

Erstellung einer walkthrough-fähigen Webapplikation unter Verwendung von Ambisonics und binauraler Auralisation

Bachelorarbeit aus Aufnahmetechnik 1, SE

Thomas Deppisch

Betreuung:

Univ.Prof. DI Dr. Alois Sontacchi

Daniel Rudrich, M.Sc.

Graz, 13. Juni 2017



institut für elektronische musik und akustik



Zusammenfassung

Im Zuge dieser Arbeit wurde eine Webapplikation entwickelt, die das virtuelle Durchschreiten einer Audioszene ermöglicht. Die Audiosignale werden dabei in Echtzeit mittels eines virtuellen Ambisonics Ansatzes spatialisiert. Diese Arbeit erläutert die Aspekte, die zur Erstellung der Applikation notwendig sind. Dazu gehören psychoakustische Grundlagen des räumlichen Hörens, mathematische Zusammenhänge von Ambisonics sowie das Konzept binauraler Auralisation. Zudem enthält die Arbeit einen Vergleich binauraler Ambisonics Dekoder, bei dem die Auswirkungen der Parameter Ambisonicsordnung, Lautsprecheranzahl und $\max \vec{r}_E$ -Optimierung anhand von vier objektiven Kriterien gezeigt werden. Die Untersuchungen erfolgen ausschließlich für gleichmäßige Lautsprecheranordnungen. Abschließend wird die Struktur und Funktionsweise der Webapplikation erklärt.

Abstract

In the course of this thesis a web application allowing a virtual walk through an audio scene is developed. The audio signals are spatialized in realtime using a virtual Ambisonics approach. This thesis outlines the necessary aspects for the creation of the application. These include the psychoacoustics of spatial hearing, the mathematics of Ambisonics and the concept of binaural auralization. In addition, this thesis contains a comparison of binaural Ambisonics decoders. The impact of parameters such as Ambisonics order, number of loudspeakers and $\max \vec{r}_E$ optimization is examined by means of four objective criterias. This examination is done for regular loudspeaker arrangements only. Finally, the structure and functionality of the web application is explained.

Inhaltsverzeichnis

1	Einleitung	4
2	Psychoakustische Grundlagen des räumlichen Hörens	5
2.1	Richtungswahrnehmung	5
2.2	Distanzwahrnehmung	9
3	Ambisonics	10
3.1	Einführung	10
3.2	Mathematische Herleitung	11
3.3	Enkodierung	15
3.4	Dekodierung	17
3.5	Schallfeldrotation	20
4	Binaurale Auralisation	21
4.1	Raumsimulation	21
4.2	Binaurale Synthese	25
5	Objektiver Vergleich binauraler Ambisonics Dekoder	27
5.1	Vergleichskriterien	27
5.2	Untersuchte Parameter und Erwartungen	30
5.3	Analyse und Auswertung für 2D und 3D Dekoder	32
6	Implementierung	47
6.1	Einführung	47
6.2	Struktur	48
7	Schlussfolgerung und Ausblick	56
A	Appendix	58
	Literaturverzeichnis	59

1 Einleitung

Im Verlauf dieser Arbeit wurde eine Webapplikation entwickelt, die dem Nutzer das virtuelle Durchschreiten einer Audioszene ermöglicht. Diese schriftliche Arbeit soll dem Leser einen Überblick über die Grundlagen verschaffen, die zur Erstellung der Applikation erforderlich sind. Dabei werden Antworten auf die folgenden Fragestellungen gegeben:

- Welche Mechanismen erlauben dem Menschen die akustische Lokalisation eines Objekts?
- Wie positioniert man ein Audioobjekt im virtuellen Raum?
- Wie lassen sich Distanz und Richtung des Objekts simulieren?
- Wie kann ein virtueller Raumeindruck vermittelt werden?
- Wie funktioniert Ambisonics und wie ermöglicht es eine möglichst präzise Reproduktion eines Schallfeldes?
- Wie verändern sich Eigenschaften des Ambisonics Dekoders aufgrund von Ambisonicsordnung, Lautsprecheranzahl und Optimierung?
- Wie kann Audioverarbeitung in einem Webbrowser erfolgen?

Der Aufbau der Arbeit gestaltet sich folgendermaßen: Kapitel 2 erläutert welche Mechanismen die Bestimmung von Richtung und Distanz eines Objekts aufgrund akustischer Eigenschaften ermöglichen. Die Herleitung der mathematischen Hintergründe von Ambisonics befindet sich in Kapitel 3. Außerdem werden Möglichkeiten der Enkodierung, Dekodierung und Schallfeldrotation aufgezeigt. In Kapitel 4 wird das Konzept binauraler Auralisation erklärt. Es werden einige Methoden zur Simulation eines Raumeindrucks und das Prinzip binauraler Synthese erläutert. In Kapitel 5 werden objektive Kriterien zum Vergleich binauraler Ambisonics Dekoder vorgestellt. Anhand dieser wird der Einfluss von Ambisonicsordnung, Lautsprecheranzahl und $\max \vec{r}_E$ -Optimierung untersucht. Kapitel 6 widmet sich der Beschreibung der Struktur und Funktionsweise der erstellten Applikation. Zuletzt zeigt Kapitel 7 wie auf Basis des Dekodervergleichs die Parameter der implementierten Dekoder festgelegt werden. Außerdem wird auf einige begrenzende Faktoren der Webapplikation hingewiesen und ein Ausblick weiterer Entwicklungsmöglichkeiten gegeben.

2 Psychoakustische Grundlagen des räumlichen Hörens

Die Lokalisation einer Schallquelle durch den Menschen beruht auf mehreren Faktoren. Die hier angeführten Faktoren basieren auf physikalischen Eigenschaften von Schall. Die Untersuchung der Wahrnehmung dieser physikalischen Eigenschaften ist Gegenstand der Psychoakustik. Eine Einführung in die Psychoakustik bietet [Moo03]. Neurophysiologische und visuelle Faktoren spielen bei der Lokalisation von Schallquellen ebenfalls eine Rolle. Diese werden hier allerdings nicht weiter behandelt.

In [Bla74] und [Bla97] werden die Eigenschaften des räumlichen Hörens untersucht. Angaben in diesem Kapitel, die nicht genauer referenziert sind, wurden diesen Quellen entnommen.

2.1 Richtungswahrnehmung

Die zur Richtungswahrnehmung relevanten Faktoren lassen sich systematisch in binaurale und monaurale Faktoren einteilen. Werden die Informationen zur Ermittlung einer Quellposition aus Unterschieden der beiden Ohrsignale ermittelt, spricht man von binauralen Faktoren. Im Gegensatz dazu spricht man von monauralen Faktoren, wenn die Lokalisationsinformation aus der Auswertung identischer Anteile beider Ohrsignale gewonnen wird.

2.1.1 Binaurale Faktoren

Binaurale Faktoren sind maßgeblich für die Richtungsbestimmung von lateralen Quellen verantwortlich.

Interaurale Laufzeitdifferenzen. Die Auswertung interauraler Laufzeitdifferenzen (*interaural time differences*, ITDs) erlaubt dem Hörer die Wahrnehmung der Richtung einer Schallwelle aufgrund ihrer verzögerten Ankunft am kontralateralen (abgewandten) Ohr im Gegensatz zum ipsilateralen (zugewandten) Ohr (vgl. Abbildung 1). Diese Laufzeitdifferenz kann ebenfalls durch eine Phasendifferenz beschrieben werden. Sobald die Wellenlänge des Signals kleiner als der Kopfdurchmesser ist, kann diese Phasendifferenz nicht mehr sinnvoll ausgewertet werden. Deswegen sind ITDs vor allem für tiefe Frequenzen bis etwa 1,6 kHz ausschlaggebend. Es ist allerdings gezeigt worden, dass auch für Frequenzen über 1,6 kHz Laufzeitdifferenzen zur Lokalisation genutzt werden. Dies erfolgt durch Auswertung der Hüllkurve eines Signals [PK15, MM02].

Interaurale Pegeldifferenzen. Durch Abschattungseffekte des Kopfes entstehen interaurale Pegeldifferenzen (*interaural level differences*, ILDs), wenn die Wellenlänge des Signals kleiner als der Kopfdurchmesser ist (vgl. Abbildung 2). Ist die Wellenlänge etwa

in der Größenordnung des Kopfdurchmessers entstehen Beugungseffekte, bei noch größeren Wellenlängen hat der Kopf keinen Einfluss auf die Ausbreitung der Welle. Daher erfolgt die Auswertung von ILDs für die Richtungswahrnehmung konträr zu den ITDs vor allem für Frequenzen ab etwa 1 kHz.

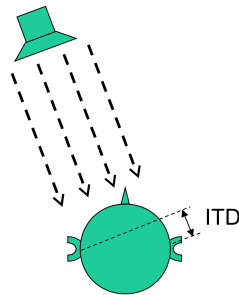
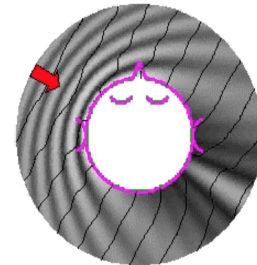


Abbildung 1 – Entstehung interauraler Laufzeitdifferenzen [Maj14].



$f = 8000 \text{ Hz}$, $\theta = 60^\circ$

Abbildung 2 – Abschattung am Kopf aufgrund einer ebenen Welle mit $f = 8 \text{ kHz}$ aus der Richtung $\theta = 60^\circ$ [Dan00]. Dunklere Bereiche kennzeichnen Schalldruckminima.

Cones of confusion. Bei Annäherung des Kopfes als Kugel können aufgrund der Symmetrie Kegelränder bestimmt werden, auf denen äquivalente ILDs und ITDs herrschen. Diese Kegel werden *cones of confusion* genannt. Mittels ILDs und ITDs kann also prinzipiell nur die Lage dieser Kegel bestimmt werden, nicht der genaue Ort der Schallquelle auf dem Kegelrand. Zur Auflösung dieser Ambiguität werden monaurale und dynamische Faktoren herangezogen.

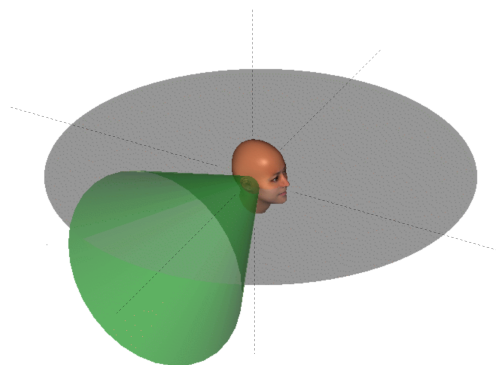


Abbildung 3 – Cone of confusion [Maj14].

Präzedenzeffekt. Bei mehreren ähnlichen Schallquellen wie sie z. B. in Räumen durch Reflexion entstehen, wird der Präzedenzeffekt wirksam: Erreicht innerhalb von etwa 2 bis 30ms das gleiche Signal verzögert aus unterschiedlichen Richtungen die Ohren einer Person, so nimmt diese nur die Richtung des ersten eintreffenden Schallsignals wahr [PK15].

2.1.2 Monaurale Faktoren

Zur Bestimmung eines Schallquellenorts in der Medianebene und zur Auflösung der Ambiguität eines *cone of confusion* werden monaurale Faktoren herangezogen. Diese manifestieren sich in richtungsabhängigen Änderungen des Amplitudengangs der Ohrsignale, vor allem für hohe Frequenzen. Die spektralen Änderungen entstehen aufgrund von Auslöschungen bzw. Verstärkungen (Interferenz) von Schallwellen kurzer Wellenlänge durch Reflexionen an Pinna (Außenohr) und Oberkörper. Die Funktion von monauralen Faktoren zur Richtungsbestimmung setzt also einen hochfrequenten Signalanteil und eine gewisse Kenntnis des Quellspektrums voraus. Die Zuordnung einer Schallquelle zu einer bestimmten Richtung in der Medianebene hängt überwiegend vom spektralen Anteil richtungsbestimmender Frequenzbänder ab. Obwohl diese Frequenzbänder individualspezifisch sind, lassen sich Zusammenhänge bei der Wahrnehmung erkennen: Lokalisation *von oben* erfolgt beispielsweise eher durch Anhebung des Frequenzbandes um 8 kHz [Bla74].

2.1.3 HRIR und HRTF

Monaurale und binaurale Faktoren spiegeln sich in der Außenohrimpulsantwort (HRIR, *head-related impulse response*) oder deren Äquivalent im Frequenzbereich, der Außenohrübertragungsfunktion (HRTF, *head-related transfer function*, vgl. Abbildung 4) wider. Die Messung einer HRIR kann durch Platzierung zweier Mikrofone in den Ohren einer Versuchsperson oder eines Kunstkopfes erfolgen. Sie ist abhängig von Quellauslenkung und Entfernung, wobei sie ab einer Entfernung von ca. 1 m bei Messung unter Freifeldbedingungen für hohe Frequenzen annähernd distanzunabhängig ist. HRTFs werden bei der binauralen Synthese zur Richtungskodierung verwendet (vgl. Kapitel 4.2). Durch unterschiedliche anthropometrische Beschaffenheit (v.a. der Pinna) entstehen individuelle Unterschiede der HRTFs bei Menschen, wodurch sich Abbildungsfehler bei der binauralen Synthetisierung mittels einer fremden oder künstlichen HRTF ergeben können.

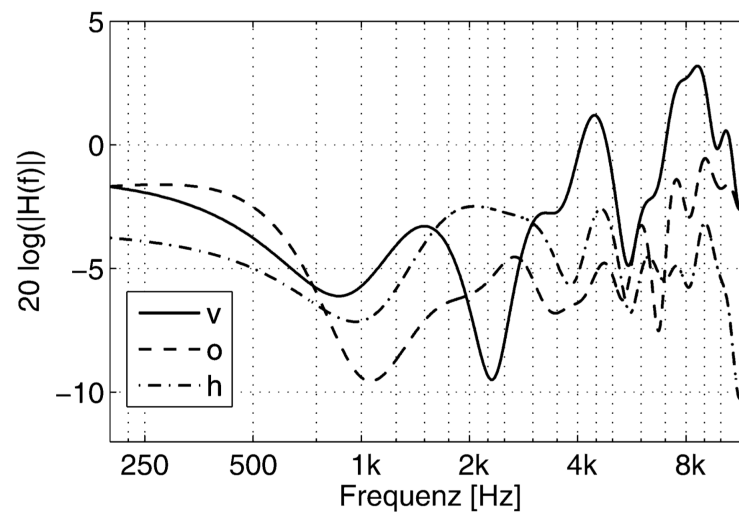


Abbildung 4 – Amplitudengänge dreier HRTFs eines Ohrs von vorne (v), oben (o) und hinten (h) [BB07].

2.1.4 Dynamische Faktoren

Bei einer Bewegung des Kopfes relativ zur Schallquelle ändern sich ITDs, ILDs und spektrale Zusammensetzung des Ohrsignals (HRTF). Die Auswertung dieser Änderungen bei sogenannten *Peilbewegungen* des Kopfes erleichtern daher die korrekte Lokalisation einer Schallquelle. Auch die Externalisation, also die Wahrnehmung einer Schallquelle außerhalb des Kopfes (im Gegensatz zur In-Kopf-Lokalisation z. B. bei gewöhnlicher Kopfhörerwiedergabe), kann dadurch stark verbessert werden [BBA13]. Die Ermöglichung von Peilbewegungen durch Headtracking-Systeme bei virtuellen Anwendungen kann die Lokalisation daher verbessern [WK99].

2.2 Distanzwahrnehmung

Die überwiegend bestimmenden Faktoren zur Distanzwahrnehmung können in drei Bereiche eingeteilt werden [BB07]: Geringe Entfernungen näher als ca. 25 cm, mittlere Entfernungen von 25 cm bis 15 m und große Entfernungen von über 15 m.

Geringe Entfernungen $r < 25$ cm. Bewegt sich eine Schallquelle in geringer Entfernung relativ zum Hörer, ändert sich die HRTF stark mit der Entfernung. Durch die Abschattung des Kopfes steigen interaurale Pegeldifferenzen für nähere Signale an. Außerdem wächst der Signalpegel mit näherer Quellposition [BB07].

Mittlere Entfernungen $25 \text{ cm} < r < 15 \text{ m}$. Bei mittlerer Quellentfernung unter Freifeldbedingungen ist die Änderung des Signalpegels ausschlaggebend für die wahrgenommene Distanz. Im Freifeld nimmt der Schalldruckpegel invers zur Entfernung ($1/r$) ab. Das entspricht einer Abnahme von 6 dB pro Entfernungsverdopplung. In halligen Räumen spielt das Intensitätsverhältnis von Direktschall zu reflektiertem Schall eine wichtige Rolle. Hier erzeugen Quellen außerhalb des Hallradius aufgrund des starken Diffusanteils einen nahezu entfernungsunabhängigen Pegel. Auch die Zeitdifferenz des Eintreffens von Direktschall zu ersten Reflexionen wird ausgewertet und trägt zu empfundener Raumgröße und Quellentfernung bei [BB07].

Große Entfernungen $r > 15 \text{ m}$. Bei Schallquellen in großen Entfernungen sorgt die Dämpfung der Luft für eine Abschwächung von hohen Frequenzen. Der mit größerer Entfernung zunehmend dumpfere Klangeindruck lässt allerdings wiederum nur Rückschlüsse auf die Entfernung zu, wenn das Quellsignal bekannt ist. Der akustische Horizont des Menschen liegt bei ca. 15 m. Einschätzungen von Entfernungen größer als 15 m sind nur aufgrund von Erfahrungswerten und visueller Unterstützung möglich [BB07].

Im-Kopf-Lokalisation. Werden Signale gewöhnlich über Kopfhörer wiedergegeben, wird die Position dieser meist im Kopf wahrgenommen. Dieser Eindruck wird auch Internalisation genannt und hängt maßgeblich mit dem Wegfall der spektralen Richtungsfilterung durch die HRTF zusammen. Auch ein nicht vorhandener Anteil an Reflexionen und die Unterdrückung der relativen Quellsbewegung bei Kopfbewegung fördern die Internalisation [BBA13]. Bei der virtuellen Synthetisierung einer Klangszene ist es daher wichtig, diese Faktoren zu berücksichtigen (vgl. Kapitel 4 und 6).

3 Ambisonics

3.1 Einführung

Ambisonics ist wie auch Wellenfeldsynthese ein holophones Verfahren, das die physikalische Reproduktion eines Schallfeldes zum Ziel hat. Die mathematischen Grundlagen von Ambisonics wurden bereits in den 70er Jahren in weiten Teilen von Michael Gerzon und unabhängig davon von Duane Cooper erarbeitet (z. B. [Ger72], [Ger77], [CS72]). Ambisonicsignale erster Ordnung (*first order Ambisonics*, FOA) werden durch die vier Kanäle W, X, Y und Z, dem sogenannten *B-Format* repräsentiert. Ambisonics höherer Ordnung (*higher order Ambisonics*, HOA) verwendet eine von der Ordnung abhängige, höhere Kanalanzahl. Ein wesentlicher Vorteil von Ambisonics gegenüber anderen Mehrkanalverfahren ist die Unabhängigkeit von En- und Dekodierung. Die Kenntnis der Lautsprecheranordnung ist somit erst bei der Dekodierung notwendig.

Der klassische Ambisonics Ansatz beruht auf der Annahme ebener Wellen, sowohl für die aufgenommene Wellenfront (Referenzwelle) als auch die reproduzierte Wellenfront. Mathematisch wird dabei die Überlagerung ebener Wellen (synthetisiertes Schallfeld) im Betrachtungspunkt mit dem Referenzschallfeld gleichgesetzt und damit Enkodier- und Dekodiervorschriften aufgestellt [BV95]. Dazu wird die homogene Helmholtzgleichung gelöst und anschließend die radiale Lösungskomponente vernachlässigt oder mittels Randwertproblem vereinfacht [Son03, Zot09].

Im Folgenden wird zur Erklärung ein neuerer Ansatz verwendet: Zuerst wird eine kontinuierliche Verteilung von Punktquellen auf der endlich weit entfernten Kugeloberfläche betrachtet. Anschließend wird die anguläre Auflösung limitiert und die Quellverteilung diskretisiert. Der sich durch die Limitierung und Diskretisierung ergebende Fehler spiegelt sich in der Größe eines optimalen Hörbereichs (*sweet area*) wider. Im Zuge der mathematischen Herleitung wird dabei die inhomogene Helmholtzgleichung gelöst und anschließend die modalen Quellstärken der kontinuierlichen und der diskretisierten Anregung verglichen. Dieser Ansatz ermöglicht eine präzisere Erklärungs- und Berechnungsmöglichkeit des sich ergebenden Resynthesefehlers [ZPF09, Zot09].

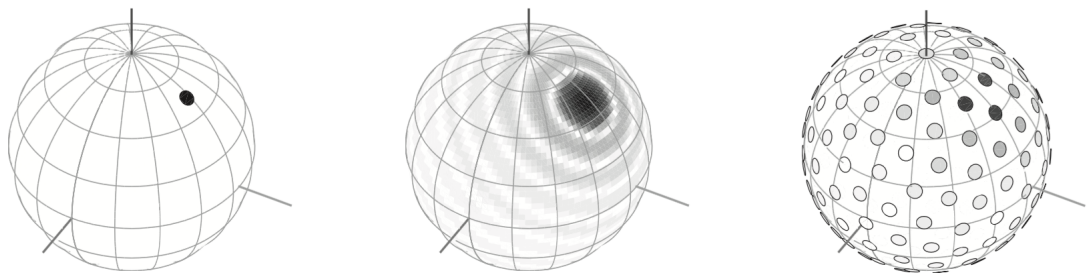


Abbildung 5 – Von links nach rechts: Kontinuierliche Punktquelle auf der Kugeloberfläche, anschließende anguläre Limitierung der Auflösung und Diskretisierung [ZF12].

3.2 Mathematische Herleitung

Schall als eine fortschreitende Welle von Druckänderungen in einem Medium kann mathematisch durch die Wellengleichung beschrieben werden. Die homogene Wellengleichung für den orts- und zeitabhängigen Schalldruck $p(\vec{r}, t)$ lautet:

$$\frac{\partial^2 p(\vec{r}, t)}{\partial t^2} = c^2 \Delta p(\vec{r}, t) \quad (1)$$

Dabei ist Δ der Laplace-Operator und c die Ausbreitungsgeschwindigkeit.

Eine Betrachtung der Wellengleichung im Frequenzbereich über Anwendung der kontinuierlichen Fouriertransformation führt in Gleichung 5 auf die zeitunabhängige Helmholtzgleichung, wobei der Faktor $k = \frac{\omega}{c}$ die Wellenzahl ist.

$$p(\vec{r}, t) = \int_{-\infty}^{\infty} p(\vec{r}, \omega) e^{i\omega t} d\omega \quad (2)$$

$$\text{Gleichung 2 in 1: } \frac{\partial^2}{\partial t^2} \left(\int_{-\infty}^{\infty} p(\vec{r}, \omega) e^{i\omega t} d\omega \right) = c^2 \Delta \left(\int_{-\infty}^{\infty} p(\vec{r}, \omega) e^{i\omega t} d\omega \right) \quad (3)$$

$$\Leftrightarrow \int_{-\infty}^{\infty} (c^2 \Delta p(\vec{r}, \omega) - (i\omega)^2 p(\vec{r}, \omega)) e^{i\omega t} d\omega = 0 \quad (4)$$

$$\Leftrightarrow \Delta p(\vec{r}, \omega) + k^2 p(\vec{r}, \omega) = 0 \quad (5)$$

Bei Betrachtung einer kontinuierlichen Verteilung von Punktquellen muss die inhomogene Helmholtzgleichung betrachtet werden:

$$\Delta p(\vec{\theta}, \omega) + k^2 p(\vec{\theta}, \omega) = -\frac{\delta(r - r_l)}{r^2} f(\vec{\theta}) \quad (6)$$

δ ist die Dirac-Verteilung. $\vec{\theta}$ ist ein Einheitsvektor abhängig von Azimutwinkel und Polwinkel, der die Betrachtung in Kugelkoordinaten ermöglicht.

$$\vec{\theta}(\varphi, \vartheta) = \begin{pmatrix} \cos(\varphi) \sin(\vartheta) \\ \sin(\varphi) \sin(\vartheta) \\ \cos(\vartheta) \end{pmatrix} \quad (7)$$

Gleichung 6 beschreibt also die Anregung der Helmholtzgleichung auf der Kugeloberfläche mit Radius r_l durch die kontinuierliche Quellstärkeverteilung $f(\vec{\theta})$.

Bevor die inhomogene Helmholtzgleichung gelöst wird, erfolgt nun die Lösung der homogenen Helmholtzgleichung. Das Ergebnis der homogenen Gleichung wird später zur Lösung des inhomogenen Problems benötigt. Bei der Lösung der homogenen Gleichung ergeben sich die Kugelflächenfunktionen und mit ihnen die Möglichkeit einer *Spherical Harmonics Transformation (SHT)*. Diese werden ebenfalls für die Lösung der inhomogenen Helmholtzgleichung benötigt. Auf eine mathematisch detaillierte Lösung wird hier verzichtet und stattdessen auf [Wil99] und [Zot09] verwiesen.

3.2.1 Lösung der homogenen Helmholtzgleichung

Die homogene Lösung kann durch das Verfahren der Variablenseparation erfolgen. Dabei wird die homogene Helmholtzgleichung in drei Differentialgleichungen (für Radius r , Azimutwinkel φ und Polarwinkel ϑ) aufgeteilt. Die Differentialgleichung für den Polarwinkel ϑ ist eine Legendresche Differentialgleichung, die mittels assoziierter Legendrefunktionen erster Art $P_n^m(\cos(\vartheta))$ gelöst werden kann. Für den Azimutwinkel φ ergibt sich eine lineare Differentialgleichung, die mit einer komplexen e-Funktion $e^{im\varphi}$ gelöst werden kann. Die dritte erhaltene Differentialgleichung für den Radius r ist eine sphärische Besselsche Differentialgleichung. Sie kann bei Betrachtung des inneren Feldes ($r < r_l$), mit den sphärischen Besselfunktionen $j_n(kr)$ gelöst werden. Bei Betrachtung des äußeren Feldes ($r > r_l$) kann die Lösung mittels sphärischer Hankelfunktionen zweiter Art $h_n^{(2)}(kr)$ ermittelt werden [Zot09].

Die Lösungen für Azimutwinkel φ und Polarwinkel ϑ können in den Kugelflächenfunktionen $Y_n^m(\vec{\theta})$ zusammengefasst werden. Damit ergibt sich als Lösung der homogenen Helmholtzgleichung [Zot09]:

$$p_h(kr, \vec{\theta}) = Y_n^m(\vec{\theta}) \cdot \begin{cases} j_n(kr) & \text{für das innere Feld } r \leq r_l \\ h_n^{(2)}(kr) & \text{für das äußere Feld } r > r_l \end{cases} \quad (8)$$

3.2.2 Kugelflächenfunktionen

In der Praxis ist häufig eine Darstellung der realwertigen Kugelflächenfunktionen ausreichend [NZDS11]. Die komplexe e-Funktion aus dem Lösungsterm für den Azimutwinkel φ wird dann durch einen Sinus- bzw. Cosinus-Term ersetzt:

$$Y_n^m(\varphi, \vartheta) = N_n^{|m|} P_n^{|m|}(\cos(\vartheta)) \cdot \begin{cases} \sin(m\varphi) & \text{für } m < 0 \\ \cos(m\varphi) & \text{für } m \geq 0 \end{cases} \quad (9)$$

Die Kugelflächenfunktionen $Y_n^m(\varphi, \vartheta)$ mit Ordnung n und Grad m bilden ein Set von orthogonalen Basisvektoren. Durch die richtige Wahl der Normierungskonstante $N_n^{|m|}$ werden diese orthonormal.

$$N_n^m = (-1)^m \sqrt{\frac{(2n+1)(2-\delta_m)(n-m)!}{4\pi(n+m)!}}, \quad \delta_m = \begin{cases} 1 & \text{für } m = 0 \\ 0 & \text{sonst} \end{cases} \quad (10)$$

In der Praxis wird häufig auf eine vollständige Normalisierung der Basisvektoren wie hier dargestellt verzichtet. Beispielsweise hat die Schmidt Seminormalisierung (SN3D, Gleichung 11) den Vorteil, dass Übersteuern von Signalanteilen höherer Ordnung verhindert wird [NZDS11].

$$N_{n,SN3D}^m = \sqrt{\frac{(2-\delta_m)(n-m)!}{4\pi(n+m)!}} \quad (11)$$

Der in Gleichung 10 abgebildete Normalisierungsfaktor N_n^m wird in der Praxis bei N3D-Normalisierung verwendet.

Bei einer Aufzählung bis zur Ordnung n ergeben sich $(n + 1)^2$ Kugelflächenfunktionen (siehe Abbildung 6). Diese Anzahl spiegelt die Anzahl erhaltener Ambisonicskanäle wider.

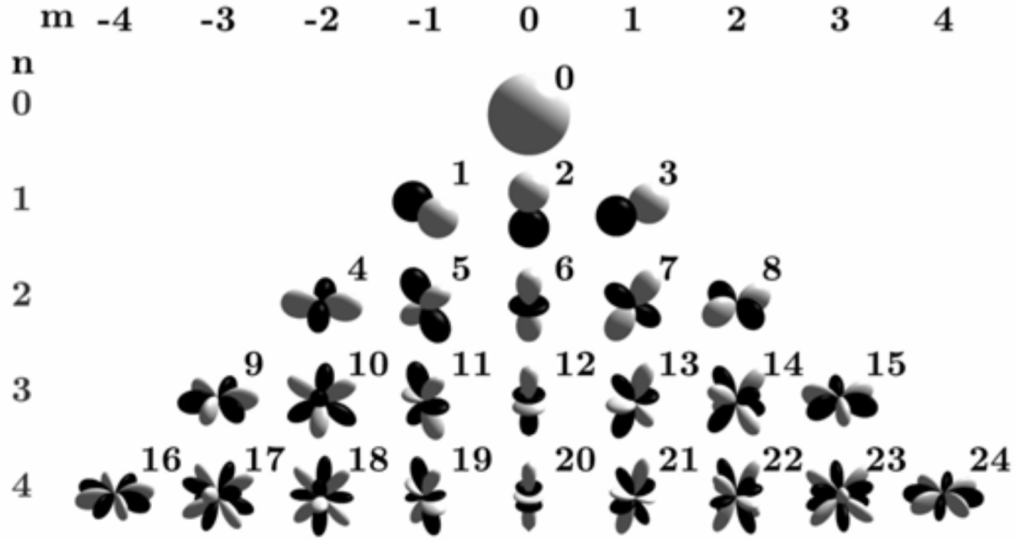


Abbildung 6 – Kugelflächenfunktionen bis zur vierten Ordnung [Kro14].

3.2.3 Spherical Harmonics Transformation

Da die Kugelflächenfunktionen ein Set von orthogonalen Basisvektoren bilden, kann eine Funktion $f(\vec{\theta})$ durch Reihenentwicklung in eine Linearkombination der Kugelflächenfunktionen zerlegt werden. Dazu werden analog zur Fouriertransformation Entwicklungskoeffizienten ϕ_{nm} benötigt. Über die Orthogonalität kann ein Transformationsintegral für die Entwicklungskoeffizienten ϕ_{nm} hergeleitet werden, sodass ein Transformationspaar aus *Spherical Harmonics Transformation* (SHT) und ihrer Inversen (ISHT) entsteht [Pom08, Zot09]. Die Integration bei der SHT erfolgt dabei über die Einheitskugel $\mathbb{S}^2$.

$$\mathcal{ISHT}\{\phi_{nm}\} = f(\vec{\theta}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \phi_{nm} Y_n^m(\vec{\theta}) \quad (12)$$

$$\mathcal{SHT}\{f(\vec{\theta})\} = \phi_{nm} = \int_{\mathbb{S}^2} f(\vec{\theta}) Y_n^m(\vec{\theta}) d\vec{\theta} \quad (13)$$

Die Repräsentation ϕ_{nm} von $f(\vec{\theta})$ in der *Spherical Harmonics Domäne* wird auch Ambisonicssignal genannt und stellt eine Betrachtung im Frequenzbereich ($\phi_{nm}(\omega)$) dar [ZPN10].

3.2.4 Lösung der inhomogenen Helmholtzgleichung

Die Lösung der inhomogenen Helmholtzgleichung kann mittels Produktansatz in radiale Lösung $R(kr)$ und anguläre Lösung $\Phi(\vec{\theta})$ aufgeteilt werden [Zot09].

$$p(kr, \vec{\theta}) = R(kr)\Phi(\vec{\theta}) \quad (14)$$

Durch SH-Transformation der Quellstärke $f(\vec{\theta})$ ergibt sich die Lösung ϕ_{nm} des angulären Anteils (Gleichung 13).

Für die Lösung des radialen Anteils ergibt sich unter Zuhilfenahme einer Greenschen Funktion eine inhomogene sphärische Besseldifferentialgleichung. Die Lösung erfolgt mittels Variation der Parameter, wobei die zwei homogenen Lösungen $j_n(kr)$ und $h_n^{(2)}(kr_l)$ zur Lösung herangezogen werden [ZPF09, Zot09].

Somit ergibt sich als Gesamtlösung für den Schalldruck $p(kr, \vec{\theta})$:

$$p(kr, \vec{\theta}) = -ik \sum_{n=0}^{\infty} \sum_{m=-n}^n \phi_{nm} Y_n^m(\vec{\theta}) \cdot \begin{cases} h_n^{(2)}(kr_l) j_n(kr) & \text{für } r \leq r_l \\ j_n(kr) h_n^{(2)}(kr_l) & \text{für } r \geq r_l \end{cases} \quad (15)$$

Diese allgemeine Lösung kann durch Annahme der Anregung durch Punktquellen vereinfacht werden. Die Darstellung einer Punktquelle an der Position $\vec{\theta}_0$ ist mittels der Delta-Verteilung δ möglich. Die Delta-Verteilung ist immer Null, außer ihr Argument ist Null. Somit erhält in der folgenden Gleichung $f(\vec{\theta})$ nur dann einen Wert ungleich Null, wenn das Produkt aus dem transponierten Richtungsvektor $\vec{\theta}^T$ mit dem Richtungsvektor der Punktquelle $\vec{\theta}_0$ (Skalarprodukt) das Ergebnis Eins liefert (für $\vec{\theta}$ siehe Gleichung 7). Dies ist nur dann der Fall, wenn die beiden Richtungsvektoren in Azimutwinkel φ und Elevationswinkel ϑ übereinstimmen (trigonometrischer Pythagoras: $\sin^2(\alpha) + \cos^2(\alpha) = 1$).

$$f(\vec{\theta}) = \delta(1 - \vec{\theta}^T \vec{\theta}_0) \quad (16)$$

$$\mathcal{SHT}\{f(\vec{\theta})\} = \phi_{nm} = \int_{\mathbb{S}^2} \delta(1 - \vec{\theta}^T \vec{\theta}_0) Y_n^m(\vec{\theta}) d\vec{\theta} = Y_n^m(\vec{\theta}_0) \quad (17)$$

3.3 Enkodierung

Die Enkodierungsvorschrift ergibt sich aus der SHT für eine Punktquelle (Gleichung 17). Um ein Monosignal $s(t)$ aus einer Richtung definiert durch $\vec{\theta}_0$ zu enkodieren, müssen die Kugelflächenfunktionen für diese Position ausgewertet werden:

$$\chi_n^m(t) = Y_n^m(\vec{\theta}_0)s(t) \quad (18)$$

Analog dazu ist eine planare Enkodierung (ohne Höheninformationen) möglich, indem das Monosignal nur zweidimensional ausgewertet wird:

$$\chi_m(t) = \Phi_m(\varphi_0)s(t) \quad (19)$$

$$\Phi_m(\varphi) = \sqrt{\frac{2 - \delta_m}{2\pi}} \cdot \begin{cases} \sin(m\varphi) & \text{für } m < 0 \\ \cos(m\varphi) & \text{für } m \geq 0 \end{cases} \quad (20)$$

Für die Anzahl L an Ambisonicskanälen für planares Ambisonics gilt $L = 2m + 1$. Betrachtet man Abbildung 6 so findet man die für die planare Enkodierung relevanten Kugelflächenfunktionen am linken und rechten Rand. Der Zusammenhang zwischen planarer und peripherer (dreidimensionaler) Enkodierung wird ersichtlich, wenn die Kugelflächenfunktionen (vgl. Gleichung 9) alternativ folgendermaßen dargestellt werden:

$$Y_n^m(\varphi, \vartheta) = (-1)^{|m|} \sqrt{\frac{(2n+1)(n-|m|)!}{2(n+|m|)!}} P_n^{|m|}(\cos(\vartheta)) \Phi_m(\varphi) \quad (21)$$

Da die menschliche Lokalisierungsfähigkeit in der Horizontale besser ausgeprägt ist als in der Vertikale [Bla97], kann in praktischen Anwendungen eine schlechtere Auflösung der Höheninformation zugunsten einer erhöhten horizontalen Auflösung erwünscht sein. Dieser Ansatz wird in *mixed-order* Ambisonics (MOA) verfolgt [Tra09, FMK⁺11].

Enkodierung von Distanz. Für die Enkodierung von Distanz werden die Ambisonicssignale im Frequenzbereich ϕ_{nm} (vgl. Gleichung 17) betrachtet. Um eine beliebige Distanz zu enkodieren, müssen die sphärischen Hankelfunktionen $h_n^{(2)}(kr_l)$ aus Gleichung 15 mit Hankelfunktionen des neuen Radius $h_n^{(2)}(kr_0)$ kompensiert werden [ZPF09]. Diese Kompensation erfolgt bereits bei der Enkodierung und wird auch als Nahfeldkompensation (*near field compensation*, NFC) bezeichnet [Dan03].

$$\phi_{nm,d} = \frac{h_n^{(2)}(kr_0)}{h_n^{(2)}(kr_l)} \cdot \phi_{nm} \quad (22)$$

Bei Betrachtung des durch die Kompensation resultierenden Amplitudengangs ergeben sich mit steigender Ordnung extrem große Verstärkungen für tiefe Frequenzen (siehe Abbildung 7). Damit ist diese Distanzenkodierungsvorschrift nicht praktikabel.

Bei zusätzlicher Miteinbeziehung des Nahfeldeffekts der nur endlich weit entfernten Lautsprecher bei der Dekodierung, können stabile NFC Filter hergeleitet werden [DM04, Dan03]. Auch bei diesen Filtern entstehen für hohe Ordnungen und geringe Distanzen große Verstärkungen der tiefen Frequenzen, die es sinnvoll zu begrenzen gilt (vgl. Abbildung 8).

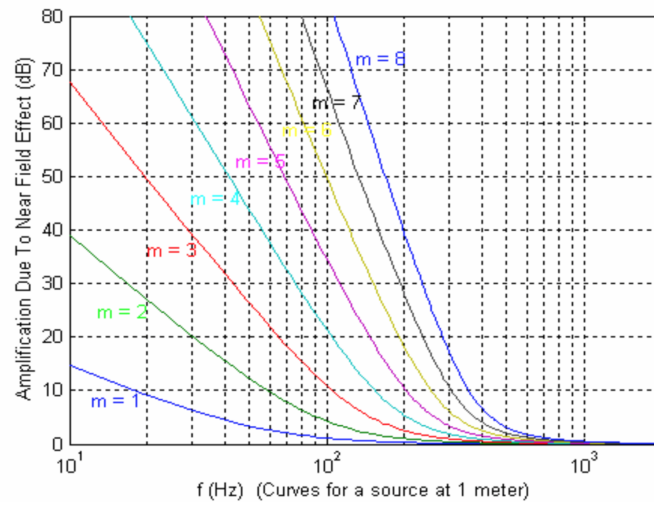


Abbildung 7 – Starke Verstärkung der tiefen Frequenzen bei Nahfeldkompensation für verschiedene Ordnungen m [DM04].

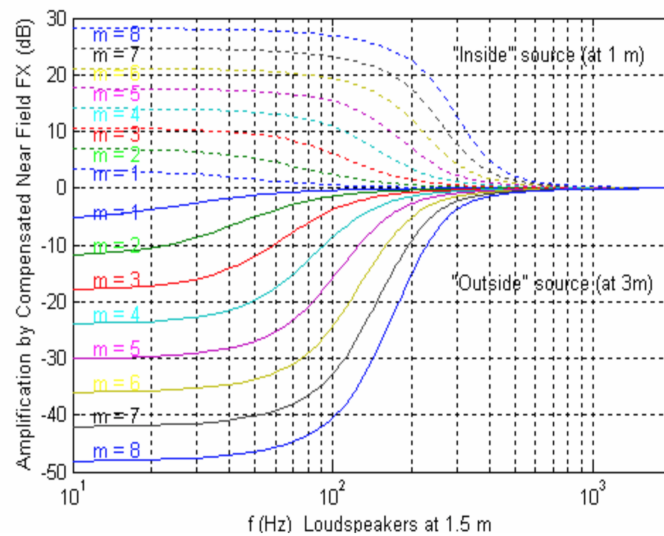


Abbildung 8 – Amplitudengänge der hergeleiteten NFC Filter unter Miteinbeziehung des Nahfeldeffekts der Lautsprecher für verschiedene Ordnungen m bei Distanz 1m ($r_0/r_l = 2/3$) und 3m ($r_0/r_l = 2$) [DM04].

3.4 Dekodierung

Eine kontinuierliche und damit nicht praktikable Dekodierungsvorschrift ergibt sich direkt aus der Reihenentwicklung der ISHT (Gleichung 12):

$$s(\vec{\theta}, t) = \sum_{n=0}^N \sum_{m=-n}^n Y_n^m(\vec{\theta}) \chi_n^m(t) \quad (23)$$

Um eine praxistaugliche Dekodierungsvorschrift für diskrete Lautsprecherpositionen abzuleiten, ist es hilfreich die En- und Dekodierung in Vektor- bzw. Matrixoperationen zu beschreiben. Dazu werden alle benötigten Kugelflächenfunktionen der Ordnung N für die Richtung $\vec{\theta}_0$ ausgewertet und in den Vektor $\vec{y}_N$ geschrieben. Nach Multiplikation des enkodierten Signals $\vec{\chi}_N(t)$ mit der Dekodermatrix $\mathbf{D}$ enthält der Vektor $\vec{s}_{ls}(t)$ die dekodierten Signale für alle Lautsprecherkanäle. Durch Multiplikation mit einer Diagonalmatrix mit den Gewichtungsfaktoren a_N wird meist eine Optimierung der Dekodierung vorgenommen.

$$\vec{s}_{ls}(t) = \mathbf{D} \text{diag}\{\vec{a}_N\} \vec{\chi}_N(t) \quad (24)$$

$$\vec{\chi}_N(t) = \vec{y}_N(\vec{\theta}_0) s(t) \quad (25)$$

Durch die Maximierung des Energievektors $\vec{r}_E$ ($\max \vec{r}_E$) werden Nebenkeulen unterdrückt, die durch die anguläre Bandbegrenzung entstehen (siehe Abbildung 5) [Fra13a]. Die Maximierung des Schnellektors erfolgt bei Nutzung eines *basic* Dekoders ohne Gewichtungsfaktoren (bzw. $a_N = 1$). Eine detaillierte Beschreibung der Vektoren und der Auswirkung ihrer Optimierung befindet sich in Abschnitt 5.1.1. Weitere Optimierungsansätze wie beispielsweise *in-Phase* sind möglich [Son03].

Um eine $\max \vec{r}_E$ -Optimierung für einen 2D Dekoder durchzuführen, müssen die Komponenten des Ambisonicssignals je nach Gesamtordnung N und Ordnung m_n der gerade betrachteten Kugelflächenfunktion mit einem Faktor a_n multipliziert werden. Dieser Faktor berechnet sich zu [Dan00]:

$$a_n = \cos \left(\frac{\pi}{2(N+1)} m_n \right) \quad (26)$$

In 3D können die Koeffizienten mit folgender Näherungsgleichung berechnet werden:

$$a_n = P_n \left(\cos \left(\frac{137,9^\circ}{N+1,51} \right) \right) \quad (27)$$

Eine Herleitung befindet sich im Appendix von [ZF12].

Für die Berechnung einer geeigneten Dekodermatrix $\mathbf{D}$ bei diskreten Lautsprecherpositionen gibt es verschiedene Ansätze. Einige davon werden im Folgenden skizziert.

3.4.1 Mode-matching

Ein mathematisch begründeter Ansatz ist das sogenannte *mode-matching*. Dabei wird ein modaler Quellstärkenabgleich (*modal source strength matching*) durchgeführt, d. h. es werden die modalen Quellstärken von kontinuierlichen und diskretisierten Anregungspositionen gleichgesetzt [Zot09]. Vorher werden die zu vergleichenden Quellstärkeverteilungen angular bandbegrenzt (*angular smoothing*). Dies entspricht dem Auflösungsverlust durch Begrenzung der Kugelflächenfunktionen ab einer gewissen Ordnung N . Die Bandbegrenzung wird im Folgenden mit dem idealen Tiefpassfilter $\mathcal{B}_N$ dargestellt (vgl. Gleichungen 16, 17).

$$\mathcal{B}_N\{\hat{f}(\vec{\theta})\} \stackrel{!}{=} \mathcal{B}_N\{f(\vec{\theta})\} \quad (28)$$

$$\Rightarrow \hat{\phi}_{nm} \stackrel{!}{=} \phi_{nm} \quad (29)$$

$\hat{f}(\vec{\theta})$ setzt sich aus einer Linearkombination der diskreten Lautsprechergains g_l und der Quellstärke einer Punktquelle zusammen (vgl. Gleichung 16):

$$\mathcal{B}_N\{\hat{f}(\vec{\theta})\} = \sum_{l=1}^L g_l \mathcal{B}_N\{\delta(1 - \vec{\theta}^T \vec{\theta}_l)\} \quad (30)$$

In Vektor-/Matrixnotation ergibt sich damit folgender Zusammenhang:

$$\vec{\phi}_N = \mathbf{Y}_N \vec{g} \stackrel{!}{=} \vec{\phi}_N \quad (31)$$

Der Vektor $\vec{g}$ enthält die gesuchten L diskreten Lautsprechergains. Der Vektor $\vec{\phi}_N$ enthält die Entwicklungskoeffizienten ϕ_{nm} bis zur Ordnung N . Die Matrix $\mathbf{Y}_N$ enthält für jede diskrete Lautsprecherposition $\vec{\theta}_l$ einen Vektor $\vec{y}_N(\vec{\theta}_l)$ mit allen Kugelflächenfunktionen Y_n^m bis zur Ordnung N :

$$\mathbf{Y}_N = (\vec{y}_N(\vec{\theta}_1), \vec{y}_N(\vec{\theta}_2), \dots, \vec{y}_N(\vec{\theta}_L)) \quad (32)$$

Gleichung 31 stellt damit den Ausgangspunkt des modalen Quellstärkeabgleichs dar. Um den Abgleich durchzuführen wird die Dekodermatrix $\mathbf{D}_N$ eingeführt, die die Gains $\vec{g}$ kontrolliert:

$$\vec{g} = \mathbf{D}_N \vec{\phi}_N \quad (33)$$

$$\Rightarrow \vec{\phi}_N = \mathbf{Y}_N \mathbf{D}_N \vec{\phi}_N \stackrel{!}{=} \vec{\phi}_N \quad (34)$$

Ein exakter Abgleich ist also möglich, wenn die Dekodermatrix $\mathbf{D}_N$ die Rechtsinverse von $\mathbf{Y}_N$ darstellt. Da $\mathbf{Y}_N$ mit den Dimensionen $(N+1)^2 \times L$ allgemein keine quadratische Matrix ist, muss die Pseudoinverse herangezogen werden:

$$\mathbf{D}_N = \mathbf{Y}_N^T (\mathbf{Y}_N \mathbf{Y}_N^T)^{-1} \quad (35)$$

Eine Einschränkung bildet die eindeutige Bestimmung der Pseudoinversen bei einer bestimmten Lautsprecheranordnung. Die Pseudoinverse liefert nur brauchbare Ergebnisse, wenn die Matrix gut konditioniert ist, d. h. das Verhältnis zwischen maximalem und minimalem Eigenwert klein ist [Son03]. Eine korrekt gestellte (*well-posed*), also eindeutige und stabile Lösung ergibt sich nur bei $L \geq (N+1)^2$ gleichmäßig angeordneten Lautsprechern [ZF12].

3.4.2 Sampling

Ein anderer Ansatz einen Dekoder zu erhalten, stellt das sogenannte *sampling* Verfahren dar. Dabei werden die Lautsprecher gains $\vec{g}$ durch Abtastung an jeder diskreten Lautsprecherposition erhalten [ZF12]:

$$\vec{g} = \mathbf{D}_N \vec{\phi}_N = \mathbf{Y}_N^T y_N(\theta_0) \quad (36)$$

$$\Rightarrow \mathbf{D}_N = \mathbf{Y}_N^T \quad (37)$$

Meist wird die Dekodermatrix noch mit dem mittleren Winkelsegment eines Lautsprechers auf der Einheitskugel normalisiert:

$$\mathbf{D}_N = \frac{4\pi}{L} \mathbf{Y}_N^T \quad (38)$$

3.4.3 Weitere Ansätze

Energy preserving Dekoder basieren auf Singulärwertzerlegung der Dekodermatrix. Dadurch wird die Beschränkung auf gleichmäßige Lautsprecheranordnungen vermieden [ZPN12].

Um eine optimale Dekodermatrix zu erhalten, ist eine möglichst gleichmäßige Lautsprecherverteilung von Vorteil. Dies ist mit sogenannten *sphärischen t-designs* möglich.

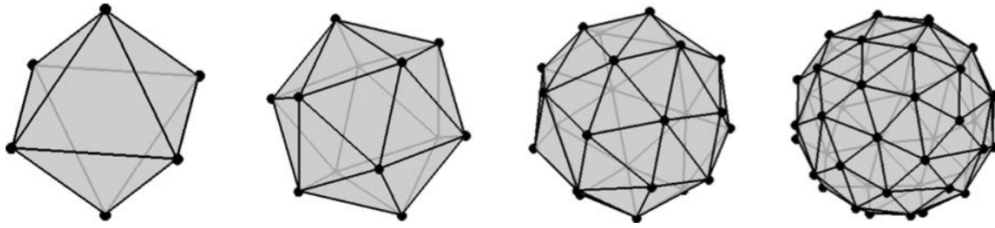


Abbildung 9 – Sphärische t-designs für $t=3$ (Oktaeder), $t=5$ (Ikosaeder), $t=7$, $t=9$ [ZF12].

Bei Einsatz einer Lautsprecherverteilung nach einem *sphärischen t-design* mit $t \geq 2N+1$ vereinfachen sich die Gleichungen der Dekodermatrizen von *sampling*, *mode-matching* und *energy preserving* Dekodern zum gleichen Resultat (Gleichung 38) [ZF12]. Es kann außerdem gezeigt werden, dass sich für *t-designs* mit $t \geq 2N+1$ unabhängig von der Richtung θ_0 ein Energievektor $\vec{r}_E$ mit konstanter Länge und eine konstante Energieverteilung ergeben [ZF12]. Eine solche Lautsprecherverteilung ist für reale Lautsprecheranordnungen allerdings nur schwer zu erreichen.

AllRAD (*All-Round Ambisonic Decoding*) Dekoder setzen *sphärische t-designs* ein, um damit Positionen für virtuelle Lautsprecher zu bestimmen. Diese virtuellen Lautsprecher werden dann mittels VBAP (*Vector Base Amplitude Panning*) auf die realen Lautsprecher abgebildet. Damit ist der *AllRAD* Dekoder für beliebige Lautsprecherverteilungen

geeignet. Die Qualität des Resultats steigt allerdings mit einer homogenen Lautsprecher-
verteilung [ZF12].

3.5 Schallfeldrotation

Durch die Orthogonalität bzw. Orthonormalität der Kugelflächenfunktionen können diese durch lineare Operationen ineinander übergeführt werden. Daher lässt sich eine Schallfeldrotation durch Multiplikation einer Rotationsmatrix mit den Ambisonicssignalen realisieren. Während bei planarem Ambisonics nur eine Rotation um die z-Achse möglich ist, lässt sich für periphones Ambisonics eine Rotation um alle drei Koordinatenachsen realisieren. Für eine Bestimmung der Rotationsmatrizen in 3D wird auf [Kro14] verwiesen. Im 2D-Fall wird die Rotationsmatrix über trigonometrische Summensätze bestimmt und hat für eine Rotation um den Winkel φ folgenden Aufbau [Son03]:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & \cos(\varphi) & -\sin(\varphi) & 0 & 0 & \dots \\ 0 & \sin(\varphi) & \cos(\varphi) & 0 & 0 & \dots \\ 0 & 0 & 0 & \cos(2\varphi) & -\sin(2\varphi) & \dots \\ 0 & 0 & 0 & \sin(2\varphi) & \cos(2\varphi) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (39)$$

4 Binaurale Auralisation

Kleiner et al. [KDS93] definieren Auralisation wie folgt:

“Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space.”

Auralisation beschreibt also die Hörbarmachung einer Quelle im virtuellen Raum im Gegensatz zur Darstellung der Quelle durch numerische Größen. Besonders im Bereich der raumakustischen Simulation findet Auralisation Anwendung zur Planung der akustischen Eigenschaften von Räumen.

Der Prozess der Auralisation kann in die drei Systemblöcke Schallgenerierung, Schallausbreitung und Schallreproduktion aufgeteilt werden [Vor14]. Die Schallgenerierung kann z. B. durch ein *trockenes* Mikrofonsignal, d. h. ein reines Quellsignal ohne Rauminformation erfolgen. Die Eigenschaften der Schallausbreitung können durch die Impulsantwort $h(n)$ eines akustischen Systems (z. B. eines Raumes) vollständig repräsentiert werden, wenn das akustische System durch ein linear-zeitinvariantes System angenähert werden kann. Die Reproduktion an einer bestimmten Position erfolgt dann durch Faltung des Quellsignals mit der Systemimpulsantwort an der gewünschten Position. Die lineare Faltung zweier Signale wird in der Praxis aufgrund des geringeren Rechenaufwands oft durch die Multiplikation der Signale im Frequenzbereich realisiert. Bei binauraler Auralisation muss zusätzlich die Außenohrübertragungsfunktion in die Berechnung miteinfließen: Die Kombination von Raumimpulsantwort und Außenohrübertragungsfunktion wird binaurale Raumimpulsantwort (*binaural room impulse response*, BRIR) genannt [KDS93].

4.1 Raumsimulation

Für die Berechnung der Raumimpulsantwort des zu simulierenden Raumes stehen wellenbasierte, geometrische und künstliche Methoden zur Verfügung (vgl. Abbildung 10) [Vor08]. Hier wird ein kurzer Überblick über einige geometrische und wellenbasierte Herangehensweisen gegeben. Für künstliche Methoden und weitere Beispiele der anderen Kategorien wird auf [Vor08, SK02] und [Zö11] verwiesen.

4.1.1 Raumsimulation nach geometrischen Prinzipien

Grundlage der geometrischen Raumakustik ist die Annahme von strahlenförmiger Schallausbreitung. Dabei können die Gesetze der Optik angewendet werden. Wird eine Schallwelle in ihrer Ausbreitung von einem Hindernis gestört, erfolgt je nach Verhältnis der Wellenlänge zur Hindernisgröße Reflexion, Beugung oder Brechung. Simulation mittels geometrischer Raumakustik basiert auf Reflexion. Die Wellenphänomene Beugung und Brechung bilden damit die Grenzen der Anwendbarkeit. Somit bieten geometrische Ansätze prinzipbedingt nur korrekte Lösungen, wenn die Wellenlänge ausreichend klein gegenüber den Hindernissen und ausreichend groß gegenüber der Oberflächenstruktur der Hindernisse ist [SS15].

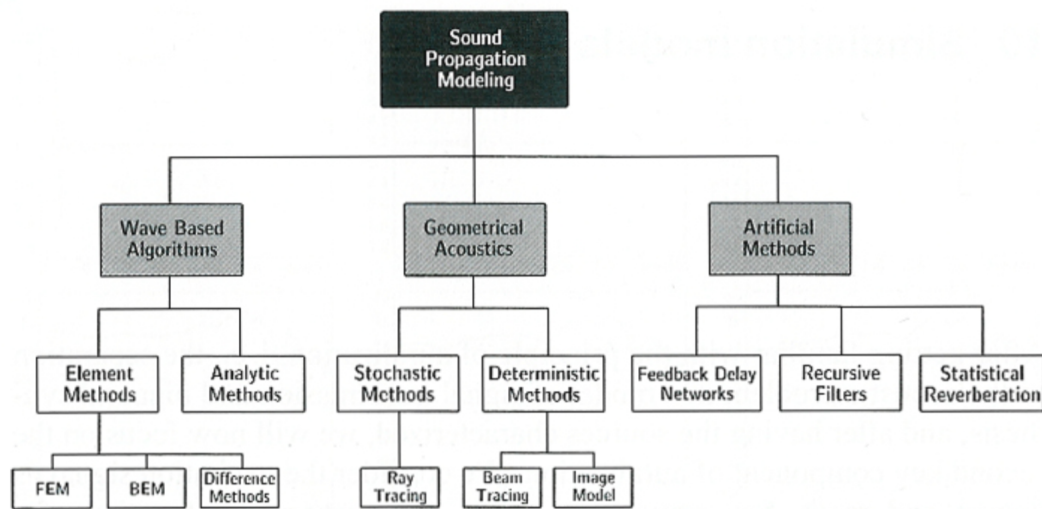


Abbildung 10 – Simulationsmodelle der Schallausbreitung [Vor08].

Spiegelquellenmethode. Durch Spiegelung einer Schallquelle an allen Raumbegrenzungsflächen ergeben sich die sogenannten Spiegelquellen. Diese repräsentieren frühe Reflexionen der Originalquelle. Werden die gespiegelten Quellen wiederholt an den Begrenzungsflächen gespiegelt, entstehen Spiegelquellen höherer Ordnung und die Anzahl an berücksichtigten Reflexionen steigt [SS15] (siehe Abbildung 11).

Wird eine Schallquelle mit Quelldirektivität eingesetzt, so muss diese Direktivität ebenso auf die Spiegelquellen übertragen werden. Bei dreidimensionaler Anwendung müssen zusätzlich Reflexionen an Decke und Boden berücksichtigt werden.

Die Spiegelquellenmethode ist nicht auf rechteckige Räume begrenzt, sie kann auf beliebige Polygone erweitert werden [Bor84].

Ray tracing. *Ray tracing*, zu deutsch Strahlenverfolgung, ist im Gegensatz zur Spiegelquellenmethode eine stochastische Methode. Dabei wird die durch eine Schallstrahlendichte ausgesendete Energieverteilung in einem Empfängervolumen ausgewertet. Die ausgesendete Strahlendichte kann dabei omnidirektional oder infolge einer Quelldirektivität richtungsgewichtet erfolgen [SS15].

Eine Schallquelle sendet zum Zeitpunkt $t = 0$ Schallstrahlen in verschiedene Richtungen aus. Jedesmal wenn einer dieser Schallstrahlen an einer Oberfläche reflektiert wird, erfolgt eine Richtungsänderung und eine Energieabnahme. Beim Durchtreten eines (z. B. kugelförmigen) Zielvolumens werden die Zeitdifferenz und die Energiedifferenz im Vergleich zum ausgesandten Schallstrahl ausgewertet (siehe Abbildung 12a).

Absorptionseigenschaften der Wände können durch Anpassung der Energieabnahme mit eingerechnet werden. Sollen spektrale Eigenschaften der Absorption berücksichtigt werden, muss die Energieabnahme für verschiedene Frequenzbänder differenziert behandelt werden. Die Auswertung wird beendet, sobald eine benutzerdefinierte Minimalenergie

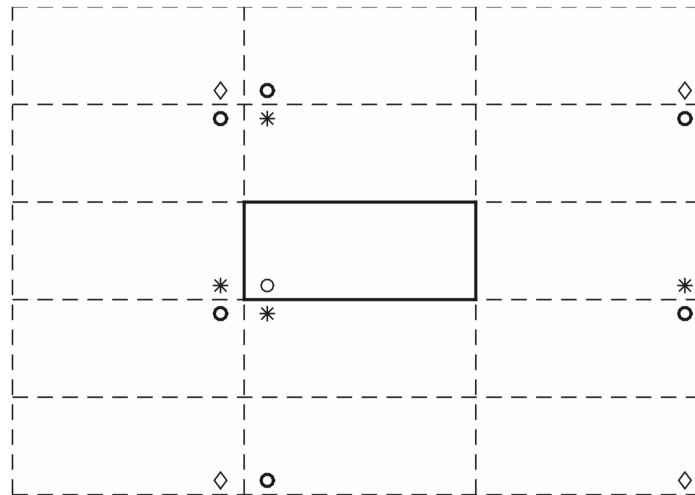


Abbildung 11 – Spiegelquellen für einen Rechteckraum. Originalquelle o, Spiegelquellen erster Ordnung *, zweiter Ordnung O und dritter Ordnung ◇ sind eingezeichnet [SS15].

oder eine Maximalzeit der Ausbreitung erreicht wird.

Auch Schallstreuung bei Reflexion kann mit in die Simulation einfließen. Dabei sind mehrere Szenarien denkbar [SS15]: In Abbildung 12b wird der reflektierte Schallstrahl in einen primären gespiegelten Strahl und mehrere gleichmäßig abgestrahlte diffuse Schallstrahlen aufgeteilt. In Abbildung 12c ist die Reflexionsrichtung abhängig von einem Streukoeffizienten der Wand. Wenn der Streukoeffizient größer als eine Zufallszahl ist, wird in eine zufällige Richtung abgestrahlt, sonst erfolgt eine gespiegelte Abstrahlung. Eine andere Methode wird als *diffuse rain* bezeichnet und ist in Abbildung 12d zu sehen. Dabei wird jede Reflexionsrichtung verfolgt, die nach einer Anzahl an Reflexionen zum Empfängervolumen führt.

Welches Verfahren angewendet wird, muss in der Praxis nach Abwägung der benötigten Genauigkeit und der zur Verfügung stehenden Rechenleistung entschieden werden.

Es existiert eine Vielzahl an Erweiterungsmöglichkeiten und Modifikationen des *basic ray tracing* Verfahrens. Auch hybride Verfahren sind möglich. Einen Überblick liefert [SS15].

4.1.2 Raumsimulation nach wellentheoretischen Prinzipien

Wellenbasierte Methoden ermöglichen die Berechnung der Raumimpulsantwort durch numerische Lösung der Wellengleichung. Diese Herangehensweise führt prinzipbedingt zu einem hohen Rechenaufwand, der mit größerem betrachteten Frequenzbereich ansteigt. Zudem übersteigt die Genauigkeit der Berechnung detaillierter Interferenzmuster mittels wellenbasierter Ansätze bei hohen Frequenzen die Frequenzauflösung des menschlichen Ohrs [SK02]. Daher werden wellenbasierte Methoden meist nur zur Berechnung der Schallausbreitung für tiefe Frequenzen herangezogen. Bei hybrider Anwendung können sie hier die Schwächen anderer Verfahren (z. B. Vernachlässigung von Beugungsef-

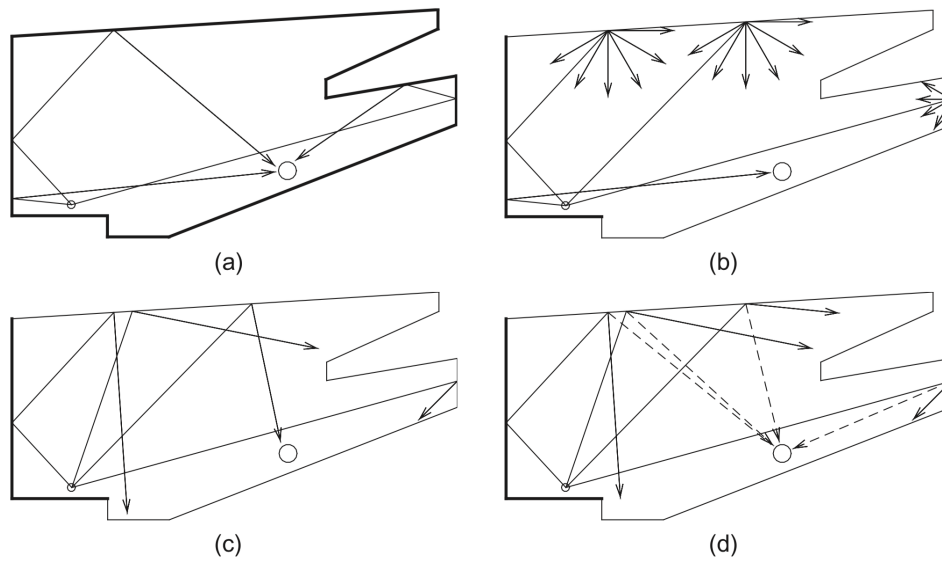


Abbildung 12 – Ray tracing mit verschiedenen Arten der Reflexion. In a) erfolgt gespiegelte Abstrahlung, in b) Aufteilung in primären Spiegelstrahl und eine Vielzahl an diffus reflektierten Strahlen. In c) wird abhängig von einem Streukoeffizienten zufällig zwischen gespiegelter und diffuser Reflexion entschieden und in d) wird die *diffuse rain* Methode angewendet [SS15].

fekten bei geometrischen Verfahren) ausgleichen.

Zu den numerischen Lösungsmethoden der Wellengleichung zählen die Methode endlicher Elemente (*finite element method*, FEM) und die Randelementmethode (*boundary element method*, BEM). Bei der FEM wird der Raum in Volumenelemente aufgeteilt, für die anschließend eine numerische Lösung der Wellengleichung gefunden wird. Bei Anwendung der BEM wird die den Raum umgebende Grenzfläche in Oberflächenelemente aufgeteilt und die numerische Lösung der Wellengleichung mit Hilfe der sich ergebenden Randbedingungen erhalten (siehe Abbildung 13). Die benötigte Genauigkeit der Unterteilung ist für beide Verfahren abhängig von der kleinsten zu simulierenden Wellenlänge, da das Samplingtheorem erfüllt sein muss [SK02].

Eine weitere wellenbasierte Methode ist die *digital waveguide mesh* Methode [SHLV99]. Ein digitaler Wellenleiter (*waveguide*) wird dabei durch ein bidirektionales Delay repräsentiert. Die Verbindung der Wellenleiter erfolgt durch Knotenpunkte, genannt *scattering junctions*. Bei Übertragung einer Welle über *scattering junctions* wird die Energie prinzipiell verlustlos an die angrenzenden Knotenpunkte verteilt. Erst durch Einführung von Gainfaktoren kleiner Eins oder Filtern wird ein Abklingen der Wellenausbreitung erreicht. Diffusion und Absorption kann durch Einbringen von Grenzbedingungen simuliert werden. Auch hierbei ist die örtliche Auflösung des Netzes ausschlaggebend für die Genauigkeit der Simulation [KHI05].

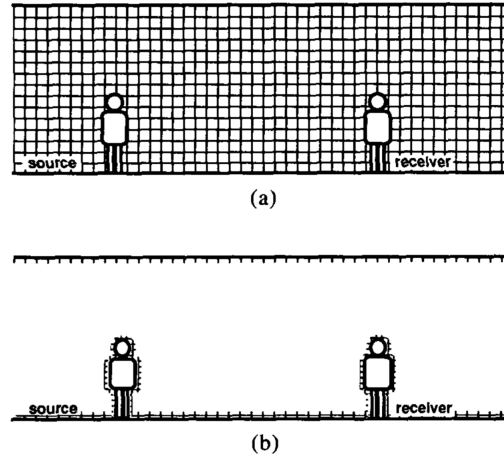


Abbildung 13 – (a) *Finite element method* mit Unterteilung des Raumes in Teilvolumen, (b) *boundary element method* mit Unterteilung der Oberfläche [KDS93].

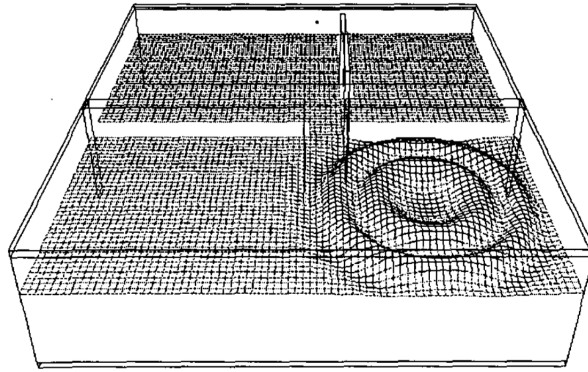


Abbildung 14 – Zweidimensionaler Ausschnitt aus einem *digital waveguide mesh* [SHLV99].

4.2 Binaurale Synthese

Binaurale Synthetisierung einer Quelle aus einer Richtung $\vec{\theta}_0$ kann prinzipiell durch Faltung des Zeitsignals der Quelle $x(t)$ mit den korrespondierenden HRIRs für das linke und rechte Ohr erfolgen [Vor14]:

$$s_l(t) = HRIR_l(\vec{\theta}_0) * x(t) \quad (40)$$

$$s_r(t) = HRIR_r(\vec{\theta}_0) * x(t) \quad (41)$$

Dies ist allerdings nur für diskrete Punkte möglich, für die eine HRIR vorhanden ist. Bei dynamischer Quellrichtungsänderung ergibt sich daher ein Problem, wenn eine Schallquelle aus einer Richtung synthetisiert werden soll, für die keine HRIR vorhanden ist.

Die klassischen Ansätze der Binauralsynthese lösen dieses Problem durch verschiedene Arten von HRIR bzw. HRTF Interpolation [HBS99,SHLV99,QdS11]. Im Folgenden wird das Problem mittels eines virtuellen Ambisonics Ansatzes gelöst [NMSH03].

Virtueller Ambisonics Ansatz. Der virtuelle Ambisonics Ansatz basiert auf der Annahme virtueller Lautsprecher: Virtuelle Lautsprecher werden im Raum platziert und das Ambisonicssignal für die Positionen $\vec{\theta}_q$ dieser m Lautsprecher dekodiert. Anschließend werden die erhaltenen Signale mit zu den Lautsprecherpositionen passenden HRIRs gefaltet. Durch Superposition werden die Kopfhörersignale für das linke und rechte Ohr $s_l(t)$ und $s_r(t)$ erhalten [NMSH03]:

$$s_l(t) = \sum_{q=1}^m HRIR_{l,q}(\vec{\theta}_q) * (e_q^T \mathbf{D}_{vls} \vec{\chi}_N(t)) \quad (42)$$

$$s_r(t) = \sum_{q=1}^m HRIR_{r,q}(\vec{\theta}_q) * (e_q^T \mathbf{D}_{vls} \vec{\chi}_N(t)) \quad (43)$$

Dabei bezeichnet $\mathbf{D}_{vls}$ die Dekodermatrix für die virtuellen Lautsprecherpositionen und $\vec{\chi}_N(t)$ einen Vektor, der die enkodierten Ambisonicssignale enthält (vgl. Gleichung 24). $\mathbf{D}_{vls}$ hat für periphones Ambisonics die Dimensionen $m \times (N + 1)^2$, sodass nach Multiplikation mit $\vec{\chi}_N(t)$ (Dimension $(N + 1)^2 \times 1$) ein Vektor entsteht, dessen Länge der Anzahl virtueller Lautsprecher m entspricht. Durch Multiplikation mit dem transponierten Einheitsvektor e_q^T wird jedes Vektorelement (des Ergebnisvektors aus $\mathbf{D}_{vls} \vec{\chi}_N(t)$) mit der korrespondierenden HRIR gefaltet.

Durch die festen Positionen der virtuellen Lautsprecher muss keine Interpolation zwischen HRIRs bzw. HRTFs erfolgen. Bewegung der Quellen bzw. des Hörers erfolgt durch Ambisonics-Enkodierung bzw. Rotation der Ambisonicssignale. Fehler bei der Schallfeldreproduktion mittels Ambisonics durch eine begrenzt große *sweet area* werden vermieden, da sich der Hörer bei dieser Technik (virtuell) immer im Koordinatenursprung befindet.

Wird von einer symmetrischen Anthropometrie des Kopfes ausgegangen, wie es z. B. bei Kunstköpfen der Fall ist, kann die Hälfte der Faltungen durch Nutzung der HRTF nur eines Ohres eingespart werden. Das Signal am rechten Ohr entspricht dann dem Signal am linken Ohr, wenn das Schallfeld an der Medianebene gespiegelt würde. Aufgrund der Symmetrie der Kugelflächenfunktionen kann eine Spiegelung an der xz-Ebene in der Ambisonicsdomäne durch Invertierung der Polarität der Kugelflächenfunktionen mit $m < 0$ (linke Seite in Abbildung 6) erfolgen [PPQ16].

5 Objektiver Vergleich binauraler Ambisonics Dekoder

5.1 Vergleichskriterien

Im Folgenden werden einige objektive Kriterien zum Vergleich binauraler Ambisonics Dekoder vorgestellt. Weitere objektive Analysemöglichkeiten beinhalten z. B. die Betrachtung der Impulsantwort im Zeitbereich [Col13] und die Untersuchung der Klangfarbe mittels binauralem Lautheitsmodell [LX16]. Eine Sammlung binauraler Hörmodelle ist in [SM13] zu finden.

5.1.1 Energie- und Schnellevektoren

Eine Möglichkeit der objektiven Evaluierung und Optimierung von Ambisonicsdekodern geht auf Gerzon zurück und beinhaltet die Auswertung des Schnellektors $\vec{r}_V$ und des Energievektors¹ $\vec{r}_E$ im Mittelpunkt des Lautsprecherarrays [Ger92]. Der Schnellektor $\vec{r}_V$ beschreibt die wahrgenommene Quellposition aufgrund des Gesamtschnellektors aller Lautsprecher im Frequenzbereich bis etwa 700 Hz. Hier erfolgt Lokalisation von lateralen Quellen vor allem aufgrund von ITDs. Im Frequenzbereich von 500 Hz bis 5000 Hz wird die Lokalisation mittels Energievektor $\vec{r}_E$ (auf Basis der ILDs) beschrieben [Ger92]. Schnelle- und Energievektor berechnen sich bei K Lautsprechern mit Gain g_k und Position $\vec{u}_k$ wie folgt:

$$\vec{r}_V = \frac{\sum_{k=1}^K g_k \vec{u}_k}{\sum_{k=1}^K g_k} \quad (44)$$

$$\vec{r}_E = \frac{\sum_{k=1}^K g_k^2 \vec{u}_k}{\sum_{k=1}^K g_k^2} \quad (45)$$

Durch die Division durch die Summe der Lautsprecher gains (bei $\vec{r}_V$) bzw. der quadrierten Lautsprecher gains (bei $\vec{r}_E$) erfolgt eine Normierung. Die Beträge von $\vec{r}_V$ und $\vec{r}_E$ werden damit maximal Eins. Der Betrag des Energievektors $\vec{r}_E$ kann nie Eins werden, wenn das Quellsignal durch mehr als einen Lautsprecher repräsentiert wird. Bei der Lokalisation einer realen Quelle unter Freifeldbedingungen stimmt die Richtung beider Vektoren überein und der Betrag beider Vektoren wird Eins.

In [Fra13b] wird gezeigt, dass der Betrag des Energievektors zudem die wahrgenommene Quellbreite gut beschreibt: Ein kürzerer $\vec{r}_E$ korreliert mit einer größeren empfundenen Quellbreite. Fluktuationen der Länge des Energievektors können ein Prädiktor für wahrgenommene Klangfarbenänderung sein [Fra13a]. Zudem kann der Energievektor laut [WFZ14] auch bei vertikaler Quellauslenkung ein Prädiktor für die Lokalisation sein. Die Analyse von Richtung und Betrag von $\vec{r}_V$ und $\vec{r}_E$ stellt damit ein erstes Kriterium zur Beurteilung eines Ambisonicsdekoders dar.

1. Genau genommen stellt der Vektor $\vec{r}_E$ eine Intensität dar. Energie ist keine gerichtete Größe und kann daher nicht mit einem Vektor dargestellt werden. Um in Übereinstimmung mit der Literatur zu bleiben, wird der Vektor hier trotzdem weiterhin als Energievektor bezeichnet.

In der *Ambisonic Decoder Toolbox*² [HBL12] und der *Higher Order Ambisonics Toolbox*³ [Pol16] für *MATLAB*⁴ werden Skripte zur Analyse von Energie- und Schnellevektor bei 3D Ambisonics Dekodern bereitgestellt. Diese Skripte werden zur Analyse von Energie- und Schnellevektor verwendet bzw. für 2D Dekoder adaptiert.

5.1.2 Lokalisation in der Horizontalebene mittels psychoakustischer Hörmodelle

Ein psychoakustisches Modell zur Simulation der Lokalisation bei lateralen Quellsignalen wurde von Lindemann entwickelt [Lin86]. Das Modell erweitert das Delayline-Modell [Jef48]. Dieses modelliert die Wahrnehmung interauraler Laufzeitdifferenzen mittels Kreuzkorrelation von Signalen innerhalb zweier Delaylines, die jeweils ein Ohr repräsentieren. Die Maxima der Kreuzkorrelation ausgewertet in Frequenzbändern, stellen den Grad der Lateralisation dar. Die Erweiterung erfolgt durch zusätzliche Berücksichtigung monauraler Faktoren. Außerdem fließen interaurale Pegeldifferenzen durch Berücksichtigung kontralateraler Hemmung in die Simulation mit ein.

Ein anderes Modell basiert auf *rate coding* und der Auswertung interauraler Phasendifferenzen [DEH11]. In [Fra14] und [SBvW16] wird allerdings gezeigt, dass für Ambisonics erster bzw. dritter Ordnung bei zentraler Hörposition die Vorhersage der Lokalisationswahrnehmung mittels Energievektor $\vec{r}_E$ besser oder gleich gut, wie mit den genannten Hörmodellen, funktioniert. Daher werden diese Modelle im Folgenden nicht berücksichtigt.

5.1.3 Lokalisation in Sagittalebene mittels psychoakustischem Hörmodell

Ein Modell zur Simulation der Lokalisationswahrscheinlichkeit in Sagittalebene wird in [BML13] vorgestellt. Es erweitert ein Modell zur Lokalisation in der Medianebene [LB02] mittels monauraler Faktoren. Die Modellierung erfolgt durch den Vergleich des einfallenden Spektrums mit gespeicherten richtungsabhängigen Anteilen von HRTFs (=DTFs, *directional transfer functions*). Nach Betrachtung der Übereinstimmung mit jeder einzelnen DTF wird die Lokalisationswahrscheinlichkeit gemittelt und eine Wahrscheinlichkeitsdichtefunktion (PDF, *probability density function*) berechnet. Die Erweiterung des Modells beinhaltet unter anderem die Berücksichtigung der Bandbreite des empfangenen Signals, Kalibrierung nach hörspezifischer Empfindlichkeit und binaurale Gewichtungsfaktoren. Das Modell berechnet außerdem den mittleren Quadrantenfehler (QE) und den mittleren lokalen polaren RMS (*root mean square*, *Effektivwert*) Fehler (PE). Der QE für eine Richtung ist der Prozentsatz an angulären Fehlern größer gleich 90°. Bei hohen QEs ist also die Wahrscheinlichkeit einer *Vorne-Hinten-* bzw. *Oben-Unten-Verwechslung* höher. Der PE berechnet sich aus Erwartungswert und Varianz des modellierten Quellrichtungswinkels innerhalb der korrekten Hemisphäre. Die angulären Fehler größer gleich 90°, die zu den QEs führen, sind also nicht miteingerechnet. Ein geringer

2. <https://bitbucket.org/ambidecodertoolbox/adt.git> (zuletzt aufgerufen am 27.05.17)

3. <https://github.com/polarch/Higher-Order-Ambisonics> (zuletzt aufgerufen am 05.06.17)

4. <http://www.mathworks.com/> (zuletzt aufgerufen am 27.05.17)

PE ist also ein Anzeichen für präzise Lokalisation.

Für dieses Modell ist eine Implementierung in der *Auditory Modelling Toolbox*⁵ [SM13] für MATLAB vorhanden.

5.1.4 Relativer spektraler HRTF-Synthesefehler

Die Differenz zwischen dem Frequenzgang des binauralen Ambisonicssystems und dem Frequenzgang der benutzten HRTF aus der gleichen Richtung bezogen auf die HRTF, ergibt einen relativen spektralen Fehler. Dieser ermöglicht die Beurteilung der spektralen Veränderung durch das Ambisonicssystem [KD15]:

$$K(\vec{\theta}_0, \omega_k) = 20 \log \left(\frac{|H_{Ambi,l}(\vec{\theta}_0, \omega_k) - HRTF_l(\vec{\theta}_0, \omega_k)|}{|HRTF_l(\vec{\theta}_0, \omega_k)|} \right) \quad (46)$$

Hierbei wird nur HRTF und Frequenzantwort des Systems an einem Ohr (in Gleichung 46 das linke Ohr) betrachtet. $H_{Ambi,l}(\vec{\theta}_0, \omega_k)$ ist der Wert der Übertragungsfunktion des Ambisonicssystems bei der Frequenz ω_k und Quellrichtung $\vec{\theta}_0$ am linken Ohr. $HRTF_l(\vec{\theta}_0, \omega_k)$ ist der Wert der HRTF des linken Ohrs bei gleicher Frequenz und Richtung. Durch Hinzunahme der Betragsfunktion repräsentiert K den Amplitudenfehler bei der Frequenz ω_k . Eine qualitative Beurteilung der Änderung der sich ergebenden Lokalisationsgenauigkeit durch den relativen spektralen HRTF-Synthesefehler ist nicht möglich. Der spektrale Fehler K lässt lediglich eine Aussage über die relative Verbesserung oder Verschlechterung des Systems zu.

5.1.5 Klangfarbenänderung durch Fluktuation der Anzahl aktiver Lautsprecher

Bei Überlagerung mehrerer gleicher, aber zeitverzögerter Signale entstehen Kammfiltereffekte [Ahn88]. In der Praxis entstehen diese z. B. bei gleichzeitigem Abspielen des gleichen Signals mit mehreren Lautsprechern, die verschiedene Distanzen zu den beiden Ohren besitzen. Dies ist bei einer sphärischen Lautsprecheranordnung bei allen Lautsprechern außerhalb der Medianebene der Fall. Daher wird in [Fra13a] die Fluktuation der Anzahl aktiver Lautsprecher bei einem kontinuierlichen Panning als Prädiktor für eine Änderung der Klangfarbe vorgeschlagen. Die Klangfarbenänderung ist dabei stärker wenn der Sprung in der Lautsprecheranzahl größer ist und wenn die Anzahl der beteiligten Lautsprecher geringer ist. Das bedeutet, dass beispielsweise ein Sprung von drei auf vier beteiligte Lautsprecher eine weniger große Klangfarbenfluktuation nach sich zieht, als ein Wechsel von einem auf zwei Lautsprecher. Auch bei binauraler Synthese mit virtuellen Lautsprechern kommen die zu Kammfiltereffekten führenden Laufzeitdifferenzen durch die Faltung der Signale mit den HRIRs zur Geltung.

5. <http://amtoolbox.sourceforge.net/> (zuletzt aufgerufen am 27.05.17)

5.2 Untersuchte Parameter und Erwartungen

Die unter Abschnitt 5.1 angeführten Kriterien werden bei 2D Dekodern und 3D Dekodern untersucht. Die untersuchten Parameter sind Ambisonicsordnung n , $\max \vec{r}_E$ -Optimierung und Anzahl der virtuellen Lautsprecher. Die virtuellen Lautsprecher werden dabei gleichmäßig auf einem Kreis (2D) oder einer Kugeloberfläche (3D) angeordnet, wie es bei virtuellen Lautsprecheranordnungen leicht möglich ist. Bei einer solchen Lautsprecheranordnung und Verwendung von $L \geq 2n + 1$ Lautsprechern (2D) bzw. einem t -design mit $t \geq 2n + 1$ (3D) sind die Dekodermatrizen für *sampling*, *mode-matching*, *energy-preserving* und auch *AllRAD* Dekoder äquivalent [ZF12, Fra13a]. Bei beiden Dekodertypen werden die *MIT Kemar* Kunstkopf HRTFs aus der SOFA-Datenbank (siehe Abschnitt 6.1) verwendet.

Da eine Begrenzung der Ordnung den Abbruch der Reihenentwicklung der Schallfeldreproduktion (Gleichung 15) bedeutet, kann das Schallfeld mit höherer Ambisonicsordnung präziser resynthetisiert werden. Ein Abbruch der Reihenentwicklung in Gleichung 15 bedeutet zugleich, dass nur eine endliche Anzahl an Besselfunktionen in das Resultat miteinfließt. In [Pol96, WA01] findet sich die Herleitung des Abbildungsfehlers des Ambisonicssystems aufgrund des Abbruchs der Reihenentwicklung bei Ordnung N (siehe Abbildung 15).

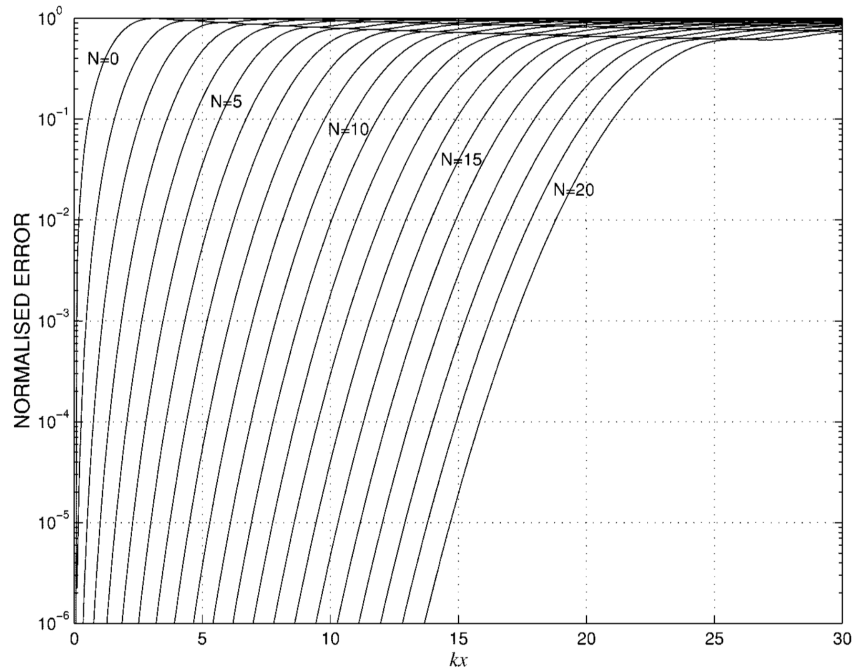


Abbildung 15 – Normalisierter Abbildungsfehler in Abhängigkeit der Ordnung N [WA01].

Der normalisierte Abbildungsfehler aufgrund des Abbruchs der Reihenentwicklung ist hierbei als Funktion von $k \cdot x$ dargestellt. Die Variable x stellt dabei die Entfernung zum Mittelpunkt und $k = 2\pi f/c$ die Wellenzahl dar. Der Abbildungsfehler steigt also mit stei-

gender Frequenz und Abstand zum Mittelpunkt.

In [Pol96, Ger92] wird gezeigt, dass der Abbildungsfehler eines 2D Ambisonicssystems bei Verwendung von $L = 2n + 1$ Lautsprechern minimiert wird. Diese Minimierung ist laut [DRP98] allerdings nur im tieffrequenten Bereich gültig. Für die Optimierung im hochfrequenten Bereich sind demnach $L \geq 2n + 2$ Lautsprecher notwendig. Für periphones Ambisonics unter Benutzung von *t-designs* wird in [ZF12] für optimale Dekodierung ein Mindestpolyedergrad von $t = 2n + 1$ ermittelt.

In [Sol08] wird allerdings für 2D Ambisonicssysteme gezeigt, dass durch Nutzung einer höheren Lautsprecheranzahl als minimal notwendig ($L > 2n + 1$) spektrale Verschlechterungen (*spectral impairments*) bei hohen Frequenzen entstehen. Diese zeigen sich besonders bei Abweichung von einer zentralen Hörposition und virtuellen Quellen zwischen Lautsprechern. Die spektralen Verschlechterungen werden in [Col13] durch Analyse der Impulsantworten und in [BH14] auch für 3D Systeme untersucht. Nach [BH14] entsteht durch Überlagerung von Kammfiltereffekten im hohen Frequenzbereich eine Tiefpasswirkung bei Verwendung von mehr als $L = 2n + 2$ (2D) bzw. $L = (2n + 1)^2$ (3D) Lautsprechern. Diese Tiefpasswirkung ist für 2D Ambisonics dritter Ordnung bei verschiedenen Lautsprecheranzahlen in Abbildung 16 dargestellt.

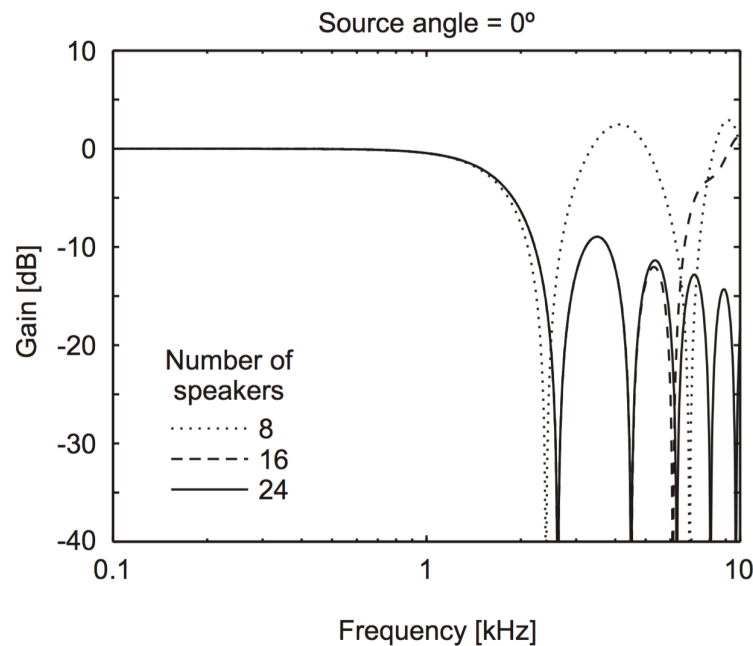


Abbildung 16 – *Spectral impairments* aufgrund einer Erhöhung der Lautsprecheranzahl bei 2D Ambisonics dritter Ordnung und 0° Quellrichtung [Col13].

5.3 Analyse und Auswertung für 2D und 3D Dekoder

5.3.1 Energie- und Schnellevektoren

2D Dekoder. Die Untersuchungen werden für die Ordnungen $n = 1, 3, 5, 8, 20$ mit $2n + 1$, $2n + 2$ und $4n$ Lautsprechern jeweils mit und ohne $\max \vec{r}_E$ -Optimierung durchgeführt. In Abbildung 17 sind die Beträge des Schnelle- und Energievektors für die untersuchten Ordnungen bei $2n + 2$ virtuellen Lautsprechern ohne $\max \vec{r}_E$ -Optimierung zu sehen. In Abbildung 18 sind ebenfalls die Beträge der Vektoren zu sehen, diesmal aber mit $\max \vec{r}_E$ -Optimierung.

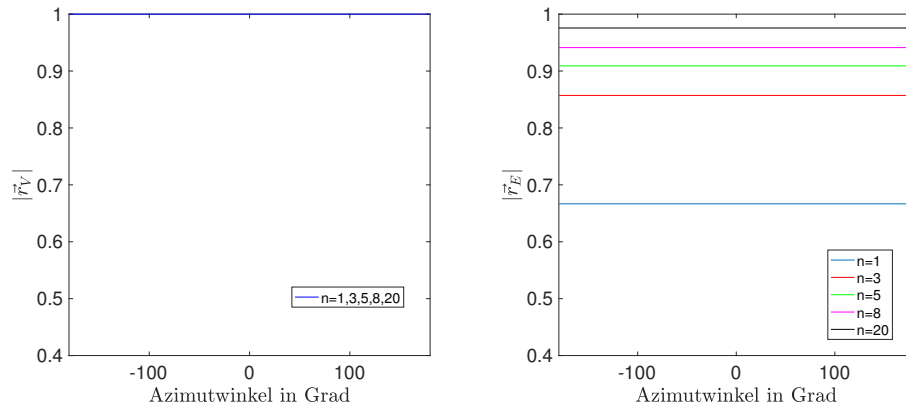


Abbildung 17 – 2D Dekoder: Beträge des Schnellevektors $\vec{r}_V$ und Energievektors $\vec{r}_E$ für die Ordnungen $n = 1, 3, 5, 8, 20$ bei $2n + 2$ virtuellen Lautsprechern ohne $\max \vec{r}_E$ -Optimierung.

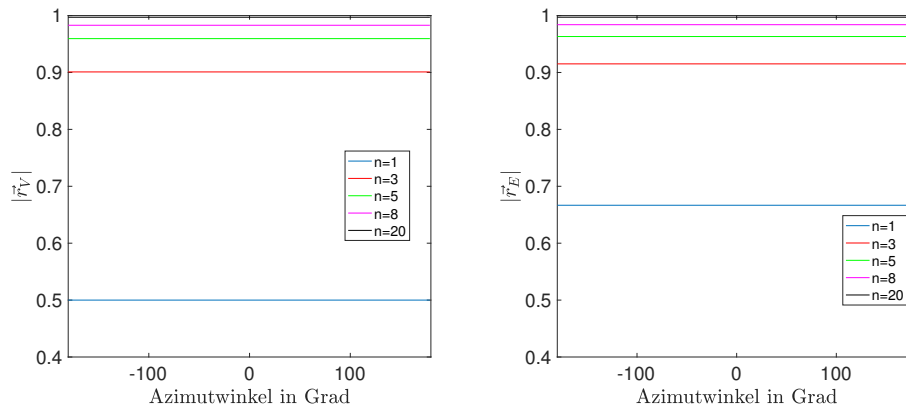


Abbildung 18 – 2D Dekoder: Beträge des Schnellevektors $\vec{r}_V$ und Energievektors $\vec{r}_E$ für die Ordnungen $n = 1, 3, 5, 8, 20$ bei $2n + 2$ virtuellen Lautsprechern mit $\max \vec{r}_E$ -Optimierung.

Dabei ist ersichtlich, dass der Betrag des Energievektors mit höherer Ambisonicsordnung

steigt. Wird keine $\max \vec{r}_E$ -Optimierung durchgeführt, so bleibt der Betrag des Schnellektors ordnungsunabhängig bei Eins. Durch $\max \vec{r}_E$ -Optimierung kann der Betrag des Energievektors vergrößert werden. Dabei verringert sich der Betrag des Schnellektors. Es ergibt sich also bei Maximierung des Energievektors eine verbesserte Lokalisation für hochfrequente Signale, auf Kosten der Lokalisation für tieffrequente Signale unter etwa 700 Hz. Bei Verwendung von $4n$ Lautsprechern ergeben sich etwa die gleichen Verhältnisse wie bei $2n + 2$ Lautsprechern (hier nicht dargestellt). Bei beiden Varianten bleiben die Beträge der Vektoren über den Winkelbereich konstant.

Bei Verwendung von nur $2n + 1$ virtuellen Lautsprechern ergibt sich eine Variation über den Winkelbereich. In Abbildung 19 sind die Beträge von $\vec{r}_V$, $\vec{r}_E$ und auch die Abweichung der Richtung beider zur Quellrichtung für $2n + 1$ Lautsprecher ohne $\max \vec{r}_E$ -Optimierung bei Ambisonics fünfter Ordnung zu sehen. Diese Richtungsabweichung wird im Folgenden auch *angulärer Fehler* genannt. In Abbildung 20 sind die gleichen Verhältnisse, allerdings mit $\max \vec{r}_E$ -Optimierung dargestellt. Zum Vergleich zeigt Abbildung 21 die gleichen Kriterien mit $\max \vec{r}_E$ -Optimierung und $2n + 2$ Lautsprechern.

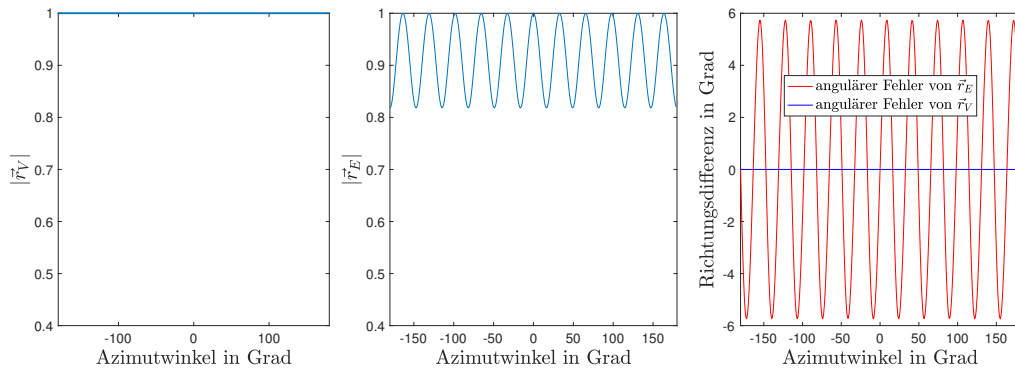


Abbildung 19 – 2D Dekoder: Beträge und anguläre Fehler von $\vec{r}_V$, $\vec{r}_E$ für $2n + 1$ Lautsprecher ohne $\max \vec{r}_E$ -Optimierung bei Ambisonics fünfter Ordnung.

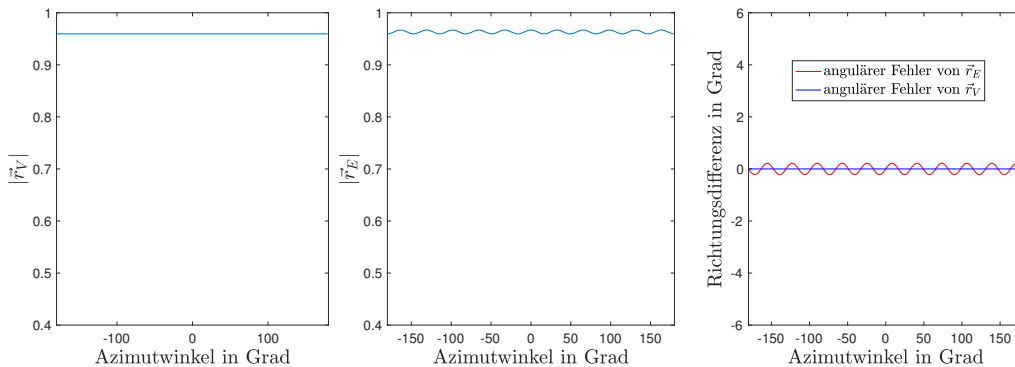


Abbildung 20 – 2D Dekoder: Beträge und anguläre Fehler von $\vec{r}_V$, $\vec{r}_E$ für $2n+1$ Lautsprecher mit $\max \vec{r}_E$ -Optimierung bei Ambisonics fünfter Ordnung.

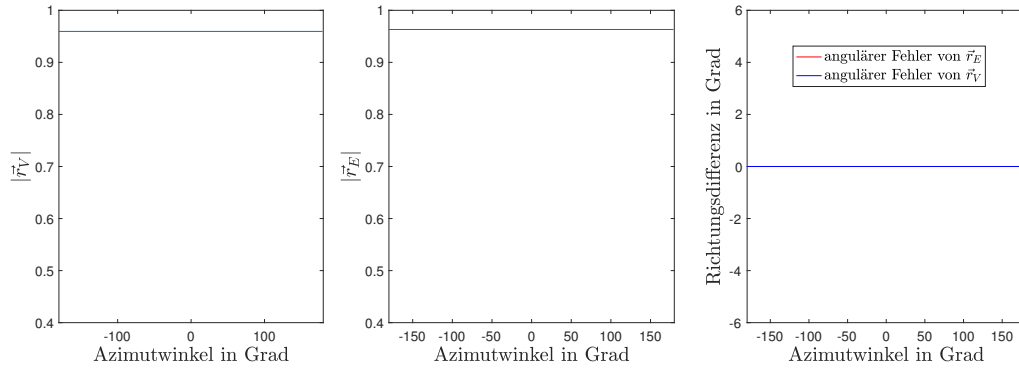


Abbildung 21 – 2D Dekoder: Beträge und anguläre Fehler von $\vec{r}_V$, $\vec{r}_E$ für $2n+2$ Lautsprecher mit $\max \vec{r}_E$ -Optimierung bei Ambisonics fünfter Ordnung.

Daraus wird deutlich, dass bei Verwendung von weniger als $2n+2$ Lautsprechern Schwankungen in der Länge und Richtung des Energievektors entstehen. Dadurch kann sich eine schwankende empfundene Quellbreite und Änderung der wahrgenommenen Quellposition ergeben. Auch Klangfarbenfluktuationen sind möglich [Fra13a]. Durch $\max \vec{r}_E$ -Optimierung können diese Effekte verringert, aber nicht verhindert werden. Allgemein nehmen die Fluktuationen mit steigender Ordnung und $\max \vec{r}_E$ -Optimierung ab.

3D Dekoder. Eine Untersuchung der gleichen Parameter wie im 2D-Fall zeigt, dass die für 2D festgestellten Verhältnisse immernoch gelten: Der Betrag des Energievektors steigt mit der Ambisonicsordnung. Ohne $\max \vec{r}_E$ ist der Betrag des Schnellektors konstant Eins, mit $\max \vec{r}_E$ sinkt er zunächst, steigt allerdings mit höherer Ordnung wieder an. Der anguläre Fehler des Energievektors sinkt mit steigender Lautsprecheranzahl und bleibt ab $t = 2n + 1$ etwa konstant. So ergibt sich bei Ambisonics fünfter Ordnung mit $\max \vec{r}_E$ bei $t = 2n - 1$ noch ein maximaler Fehler der Quellrichtung zur Richtung von $\vec{r}_E$ von etwa $0,54^\circ$. Bei $t = 2n$ sinkt dieser maximale anguläre Fehler auf etwa $0,06^\circ$ und ab $t \geq 2n + 1$ bleibt er konstant bei etwa $1,5 \cdot 10^{-6}^\circ$.

Auch die Fluktuation des Betrags des Energievektors geht ab $t = 2n + 1$ deutlich zurück. Bei Ambisonics fünfter Ordnung und $t = 2n - 1$ liegt der Unterschied von maximaler zu minimaler Länge des Energievektors noch bei etwa 0,011 bzw. $-19,6\text{dB}$. Bei $t = 2n$ liegt der Unterschied bei ca. 0,002 ($-27,0\text{dB}$) und bei $t = 2n + 1$ fällt der Fehler deutlich auf etwa $3,3 \cdot 10^{-14}$ ($-134,8\text{dB}$). Erhöht man die Ordnung bei gleichbleibender Lautsprecherzahl wird ersichtlich, dass nicht die maximale Ordnung, sondern ein t -design mit $t \geq 2n + 1$ die besten Ergebnisse erzielt: Bei Ambisonics siebter Ordnung und Verwendung eines 11 -designs (erfüllt $t = 2n + 1$ für Ambisonics fünfter Ordnung) ergibt sich ein deutlich größerer maximaler angulärer Fehler ($2,25^\circ$) und auch eine deutlich größere Fluktuation in der Länge des Energievektors von maximal 0,032 ($-14,9\text{dB}$), als bei Verwendung von fünfter Ordnung mit gleicher Lautsprecherzahl.

Diese Verhältnisse werden in den folgenden Abbildungen beispielhaft aufgezeigt. In Abbildung 22 sind die Beträge und angulären Fehler des Energie- und Schnellektors bei

fünfter Ordnung und $t = 2n = 10$ zu sehen. Dabei ist der geringe angularer Fehler des Energievektors von maximal etwa $0,06^\circ$ zu erkennen. Die roten Punkte in den Abbildungen repräsentieren die Positionen der virtuellen Lautsprecher. Abbildung 23 zeigt ebenfalls den Dekoder bei fünfter Ordnung, allerdings für $t = 2n + 1 = 11$. Bei einem t -design mit Polyedergrad 11 kommen 70 virtuelle Lautsprecher zum Einsatz. Bei Verwendung der gleichen Anzahl an Lautsprechern und Erhöhung auf siebte Ordnung Ambisonics ergibt sich der in Abbildung 24 erkennbare, deutlich größere angularer Fehler des Energievektors (maximal etwa $2,25^\circ$). Die Beträge von $\vec{r}_V$ und $\vec{r}_E$ sind im Vergleich zur fünften Ordnung angestiegen. Allerdings sind im Betrag des Energievektors bei siebter Ordnung in Abbildung 24 bereits leichte Fluktuationen erkennbar.

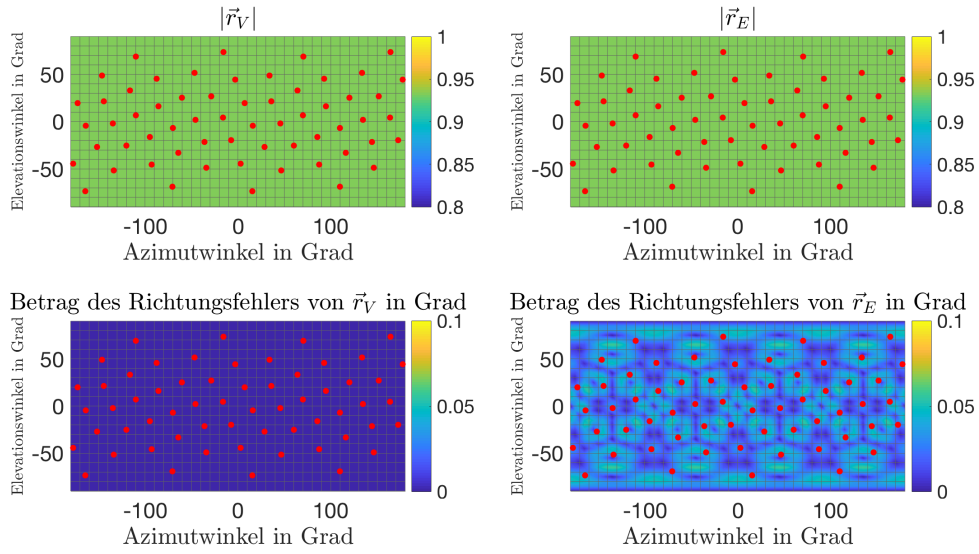


Abbildung 22 – 3D Dekoder: Beträge und angularer Fehler von $\vec{r}_V$, $\vec{r}_E$ für $t = 2n = 10$ mit $\max \vec{r}_E$ -Optimierung bei Ambisonics 5. Ordnung. Die roten Punkte kennzeichnen die Positionen der virtuellen Lautsprecher.

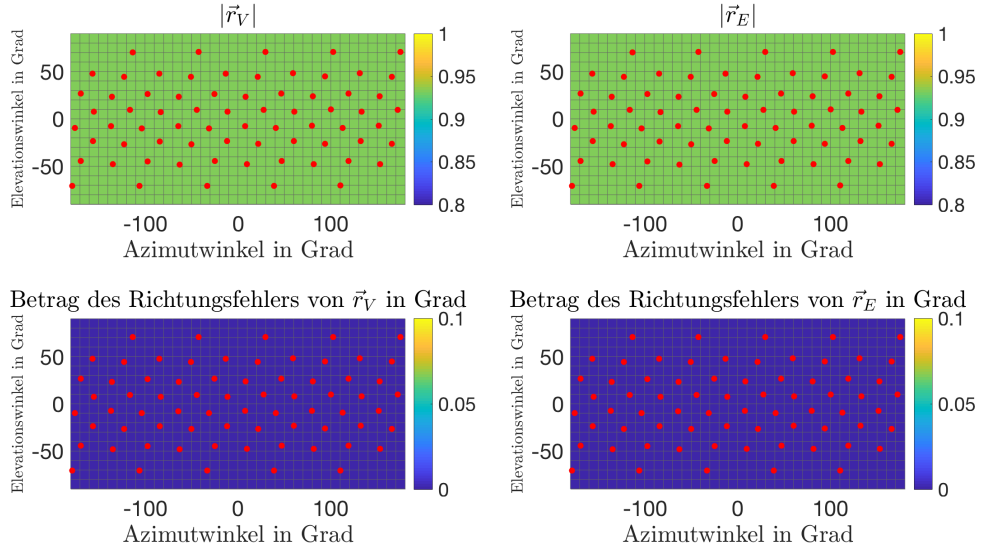


Abbildung 23 – 3D Dekoder: Beträge und anguläre Fehler von $\vec{r}_V, \vec{r}_E$ für $t = 2n+1 = 11$ mit $\max \vec{r}_E$ -Optimierung bei Ambisonics 5. Ordnung. Die roten Punkte kennzeichnen die Positionen der virtuellen Lautsprecher.

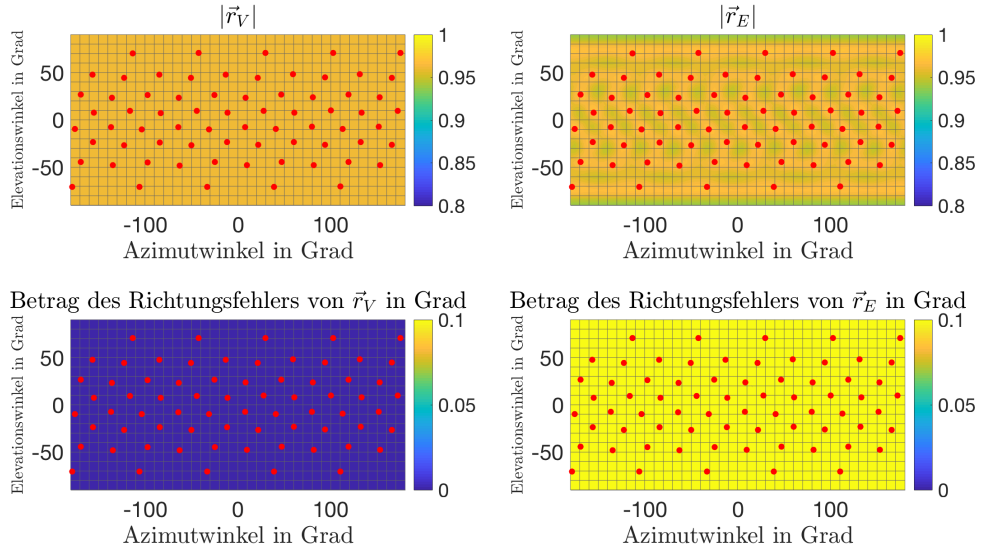


Abbildung 24 – 3D Dekoder: Beträge und anguläre Fehler von $\vec{r}_V, \vec{r}_E$ für $t = 11$ mit $\max \vec{r}_E$ -Optimierung bei Ambisonics 7. Ordnung. Die roten Punkte kennzeichnen die Positionen der virtuellen Lautsprecher.

5.3.2 Lokalisation in der Medianebene mittels psychoakustischem Hörmodell

Das in [BML13] vorgestellte Modell zur Analyse von Lokalisationswahrscheinlichkeiten in Sagittalebene wird zur Evaluierung der 3D Dekoder in der Medianebene verwendet. Dabei werden die virtuellen Lautsprecher signale mit der DTF des Hörers *NH58* gefaltet und anschließend die Lokalisationsleistung des Hörers *NH58* ausgewertet. Lokalisationsfehler aufgrund einer nicht individuellen HRTF werden im Ergebnis also nicht dargestellt. Das Modell vergleicht intern die Impulsantwort am Eingang mit den DTFs eines Referenzhörers (hier *NH58*). Werden die reinen DTFs des Referenzhörers in das Modell geschickt, ergibt sich eine Abbildung der maximalen Lokalisationsgenauigkeit, die im Modell für diesen Hörer erreichbar ist. Diese maximale Lokalisationsgenauigkeit zeigt die linke Seite in Abbildung 25. Auf der Abszisse ist der Elevationswinkel der Quelle zu sehen. Die Ordinate beschreibt den modellierten wahrgenommenen Elevationswinkel. Die Lokalisationswahrscheinlichkeit ist mit der Farbskala dargestellt. Hellere Bereiche sind Bereiche höherer Lokalisationswahrscheinlichkeit. Wie in Abbildung 25 zu sehen, ergibt sich bei maximaler Lokalisationsgenauigkeit also eine helle Diagonale.

Abbildungen 25 rechts und 26 zeigen die Lokalisationswahrscheinlichkeiten bei Benutzung des Ambisonicssystems dritter, fünfter und achter Ordnung und einem *t-design* mit $t = 2n + 1$. Dabei ist erkennbar, dass die Wahrscheinlichkeit der Lokalisation einer Quelle aus der richtigen Richtung bei Erhöhung der Ordnung von fünf auf acht steigt. Bei Vergleich der Lokalisationswahrscheinlichkeiten von dritter und fünfter Ordnung kann aufgrund der Abbildung keine genaue qualitative Aussage gemacht werden. Bei Betrachtung des mittleren Quadrantenfehlers (QE) und des mittleren polaren RMS Fehlers (PE) zeigt sich allerdings vor allem für den Quadrantenfehler eine leichte Verbesserung bei fünfter Ordnung im Gegensatz zu dritter Ordnung (vgl. Tabelle 1). Grundsätzlich wird in der Tabelle die Verbesserung der Lokalisationspräzision durch geringere PE's und die Verringerung der Wahrscheinlichkeit von *Vorne-Hinten-* bzw. *Oben-Unten-Verwechslung* durch geringe QE's mit höherer Ordnung deutlich. Bei Ambisonics zehnter Ordnung ist allerdings entgegen dem Trend eine leichte Erhöhung des mittleren Quadrantenfehlers zu sehen.

Ordnung	1	3	5	8	10
mittlerer Quadrantenfehler, QE (%)	32,7	27,3	23,8	18,7	19,3
mittlerer lokaler RMS Fehler, PE (°)	44,7	39,9	39,2	36,0	35,4

Tabelle 1 – Mittlerer Quadrantenfehler und mittlerer lokaler RMS Fehler bei Elevationswinkeln von -40 bis 220° bei 3D Ambisonics mit verschiedenen Ordnungen und einem $2n + 1$ -design (mit $\max \vec{r}_E$ -Optimierung).

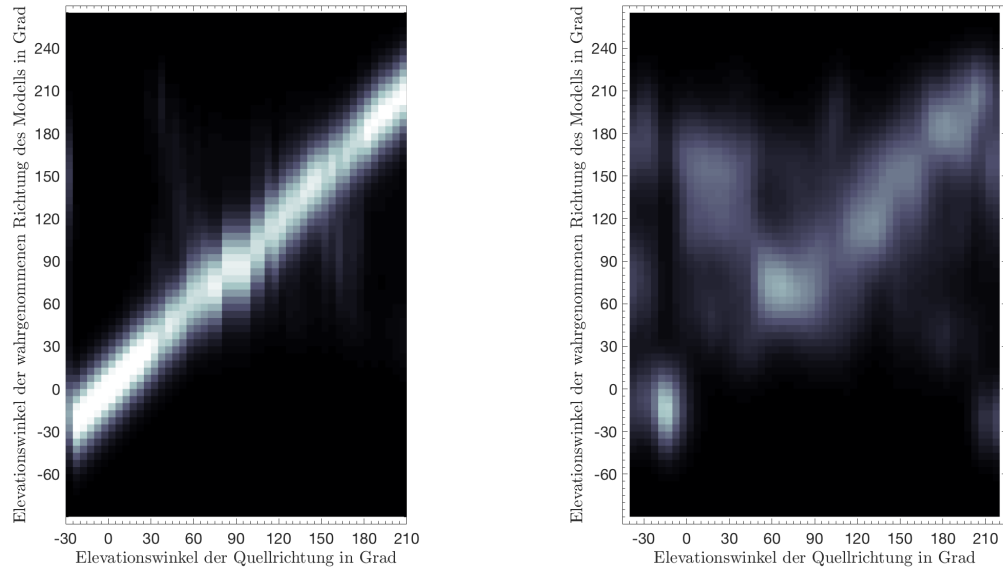


Abbildung 25 – V.l.n.r.: Maximale Lokalisationsgenauigkeit des Referenzhörers im Modell und Lokalisationswahrscheinlichkeit des Referenzhörers im Modell für Ambisonics 3. Ordnung bei $t = 2n + 1 = 7$, mit $\max \vec{r}_E$ -Optimierung. Hellere Bereiche sind Bereiche höherer Lokalisationswahrscheinlichkeit.

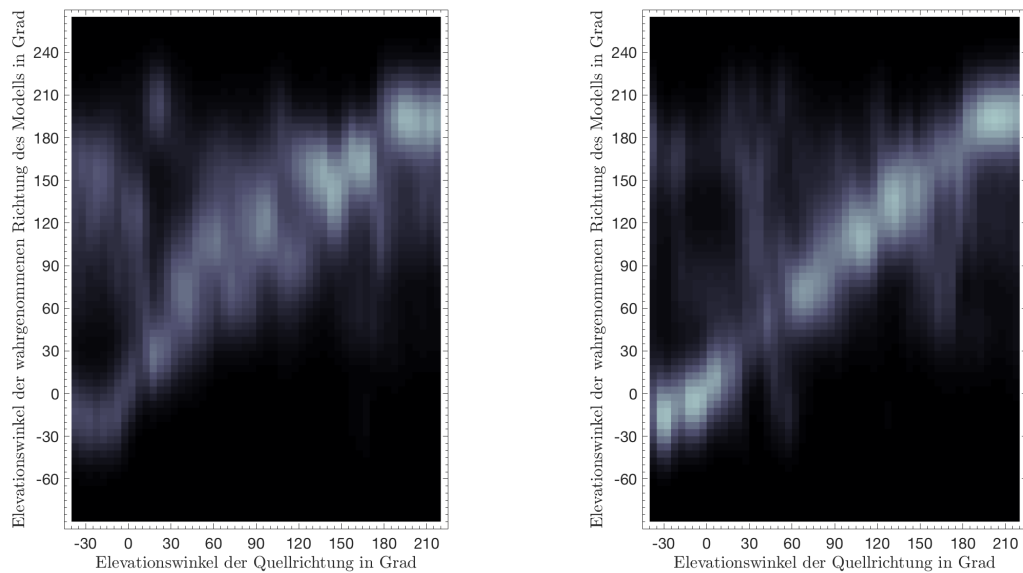


Abbildung 26 – V.l.n.r.: Lokalisationswahrscheinlichkeit des Referenzhörers im Modell für Ambisonics 5. Ordnung bei $t = 2n + 1 = 11$ und Ambisonics 8. Ordnung bei $t = 2n + 1 = 17$, mit $\max \vec{r}_E$ -Optimierung. Hellere Bereiche sind Bereiche höherer Lokalisationswahrscheinlichkeit.

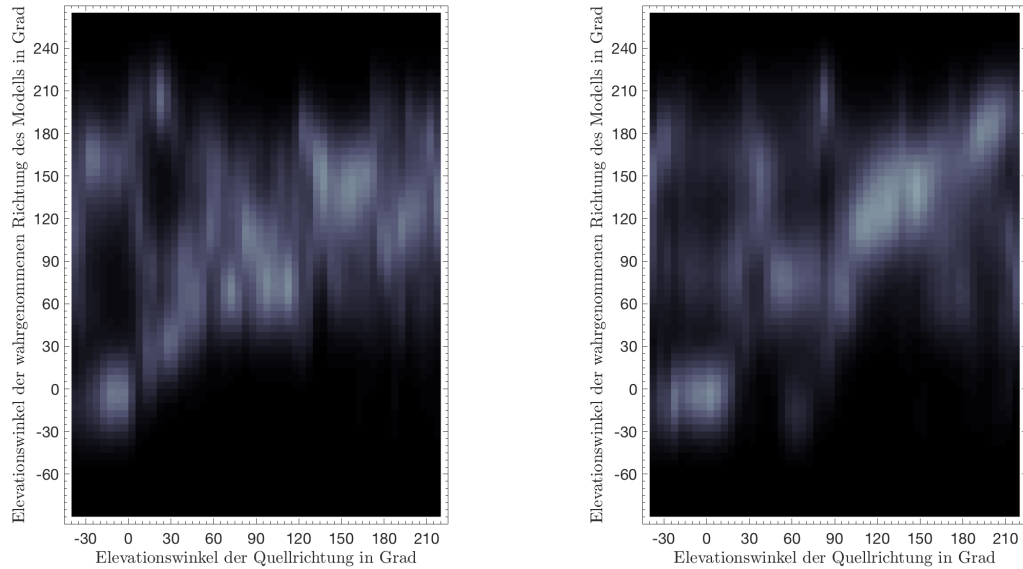


Abbildung 27 – V.l.n.r.: Lokalisationswahrscheinlichkeit des Referenzhörers im Modell für Ambisonics 5. Ordnung bei $t = 2n + 1 = 11$ ohne $\max \vec{r}_E$ -Optimierung und bei $t = 17$ mit $\max \vec{r}_E$ -Optimierung. Hellere Bereiche sind Bereiche höherer Lokalisationswahrscheinlichkeit.

Abbildung 27 links zeigt im Vergleich zu Abbildung 26 links die Verschlechterung der Lokalisation unter Verwendung des gleichen Dekoders ohne $\max \vec{r}_E$ -Optimierung.

Tabelle 2 zeigt QE und PE für Ambisonics fünfter Ordnung mit $\max \vec{r}_E$ -Optimierung und der Verwendung von t -designs verschiedenen Grades. Daraus ist zu erkennen, dass die Fehler PE und QE mit höherem Grad des t -designs bzw. höherer Lautsprecheranzahl eher steigen.

Grad des t-designs	8	10	11	14	17	21
mittlerer Quadrantenfehler, QE (%)	23,2	28,5	23,8	27,2	29,3	30,1
mittlerer lokaler RMS Fehler, PE (°)	39,0	39,4	39,2	40,4	40,9	40,1

Tabelle 2 – Mittlerer Quadrantenfehler und mittlerer lokaler RMS Fehler bei Elevationswinkeln von -40 bis 220° bei 3D Ambisonics fünfter Ordnung und t -designs verschiedenen Grades (mit $\max \vec{r}_E$ -Optimierung).

Im Vergleich zu Abbildung 26 links ist in Abbildung 27 rechts ebenfalls der Ambisonics Dekoder bei fünfter Ordnung allerdings mit einem t -design von $t = 17$ dargestellt. Darin ist ebenfalls die tendenziell schlechtere Lokalisationsleistung erkennbar. Bei $t = 2n + 1 = 11$ sind PE und QE relativ gering. Für $t > 2n + 1$ steigen die Fehler an. Für $t = 10$ ist ein Anstieg von PE und QE zu erkennen, für $t = 8$ allerdings wieder eine Abnahme.

Die in Tabelle 2 dargestellte Zunahme des QE und PE durch Erhöhung der Lautsprecher-

anzahl kann mit den in Abschnitt 5.2 beschriebenen spektralen Verschlechterungen durch Kammfiltereffekte erklärt werden. Die Kammfiltereffekte betreffen besonders den oberen Frequenzbereich, in dem sich auch die zur Lokalisation in Sagittalebene wichtigen monauralen Merkmale befinden.

5.3.3 Relativer spektraler HRTF-Synthesefehler

In Abbildungen 28 und 29 sind Frequenzgang des jeweiligen binaural Ambisonic Dekoders, Frequenzgang der verwendeten HRTF und der relative spektrale Fehler dargestellt. Abbildung 28 zeigt den 2D-Dekoder bei fünfter Ordnung mit $2n + 2$ Lautsprechern in der Horizontalebene in 5° Schritten von 0 bis 360° . Bei den gleichen Einstellungen liefert der 3D-Dekoder in der Horizontalebene ähnliche Resultate (hier nicht dargestellt). Abbildung 29 zeigt den 3D-Dekoder bei fünfter Ordnung und $t = 2n + 1$ in der Medianebene in 10° Schritten von -40 bis 220° Elevation. Die Winkelauflösung wird auf 10° angepasst, da die benutzte HRTF keine höhere Auflösung in der Medianebene bereitstellt. Nach Analyse der Auswirkung der Parameter auf den spektralen Fehler wurde festgestellt, dass für 2D sowie 3D Dekoder in der Horizontal- sowie Medianebene die gleichen Aussagen gelten. Daher erfolgt die gleichzeitige Betrachtung von 2D und 3D.

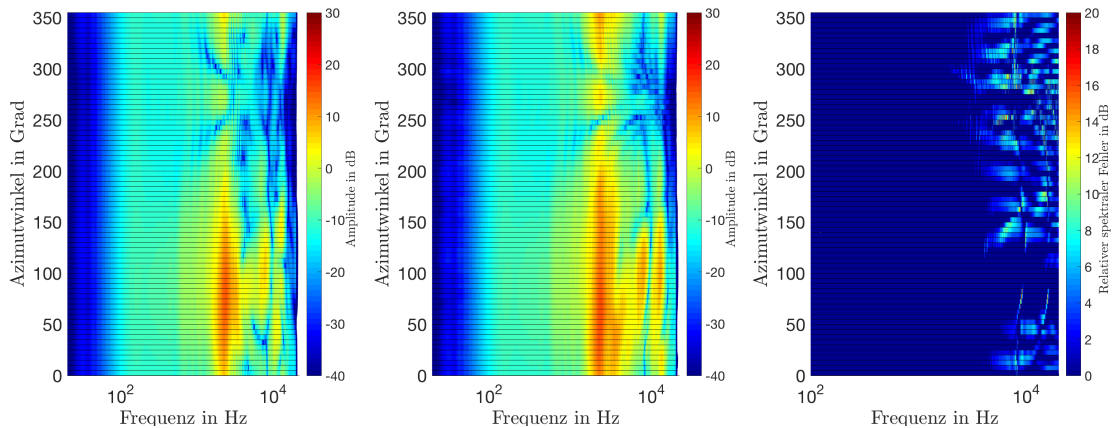


Abbildung 28 – V.l.n.r.: Frequenzgang des 2D binaural Ambisonic Dekoders (Ordnung 5, $2n + 2$ Lautsprecher, $\max \vec{r}_E$), Frequenzgang der verwendeten HRTFs und der relative spektrale Fehler jeweils in 5° Schritten über einen Azimut-Winkelbereich von 0 bis 360° .

Abbildungen 28 rechts und 30 zeigen, dass der relative spektrale Fehler mit einer höheren Ambisonicsordnung sinkt. Dabei verschieben sich die Fehler mehr und mehr zu höheren Frequenzen hin. Bei 2D und Ambisonics 20. Ordnung befinden sich die spektralen Fehler beispielsweise schon überwiegend über 10 kHz (Abbildung 31 rechts). Eine $\max \vec{r}_E$ -Optimierung verringert den spektralen Fehler erheblich (Abbildung 28 rechts im Vergleich zu Abbildung 31 links). Eine höhere Anzahl an Lautsprechern verringert ebenfalls den relativen spektralen Fehler (Abbildungen 29 rechts und 32). Dies kann damit erklärt werden, dass an jedem virtuellen Lautsprecher eine HRTF eingesetzt wird und somit die

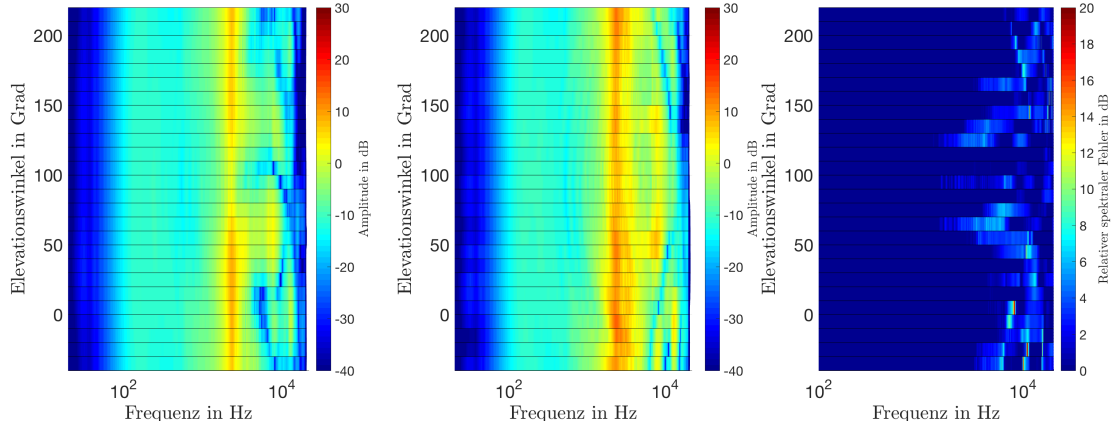


Abbildung 29 – V.l.n.r.: Frequenzgang des 3D binaural Ambisonic Dekoders (Ordnung 5, $t = 2n + 1$, $\max \vec{r}_E$), Frequenzgang der verwendeten HRTFs und der relative spektrale Fehler jeweils in 10° Schritten in der Medianebene über einem Elevations-Winkelbereich von -40 bis 220° .

Punkte, zwischen denen das Ambisonicssystem interpoliert, dichter zusammenliegen. Die Verringerung des Fehlers aufgrund einer höheren Anzahl an Lautsprechern ist bei 3D mit t -designs $t \geq 2n$ nur gering. Allgemein ergeben sich in der Horizontalebene am kontralateralen Ohr (um 270°) und in der Medianebene um 90° die größten Fehler.

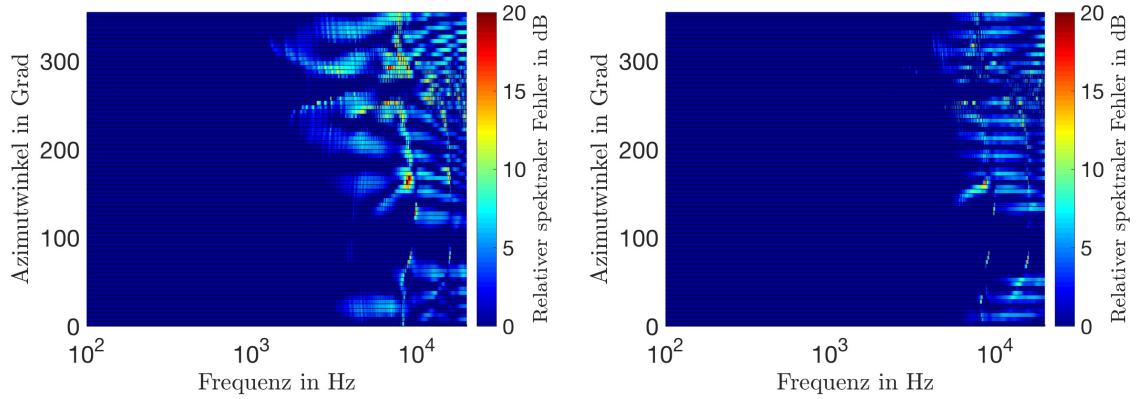


Abbildung 30 – V.l.n.r.: Relativer spektraler Fehler des 2D Dekoders für Ambisonics 3. Ordnung und 8. Ordnung mit jeweils $2n + 2$ Lautsprechern, mit $\max \vec{r}_E$ -Optimierung.

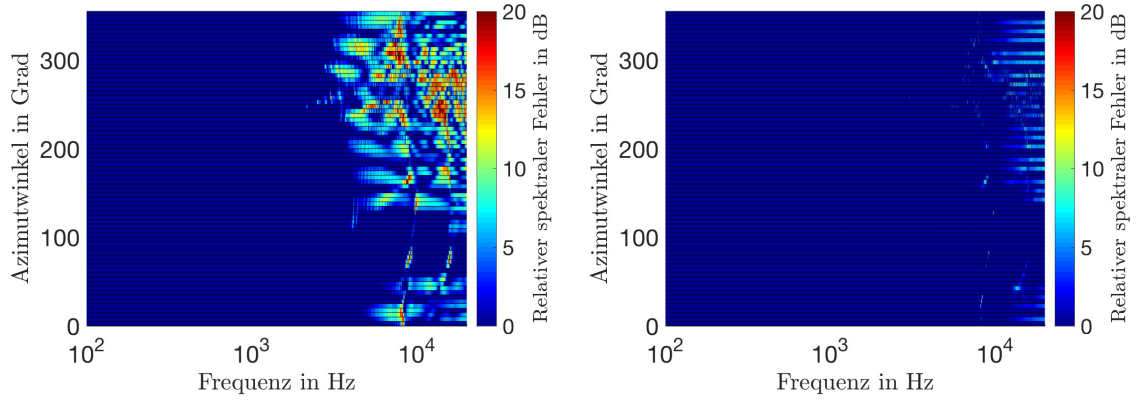


Abbildung 31 – V.l.n.r.: Relativer spektraler Fehler des 2D Dekoders für Ambisonics 5. Ordnung mit $2n + 2$ Lautsprechern ohne $\max \vec{r}_E$ -Optimierung und für Ambisonics 20. Ordnung mit $2n + 2$ Lautsprechern mit $\max \vec{r}_E$ -Optimierung.

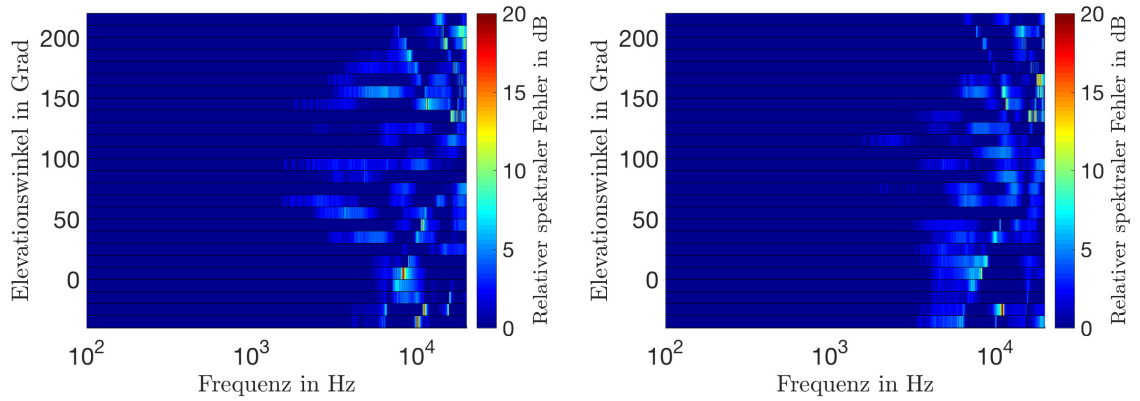


Abbildung 32 – V.l.n.r.: Relativer spektraler Fehler des 3D Dekoders in der Medianebene für Ambisonics 5. Ordnung mit $t = 2n$ und $t = 2n + 2$, mit $\max \vec{r}_E$ -Optimierung.

5.3.4 Klangfarbenänderung durch Fluktuation der Anzahl aktiver Lautsprecher

2D Dekoder. In Abbildung 33 ist der 2D Dekoder mit $\max \vec{r}_E$ -Optimierung bei dritter, fünfter und achter Ordnung mit je $2n + 2$ Lautsprechern dargestellt. Neben den dargestellten Ordnungen wurde für die gleichen Verhältnisse noch 11. und 20. Ordnung untersucht. Bei allen untersuchten Ordnungen sind grundsätzlich drei Lautsprecher aktiv, nur in der Mitte zwischen zwei Lautsprechern geht die Anzahl aktiver Lautsprecher auf zwei zurück. In den hier dargestellten Abbildungen ist dieser Rückgang nur sehen, wenn genau in der Mitte zwischen zwei Lautsprechern die Anzahl aktiver Lautsprecher berechnet wurde. Dies ist aufgrund der auf 5° begrenzten Winkelauflösung nur bei fünfter bzw. achter Ordnung der Fall (Abbildung 33). In [Son03] wird gezeigt, dass sich die Verteilung der Lautsprecher gains im 2D Fall mit Hilfe einer *angular sinc function* (ASINC) berechnen lässt. Die $\max \vec{r}_E$ -Optimierung hat eine Verbreiterung der Hauptkeule und Dämpfung der Nebenkeulen dieser Funktion zur Folge. Wenn sich die virtuelle Quelle zwischen zwei Lautsprechern befindet, deckt die Hauptkeule nur diese beiden Lautsprecher ab, alle anderen Lautsprecher erhalten ein deutlich gedämpftes Signal. Die Fluktuationen und damit die Klangfarbenänderung (nach [Fra13a]) ist damit relativ gering und bleibt mit Erhöhung der Ordnung konstant. Bezogen auf die Gesamtanzahl geht der Anteil aktiver Lautsprecher mit höherer Ordnung zurück. So ergeben sich für dritte Ordnung Ambisonics zwischen 25-37,5% aktive Lautsprecher, für fünfte Ordnung zwischen 16,7-25% und für achte Ordnung 11,1-16,7%.

In Abbildung 34 (links) ist ebenfalls der Dekoder fünfter Ordnung mit $2n + 1$ Lautsprechern zu sehen (vgl. Abbildung 33 Mitte), diesmal allerdings ohne $\max \vec{r}_E$ -Optimierung. Dabei ist ein deutlicher Anstieg in der Fluktuation der Lautsprecheranzahl zu sehen und damit ein Anstieg der Klangfarbenfluktuation zu erwarten [Fra13a].

Abbildungen 33 (Mitte) und 34 (Mitte und rechts) zeigen wiederum den Dekoder fünfter Ordnung mit $\max \vec{r}_E$ -Optimierung, allerdings mit 12, 14 und 18 Lautsprechern. Dabei ist grundsätzlich ein Anstieg der Anzahl aktiver Lautsprecher bei Erhöhung der Lautsprecheranzahl zu sehen. Trotzdem schwankt die Anzahl aktiver Lautsprecher über den Winkelbereich immer nur um maximal einen Lautsprecher. Nach [Fra13a] ist daher davon auszugehen, dass die Klangfarbenfluktuation mit einer höheren Anzahl an Lautsprechern abnimmt. Der prozentuale Anteil aktiver Lautsprecher ist mit Lautsprecheranzahlen größer als $2n + 1$ eher höher: Für fünfte Ordnung Ambisonics ergibt sich bei 12 Lautsprechern ein Bereich aktiver Lautsprecher zwischen 16,7-25%, bei 14 Lautsprechern zwischen 25-33,3% und bei 18 Lautsprechern zwischen 22,2-27,8%.

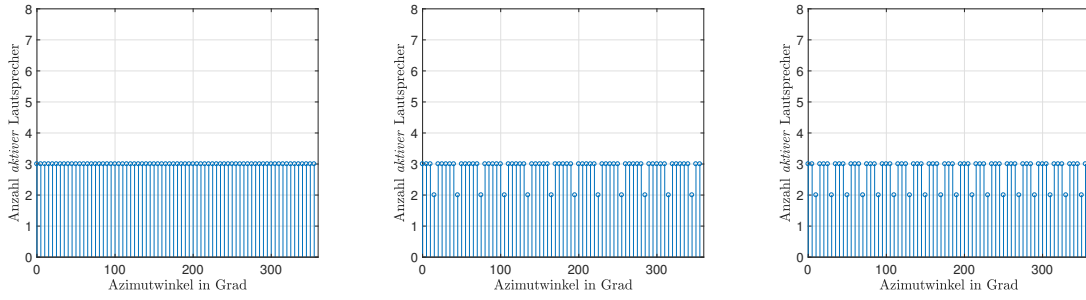


Abbildung 33 – V.l.n.r.: Anzahl *aktiver* Lautsprecher des 2D Dekoders bei Ambisonics 3., 5. und 8. Ordnung, $2n + 2$ Lautsprecher, mit $\max \vec{r}_E$.

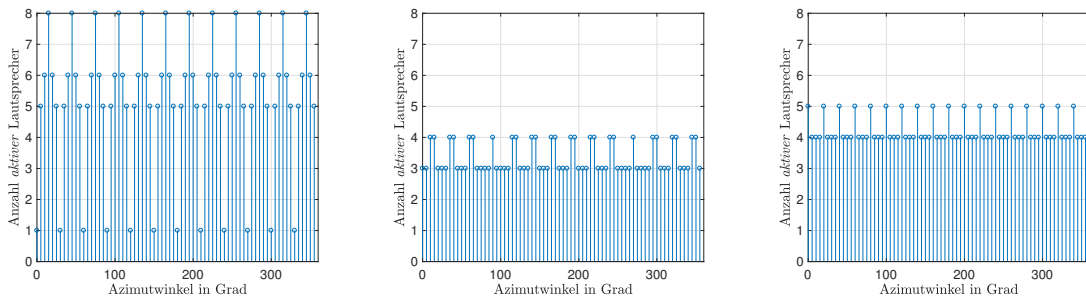


Abbildung 34 – V.l.n.r.: Anzahl *aktiver* Lautsprecher des 2D Dekoders bei Ambisonics 5. Ordnung. Ohne $\max \vec{r}_E$ mit 12 Lautsprechern bzw. mit $\max \vec{r}_E$ und 14 sowie 18 Lautsprechern.

3D Dekoder. In Abbildungen 35-37 ist die Anzahl aktiver Lautsprecher für 3D Dekoder dritter, fünfter und achter Ordnung zu sehen. Im Gegensatz zum 2D Dekoder mit $2n + 2$ Lautsprechern steigt die Anzahl aktiver Lautsprecher bei 3D und einem $t = 2n + 1$ -design tendenziell mit der Ambisonicsordnung. Untersucht werden die Ordnungen drei, fünf, acht und zehn. Bei Ordnung zehn sind je nach Quellposition zwischen acht und dreizehn Lautsprechern aktiv (hier nicht dargestellt). Es zeigt sich, dass die Anzahl aktiver Lautsprecher für die untersuchten Ordnungen über den gesamten Winkelbereich (Azimut und Elevation) maximal um 4-5 Lautsprecher fluktuiert. Die Klangfarbenfluktuation nach [Fra13a] geht mit Erhöhung der Ambisonicsordnung daher eher zurück. Prozentual gesehen ergibt sich bei dritter Ordnung Ambisonics ein Anteil aktiver Lautsprecher zwischen 16,7-33,3%, bei fünfter Ordnung zwischen 10-17,1% und bei achter Ordnung zwischen 4,5-7,8%. Der Anteil aktiver Lautsprecher bezogen auf die Gesamtanzahl sinkt also mit höherer Ambisonicsordnung.

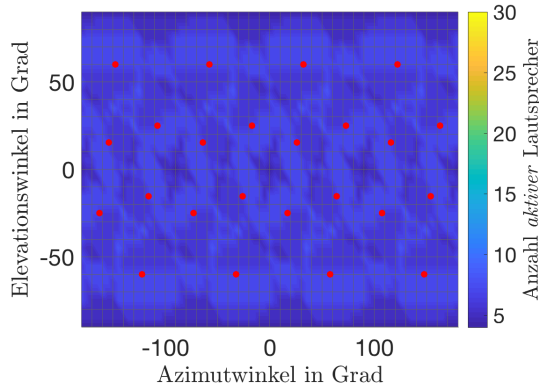


Abbildung 35 – 3D Dekoder: Anzahl *aktiver* Lautsprecher bei Ambisonics 3. Ordnung, $t = 2n + 1$, mit $\max \vec{r}_E$. Bereich *aktiver* Lautsprecher: [4,8]. Die roten Punkte kennzeichnen die Positionen der virtuellen Lautsprecher.

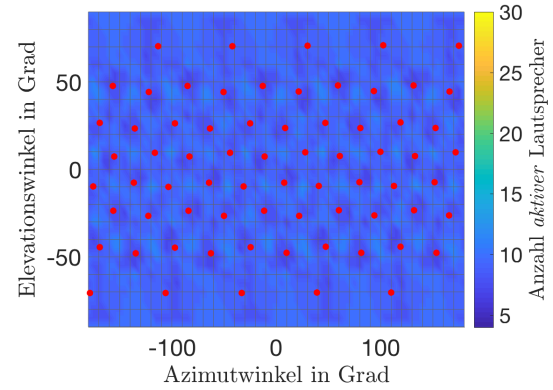


Abbildung 36 – 3D Dekoder: Anzahl *aktiver* Lautsprecher bei Ambisonics 5. Ordnung, $t = 2n + 1$, mit $\max \vec{r}_E$. Bereich *aktiver* Lautsprecher: [7,12]. Die roten Punkte kennzeichnen die Positionen der virtuellen Lautsprecher.

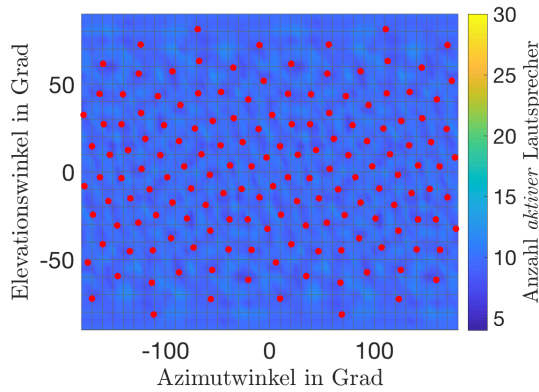


Abbildung 37 – 3D Dekoder: Anzahl *aktiver* Lautsprecher bei Ambisonics 8. Ordnung, $t = 2n + 1$, mit $\max \vec{r}_E$. Bereich *aktiver* Lautsprecher: [7,12]. Die roten Punkte kennzeichnen die Positionen der virtuellen Lautsprecher.

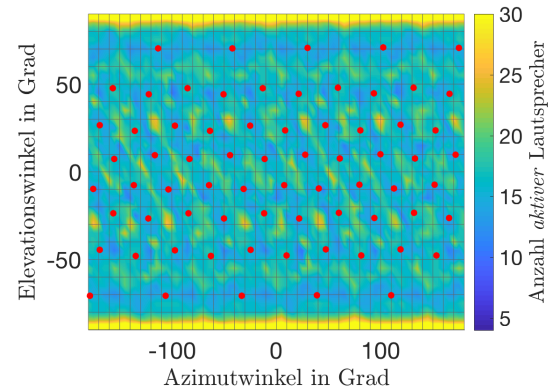


Abbildung 38 – 3D Dekoder: Anzahl *aktiver* Lautsprecher bei Ambisonics 5. Ordnung, $t = 2n + 1$, ohne $\max \vec{r}_E$. Bereich *aktiver* Lautsprecher: [10,40]. Die roten Punkte kennzeichnen die Positionen der virtuellen Lautsprecher.

Abbildung 38 zeigt analog zum 2D-Fall eine deutliche Zunahme der Fluktuation aktiver Lautsprecher und damit auch der Klangfarbe, wenn keine $\max \vec{r}_E$ -Optimierung angewendet wird. Starke Fluktuationen treten besonders bei vertikaler Quellbewegung in den Randbereichen um $\pm 90^\circ$ Elevationswinkel auf.

In den Abbildungen 39 und 40 ist im Vergleich zu Abbildung 36 eine Erhöhung der Lautsprecheranzahl bei gleichbleibender Ambisonicsordnung dargestellt. Wie im 2D-Fall erhöht sich mit der Lautsprecheranzahl auch die Anzahl aktiver Lautsprecher. Die Anzahl aktiver Lautsprecher fluktuiert nun allerdings um sieben (Abbildungen 39 und 40),

statt um sechs Lautsprecher (Abbildung 36). Da die Anzahl aktiver Lautsprecher dabei deutlich höher ist, nimmt die Klangfarbenfluktuation nach [Fra13a] mit steigender Lautsprecheranzahl eher ab. Prozentual gesehen bleibt der Anteil aktiver Lautsprecher relativ konstant: Bei $t = 11$ ergibt sich ein Anteil aktiver Lautsprecher zwischen 10-17,1%, bei $t = 17$ zwischen 10,9-14,7% und bei $t = 21$ zwischen 11,7-14,2%.

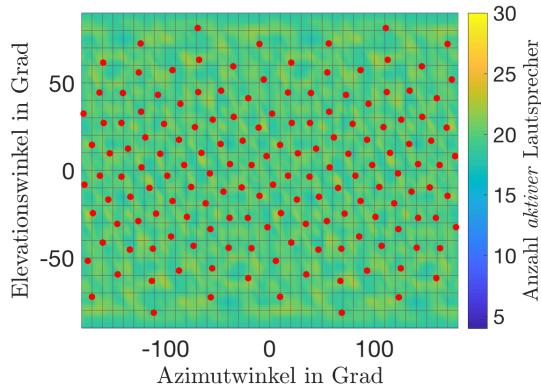


Abbildung 39 – 3D Dekoder: Anzahl *aktiver* Lautsprecher bei Ambisonics 5. Ordnung, $t = 17$, mit $\max \vec{r}_E$. Bereich *aktiver* Lautsprecher: [17,23]. Die roten Punkte kennzeichnen die Positionen der virtuellen Lautsprecher.

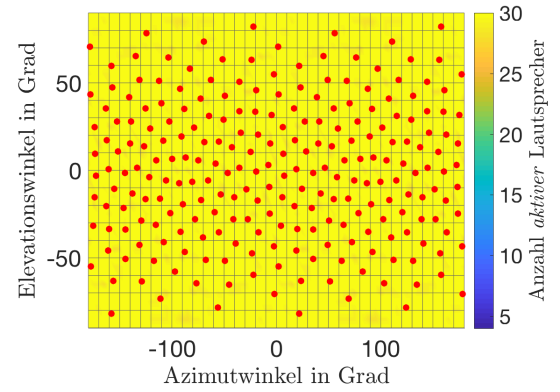


Abbildung 40 – 3D Dekoder: Anzahl *aktiver* Lautsprecher bei Ambisonics 5. Ordnung, $t = 21$, mit $\max \vec{r}_E$. Bereich *aktiver* Lautsprecher: [28,34]. Die roten Punkte kennzeichnen die Positionen der virtuellen Lautsprecher.

6 Implementierung

6.1 Einführung

In diesem Kapitel wird aufbauend auf den theoretischen Grundlagen der vorherigen Kapitel die Implementierung einer Webapplikation beschrieben, die eine akustisch nachvollziehbare virtuelle Bewegung (*walkthrough*) durch eine reale akustisch aufgenommene oder virtuell zusammengestellte Szene ermöglicht. Die Methode und technische Umsetzung zur Reproduktion einer Audioszene wurden am Institut für Elektronische Musik und Akustik (Universität für Musik und darstellende Kunst, Graz) in Kooperation mit Huawei entwickelt und sind patentrechtlich geschützt. Das Patent trägt den Namen: *Interactive Plenaudio Sound Scene Walkthrough Audio System*.

Als Programmiersprache für die Webapplikation wurde JavaScript gewählt. JavaScript ermöglicht zusammen mit HTML und CSS das Erstellen dynamischer Webseiten. JavaScript unterstützt objektorientierte Programmierparadigmen und ist im Allgemeinen eine clientseitige Programmiersprache, d. h. JavaScript-Programme werden durch den Browser des Nutzers interpretiert und nicht von einem Server. Dies hat den Vorteil der Auslagerung der Rechenlast auf den Rechner des Nutzers. Durch die clientseitige Interpretation ist es dem Nutzer prinzipiell möglich den Quelltext des Programms einzusehen.

Zur Audiosynthesierung und -verarbeitung im Browser steht die *Web Audio API*⁶ (*API* = *application programming interface*, zu deutsch *Programmierschnittstelle*) zur Verfügung. Bei der Entwicklung von Applikationen mit der *Web Audio API* (WAA) werden verschiedene *AudioNodes* innerhalb eines *AudioContext* modular verbunden (vergleichbar mit einem Modulsynthesizer). Durch Verbindung mit dem *DestinationNode* werden die Audiosignale an die Audio Hardware des Rechners angekoppelt.

Zur Nutzung von Ambisonics mittels JavaScript steht seit Juni 2016 die quelloffene (*open source*) Programmbibliothek *JSAmbisonics*⁷ [PPQ16] zur Verfügung. Darin sind unter anderem Encoder, Dekoder und Rotationsmatrizen für periphere Ambisonics und die Möglichkeit der Einbindung beliebiger HRTFs des SOFA⁸ (*Spatially Oriented Format for Acoustics*) Formats implementiert. *JSAmbisonics* nutzt intern ebenfalls die Struktur der WAA.

Das SOFA-Format ermöglicht die Repräsentation von räumlichen Audiodaten wie HRTFs und BRIRs. Gleichzeitig zielt die Initiative auf die Sammlung der weltweit frei verfügbaren Datensätze dieser Art in einem einheitlichen Format ab. Die SOFA-Datenbank ist frei zugänglich⁹.

6. <https://www.w3.org/TR/webaudio/> (zuletzt aufgerufen am 19.05.17)

7. <https://github.com/polarch/JSAmbisonics> (zuletzt aufgerufen am 19.05.17)

8. https://www.sofaconventions.org/mediawiki/index.php/Main_Page (zuletzt aufgerufen am 19.05.17)

9. <http://sofacooustics.org/data/> (zuletzt aufgerufen am 19.05.17)

6.2 Struktur

Die Struktur des erstellten Programms ist in vereinfachter Form in Abbildung 41 dargestellt.

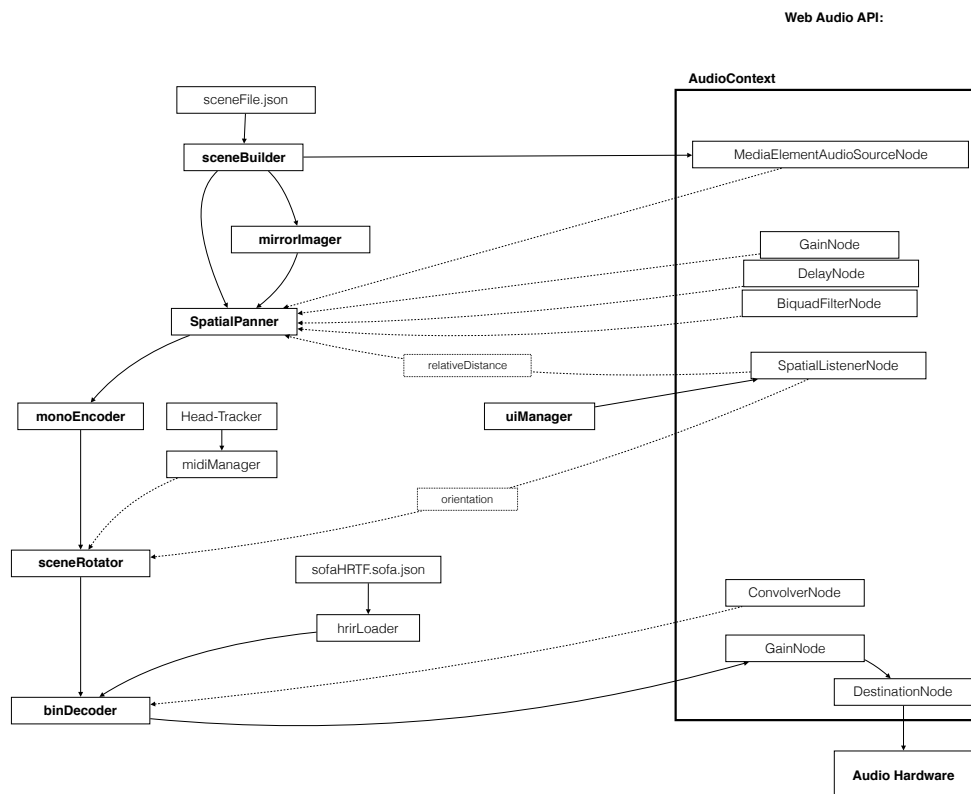


Abbildung 41 – Struktur des erstellten Programms.

6.2.1 Scene File

Um eine virtuelle Audioszene zu erstellen wird zuerst ein *Scene File* eingelesen (vgl. Abbildung 41). Ein Beispiel für ein valides *Scene File* ist im Appendix zu finden. Für das Format des *Scene Files* wurde das JSON-Format¹⁰ (*JavaScript Object Notation*) gewählt. Das JSON-Format ermöglicht die automatische Erstellung von JavaScript-Objekten auf Grundlage der eingelesenen *.json*-Datei. Dazu muss die durch das JSON-Format vorgegebene Syntax genau eingehalten werden. Das *Scene File* enthält Metadaten, die für die Erstellung der Szene notwendig sind. Die Daten werden in drei Typen eingeteilt: *room*, *mono* und *fourChannelArray*. Parameter des Typs *room* beschreiben Eigenschaften des virtuellen Raumes. Parameter des Typs *mono* und *fourChannelArray* beschreiben Eigenschaften einer einkanaligen bzw. vierkanaligen Audioquelle. Letztere wird in der Praxis

10. <http://www.json.org/> (zuletzt aufgerufen am 02.06.17)

durch Aufnahme mit einem möglichst koinzidenten vierteiligen Mikrofonarray erhalten. Die Kapseln des Mikrofonarrays sind dabei in verschiedene Raumrichtungen ausgerichtet und bilden damit Schalleinfall aus vier Richtungen ab. Es wird davon ausgegangen, dass Vierkanalquellen im Gegensatz zu Monoquellen (*Spot-Mikrofone*) keine einzelnen Instrumente abbilden, sondern ein Raumsignal aufzeichnen. Daher werden Signale von Vierkanalquellen teilweise anders verarbeitet (siehe Abschnitt 6.2.4).

Im Folgenden werden die möglichen Parameter im *Scene File* aufgezählt und kurz erklärt. Parameter die mit einem Asterisk (*) gekennzeichnet sind, besitzen die in eckigen Klammern angegebenen Default-Werte und müssen daher nicht zwangsläufig durch den Nutzer spezifiziert werden.

- Raumparameter:
 - *name**: Raumname, *[undefined]*.
 - *width*: Raumbreite.
 - *length*: Raumlänge.
 - *height*: Raumhöhe.
 - *listenerStart**: Startposition des virtuellen Hörers, *x*: *[1]*, *y*: *[1]*, *z*: *[0]*.
- Parameter einer Monoquelle:
 - *name*: Quellname.
 - *position*: Quellposition, *z*: *[0]*.
 - *gain**: Gain der Quelle, Wertebereich *[0, 1]*. *[1]*.
 - *directivity**: Quelldirektivität, Wertebereich *[0, 1]*. Dabei wird zwischen Kugel (*directivity=1*), Niere (*directivity=0,5*) und Acht (*directivity=0*) interpoliert. *[1]*
 - *orientation**: Quellorientierung (Azimutwinkel *azim*, Elevationswinkel *elev*), Wertebereich *[−180, 180]*. *azim*: *[0]*, *elev*: *[0]*.
 - *file*: Mono-Audiodatei, relative Pfadangabe.
 - *NFC**: Nahfeldkompensation (*ein=1* oder *aus=0*), *[0]*.
 - *distGain**: Distanzgain (vgl. Abbildung 43), Parameter: *exp*, *zeroGain*, *refDist*. Der Distanzgain wird mit folgender Formel berechnet, wobei *r* die Distanz von Hörer zu Quelle darstellt:

$$distGain = \begin{cases} zeroGain + (refDist^{-exp} - \frac{zeroGain}{refDist})r & \text{für } r \leq refDist \\ \frac{refDist}{r^{exp}} & \text{für } r > refDist \end{cases} \quad (47)$$

exp: *[1.4]*, *zeroGain*: *[1]*, *refDist*: *[1]*.

- Parameter einer Vierkanalquelle:
 - *name*: Quellname.
 - *center**: Mittelpunkt des Quellarrays, z: *[0]*.
 - *centerDistance**: Abstand der virtuellen Lautsprecher zum Mittelpunkt, *[0.5]*.
 - *gain**: Gain der Quelle, Wertebereich *[0, 1]*. *[1]*.
 - *directivity**: Quelldirektivität, Wertebereich *[0, 1]*. Interpolation zwischen Kugel (*directivity=1*), Niere (*directivity=0,5*) und Acht (*directivity=0*). *[1]*
 - *file*: Audiodatei mit vier Kanälen, relative Pfadangabe.
 - *channelMapping**: Zuordnung der Audiospuren zum korrespondierenden Lautsprecher. *speaker1* bis *speaker4* bezeichnen die virtuellen Lautsprecher beginnend rechts vom Mittelpunkt. Gezählt wird gegen den Uhrzeigersinn. *speaker1: [1], speaker2: [2], speaker3: [3], speaker4: [4]*.
 - *rotation**: Drehung des Lautsprecherarrays (Azimutwinkel *azim*, Elevationswinkel *elev*), Wertebereich *[−180, 180]*. *azim: [0], elev: [0]*.
 - *distGain**: Distanzgain (vgl. Abbildung 43), Parameter: *exp*, *zeroGain*, *refDist* (siehe Gleichung 47). *exp: [1.4], zeroGain: [0], refDist: [1]*.

6.2.2 SceneBuilder

Der *sceneBuilder* liest das *Scene File* ein (vgl. Abbildung 41) und setzt alle Variablen, die den virtuellen Raum beschreiben. Beim Einlesen wird das *Scene File* auf mögliche Fehler überprüft und die Default-Werte gesetzt. Außerdem wird für jede Audiospur eine *SpatialPanner*-Instanz erstellt. Diese *SpatialPanner*-Instanz repräsentiert eine Schallquelle im virtuellen Raum. Daher werden die im *Scene File* spezifizierten Attribute wie Position, Gain und Direktivität den einzelnen Instanzen zugewiesen. Zudem wird pro Monoquelle bzw. Vierkanalquelle ein *MediaElementAudioSourceNode* der WAA erzeugt. Das Abspielen von Audiodateien im Browser erfolgt mittels *HTMLMediaElements*. Das Erzeugen eines *MediaElementAudioSourceNodes* mit einem *HTMLMediaElement* als Argument bewirkt, dass die Audiosignale nicht direkt an die Audiohardware des Computers, sondern über die modular aufbaubare Struktur der WAA gesendet werden.

6.2.3 MirrorImager

Der *mirrorImager* erstellt, wenn aktiviert, je nach Einstellung für jeden *SpatialPanner* entweder vier oder zehn weitere *SpatialPanner*-Instanzen (vgl. Abbildung 41). Diese Instanzen repräsentieren Spiegelquellen. Da der Rechenaufwand stark mit der Anzahl an Spiegelquellen ansteigt, wird nur eine geringe Anzahl an Spiegelquellen erstellt. In der ersten von zwei angebotenen Einstellmöglichkeiten werden vier Spiegelquellen erster

Ordnung erzeugt. Dies sind die Spiegelquellen, die durch Spiegelung an den Raumwänden entstehen. Bei Nutzung der zweiten Einstellmöglichkeit werden zusätzlich Spiegelquellen erster Ordnung über und unter dem Raum (Spiegelung an Boden und Decke) und vier Spiegelquellen zweiter Ordnung in der Ebene erzeugt. Damit werden bei der zweiten Einstellmöglichkeit die Spiegelquellen in allen acht in der Ebene angrenzenden Spiegelräumen (siehe Abbildung 11) sowie über und unter dem Raum erzeugt. Dies ergibt eine Gesamtanzahl von zehn Spiegelquellen pro Quelle bei der zweiten Einstellmöglichkeit.

6.2.4 SpatialPanner

Die *SpatialPanner*-Instanzen steuern alle Elemente, die für die Echtzeitvirtualisierung und -spatialisierung notwendig sind. Je nach Art der Audioquelle (Monoquelle, Vierkanalquelle, Spiegelquelle) ergeben sich dabei unterschiedliche Signalwege (vgl. Abbildung 42).

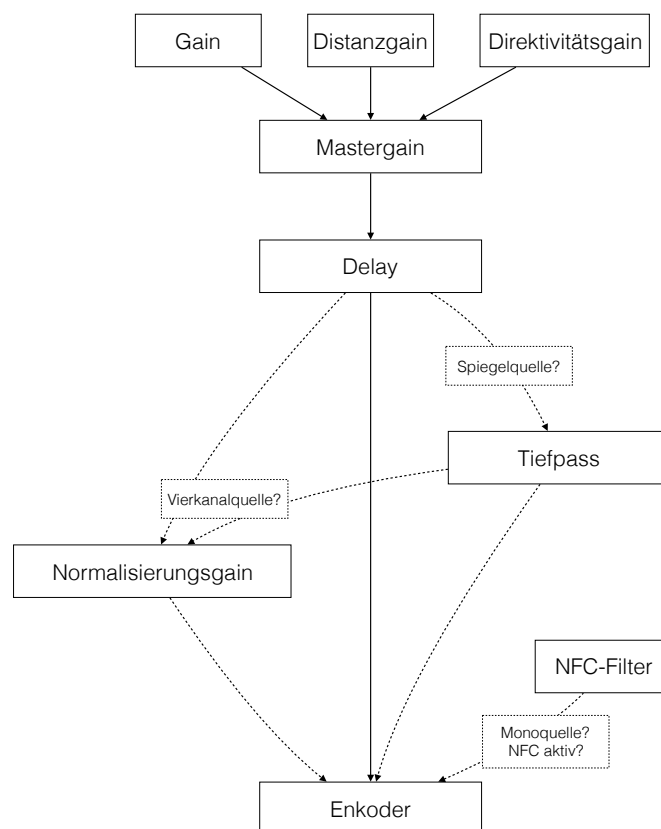


Abbildung 42 – Interne Struktur einer *SpatialPanner*-Instanz.

Bei allen Quelltypen durchläuft das Signal zuerst einen *GainNode*, dessen Gesamtgain sich aus Gain (im *Scene File* definiert), Distanzgain und Direktivitätsgain zusammensetzt. Der Distanzgain berechnet sich nach Gleichung 47. Bei Nutzung der Default-Einstellung nimmt er mit größerer Quell-Hörer-Entfernung mit $1/r^{1,4}$ ab (vgl. Abbildung 43). Dies entspricht zwar einer unrealistischen überproportionalen Abnahme im Vergleich zu Freifeldbedingungen, hat sich aber für die virtuelle Szene unter subjektiver Einschätzung als geeignet herausgestellt. Für Monoquellen bleibt der Distanzgain bei Distanzen kleiner Eins konstant auf Eins, für Vierkanalquellen fällt er für Distanzen zwischen Eins und Null linear auf Null ab (vgl. Abbildung 43). Diese unterschiedliche Behandlung hat mit der Annahme zu tun, dass Vierkanalquellen keine einzelne reale Schallquelle repräsentieren. Die Abnahme der Lautstärke für kleine Distanzen verhindert, dass der Hörer an dieser Position eine Quelle lokalisiert. Ziel ist es, mit den Vierkanalquellen ein den Hörer einhüllendes Schallfeld mit präziser Lokalisation zu schaffen.

Der Direktivitätsgain einer Quelle wird mit dem Parameter *directivity* und der folgenden Gleichung bestimmt, wobei d der Wert der *directivity* und φ der Winkel zwischen Hörer und Quelle ist. φ wird dabei für den dreidimensionalen Raum berechnet, d. h. es ergibt sich eine dreidimensionale Richtcharakteristik der virtuellen Lautsprecher.

$$dirGain = d + (1 - d)\cos(\varphi) \quad (48)$$

Somit ergibt sich für $d = 1$ eine Kugel, für $d = 0,5$ eine Niere und für $d = 0$ eine Acht als Abstrahlcharakteristik. Für andere Werte wird zwischen diesen Charakteristiken interpoliert. Bei Distanzen kleiner Eins wird zwischen der eingestellten Quelldirektivität und einer kugelförmigen Abstrahlung interpoliert. Damit werden Unstetigkeiten beim Durchschreiten einer Quelle vermieden.

Als nächstes durchläuft das Signal einen *DelayNode*. Dieser verzögert das Audiosignal der Quelle abhängig von der Distanz zum Hörer. Die Verzögerung Δt in Sekunden berechnet sich dabei mit:

$$\Delta t = \frac{r}{343 \frac{m}{s}} \quad (49)$$

Der Wert des Nenners entspricht der Schallgeschwindigkeit in Luft bei $20^\circ C$. Da sich der Wert der Verzögerung mit der relativen Quell-Hörer-Entfernung ändert, wird bei schneller Bewegung des Hörers der Doppler-Effekt hörbar: Durch die dynamische Änderung der Verzögerung entsteht eine Anhebung bzw. Abnahme der Tonhöhe (*pitch*).

Handelt es sich bei der betrachteten Quelle um eine Spiegelquelle, so erreicht das Signal als nächstes einen *BiquadFilterNode*, der als Tiefpass mit Grenzfrequenz von $f_{0,tp} = 4kHz$ bzw. $f_{0,tp} = 6kHz$ für Spiegelquellen erster bzw. zweiter Ordnung ausgeführt ist. Diese Tiefpassfilterung repräsentiert die Absorption von hohen Frequenzen bei Auftreffen des Schalls auf einer Wand.

Handelt es sich bei der Quelle um eine Vierkanalquelle, erfolgt vor der Enkodierung noch eine Normalisierung. Da die Vierkanalquellen eine Abtastung des Raumes repräsentieren, soll die gesamte von den Vierkanalquellen abgestrahlte Schallintensität im Raum konstant sein. D. h. bei Schwankung der Dichte der Vierkanalquellen in einem Raumbereich soll sich die Gesamtintensität dieser Quellen nicht ändern. Dazu wird ein weiterer *GainNode*

verwendet, dessen Wert sich bei K virtuellen Lautsprechern folgendermaßen berechnet:

$$gain_{norm} = \frac{1}{\sqrt{\sum_{k=1}^K g_k^2}} \quad (50)$$

Zuletzt erreicht das Signal den Encoder. Zur Laufzeit des Programms werden die zum Hörer relativen Azimut- und Elevationswinkel der Quellen in den *SpatialPanner*-Instanzen berechnet und an den Encoder übergeben. Kommt der virtuelle Hörer einer Quelle näher als eine Minimaldistanz, so werden alle Kanäle außer des ersten Ambisonicskanals (*W-Kanal*) mit sinkendem Hörer-Quell-Abstand gegen Null interpoliert. Dieser Minimalabstand ist im Programm auf 0,3 festgelegt. Durch die Interpolation werden Sprünge in der spatialisierten Richtung einer Quelle vermieden. Steht der Hörer direkt auf einer Quelle, so ist nur noch der *W-Kanal* aktiv und das Quellsignal erklingt ungerichtet.

Falls es sich bei der betrachteten Quelle um eine Monoquelle handelt und Nahfeldkompensation (NFC) aktiviert ist, werden die enkodierten Signale als nächstes durch *Biquad-FilterNodes* geschickt. Darin sind *Biquad Low Shelf Filter* implementiert. Der Gain der Filter berechnet sich in Übereinstimmung mit [DM04] (Abbildung 8) mit:

$$gain_{LowShelf} = N \cdot 20 \log \left(\frac{R}{r} \right) \quad (51)$$

Darin bezeichnet N die Ambisonicsordnung, R eine Referenzdistanz und r die Distanz von Hörer zu Quelle. Für Distanzen kleiner als die Referenzdistanz ergibt sich somit ein positiver Gain, für größere Distanzen ein negativer Gain. Als Referenzdistanz wird $R = 3$ gewählt. Der Gain wird auf einen Bereich von $[-3dB, +10dB]$ begrenzt.

Die Grenzfrequenz $f_{0,ls}$ der Filter ist ebenfalls distanzabhängig und berechnet sich zu:

$$f_{0,ls} = N \cdot \left(f_{min} + (f_{max} - f_{min}) \frac{R - r}{R - r_{min}} \right) \quad (52)$$

Dabei werden $f_{max} = 150Hz$, $f_{min} = 60Hz$ und $r_{min} = 0,5$ gewählt. Es ergibt sich somit eine Interpolation von $f_{0,ls}$ zwischen $N \cdot 60Hz$ und $N \cdot 150Hz$ wenn sich der Hörer zwischen der Referenzdistanz R und r_{min} befindet. Für Distanzen $r > R$ und $r < r_{min}$ wird $f_{0,ls}$ auf $f_{0,ls} = N \cdot 60Hz$ bzw. $f_{0,ls} = N \cdot 150Hz$ begrenzt. Dies stellt nur eine grobe Annäherung an die Filter aus [DM04] (Abbildung 8) dar. Die erhaltenen Filterkurven sind allerdings durch die Beschränkung der Wertebereiche für den Gain und die Grenzfrequenz stabil und praktikabel. Durch die stark vereinfachte Berechnung von $f_{0,ls}$ ergibt sich außerdem ein geringerer Rechenaufwand als er bei Verwendung der theoretisch berechneten Filterkurven möglich wäre.

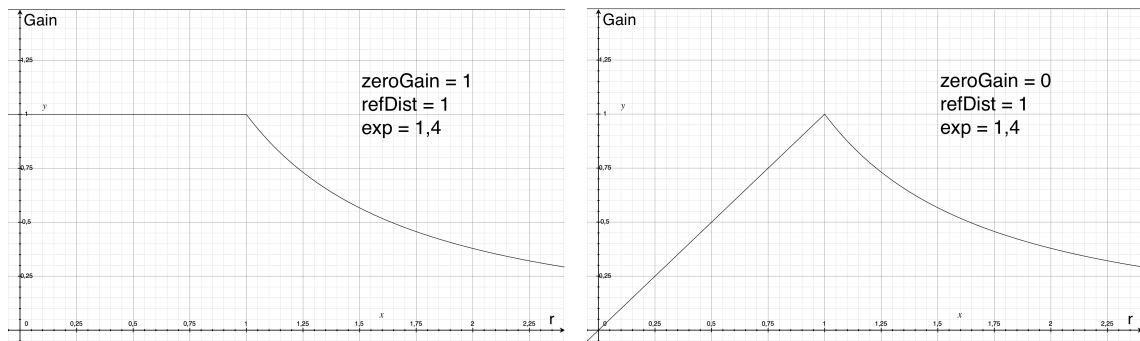


Abbildung 43 – V.l.n.r.: Default-Einstellungen der Distanzgainfunktion für eine Monoquelle bzw. eine Vierkanalquelle.

6.2.5 MonoEncoder, SceneRotator, HrirLoader und BinDecoder

MonoEncoder, *SceneRotator*, *HrirLoader* und *BinDecoder* sind in dieser Reihenfolge für die Enkodierung (vgl. Abschnitt 3.3), Schallfelddrehung (Abschnitt 3.5), das Einlesen von SOFA-HRTFs und die binaurale Dekodierung (Abschnitte 3.4 und 4.2) der Signale zuständig (vgl. Abbildung 41). In *JSAmbisonics* werden all diese Strukturen für periphones Ambisonics zur Verfügung gestellt. Für die Nutzung von 2D Ambisonics und 2D Ambisonics in Kombination mit erster Ordnung 3D werden die Elemente aus *JSAmbisonics* erweitert. Dies ermöglicht die Einsparung von Rechenaufwand trotz hoher horizontaler Auflösung. Wird kein vollständig periphones Ambisonics verwendet, ist prinzipbedingt nur die Schallfelddrehung um die z-Achse möglich. Für die Dekodierung werden in 2D $L = 2n + 2$ Lautsprecher und in 3D t -designs mit $t = 2n + 1$ verwendet. Die virtuellen Lautsprecher werden dabei entweder radialsymmetrisch auf einem Kreis (2D) oder auf den Gitterpunkten des t -designs (3D) verteilt. Es erfolgt jeweils eine $\max \vec{r}_E$ -Optimierung. Da die *Web Audio API* derzeit maximal 32 Kanäle innerhalb eines *AudioNodes* unterstützt, ergibt sich eine Limitierung der maximalen Ambisonicsordnung. Daher sind derzeit maximal periphones Ambisonics vierter Ordnung und planares Ambisonics fünfzehnter Ordnung möglich.

6.2.6 MidiManager

Der *MidiManager* bietet die Nutzung eines *MrHeadTrackers*¹¹ [RBR⁺17] (vgl. Abbildung 41) zur Verfolgung von Kopfbewegungen. Der *MrHeadTracker* ist ein kostengünstiger, quelloffener (*open source*) Headtracker. Die Daten werden dabei mittels USB über das MIDI-Protokoll (*Musical Instrument Digital Interface*) an den Browser geschickt. Zur Implementierung wird die *Web Midi API*¹² benutzt. Die *Web MIDI API* ist derzeit nur in den Browsern *Chrome* und *Opera* implementiert. Damit ist die Nutzung des Headtrackers derzeit auch nur mit einer aktuellen Version dieser Browser möglich.

11. <https://git.iem.at/DIY/MrHeadTracker> (zuletzt aufgerufen am 09.06.17)

12. <https://www.w3.org/TR/webmidi/> (zuletzt aufgerufen am 09.06.17)

6.2.7 UIManager

Der *UIManager* (*user interface manager*) (vgl. Abbildung 41) ermöglicht die graphische Darstellung der Szene und die Steuerung durch den Anwender. Abhängig von den im *Scene File* spezifizierten Raumdimensionen wird eine zweidimensionale Bewegungsfläche erstellt. In dieser kann der virtuelle Hörer frei bewegt werden. Neben dem virtuellen Hörer werden auch die virtuellen Lautsprecher mit ihrer spezifizierten Orientierung graphisch dargestellt. Zudem bietet die Benutzeroberfläche dem Anwender die Möglichkeit der Einstellung einiger Parameter wie Ordnung, 2D/3D, HRTF Set, Lautstärke und Azimut-/Elevationsblickwinkel des virtuellen Hörers. Außerdem kann zwischen bereitgestellten Audioszenen umgeschaltet werden. Die graphische Benutzeroberfläche der Webapplikation ist in Abbildung 44 dargestellt.

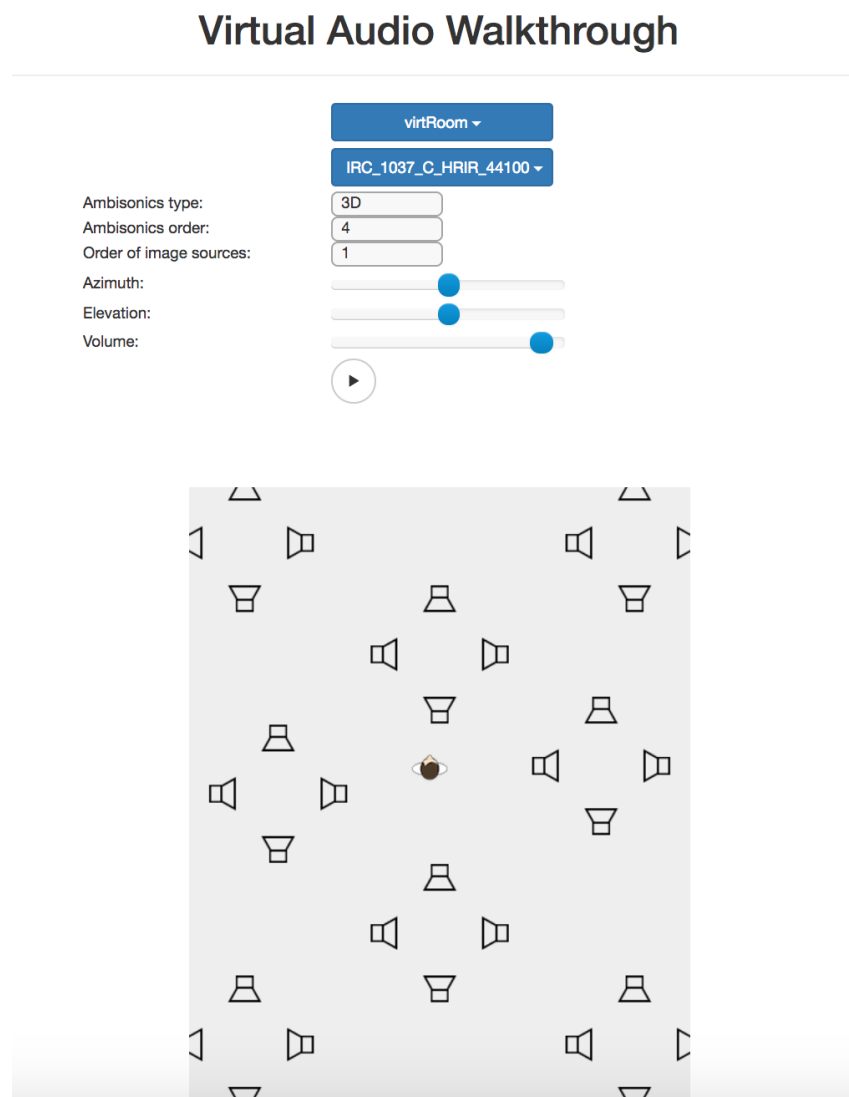


Abbildung 44 – Graphische Benutzeroberfläche der Webapplikation.

7 Schlussfolgerung und Ausblick

Dekodervergleich. Bei allen der in Kapitel 5 untersuchten Kriterien werden bessere Ergebnisse mit $\max \vec{r}_E$ -Optimierung erreicht. Das Weglassen dieser Optimierung ist nur in Spezialfällen (z. B. Dualband Dekoder) zu empfehlen. Da die Implementierung der $\max \vec{r}_E$ -Optimierung nur eine Matrixmultiplikation mit annäherbaren Gewichtungsfaktoren (Gleichungen 26, 27) beinhaltet, ist bei heutigen Computern dadurch auch kein ausschlaggebender Anstieg des Rechenaufwands zu befürchten.

Die beste Wahl der Lautsprecheranzahl ist auf Basis der hier untersuchten Kriterien nicht eindeutig auszumachen. Zwar erzielen nach Betrachtung des relativen spektralen Fehlers und der Fluktuation der Anzahl aktiver Lautsprecher mehr Lautsprecher bessere Ergebnisse, allerdings verschlechtert sich dadurch bei Untersuchung mittels Modell auch die Lokalisation in der Medianebene. Die Entscheidung für die in Abschnitt 6.2.5 genannte Lautsprecheranzahl ist unter Berücksichtigung des benötigten Rechenaufwandes und den Empfehlungen in der Literatur (z. B. [ZF12], [BH14]) gefallen.

Die Auswahl der Ambisonicsordnung wird der anwendenden Person überlassen. Dadurch kann die verfügbare Rechenleistung optimal ausgenutzt werden und zwischen verschiedenen Ordnungen verglichen werden.

Prinzipiell ist fraglich, wie gut die untersuchten Kriterien Auswirkungen auf die Wahrnehmung von Lokalisation und Klang beschreiben. Nur unter Miteinbeziehung von Hörversuchen kann eine eindeutige Evaluierung der Qualität des Resultats erfolgen.

Implementierung. Das vorgestellte Programm ermöglicht nach subjektiven Tests eine gute Lokalisation in virtuellen Audioszenen, wenn genügend Rechenleistung zur Verfügung steht. Mit steigender Anzahl von in Echtzeit spatialisierten Audiokanälen kann die dem Programm zur Verfügung stehende Rechenleistung auf durchschnittlich leistungsfähigen Computern schnell unzureichend sein. Vor allem durch Aktivierung von Spiegelquellen erfolgt ein deutlicher Anstieg an Audiokanälen. Auch eine Erhöhung der Ambisonicsordnung erhöht den Rechenaufwand. Gerade diese beiden Parameter ermöglichen allerdings einen realistischen Raumeindruck und eine verbesserte Lokalisation.

Im neuesten Entwurf der *Web Audio API*¹³ wird ein *Audio Worklet Interface* zur Einbindung von selbsterstellten JavaScript *AudioNodes* vorgestellt. Das *Audio Worklet Interface* soll die synchrone Ausführung des JavaScript Codes innerhalb der modularen Struktur der *Web Audio API* ermöglichen. Da das *Audio Worklet Interface* Audiosignale innerhalb des echtzeitfähigen *rendering threads* verarbeitet, könnte die Einbindung des *SpatialPanners* in eine *AudioWorkletNode* in der Zukunft eine Verbesserung der Recheneffizienz erzielen. Derzeit ist das *Audio Worklet Interface* allerdings noch in keinem Webbrowser implementiert. Auch eine Untersuchung und Optimierung des Programms bezüglich der Ausnutzung mehrerer Prozessorkerne könnte eine Verbesserung der Recheneffizienz ermöglichen.

Das Programm wurde bisher nur lokal getestet. Bei Einbindung in eine Webseite muss bei großen Szenen eine große Datenmenge auf den Anwendercomputer geladen werden. Die Geschwindigkeit der Internetverbindung des Endnutzers kann daher eine weitere Limitie-

13. <https://webaudio.github.io/web-audio-api/> (zuletzt aufgerufen am 12.06.17)

rung bedeuten. Diese beiden Problempunkte werden in der Zukunft durch den allgemeinen Anstieg von Computerrechenleistung bzw. der Weiterentwicklung von Webbrowsern und dem Ausbau schneller Internetverbindungen zunehmend weniger gravierend sein. Auch eine Unterstützung der *Web MIDI API* durch mehr Webbrowser wird für die Zukunft erwartet.

Die Objekte *monoEncoder*, *sceneRotator*, *hrirLoader* und *binDecoder* der quelloffenen Programmbibliothek *JSAmbisonics* wurden zur Realisierung von 2D-Spatialisierung (mit optionaler erster Ordnung 3D) erweitert. Eine Implementierung dieser Objekte zur Ermöglichung beliebiger gemischter Ordnungen und anschließende Einbindung in *JSAmbisonics* ist geplant.

A Appendix

Beispiel Scene File

```
[{
  "type": "room",
  "width": 4.5,
  "length": 5.5,
  "height": 4,
  "listenerStart": {"x":2,"y":1}
},{
  "type": "mono",
  "name": "Noise",
  "position": {"x":1,"y":1,"z":1},
  "gain": 0.8,
  "orientation": {"azim":90,"elev":-45},
  "distGain": {"exp": 1.4,"zeroGain": 1,"refDist": 1},
  "file": "sounds/noise.wav"
},{
  "type": "mono",
  "name": "Sin",
  "position": {"x":1,"y":4},
  "NFC": 1,
  "directivity": 0.5,
  "orientation": {"azim":-90,"elev":0},
  "file": "sounds/sin200.wav"
},{
  "type": "fourChannelArray",
  "name": "Oktava",
  "center": {"x":4,"y":4},
  "centerDistance": 0.5,
  "directivity": 0.5,
  "file": "sounds/oktava1.ogg",
  "channelMapping": {"speaker1":1,"speaker2":2,"speaker3":3,"speaker4":4}
}]
```

Listing 1 – Beispiel für ein valides sceneFile.json

Literatur

- [Ahn88] W. Ahnert, “Comb-filter distortions and their perception in sound reinforcement systems,” in *preprint 2565, 84th Conv. Audio Eng. Soc.*, 1988.
- [BB07] J. Blauert and J. Braasch, “Räumliches Hören,” in *Handbuch der Audiotechnik*, S. Weinzierl, Ed. Springer, 2007.
- [BBA13] O. Brimijoin, A. Boyd, and M. Akeroyd, “The contribution of head movement to the externalization and internalization of sounds,” *PLoS One*, vol. 8, no. 12, 2013.
- [BH14] E. Benjamin and A. Heller, “Assessment of ambisonic system performance using binaural measurements,” in *Paper 9129, 137th Conv. Audio Eng. Soc.*, 2014.
- [Bla74] J. Blauert, *Räumliches Hören*. S. Hirzel Verlag, Stuttgart, 1974.
- [Bla97] ———, *Spatial Hearing: The psychophysics of human sound localization*. MIT Press, 1997.
- [BML13] R. Baumgartner, P. Majdak, and B. Laback, “Assessment of sagittal-plane sound localization performance in spatial-audio applications,” in *The Technology of Binaural Listening*, J. Blauert, Ed. Springer, 2013.
- [Bor84] J. Borish, “Extension of the image model to arbitrary polyhedra,” *Journal of the Acoustical Society of America*, vol. 75, 1984.
- [BV95] J. Bamford and J. Vanderkooy, “Ambisonic sound for us, an analysis of imaging in ambisonics, stereo and dolby surround systems,” in *preprint 4138, 99th Conv. Audio Eng. Soc.*, 1995.
- [Col13] T. Collins, “Binaural ambisonic decoding with enhanced lateral localization,” in *Paper 8878, 134th Conv. Audio Eng. Soc.*, 2013.
- [CS72] D. Cooper and T. Shiga, “Discrete-matrix multichannel stereo,” in *preprint 849, 42nd Conv. Audio Eng. Soc.*, 1972.
- [Dan00] J. Daniel, “Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia,” Thèse de doctorat de l’Université Paris 6, 2000.
- [Dan03] ———, “Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonics format,” in *AES 23rd International Conference on Signal Processing in Audio Recording and Reproduction*, 2003.
- [DEH11] M. Dietz, S. Ewert, and V. Hohmann, “Auditory model based direction estimation of concurrent speakers from binaural signals,” *Speech Communication*, Vol. 53, No. 5, 2011.
- [DM04] J. Daniel and S. Moreau, “Further study of sound field coding with higher order ambisonics,” in *116th Conv. Audio Eng. Soc.*, 2004.
- [DRP98] J. Daniel, J.-B. Rault, and J.-D. Polack, “Ambisonic encoding of other audio formats for multiple listening conditions,” in *preprint, 105th Conv. Audio Eng. Soc.*, 1998.

- [FMK⁺11] S. Favrot, M. Marschall, J. Käsbach, J. Buchholz, and T. Weller, “Mixed-order ambisonics recording and playback for improving horizontal directivity,” in *131st Conv. Audio Eng. Soc.*, 2011.
- [Fra13a] M. Frank, “Phantom sources using multiple loudspeakers in the horizontal plane,” Dissertation, Universität für Musik und darstellende Kunst Graz, Graz, A, 2013.
- [Fra13b] —, “Source width of frontal phantom sources: Perception, measurement, and modeling,” *Archives of Acoustics*, vol. 38, no. 3, 2013.
- [Fra14] —, “Localization using different amplitude-panning methods in the frontal horizontal plane,” in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, 2014.
- [Ger72] M. Gerzon, “Periphony: With-height sound reproduction,” in *2nd Conv. of the Central Europe Section of the Audio Eng. Soc.*, 1972.
- [Ger77] —, “Design of ambisonic decoders for multispeaker surround sound,” in *58th Conv. Audio Eng. Soc.*, 1977.
- [Ger92] —, “General metatheory of auditory localisation,” in *Preprint 3306, 92nd Conv. Audio Eng. Soc.*, 1992.
- [HBL12] A. Heller, E. Benjamin, and R. Lee, “A toolkit for the design of ambisonic decoders,” in *Linux Audio Conference*, 2012.
- [HBS99] K. Hartung, J. Braasch, and S. J. Sterbing, “Comparison of different methods for the interpolation of head-related transfer functions,” in *AES 16th International Conference on Spatial Sound Reproduction*, 1999.
- [Jef48] L. Jeffress, “A place theory of sound localization,” *J. Comp. Physiol. Psychol.*, vol. 41, 1948.
- [KD15] G. Kearney and T. Doyle, “Height perception in ambisonic based binaural decoding,” in *Paper 9423, 139th Conv. Audio Eng. Soc.*, 2015.
- [KDS93] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, “Auralization - an overview,” *J. Audio Eng. Soc.*, Vol. 41, No. 11, 1993.
- [KHI05] M. Karjalainen, P. Huang, and J. O. S. III, “Digital waveguide networks for room response modeling and synthesis,” in *Paper 6394, 118th Conv. Audio Eng. Soc.*, 2005.
- [Kro14] M. Kronlachner, “Spatial transformations for the alteration of ambisonic recordings,” Master’s Thesis, Universität für Musik und darstellende Kunst Graz, Graz, A, 2014.
- [LB02] E. Langendijk and A. Bronkhorst, “Contribution of spectral cues to human sound localization,” *Journal of the Acoustical Society of America*, vol. 112, 2002.
- [Lin86] W. Lindemann, “Extension of a binaural cross-correlation model by contralateral inhibition,” *Journal of the Acoustical Society of America*, vol. 80, 1986.
- [LX16] Y. Liu and B. Xie, “Analysis on the timbre of horizontal ambisonics with different decoding methods,” in *Paper 9677, 141th Conv. Audio Eng. Soc.*, 2016.

- [Maj14] P. Majdak, “Vorlesungsskript Algorithmen in Akustik und Computermusik,” 2014, abgerufen am 12.05.17, <http://piotr.majdak.com/alg/VO/spatial1.pdf>.
- [MM02] E. Macpherson and J. Middlebrooks, “Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited,” *Journal of the Acoustical Society of America*, vol. 111, no. 5, 2002.
- [Moo03] B. Moore, *An Introduction to the Psychology of Hearing*. Academic Press, 2003.
- [NMSH03] M. Noisternig, T. Musil, A. Sontacchi, and R. Höldrich, “3d binaural sound reproduction using a virtual ambisonic approach,” in *IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, 2003.
- [NZDS11] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, “Ambix - a suggested ambisonics format,” in *Ambisonics Symposium*, Lexington, KY, USA, 2011.
- [PK15] V. Pulkki and M. Karjalainen, *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics*. Wiley, 2015.
- [Pol96] M. Poletti, “The design of encoding functions for stereophonic and polyphonic sound systems,” *J. Audio Eng. Soc.*, Vol. 44, No. 11, 1996.
- [Pol16] A. Politis, “Microphone array processing for parametric spatial audio techniques,” Dissertation, Aalto University, Finland, 2016.
- [Pom08] H. Pomberger, “Angular and radial directivity control for spherical loudspeaker arrays,” Diploma Thesis, Universität für Musik und darstellende Kunst Graz, Graz, A, 2008.
- [PPQ16] A. Politis and D. Poirier-Quinot, “Jsambisonics: A web audio library for interactive spatial sound processing on the web,” 2016.
- [QdS11] M. Queiroz and G. H. M. de Sousa, “Efficient binaural rendering of moving sound sources using hrtf interpolation,” *Journal of New Music Research*, Vol. 40, No. 3, 2011.
- [RBR⁺17] M. Romanov, P. Berghold, D. Rudrich, M. Zaunschirm, M. Frank, and F. Zotter, “Implementation and evaluation of a low-cost head-tracker for binaural synthesis,” in *Paper 9689, 142nd Conv. Audio Eng. Soc.*, 2017.
- [SBvW16] P. Stitt, S. Bertet, and M. van Walstijn, “Extended energy vector prediction of ambisonically reproduced image direction at off-center listening positions,” *J. Audio Eng. Soc.*, Vol. 64, No. 5, 2016.
- [SHLV99] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, “Creating interactive acoustic virtual environments,” *J. Audio Eng. Soc.*, Vol. 47, No. 9, 1999.
- [SK02] P. Svensson and U. Kristiansen, “Computational modelling and simulation of acoustic spaces,” in *AES 22rd International Conference on virtual, synthetic and entertainment audio*, 2002.
- [SM13] P. Sondergaard and P. Majdak, “The auditory modeling toolbox,” in *The Technology of Binaural Listening*, J. Blauert, Ed. Springer, 2013.

- [Sol08] A. Solvang, "Spectral impairment for two-dimensional higher order ambisonics," *J. Audio Eng. Soc.*, Vol. 56, No. 4, 2008.
- [Son03] A. Sontacchi, "Dreidimensionale Schallfeldreproduktion für Lautsprecher- und Kopfhöreranwendungen," Dissertation, Technische Universität Graz, Graz, A, 2003.
- [SS15] L. Savioja and P. Svensson, "Overview of geometrical room acoustic modeling techniques," *Journal of the Acoustical Society of America*, vol. 138, 2015.
- [Tra09] C. Travis, "A new mixed-order scheme for ambisonic signals," in *Ambisonics Symposium*, Graz, A, 2009.
- [Vor08] M. Vorländer, *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer, 2008.
- [Vor14] —, "Virtual acoustics," *Archives of Acoustics*, vol. 39, no. 3, 2014.
- [WA01] D. Ward and T. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on speech and audio processing*, Vol. 09, No. 06, 2001.
- [WFZ14] F. Wendt, M. Frank, and F. Zotter, "Panning with height on 2, 3, and 4 loudspeakers," in *2nd Int. Conf. on Spatial Audio (ICSA)*, 2014.
- [Wil99] E. G. Williams, *Fourier Acoustics*. Academic Press, 1999.
- [WK99] F. Wightman and D. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *Journal of the Acoustical Society of America*, vol. 105, 1999.
- [ZF12] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *J. Audio Eng. Soc.*, Vol. 60, No. 10, 2012.
- [Zot09] F. Zotter, "Analysis and synthesis of sound-radiation with spherical arrays," Dissertation, Universität für Musik und darstellende Kunst Graz, Graz, A, 2009.
- [ZPF09] F. Zotter, H. Pomberger, and M. Frank, "An alternative ambisonics formulation: Modal source strength matching and the effect of spatial aliasing," in *Paper 7740, 126th Conv. Audio Eng. Soc.*, 2009.
- [ZPN10] F. Zotter, H. Pomberger, and M. Noisternig, "Ambisonic decoding with and without mode-matching: A case study using the hemisphere," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, 2010.
- [ZPN12] —, "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, Vol.98, 2012.
- [Zö11] U. Zölzer, Ed., *DAFX, Digital Audio Effects, Second Edition*. Wiley, 2011.