliegenden Song "replace me" sind die Harmoniestimmen außerdem mit sehr viel Luft gesungen. Aus diesem Grund habe ich den Ansatz verfolgt, die Harmonien so zu mischen, als ob sie von hinten in die Ohren geflüstert werden. Dafür habe ich die Harmoniestimmen zwischen ± 90 - 130 Grad positioniert und mithilfe einer Nahfeld-HRTF binauralisiert. Um den "Flüstereffekt" zu verstärken kann man den Frequenzbereich um 1,5 kHz etwas anheben.

Chöre positioniere ich gerne rund um die Hörerin oder den Hörer herum, da bei dynamischer Binauralsynthese so das Maß an Einhüllung gesteigert werden kann. Dabei präferiere ich die tieferen Stimmen im hinteren Bereich und die höheren Stimmen im Frontalbereich. Bei statischer Binauralsynthese ergibt sich beim Maß an Umhüllung kein großer Unterschied, ob die tiefen Stimmen breit hinten (± 135 Grad) oder breit vorne (± 45 Grad) positioniert sind (vgl. Hörbeispiel Musik 14).

Kontraste

Musik lebt von Kontrasten. Diese können jeglicher Art sein, unter anderem anhand der Umhüllung. Bei Stereoproduktionen wird der Chorus meist breiter als der Verse gemischt, um das Arrangement zu unterstützen. Bei 3D Audio lässt sich dieser Effekt verstärken und es besteht die Möglichkeit einen Mix, an einem Punkt des Arrangements zusammenfallen, beziehungsweise aufgehen zu lassen.

Dies geschieht zu aller erst anhand der Positionierung der Hörereignisse. Durch eine Automation des Hallanteils kann der Effekt noch verstärkt werden und die Hörereignisse werden als näher wahrgenommen. Zur weiteren Steigerung können die Hallanteile, welche aus anderen Richtungen als das Hörereignis kommen per Automation zurückgefahren werden, sodass nur noch Schallanteile aus Richtung des Hörereignisses kommen. (vgl. Hörbeispiel Musik 15)

Entgegengesetzt kann man ein Arrangement natürlich auch aufgehen lassen. Dies kann wieder durch die Positionierung der Hörereignisse und durch den Hallanteil geschehen. In diesem konkreten Hörbeispiel lasse ich dafür die Stimmgruppen Alt 1 und Alt 2 aus den Bereichen ± 45 Grad in Richtung ± 130 Grad wandern. Zusätzlich setzten die Bässe, welche davor im vorderen Bereich verteilt waren, von hinten ein um die Umhüllung zu steigern.

Der Hall, welcher zwischenzeitlich nur von vorne kam, kommt wieder aus allen Richtungen und der generelle Hallanteil wird mit den Tonartwechseln gesteigert.

Im letzten Schlussakkord kommen zusätzlich noch zwei Subbässe hinzu, welche auf voller Breite innerhalb des Kopfes positioniert sind, um die größtmögliche Einhüllung

zu erreichen (vgl. Hörbeispiel Musik 16).

5.3 Jingles

Jingles sind Verpackungselemente, welche dazu dienen, Sendungen und Rubriken anzukündigen, Programmteile zu verbinden, Moderator*innen und Radiofrequenzen zu bewerben und Image-Slogans zu transportieren. [Buc, 2017]

Sie definieren also neben den eigentlichen Sendungsinhalten den Hörfunksender und dessen Zielgruppe.

Jingles eignen sich um das Potenzial der Binauraltechnik zu demonstrieren und dem Hörer deren Mehrwert aufzuzeigen. Dabei kann bewusst auf bereits bekannte Elemente zurückgegriffen werden, indem man zum Beispiel ein Hörereignis um den Kopf der Hörerin oder des Hörers rotieren lässt (vgl. Hörbeispiel Jingles 01). Diesen Effekt kennen viele Hörer*innen aus "8D Audio"-Mischungen, bei welchen eine Stereomischung um den Kopf der Hörer*innen rotiert wird. Außerdem kann durch Soundeffekte oder Ähnliches das mögliche Maß an Umhüllung dargestellt werden (vgl. Hörbeispiel Jingles 02).

Durch eine Art "3D Audio Trailer" kann den Hörer*innen der Unterschied zwischen der sonstigen Hörereignisrichtung innerhalb des Kopfes zu verschiedenen externalisierten Hörereignisorten demonstriert werden (vgl. Hörbeispiel Jingles 03). Auf diese Weise wird die Aufmerksamkeit der Hörer*innen auf externalisierte Hörereignisse gelenkt, wodurch die Wahrscheinlichkeit steigt, dass sie als solche wahrgenommen werden.

Wie bei allen anderen Hörfunkformaten müsste die Dynamik der Jingles sehr weit eingeschränkt werden um das typische Klangideal von Hörfunksendungen zu erreichen.

6 Fazit

Durch die Anwendung von Binauraltechnik ergibt sich für die Gestaltung auditiver Medien eine Vielzahl neuer Möglichkeiten. Dabei können binaurale Gestaltungsmittel bewusst eingesetzt werden, um die kreative Intention eines Musikstückes oder einer Geschichte zu steigern und deren Immersionsgrad zu erhöhen.

Bei Sendungsformaten wie Gesprächsrunden oder klassischer Musik wird eine möglichst realitätsnahe Übertragung angestrebt. Hierbei bietet die Binauraltechnik durch die Darstellbarkeit externalisierter Hörereignisse die Möglichkeit, ein Hörerlebnis zu schaffen, welches dem natürlichen Hören um ein weites näher kommt als bei der Wiedergabe von Stereoinhalten über Kopfhörer.

Gleichzeitig bricht eine Darstellung von rein externalisierten Hörereignissen mit den Hörgewohnheiten eines Großteils der Hörer*innen. Daher verfolgte ich in dieser Arbeit den Ansatz die Binauraltechnik als eine Art Erweiterung von Mono und Stereo zu sehen, was zusätzliche Gestaltungsmöglichkeiten schafft.

Bei statischer und nicht-individualisierter Binauralsynthese ist die "Front-Back-Confusion" weiterhin die größte Herausforderung. Selbst bei meinen eigenen Mischungen fiel mir auf, dass abhängig vom verwendeten Kopfhörer, der Tageszeit und der emotionalen Stimmung, die Position des selben Hörereignisses und dessen Maß an Externalisierung stark variiert. Dies lässt sich in gewissem Maße durch das Hinzufügen von Räumlichkeit kompensieren, was wiederum mit Hörgewohnheiten bricht und die Gestaltungsmöglichkeiten einschränkt.

Aus diesem Grund entsteht für mich ein entscheidender Mehrwert erst beim Einsatz von dynamischer Binauralsynthese. Durch "Peilbewegungen" und das bewusste Umhören wird die "Front-Back-Confusion" fast vollständig gelöst. Zudem relativieren sich gestalterische Nachteile wie die Verzerrung des dreidimensionalen Raumes bei der binauralen Wiedergabe.

Die Grundproblematik der Inkompatibilität zu Lautsprechersystemen bleibt weiterhin bestehen. Aber dadurch, dass immer mehr Medien über Internet-Streams konsumiert werden, besteht die technische Möglichkeit, verschiedene Mischungen einer Sendung anzubieten.

6.1 Ausblick

Durch objektbasierte Audioformate wie "Dolby Atmos Home" oder "MPEG-H" ist eine technische Möglichkeit gegeben, dynamische und individualisierbare Binauralmischungen zu übertragen, da die Kopfhörer- beziehungsweise Lautsprechersignale erst im Abspielgerät berechnet werden.

Dabei gilt es zu beachten, dass eine dreidimensionale Lautsprechermischung zwar technisch auf eine binaurale Wiedergabe übertragbar ist, deren auditive Wirkung jedoch sehr wohl verschieden ist. Um das volle Potenzial der Binauraltechnik ausschöpfen zu können, ist meiner Meinung nach eine explizite Binauralmischung unumgänglich, die zum Beispiel auch ermöglicht, Hörereignisorte innerhalb des Kopfes darzustellen.

7 Hörbeispielverzeichnis

Effekte

Effekte 01

Zuerst ist eine Sängerin zu hören die bei 0 Grad binauralisiert wurde.

Anschließend wurde deren Stimme mit einem Choruseffekt summiert und das resultierende Stereosignal bei ± 30 Grad binauralisiert.

Im dritten Durchlauf ist das Direktsignal bei 0 Grad binauralisiert, sowie der Chorus-effekt bei ± 30 Grad.

Effekte 02

Zuerst ist eine Sängerin innerhalb des Kopfes zu hören.

Im zweiten Durchlauf ist die Stimme mit einem Choruseffekt versehen.

Im dritten Durchlauf bleibt die Stimme im Kopf und der Choruseffekt ist bei ± 30 Grad binauralisiert.

Effekte 03

Eine Sängerin ist innerhalb des Kopfes zu hören mit binauralisiertem Nachhall.

Ab Sekunde 08 kommen binauralisierte Reflexionen des IEM RoomEncoders dazu.

Ab Sekunde 17 sind statt dem RoomEncoder binauralisierte Reflexionen eines Stereoreverbs zu hören.

Ab Sekunde 25 ist die Stimme wieder ohne Reflexionen zu hören.

Effekte 04

Gleiches Szenario wie bei Effekte 3, nur im Kontext einer BigBand-Aufnahme.

Effekte 05

Es ist das Ende eine Chorstückes zu hören, bei dem die Stimmen erst mit Hilfe des IEM FdnReverbs verhallt wurden, dann mit einem Ambisonics Faltungshall und anschließend mit einem binauralisierten Surroundhall.

Effekte 06

Die gleiche Stelle des Chorstückes, einmal nur mit Hall aus der Horizontalebene, dann auch mit Hall von oben und anschließend mit mit einer Anhebung der oberen Mitten beim Hall von oben.

Effekte 07

Eine Sängerin ist zu hören welche nicht binauralisiert wurde. Dazu kommt ein binauralisiertes Slap-Delay.

Effekte 08

Ein Schnipsen ist zu hören, welches in drei verschiedene Ping-Pong Delays mit unterschiedlicher Verzögerung führt. Diese sind wie folgt positioniert:



Abbildung 7.1

Effekte 09

Das Schnipsen führt nur in das vordere Delay, welches ins mittlere führt, das wiederum ins hintere Delay führt.

Sprache und Hörspiel

Sprache 01

Es ist ein Sprecher zu hören welcher erst bei 0 Grad und anschließend bei 180 Grad binauralisiert ist.

Sprache 02

Gleiches Szenario, nur dass neben dem trockenen Sprecher auch räumliche Reflexionen zu hören sind.

Sprache 03

Zuerst ist ein Sprecher und dessen Reflexionen zu hören der bei 0 Grad binauralisiert ist.

Danach ist der Sprecher nicht binauralisiert, aber es sind noch binauralisierte Reflexionen zu hören.

Anschließend ist der Sprecher in der Kopfmittle ohne Reflexionen zu hören.

Sprache 04

Ein Ausschnitt aus einer Talksendung ist zuerst in Stereo zu hören, dann binauralisiert und dann binauralisiert inklusive Reflexionen.

Sprache 05

Eine Diskussion ist zu hören, bei der durcheinander geredet wird. Erst in Mono, dann in Stereo, dann nah beieinander binauralisiert und anschließend weiter entfernt voneinander.

Hörspiel 01

Eine exemplarische fiktive Soundszene.

Hörspiel 02

Erst ist eine Kunstkopf-Grundatmo zu hören. Diese wird erst durch ein Fahrgeräusch und anschließend mit einem Scheibenwischergeräusch und Regengeräusch ergänzt.

Hörspiel 03

Das Fahrgeräusch wird abwechselnd mithilfe des IEM DirectivityShaper und dem IEM Multienncoder binauralisiert.

Hörspiel 04

Das Regengeräusch ist abwechselnd von vorne- und hinten-oben und nur von vorne-oben zu hören.

Hörspiel 05

Eine Stereoaufnahme eines vorbeifahrenden Autos wurde in seitlicher Richtung binauralisiert.

Hörspiel 06

Die Stimme der Protagonistin ist abwechselnd mit und ohne räumliche Reflexionen zu hören.

Hörspiel 07

Ein direkter Vergleich von einem simulierten Telefon am Ohr und der Kunstkopfaufnahme eines Handys an dessen Ohr.

Musik

Musik 01

Ein Vergleich von (1) AB-Stereofonie, (2) Kunstkopfaufnahme und (3) binauralisiertem ORTF 3D anhand einer Orgelaufnahme.

Musik 02

Ein Vergleich der Binauralisierung des ORTF 3D anhand einer (1) HRIR und einer (2) BRIR.

Am Ende wird zudem der Schlussakkord und dessen Nachhall verglichen.

Musik 04

1. Durchlauf: Kunstkopf
2. Durchlauf: Kunstkopf + AB
3. Durchlauf: ORTF 3D
4. Durchlauf: ORTF 3D + AB

Musik 05

Demonstration von Mono-, Stereo- und Binauralelementen einer BigBand-Mischung.

Musik 06

Vergleich von Stereo-Piano und binauralisiertem Piano im BigBand-Kontext.

Sekunde 0-6: Stereo

Sekunde 6-12: binauralisiert bei ± 30 Grad

Sekunde 12-18: Stereo

Ab Sekunde 18: Binaural

Musik 07

Vergleich von räumlicher Wahrnehmung, abhängig von der Positionierung der Gitarren.

Musik 08

Vergleich der räumlichen Einbettung der unbinauralisierten Gesangsstimme erst ohne, dann mit Reflexionen und Slap-Delay.

Musik 09

Vergleich von Stereo-Overheads und binauralisierten Overheads.

Sekunde 0-17: Stereo

Sekunde 17-33: binauralisiert bei ± 30 Grad

Sekunde 33-50: binauralisiert bei ± 60 Grad

Sekunde 50-56: Stereo

Ab Sekunde 56: binauralisiert bei ± 60 Grad

Musik 10

Vergleich der räumlichen Wahrnehmung anhand der Positionierung des Surroundhalls.

Sekunde 0-7: L/R bei ± 30 Grad und Ls/Rs bei ± 110 Grad

Sekunde 7-14: L/R bei ± 45 Grad und Ls/Rs bei ± 135 Grad

Ab Sekunde 14 kommt zu letzterer Positionierung noch ein Stereohall bei ± 20 Grad dazu

Musik 11

Vergleich des Maßes an Umhüllung abhängig von der Positionierung der Bläser.

1. Durchlauf: Bläser sind im Bereich von ± 30 Grad positioniert.
2. Durchlauf: Bläser sind im Bereich von ± 45 Grad positioniert.
3. Durchlauf: Bläser sind rundherum positioniert.

Musik 12

Vergleich des Maßes an Umhüllung abhängig der Positionierung der Einzelstimmen eines Chores.

1. Durchlauf: Alles Stimmen sind im Bereich von ± 45 Grad positioniert.
2. Durchlauf: Die Stimmen sind rundherum positioniert.

Musik 13

Vergleich von binauralisierten und Stereo-Gesangsdopplungen.

1. Durchlauf: Stereo
2. Durchlauf: binauralisiert bei ± 45 Grad

Musik 14

Vergleich des Maßes an Umhüllung abhängig der Positionierung der Backing-Chöre.

1. Durchlauf: Die vier Stimmen sind im Bereich von ± 30 Grad positioniert.
2. Durchlauf: Die vier Stimmen sind bei $45, 135, -135$ und -45 Grad positioniert.

Musik 15

Vergleich des räumlichen Kontrastes anhand von Automationen des Halls.

1. Durchlauf: Kontrast anhand des Arrangements
2. Durchlauf: Verstärkung des Kontrasts durch Lautstärkenautomation des Halls
3. Durchlauf: Zusätzlich zur Lautstärkenautomation werden die hinteren und oberen Hallanteile weggenommen.

Musik 16

Erhöhung des Maßes an Umhüllung anhand von Hallautomation und Automation der Schallereignispositionen.

Jingles

Jingles 01

Skizzierte Nachrichten-Jingle

Jingles 02

Skizzierte Wetter-Jingle

Jingles 03

Beispielhafter 3D-Radio-Trailer.

Literaturverzeichnis

- [Buc, 2017] (2017). Radio-journalismus: Ein handbuch für ausbildung und praxis im hörfunk. Online-Ressource (XIII, 650 Seiten 29 Abb, online resource).
- [Albert S. Bregman, 1990] Albert S. Bregman, P. A. A. (1990). Demonstrationsto accompany bregman's auditory scene analysis. *MIT Press*.
- [Buff, 2020] Buff, H.-M. (2020). *Überall - Musikproduktion in 3D Audio für Kopfhörer*. Ebner Media Group GmbH Co. KG.
- [DearRealityGmbH, 2022a] DearRealityGmbH (2022a). Dearvr monitor v1.11.0.
- [DearRealityGmbH, 2022b] DearRealityGmbH (2022b). Dearvr pro user manual v1.10.0.
- [Dickreiter et al., 2014] Dickreiter, M., Dittel, V., Hoeg, W., and Wöhr, M., editors (2014). *Handbuch der Tonstudiotchnik*. De Gruyter Saur.
- [Faller et al., 2015] Faller, C., Menzer, F., and Tournery, C. (2015). Binaural audio with relative and pseudo head tracking. In *Audio Engineering Society Convention 138*.
- [Fonseca, 2020] Fonseca, N. (2020). Sound particles - reference manual - v2.1.
- [Francis, 2015] Francis, G. (2015). Auf gehts... ein reaper benutzerhandbuch. 4.76.
- [Frank and Zotter, 2018] Frank, M. and Zotter, F. (2018). Simple reduction of front-back confusion in static binaural rendering.
- [Genelec,] Genelec. *Genelec Aural ID System Operating Manual*.
- [IEM, 2022] IEM (2022). Plug-in descriptions.
- [ITU, 2022] ITU (2022). Note: This itu-r report was approved in english by the study group under the procedure detailed in resolution itu-r 1.

- [Lindau et al., 2010] Lindau, a., estrella, j., and weinzierl, s. (2010). individualization of dynamic binaural synthesis by real time manipulation of itd. *journal of the audio engineering society*.
- [McCormack and Politis, 2019] McCormack, L. and Politis, A. (2019). Audio engineering society convention e-brief 111.
- [Melchior, 2021] Melchior, P. D. F. (2021). Vorleseungsfolien 3d audiosysteme.
- [WavesAudioLtd,] WavesAudioLtd. Virtual mix room user guide.
- [Weinzierl, 2008] Weinzierl, S. (2008). *Handbuch der Audiotechnik*. Springer, Berlin and Heidelberg.
- [Zotter and Frank, 2019] Zotter, F. V. and Frank, M. V. (2019). Ambisonics: a practical 3d audio theory for recording, studio production, sound reinforcement, and virtual reality. 1 Online-Ressource (XIV, 210 Seiten).