

Dreidimensionale Schallfeldreproduktion für Lautsprecher- und Kopfhöreranwendungen

Dissertation

vorgelegt an der Technischen Universität Graz

von

Alois Sontacchi

Graz, April 2003

Vorwort

Der Versuch ein Schallfeld zu reproduzieren ist kein neuartiger Gedanke. Vielmehr wurde dieses Ziel schon kurz nach der Realisierung der ersten elektroakustischen Wandler angestrebt. Die wesentlichen Problemstellungen waren und sind die Kodierungsvorschrift der einzelnen Kanäle und die Anzahl der dafür benötigten Übertragungskanäle. Durch die technischen Entwicklungen der letzten Jahrzehnte stellen Rechenleistung und Speicherbedarf nur begrenzt ein Limit für Systemlösungen dar. Die Entwicklung geht daher in Richtung einer komplexen Signalverarbeitung, in der ein bestmöglicher „Trade off“ zwischen Rechenkomplexität, Speicheranforderung, Kosten und Qualität angestrebt wird. Ziele der heutigen Aufnahme- und Wiedergabetechnik sind einerseits die Schaffung einer möglichst naturgetreuen Reproduktion einer „akustischen Wirklichkeit“, andererseits die Schaffung virtueller Klangräume, die unabhängig von einem konkreten realen Raum zumindest in einem beschränkten Bereich des tatsächlichen Abhörortes entstehen sollen.

Bei der Reproduktion von Schallfeldern beziehungsweise bei der Entwicklung von 3D Beschallungssystemen werden die Attribute der Schallreize, die in natürlichen Hörsituationen auftreten, modifiziert und/oder durch zusätzliche Eigenschaften ergänzt, damit eine gezielte Steuerung der räumlichen Wahrnehmung des Hörenden möglich wird. Daher spielen neben den messbaren physikalischen Größen die psychoakustischen Erwägungen über die menschliche Lokalisationseigenschaften eine wesentliche Rolle.

Aufgrund der Entwicklungen auf dem Gebiet der virtuellen Raumakustik, die im Bereich der Binauraltechnik mit Kopfhörerwiedergabe schon zufriedenstellend angewendet wird, stellt sich die Frage, ob auch durch mehrkanalige Lautsprecherbeschallung und entsprechende Signalverarbeitung ein ebenso überzeugendes Ergebnis erreicht werden kann. In enger Beziehung zu einem solchen Ansatz stehen Anwendungen auf dem Gebiet der „Aktiven Lärmreduzierung“ (*Active Noise Control*, ANC), die aber hier nicht vertieft werden sollen. Weiter ähnliche Anwendungsbereiche entstehen durch die geänderten Anforderungen beim Bau von Veranstaltungsräumlichkeiten. Die bisher separierte und spezialisierte Verwendung von Räumlichkeiten ist heute schon aus Platz- bzw. Kostengründen nicht mehr denkbar. Dadurch setzt sich mehr und mehr der Bau von Mehrzweckhallen bzw. -sälen durch. Das dabei auftretende Problem stellt auch hier die gewünschte variable Raumakustik dar. Einerseits wird ein Chor oder Ensemble einen Raum mit langer Nachhallzeit bevorzugen (tragender Raum), andererseits werden für einen Redner oder für modernes Musiktheater eher kurze Nachhallzeiten bevorzugt, weil dadurch die Wortverständlichkeit weniger gestört wird. Bei dieser Anwendung werden alle Schallquellen im Primärraum mittels räumlich verteilter Mikrofone aufgenommen. Durch eine entsprechende Signalverarbeitung wird diesen ein neuer Raumeindruck aufgeprägt (Sekundärraum), der mittels Lautsprecher wieder in den Primärraum rückgeführt wird. Die Charakteristika (d.s. Nachhallzeit, Energiedichte, erste Reflexionen,...) des dabei neu entstandenen Raumeindrucks lassen sich durch Veränderung des Sekundärraumes weitgehend unabhängig voneinander einstellen und steuern.

In dieser Arbeit wird sowohl auf die Lösungsansätze von holophonen als auch von transauralen Systemen eingegangen, wobei der Schwerpunkt auf holophonen Lösungen liegt. Es werden die Vor- und Nachteile regelungstechnischer Lösungen und verschiedener Kodierungsvorschriften untersucht.

Zur Struktur dieser Arbeit

Am Beginn werden die grundlegenden Theorien und Beschreibungsgrößen des Richtungs- und Entfernungshören beschrieben. Des Weiteren werden die Prinzipien von Beschallungssystemen kurz skizziert. Im zweiten Kapitel wird auf die zwei Beschallungsansätze „Ambisonic“ und Wellenfeldsynthese (*Wave Field Synthesis*, WFS) detailliert eingegangen. Sie bilden die Grundlage für die entwickelten und untersuchten Systeme, die in den folgenden Kapitel vorgestellt werden. In Kapitel 3 werden Lautsprecherapplikationen behandelt, die aus einer Verbindung von Ambisonic und WFS entwickelt wurden. Es werden verschiedene theoretische Lösungen zur Kompensation der Reflexionen des Wiedergaberaumes vorgestellt. Darüber hinaus werden Modelle zur Distanzkodierung und zur Erzeugung von virtuellen Räumen und 3D Hall gezeigt. Im vierten Kapitel werden Kopfhörerapplikationen vorgestellt. Aufbauend auf früheren Arbeiten am IEM wurde eine Gruppe von Algorithmen entwickelt, die eine *Toolbox* für virtuelle zeitvariante Binauralsysteme bilden. Hörversuche sollten Aufschluss über die Qualität der vorgestellten Systeme geben. Sie sind in Kapitel 5 dokumentiert. Die Arbeit wird abschließend in Kapitel 6 zusammengefasst.

Kurzfassung

In dieser Arbeit werden neue Lautsprecher- und Kopfhöreranwendungen für dreidimensionale Schallfeldreproduktion vorgestellt. Diese Anwendungen basieren auf der erweiterten Ambisonic-Theorie und der Wellenfeldsynthese. Während Ambisonic als ein mathematisch motiviertes Verfahren bezeichnet werden kann, handelt es sich bei der Wellenfeldsynthese um eine physikalische Modellierung des Schallfeldes.

Als bereits realisierte Lautsprecherapplikation für 3D-Wiedergabe virtueller Quellen in der oberen Hemisphäre wird der IEM CUBE präsentiert. Zur Erzeugung einer virtuellen Raumakustik werden die frühen realen Raumreflexionen, basierend auf dem Ambisonic-Ansatz und den gemessenen direktionalen Raumimpulsantworten, kompensiert und eine Distanzkodierung virtueller Quellen vorgestellt. Objektive Größen zur Beurteilung der synthetisierten Schallfelder werden eingeführt.

Ein zeitvariantes Binauralsystem wird diskutiert, welches im Gegensatz zu früheren Ansätzen nicht die Außenohrimpulsantworten (HRIRs) interpoliert, sondern das Schallfeld modifiziert. Zur objektiven Beschreibung der Lokalisation und der Lokalisationsunschärfe wird ein Fehlermaß definiert, das auf der psychoakustisch bewerteten Abweichung der durch das Wiedergabesystem verfälschten HRIRs zu den Referenzkurven basiert. Eine Optimierung des Binauralsystems sowie die Verifikation des Fehlermaßes erfolgt anhand von Hörversuchen.

Abstract

In this thesis new applications for three-dimensional sound field reproduction via loudspeakers and headphones are presented. These applications are based on the extended ambisonic theory and the wave field synthesis. While ambisonic can be considered as a mathematically motivated approach, wave field synthesis is related to physical modelling of the sound field.

The IEM CUBE is presented as a state-of-the-art multi-channel loudspeaker application for 3D reproduction of virtual sound sources in the upper hemisphere. As virtual room acoustics systems require the compensation of early reflections in the real room, a new concept for dereverberation based on the ambisonic approach and measured directional room impulse responses is proposed. Furthermore distance coding for virtual sound sources is presented. The reproduction quality is assessed by objective sound field indicators.

A comprehensive time-variant binaural system is proposed utilizing virtual sound field modification instead of interpolating head related impulse responses (HRIRs) like in conventional implementations. An objective error model is proposed for the description of localisation and localisation blur. This model is based on the psychoacoustically evaluated deviation of the system HRIRs to the reference curves. Various parameter settings of the binaural system as well as the properties of the error model are subjectively evaluated by listening tests.

Inhaltsverzeichnis

Vorwort

Zur Struktur dieser Arbeit

Kurzfassung

Abstract

1	Grundlagen	1
1.1	Grundlagen des räumlichen Hörens	2
1.1.1	Die Richtungswahrnehmung	3
1.1.2	Die Entfernungswahrnehmung	8
1.1.3	Peilbewegungen – spontane Drehbewegungen	10
1.2	Prinzipien der Beschallungstechnik	12
1.2.1	Stereofone bzw. zweikanalige Systeme	13
1.2.2	Matrizierung und Panning bei Lautsprecherwiedergabe	16
1.2.3	Diskrete Surround Systeme [3/2, (3/4) bzw. 5.1, (7.1)]	24
1.2.4	Holophone Systeme	25
2	Theoretische Ansätze	26
2.1	Holophonie – Bounded Surface Control (BSC)	27
2.2	Ambisonic	29
2.2.1	Kodierungs- und Dekodierungsvorschriften für Ambisonic	30
2.2.2	Optimierungskriterien für Dekodierungsvorschriften	39
2.2.3	Eigenschaften der Ambisonic Kodierung und Alternativen	43
2.2.4	Untersuchung der Fehler bei idealen Wiedergabebedingungen	44
2.3	Wellenfeldsynthese	49
2.3.1	Kirchhoff-Helmholtz Integral (KHI)	51
2.3.2	Rayleigh Integral	52
2.3.3	Syntheseoperatoren	54
2.3.3.1	Der 3D Syntheseoperator	54
2.3.3.2	Der 2½ D Syntheseoperator	54
2.3.3.3	Syntheseoperatoren für gekrümmte Linienformen	58
2.3.3.4	Einbindung der Abstrahlcharakteristik der Lautsprecher	59
2.3.4	Artefakte	60
2.3.4.1	Durch endliche Abmessungen	60
2.3.4.1	Durch die Diskretisierung (Räumliches Aliasing)	62
2.4	Verknüpfung von Ambisonic und Wellenfeldsynthese	64

3	Lautsprecherapplikationen	68
3.1	Der IEM-CUBE	69
3.1.1	Wahl des Renderingalgorithmus	69
3.1.2	Anforderung an das Beschallungssystem	70
3.1.3	Module	71
3.1.4	Wahl der Lautsprecheranordnung	72
3.2	Objektive Beurteilungsgrößen für Schallfelder	75
3.2.1	Schnelle-Methode	76
3.2.2	Schallintensitäts-Methode	77
3.2.3	Die komplexe Energieflussgeschwindigkeit u	78
3.2.4	Grafische Repräsentationsbeispiele	79
3.3	Kompensation von Raumreflexionen	85
3.3.1	durch Messungen im Ambisonic Bereich	85
3.3.1.1	Herleitung für ein 2D-System mit Ambisonic 1.Ordnung	85
3.3.2	durch Messung an der Oberfläche, zeitinvariant.	88
3.3.2.1	Sensoranordnung	89
3.3.2.2	Bestimmung der Kompensationsfilter	90
3.3.2.3	Lösung im Frequenzbereich	92
3.3.2.4	Lösung im Zeitbereich	96
3.3.2.5	Adaptive Lösung	97
3.3.3	durch Messung an einer Oberfläche, zeitvariant.	101
3.4	Dreidimensionaler Nachhall	103
3.4.1	Modellierung der Raumimpulsantworten	103
3.4.2	Einfaches funktionales Modell für 3D-Hall	106
3.4.3	Physikalisch modellierter 3D-Hall	107
3.4.4	Gestaltbare Raumakustik	108
3.5	Distanzkodierung virtueller Quellen	110
3.5.1	Ansatz	111
3.5.2	Anzahl und Anordnung der virtuellen Lautsprecher	111
3.5.3	Berechnung der Lautsprechersignale Q_m mit Hilfe der WFS	114
3.5.4	Optimierung der Lautsprechersignale mit Hilfe des LMS Ansatzes	116
3.5.5	Objektive Betrachtung der erzeugten Schallfelder	118
4	Kopfhörerapplikationen	121
4.1	Berechnung eines binauralen Kopfhörersignals	123
4.1.1	Rotationsmatrizen für den 3-dimensionalen Fall erster Ordnung	125
4.2	Implementierung eines 2D-Systems, 3.Ordnung	126
4.2.1	Positionierung der Lautsprecher und Filterreduktion	127
4.2.2	Abbildungsfehler und Optimierung des Dekoders	131
4.3	Erweiterungen des zeitvarianten Binauralsystems	133
4.3.1	Die Rotation des Schallfeldes	133
4.3.2	Die akustische Linse	134
4.3.3	Translatorische Bewegung	138
4.3.4	Verbesserte Lokalisation in der Blickrichtung	142
4.3.5	Verräumlichung	144

5	Hörversuche	158
5.1	Objektives Fehlermaß	159
5.1.1	Berechnung von Lokalisationsfehlern	159
5.1.2	Berechnung der interauralen Laufzeitdifferenz	160
5.1.3	Lokalisation aufgrund interauraler Laufzeitdifferenzen	161
5.1.4	Berechnung der interauralen Pegeldifferenz	163
5.1.5	Lokalisation aufgrund interauraler Pegeldifferenzen	163
5.1.6	Zusammenführung der Lokalisationsparameter	163
5.1.7	Lokalisationsunschärfe	164
5.1.8	Mittlere Fehler	165
5.2	Objektive Analyse der Systemparameter	166
5.2.1	Ordnung des Ambisonic-Systems	166
5.2.2	Anzahl und Anordnung der Lautsprecher	166
5.2.3	Filterlänge	168
5.2.4	Gewichtung der höchsten Ordnung	169
5.2.5	Systeme 3. und 4.Ordnung mit Kaiserfenster	169
5.2.6	Erweiterungsmodul für das objektive Maß	170
5.3	Hörversuche zu den ausgewählten Systemen	171
5.3.1	Versuchsdesign	171
5.3.1.1	Versuchspersonen	172
5.3.1.2	Testsignale	173
5.3.1.3	Versuchsaufbau und Hardware	173
5.3.1.4	Versuchsablauf	175
5.3.2	Auswertung der Hörversuche	175
5.3.3	Interpretation der Ergebnisse	177
6	Zusammenfassung	183
	Literaturverzeichnis	187

Anhang

Kapitel 1

Grundlagen

Grundlagen des räumlichen Hörens

Prinzipien der Beschallungstechnik

1.1. Grundlagen des räumlichen Hörens

Die Auswertung der beiden Ohrsignale¹, das Zusammenwirken beider Ohren, führt zu einem akustischen räumlichen Eindruck. Aus der Kombination von Richtungs- und Entfernungswahrnehmung resultiert die Lokalisation einer natürlichen Schallquelle. Der englische Physiker John Strutt (Lord Rayleigh) formulierte 1907 die sogenannte „Duplex Theorie“ [Strutt, 1907], welche besagt, dass die Pegel- und Laufzeitunterschiede zwischen dem ipsilateralen (dem Schallereignis zugewandten) und dem kontralateralen (dem Schallereignis abgewandten) Ohr zur Lokalisation beitragen. Diese ergeben sich aufgrund der Einfallsrichtung der Schallereignisses und des Verhältnisses der Wellenlänge zu den geometrischen Kopfabmessungen. Die wichtigen Kenngrößen und Eigenschaften des menschlichen Gehörs bezüglich der Wahrnehmung werden durch psychoakustische Experimente ermittelt. Grundlegende Untersuchungen zu Lokalisationseigenschaften² wurden von Jens Blauert [Blauert, 1997] durchgeführt.

Als Koordinatensystem wird ein kopfbezogenes sphärisches Koordinatensystem verwendet, das gegenüber Kopfbewegungen invariant ist. Es wird angenommen, dass die Lokalisation von Schallquellen in der Horizontalebene (xy-Ebene, *Azimutlokalisation*) durch Unterschiede der eintreffenden Schallsignale am ipsi- und kontralateralen Ohr (*interaural differences*) verursacht wird. Die Lokalisation von Schallquellen, die sich nicht in der Horizontalebene befinden (*Elevationslokalisation*), werden durch zusätzliche Auswertung der spektralen Unterschiede (*spectral cues*) der beiden Ohrsignale ermöglicht. Die physikalischen Größen, die die genannten Lokalisationsunterschiede verursachen, sind durch die akustischen Übertragungspfade von einer Schallquelle bis zum linken bzw. rechten Gehörgang / Trommelfell beschrieben. Diese Paare von akustischen Übertragungstrecken können gemessen werden und sind als Außenohrübertragungsfunktionen (*Head-Related Transfer Functions*, HRTFs) bekannt. Darüber hinaus ist auch eine Schätzung der Entfernung von Schallquellen möglich. Bei Unsicherheit in der Beurteilung einer Quellenrichtung werden meist kleine Kopfbewegungen (Peilbewegungen als *dynamic cues*) durchgeführt. Im Folgenden wird auf die wesentlichen Größen bei der Lokalisation von Schallquellen eingegangen.

¹ Der Begriff „Ohrsignale“ wurde durch Jens Blauert [Blauert, 1974] geprägt und bezeichnet die Schallsignale in den Gehörgängen. Sie sind die wesentlichsten Eingangssignale einer (Versuchs-) Person bezüglich des räumlichen Hörens.

² Unter Lokalisation versteht man das Feststellen der Elevation (Erhebung in der Medianebene), des Azimut (Auslenkung in der Horizontalebene) und der Distanz der Schallquelle vom Koordinatenursprung.

1.1.1. Die Richtungswahrnehmung

Die Richtungswahrnehmung ist ein komplexer Prozess, der zum Teil durch die Analyse des eintreffenden Schalls auf der Basiliarmembran, aber vor allem in der darauffolgenden Weiterverarbeitung im Gehirn erfolgt. Die Fähigkeit zur Lokalisation ist zum einen stark verknüpft mit der visuellen Wahrnehmung und zum anderen auch durch das erlernte Unterscheiden von Schallreizen und somit von den Erfahrungen. Dadurch ist es auch dem *monauralen* (einohrigen) Hörer teilweise möglich unterschiedliche Schalleinfallrichtungen zu erkennen. Die Beurteilung der Schallereignisrichtung beim *binauralen* (beidohrigen) Hören kann aufbauend auf die Anordnung und der Beschaffenheit der beiden Ohren im wesentlichen mit drei Größen beschrieben werden.

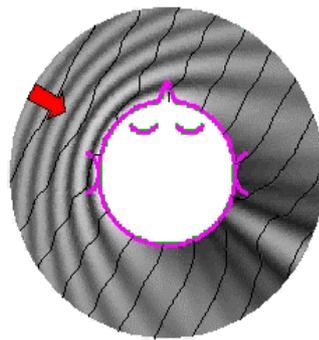
- Interaurale Pegeldifferenzen (*Interaural Level Differences, ILD*)
- Interaurale Zeitdifferenzen (*Interaural Time Differences, ITD*)
- Spektrale Unterschiede (*Spectral Cues*)

Richtungshören in der Horizontalebene – interaurale Unterschiede

In der Horizontalebene tragen vor allem die interauralen Unterschiede wie ILD und ITD zur Richtungsbestimmung bei. Diese beiden Größen werden meist bei Annahme einer ebenen Welle - Schallquelle im Unendlichen - in der Freifeldsituation beschrieben.

Interaurale Pegeldifferenzen

Die Pegeldifferenzen zwischen kontra- und ipsilateralen Ohr werden aufgrund von Beugungs- und Abschattungserscheinungen verursacht. Trifft eine Schallwelle auf den Kopf, so stellt dieser ein Hindernis dar. Ist der Kopf klein im Vergleich zu Wellenlänge, so werden um dieses Hindernis herum kaum unterschiedliche Druckamplituden auftreten (Beugung). Ist das Hindernis ‚Kopf‘ groß im Vergleich zu Wellenlänge, so findet an der Vorderseite eine Reflexion statt, seitlich wird die Schallwelle ungehindert passieren und an der Hinterseite des Hindernisses entsteht ein Abschattung (*head-shadow effect*), siehe Abbildung 1.1. Die Pegeldifferenzen der im freien Schallfeld auftretenden Ohrsignale sind somit nicht nur vom Einfallswinkel sondern auch von der Frequenz des Schallreizes abhängig. Sie werden hauptsächlich im Frequenzbereich über 1,5 kHz zur Bestimmung der Einfallrichtung herangezogen.



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

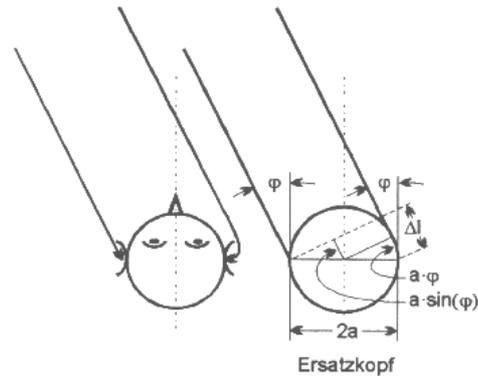


Abb. 1.1: Beugungs- und Abschattungserscheinungen am Kopf als „Kugelmodell“ für eine ebene Schallwelle aus der Richtung von $\theta = 60^\circ$ mit 8 kHz [Daniel, 2000].

Abb. 1.2: Wegdifferenzen einer ebenen Schallwelle beim Auftreffen auf den zur Kugel stilisierten Kopf [Graber, 1997].

Interaurale Zeitdifferenzen

Schallwellen, die seitlich auf den Kopf auftreffen, weisen am kontralateralen Ohr eine Laufzeitverzögerung gegenüber ipsilateralen Ohr auf, die durch die längere Wegstrecke verursacht ist. Wird der Kopf als ein Kugel modelliert, ergibt sich aufgrund der Geometrie für die Laufzeitdifferenz ITD folgende näherungsweise Abhängigkeit (Gl. 1.1 und Abbildung 1.2) vom Einfallswinkel φ (wobei φ in [rad]):

$$ITD = \frac{a}{c} \cdot (\varphi + \sin \varphi) \quad \text{mit } -\frac{\pi}{2} \leq \varphi \leq +\frac{\pi}{2} \quad (1.1)$$

Wobei a der Kugelradius („Kopfradius“) und c die Schallgeschwindigkeit (ca. 340 m/s) ist. Die Auswertung der Laufzeitdifferenz ist ebenfalls nur für einen begrenzten Frequenzbereich gültig. Eine eindeutige Bestimmung der Einfallsrichtung aufgrund der Laufzeit- bzw. Phasendifferenz ist nur in jenem Frequenzbereich möglich, für den die halbe Wellenlänge größer als die maximale Wegstrecke ist. Somit sind die interauralen Laufzeitdifferenzen bis zu einer Frequenz von 1,5 kHz wirksam. Darüber hinaus werden die Phasenverschiebungen nicht mehr eindeutig abgebildet.

ILD und IDT ergänzen einander in ihrer Wirksamkeit und überstreichen den gesamten hörbaren Frequenzbereich.

Zur Untersuchung der Richtungswahrnehmung werden unterschiedliche Stimuli verwendet. Dazu gehören breitbandige Signale wie Musik, Sprache und Rauschen, impulshafte Geräusche, aber auch synthetische, schmalbandige Klänge. Die konkrete Wahl des Stimulus beeinflusst die Fähigkeit zur Richtungswahrnehmung in der Horizontalebene zum Teil

erheblich. So ergeben sich vor allem unterschiedliche Werten für die Lokalisationsunschärfe¹. Sie ist ein Parameter für das akustische Auflösungsvermögen des Menschen. Eine Zusammenfassung dieser Werte findet sich bei [Blauert, 1997]. Die starke Streuung erklärt sich aus der unterschiedlichen Bandbreite der Schallreize, durch deren Vertrautheit (Sprache ist dem Menschen vertrauter als Sinustöne oder schmalbandiges Rauschen), zum Teil jedoch auch durch die verschiedenen Meßmethoden.

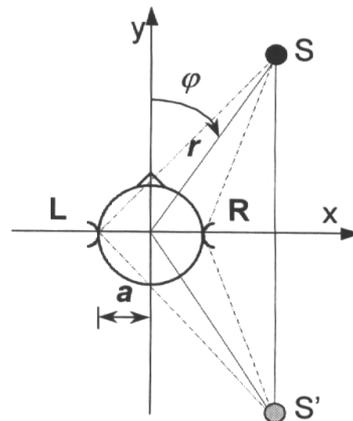


Abb. 1.3: Schallquelle S und S' weisen am Kugelmodell gleiche ILD und ITD auf (vgl. [Täschl, 2000]).

Die Zeit- und Pegelunterschiede können nicht als ausreichend für allseitige Lokalisation betrachtet werden, da es ausgehend von der Symmetrie des Kugelmodells unterschiedliche Richtungen mit gleicher ILD und ITD gibt (siehe Abbildung 1.3). Da diese Punkte auf einer Kugeloberfläche zu liegen kommen, wird dieses Ambiguitätsphänomen auch „*cone of confusion*“ genannt.

Richtungshören in der Medianebene – spektrale Unterschiede

Bei Schallereignissen in der Medianebene fehlen die interauralen Unterschiede wie ILD und IDT. Es müssen daher monaurale Merkmale existieren, die zur Bewertung der Richtung herangezogen werden können. Monaurale Faktoren beschreiben Änderungen, die an beiden Ohren gleich sind. Die Änderung der Schallquellenposition führt zu einer Änderung des Amplitudengangs der Außenohrübertragungsfunktion vorwiegend im hochfrequenten Bereich. Diese spektralen Unterschiede ergeben sich durch Reflexionen von Schallen mit kurzer Wellenlänge am Oberkörper (*torso*) und am Außenohr (*pinna*). Dadurch kommt es zu konstruktiven und destruktiven Interferenzerscheinungen. Die Auslöschungen sind dabei wesentlich stärker ausgeprägt als die Überhöhungen. Es wird davon ausgegangen, dass der

¹ Blauert definiert in [Blauert, 1974] die Lokalisationsunschärfe als: „... die kleinste Änderung eines bestimmten Merkmals oder bestimmter Merkmale des Schallereignisses oder eines anderen, mit dem Hörereignis korrelierten Ereignisses, die gerade zu einer Ortsänderung des Hörereignisses (z.B. bezüglich Richtung und/oder Entfernung) führt.“

Es sei hier darauf hingewiesen, dass die Lokalisationsunschärfe von vielen Parametern abhängt; vor allem von der Schalleinfallrichtung, der Signalbeschaffenheit (wie Art des Signals, Frequenz und Dauer), aber auch vom Alter der Versuchspersonen.

Frequenzbereich der ersten Auslöschung (*pinna notch*) wesentlich für Bestimmung der Elevation verantwortlich ist. Ein breitbandiges Quellspektrum sowie ausreichende Kenntnis der spektralen Beschaffenheit der Schallquelle sind demnach Voraussetzung, um diesen Faktor zur Positionswahrnehmung nutzen zu können (vgl. Untersuchungen mit bekannten und verzerrten Quellspektren [Wightman, Kistler, 1989a]).

Lokalisation in der Medianebene lässt sich teilweise durch richtungsbestimmende Frequenzbänder beschreiben. Die Idee der richtungsbestimmenden Bänder resultiert aus einer Beobachtung von Blauert (vgl. [Blauert, 1969]). Wenn ein einziger Lautsprecher in der Medianebene einen Ton bzw. Schmalbandrauschen mit ansteigender Frequenz (200 – 16kHz) abstrahlt, vollführt die wahrgenommene Schallquelle eine Wanderbewegung (siehe Abbildung 1.4). Die Richtung des Hörereignisses bewegt sich mehrmals auf einer Bahn über den Kopf zwischen vorne, oben und hinten.

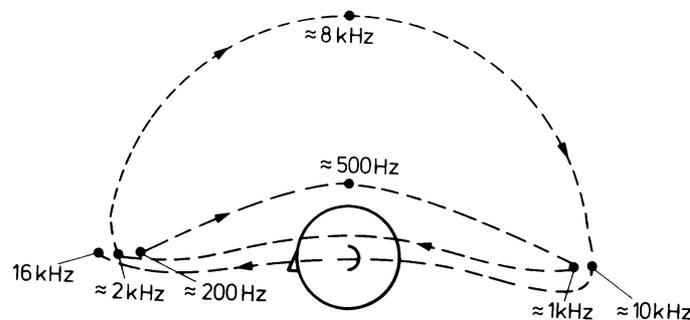


Abb. 1.4: Bahn der wahrgenommenen Hörereignisrichtung als Funktion der Frequenz bei Abstrahlung von Schmalbandrauschen aus einer beliebigen Richtung der Medianebene (Versuchsperson, Kopf fest, stark schematisiert) vgl. [Blauert, 1974].

Schallereignisse in der Medianebene werden demnach entsprechend ihrem Spektrum geortet. So ordnet man z.B. Schallsignale mit einer Frequenz unter 500 Hz und um 3 kHz die Schallquellenrichtung „vorne“, jenen um 8 kHz „oben“, und Schallsignale mit Frequenzen bei 1 kHz die Schallquellenrichtung „hinten“ zu (siehe Abbildung 1.5).

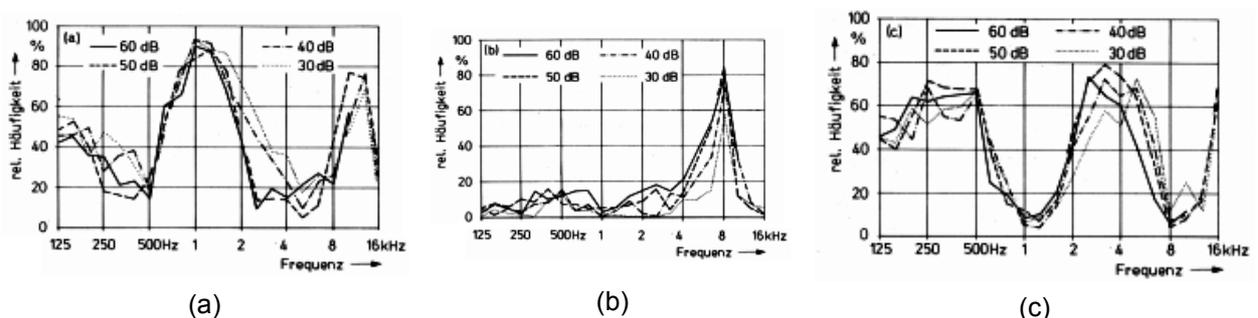


Abb. 1.5: Richtungsbestimmende Frequenzbänder (vgl. [Blauert, 1974])

Relative Häufigkeiten der Antwort „h“ (hinten), 20 Versuchspersonen, bei jeder Terzmittenfrequenz jeweils einmal von vorn und einmal von hinten beschallt, b) wie a), jedoch Antwort „o“ (oben), c) wie a), jedoch Antwort „v“ (vorne).

In der Medianebene wird wesentlich schlechter als in der Horizontalebene aufgelöst. Die größten Unschärfen treten über dem Kopf auf. Auralisationssysteme, die das Konzept der richtungsbestimmenden Bänder als einzige Ankergröße für die Positionierung von virtuellen Quellen in der Medianebene verwenden, führen zu keinem zufriedenstellenden Ergebnis. Darüber hinaustreten dadurch eine artifizielle Klangfärbung auf.

Außenohrübertragungsfunktionen

Seit 1960 wurde verstärktes Augenmerk auf die Pinna und deren Auswirkungen auf die Richtungswahrnehmung gelegt. Arbeiten von Batteau [Batteau, 1967], Hebrank [Hebrank, Wright, 1974], Wright [Wright et al., 1974] beschäftigten sich unter anderem dahingehend und hatten eine Flut weiterführender Untersuchungen zur Folge. Es zeigte sich, dass die Filterung, welche durch die Pinna entsteht, besonders für die Lokalisation in der Medianebene und für die Vorne- / Hinten- Unterscheidung bedeutend ist, da hierbei nur minimale interaurale Differenzen auftreten. Bedingt durch die Anordnung der beiden Ohren und die Geometrie von Ohrmuschel, Kopf und Torso ergeben sich abhängig von der Schalleinfallrichtung Unterschiede bei der Übertragung von einer Schallquelle zu den Ohren. Diese Abhängigkeiten werden durch die kopfbezogenen Außenohrübertragungsfunktionen¹ (*Head-Related Transfer Functions*, HRTFs) beschrieben. Die HRTF beschreibt den Schalldruck im Koordinatenursprung bei Abwesenheit des Kopfes im Verhältnis zum Druck am Trommelfell und ist außer vom Einfallswinkel der Schallquelle auch von derer Entfernung abhängig.

Die binauralen HRTFs kann man sich als entfernungs- und frequenzabhängige Amplituden- und Zeitverzögerungsdifferenzen vorstellen. Sie sind somit Funktionen in 4 Variablen (Azimut, Elevation, Entfernung, Frequenz). Kleine Unterschiede zwischen linker und rechter HRTF sind noch zusätzlich zu berücksichtigen, da die beiden Ohren nicht exakt gleich sind. Insgesamt führen also die asymmetrischen, komplex geformten Pinnae zu einem eindeutigen Set von kleinsten Zeitverzögerungen, Resonanz- und Beugungserscheinungen, das in eine eindeutige HRTF für jede Schallquellenposition übertragen wird.

Werden die HRTF bei Freifeldbedingungen (schalltoter Raum) gemessen und die Entfernung zwischen Schallquelle und Ohren ist größer als ein Meter, dann befindet sich die Quelle für höhere Frequenzen (> 500Hz) im Fernfeld. Die gemessenen HRTF sind somit von der Entfernung unabhängig. Dadurch fehlen sämtliche Informationen über die Quellentfernung und den umgebenden Raum. Der direkte Einsatz dieser HRTFs zur Spatialisierung von Quellen über Kopfhörer führt damit zu Wahrnehmungen, die meist als „zu nahe am Kopf“ oder sogar mit „im Kopf“ (Im-Kopf-Lokalisation, *Inside-Head Localisation*) beschrieben werden. Die Verwendung von HRTF-Kurven aus nicht reflexionsarmen Räumen führt jedoch

¹ In der Literatur finden sich auch andere Bezeichnungen, wie *Head Transfer Function* (HTF), *Pinna Transform*, *Outer Ear Transfer Function* (OETF) oder *Directional Transfer Function* (DTF), die diese Beziehungen beschreiben.

zu sehr langen Filterimpulsantworten. Darüber wäre es schwierig einen anderen virtuellen Raum zu erzeugen.

Zahlreiche Untersuchungen an Menschen und Kunstköpfen (vgl. gute Übersicht [Møller et al., 1995]) führten zu standardisierten HRTF. Diese standardisierten Übertragungsfunktionen stimmen somit nur für den „mittleren“ Menschen d.h. die individuellen HRTF können durchaus stark davon abweichen (vgl. [Wenzel et al., 1993], [Pralong, Carlile, 1996]). Im Allgemeinen treten in der Horizontalebene nur geringe Abweichungen auf (vgl. [Wightman, Kistler, 1989a und 1989b]), in der Medianebene scheint diese Standardisierung jedoch nur bedingt sinnvolle Ergebnisse zu erzielen¹.

Für die Anwendungen in den folgenden Kapiteln wurden als Ausgangsmaterial für die Außenohrübertragungsfunktionen die von Gardner und Martin [Gardner, Martin, 1994] gemessenen standardisierten KEMAR-Kurven² verwendet.

1.1.2. Die Entfernungswahrnehmung

Die Fähigkeit zur Entfernungs- bzw. Distanzwahrnehmung ist durch viele Faktoren bedingt. Zu den wichtigsten Parameter zählen die Lautstärkenverhältnisse (Pegelpportionen), die Vertrautheit der Schallreize und das Energieverhältnis von Direktschall zum reflektierten Schall. Des weiteren sind die Schallfeldausprägung, die Klangfarbenunterschiede (Richtcharakteristik und Luftabsorption) und raumakustischen Aspekte von Bedeutung.

Abhängig vom Abstand zur Schallquelle sind verschiedene Merkmale ausschlaggebend für die Entfernung zum Hörereignisses. Im Folgenden sind drei Untersuchungsergebnisse bei der Darbietung von Schallreizen über Kopfhörer aufgelistet (vgl. [Blauert, 1997]).

- 1) Bei mittleren Schallquellenentfernungen (Punktquelle, zeitlich stationäres Signal, ca. 3 – 15m) bestimmt der Schalldruckpegel³ der Ohrsignale die Entfernung des Hörereignisses. Dieser sinkt im freien Schallfeld um 6dB pro Entfernungsverdopplung: „ $1/r$ – Gesetz“
- 2) Bei Schallquellenentfernungen größer als 15m muss zusätzlich zum $1/r$ – Gesetz die frequenzabhängige Dämpfung der Luft berücksichtigt werden. Hohe Frequenzen werden stärker gedämpft als tiefe. Außer dem Schalldruckpegel der Ohrsignale ändert sich auch das Spektrum mit der Entfernung.
- 3) Bei kleinen Schallquellenentfernungen ($< 3m$) treten spektrale Einfärbungen aufgrund der Krümmung der eintreffenden Wellenfronten am Kopf und Interaktion der Welle mit der Ohrmuschel auf.

¹ Bei Abweichungen zwischen der individuellen und der standardisierten HRTF werden von den Probanden meist Mängel bzw. ein Fehlen von Präsenz geschildert. Quellen in der Nähe der Medianebene (Azimut 0°) werden zusätzlich meist im Kopf anstatt außerhalb lokalisiert.

² KEMAR steht für *Knowles Electronics Manikin for Auditory Research*. Diese Daten finden sich unter: <ftp://sound.media.mit.edu/pub/Data/KEMAR>.

³ Mit Schalldruckpegel ist hier der breitbandig gemessene Effektivwertpegel gemeint.

Pegelerhältnisse

Allgemein gilt: „Was lauter ist, erscheint näher“. Die Intensität (bzw. die Lautheit) einer Quelle ist der primäre Faktor für die Entfernungsschätzung, wenn man sich in einer Freifeldsituation (keine Reflexionen) befindet. Die Intensität nimmt bei doppelter Entfernung um 6dB ab. Wenn zwei Quellen die gleiche Schalleistung abgeben, wird die nähere als lauter empfunden. Dieser Sachverhalt wird durch das „*inverse square law*“ (Gl. 1.2) beschrieben [Begault, 1994].

$$\frac{I_1}{I_2} = \frac{r_2^2}{r_1^2} \quad (1.2)$$

I_1 und I_2 sind jeweils die Intensität der beiden Quellen am Ohr, und r_1 , r_2 sind die Entfernungen zu den beiden Quellen.

Die Kenntnis über die Charakteristik der Schallquelle ist jedoch auch hier von Bedeutung. Bei Sprache führen unterschiedliche Sprechformen zu verschiedenen Entfernungseinschätzungen unabhängig vom dargebotenen Pegel. So wird Flüstern als näher empfunden als Schreien. Diese Wahrnehmung wird durch Gelerntes, Vertrautes geleitet.

In halligen Räumen wird die Wahrnehmung der Entfernung stark durch die Reflexionen bzw. die Energie des Nachhalls beeinflusst. Außerhalb des Hallradius führen Quellen unabhängig von ihrer Entfernung zu gleichen Pegel, da der praktisch entfernungsunabhängige Diffusanteil dominiert. Das R/D-Verhältnis (Energie des reflektierten Schall zu Energie des direkten Schall) ist von Bedeutung. Betrachtet man die zeitliche Abfolge vom direkten Schallereignis und den ersten direkten Reflexionen, so ist deren Versatzzeit nicht nur Maß für die wahrgenommene Raumgröße und Raumform, sondern auch ein Maß für die Entfernung zur Schallquelle. Die Versatzzeit ist bei geringer Entfernung maximal und nimmt mit zunehmender Entfernung von der Quelle ab.

Entfernungswahrnehmung aufgrund interauraler Unterschiede

Je weiter man von der Quelle entfernt ist, desto eher entspricht das sich ausbreitende Schallfeld einer ebenen Welle. Nähert man sich einer Quelle, wird das Schallfeld stärker gekrümmt und entspricht somit einem deutlich wahrnehmbaren Nahfeld. Dieses Phänomen ist von der Frequenz und von der Schallquellenausdehnung abhängig. Translatorische Kopfbewegungen führen zu Azimutänderungen, deren Stärke abhängig von der Entfernung zur Quelle sind [Duda, 1996]. Bei nahen Quellen führt eine kleine Relativbewegung zu großen Azimutänderungen. Bewegt sich eine Quelle nahe an den Kopf heran so steigt die ILD

zusätzlich an, da der Druck invers proportional mit der Entfernung abnimmt¹. Die absoluten Entfernungsunterschieden zwischen den beiden Ohren und der Quelle können nicht mehr vernachlässigt werden.

Entfernungswahrnehmung aufgrund monauraler Phänomene

Dazu wurden zahlreiche Untersuchungen von Blauert [Blauert, 1997] und Begault [Begault, 1994] durchgeführt. Diese Ergebnisse können wie folgt zusammengefasst werden. Das Quellspektrum wird in Abhängigkeit von der Entfernung durch atmosphärische Bedingungen (Wind, Regen,...), molekulare Absorption der Luft (Höhenabfall über weite Luftstrecken), Krümmung der Wellenfront (Richtcharakteristik von Instrumenten,...), Luftfeuchtigkeit, Temperatur, etc. modifiziert.

1.1.3. Peilbewegungen

Wie bereits eingangs erwähnt spielt die visuelle Komponente und die Erfahrung bei der Lokalisation einer Schallquelle eine wesentliche Rolle. Fehlt die visuelle Komponente oder ist die Art der Schallquelle unbekannt (z.B. synthetische Klänge), kommt es häufig zu Unsicherheiten bei der Richtungsbestimmung. Wenn interaurale und monaurale Merkmale zu mehrdeutigen bzw. ungenauen Ergebnissen in der Richtungswahrnehmung führen, werden meist unbewusste spontane Drehbewegungen, sogenannte Peilbewegungen, ausgeführt. Durch diese Drehbewegungen werden die interauralen Unterschiede in Abhängigkeit von der wahren Quellrichtung kleiner oder größer, was mit dem *Cone of Confusion* erklärt werden kann.

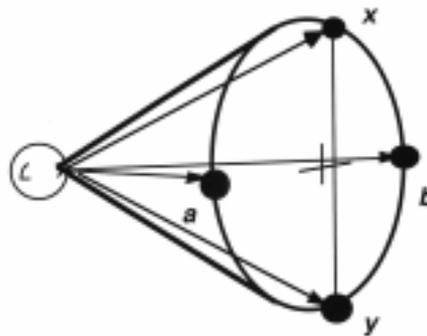


Abb. 1.6: Cone of Confusion [Begault, 1994].

Betrachtet man im Kugelmodell, also ohne Pinna-Einfluss, nur die IID und ITD, so sind diese nicht einer eindeutigen Quellrichtung zuordnen (siehe Abbildung 1.6). Ein Schallereignis am Ort a erzeugt im Kugelmodell dieselben interauralen Unterschiede wie jenes am Ort b. Dies nennt man die Vorne – Hinten - Ambiguität. Ebenso verhält es sich mit den Richtungen x und y, an denen Schallquellen die Elevationsambiguität hervorrufen können. Durch

¹ Vgl. Kugelstrahler: 2-fache Entfernung hat eine Druckabnahme von 6dB zur Folge.

Berücksichtigung der Änderungen von ITD, ILD und spektralen Merkmalen während der Drehbewegungen wird die mehrdeutige Ortung aufgelöst. Bei der Entwicklung von hochqualitativen 3D Beschallungssystemen muss diese Eigenschaft des natürlichen Hörens berücksichtigt werden. Da bei herkömmlichen transauralen Systemen das Beschallungssystem für eine fixierte Kopfposition ausgelegt ist, erhält der Hörer durch Kopfbewegungen inkonsistente Informationen über die Quellrichtung. Noch problematischer ist die Darbietung über Kopfhörer, da sich hier bei Drehungen des Kopfes die Ohrsignale nicht ändern.

Gegenüber der statischen Hörsituation bewirken die Peilbewegungen eine dynamische Hörsituation und liefern damit dem Gehör einen größeren Informationsgehalt über das Schallfeld. Untersuchungen von Thurlow et al. zeigen bereits führt [Thurlow und Runge, 1967], [Thurlow, Mangels und Runge, 1967], dass Kombinationen von Kipp- und Neigebewegungen mit Drehbewegungen bei Versuchspersonen beobachtet wurden, denen Schallreize aus unterschiedlichen Richtungen dargeboten wurden. Versuche von Mackensen et al. [Mackensen, Reichenauer und Theile, 1998] mit einem drehbaren Kunstkopf, bestätigen die verbesserten Lokalisationseigenschaften. Dabei wurde an der Versuchsperson ein „Head-Tracker“-System angebracht, welches abhängig von den Drehbewegungen des Kopfes der Versuchsperson die Orientierung des Kunstkopfes änderte. Die Ergebnisse zeigen die gravierenden Unterschiede zwischen der Lokalisation mit einem starren und mit zeitvariant geführten Kunstkopf. Die Vorne-Hinten-Ambiguität verschwindet vollständig beim Hörversuch mit nachgeführten Kunstkopf in normalen Räumen¹ und werden selbst in akustischen Extrembedingungen noch stark reduziert². Wichtige Qualitätsparameter bei diesen Versuchen und in Folge bei Entwicklungen von zeitvarianten 3D Kopfhörerapplikationen stellen die Latenzzeit des Gesamtsystems und die erreichbare Winkelgenauigkeit dar. Bei zu großen Latenzzeiten folgt das System den Kopfdrehbewegungen zu langsam. Daher werden die Schallquellen nicht stabil an einem Ort abgebildet. Untersuchungen von Felderhoff et al. [Felderhoff, Reichenauer und Theile, 1998] sowie von Wenzel [Wenzel, 1997] zeigen, dass Latenzzeiten des Gesamtsystems unter 85 ms zu keinem wahrnehmbaren „Nachführen“ des abgebildeten Schallfeldes. Als Winkelauflösung ist 1° ausreichend.

Abschließend sei noch erwähnt, dass der Doppler-Effekt bei bewegten Schallquellen auftritt. Darunter versteht man die Veränderung der Tonhöhe in Abhängigkeit von der Entfernungsänderung.

¹ Unter der Bezeichnung „normaler Raum“ versteht sich ein Abhörraum, der dem [EBU Tech. 3276] - Standard genügt

² Die durchgeführten Aufnahmen in einem reflexionsarmen Raum werden als Aufnahmen unter akustischen Extrembedingungen bezeichnet.

1.2 Prinzipien der Beschallungssysteme

Historischer Rückblick

Bis 1930 gab es nur Monoaufzeichnungen und Monowiedergaben. Erst im Zuge der Filmentwicklung zu Beginn der 30iger Jahre wurde der Mehrkanalton eingeführt. 1931 patentierte Blumlein das erste Aufnahme- und Wiedergabesystem für zwei Kanäle, welches heute zur Intensitätsstereophonie gezählt wird. Steinberg und Snow entwickelten 1934 eine Aufnahmetechnik mit drei räumlich verteilten Mikrofonen, deren Signale für eine Wiedergabe mit drei Lautsprechern (Links ,Mitte und Rechts) konzipiert war. Mit De Boer begann um 1940 die genaue Untersuchung stereofoner Abbildung. Um 1970 werden verschiedene Mehrkanal-Aufzeichnungstechniken vorgestellt, wobei es sich hier hauptsächlich um passive Matrixsysteme mit Coder und Decoder handelt (zB Quadrophonie). Die Anzahl der Übertragungskanäle war durch die gängigen Tonträger wie die Schallplatte weiterhin auf zwei begrenzt. Durch das Aufkommen der Digitaltechnik und der damit verbundenen Signalverarbeitungstechnik wurden in den 80er Jahren auch aktive (zeitvariante) Matrixsysteme (z.B. Dolby Pro Logic) realisierbar. Eine Weiterentwicklung der Speichermedien – von der CD zur DVD - und der Rechenleistungen (Zugriffszeiten) ermöglichen heute eine größere Anzahl von Übertragungskanäle.

Transaurale und Holophone Beschallungssysteme

Wie bereits eingangs erwähnt, resultieren aus der Kombination binauraler und monauraler Merkmale der Ohrsignale die Ortung von natürlichen Schallquellen. Der Zusammenhang zwischen der Richtungswahrnehmung und der ILD, ITD bzw. den spektralen Unterschieden bilden die Grundlage für viele Beschallungssysteme und Beschallungsprinzipien. Zur natürlichen Wahrnehmung der reproduzierten bzw. synthetisierten Schallfelder müssen diese Schallfelder entsprechende Kohärenz und Homogenität aufweisen.

Unter Kohärenz versteht man jenen Sachverhalt, dass die Schallquellenposition bei Veränderung der Abhörposition, d. h. durch Kopfbewegungen, stabil bleibt. Dabei ist von besonderer Bedeutung, wie groß der Bereich ist, in dem diese Forderung erfüllt ist . Im Allgemeinen wird dieser Bereich als idealer Abhörbereich (Abhörpunkt, *sweet spot*) bezeichnet.

Ein System wird dann als homogen bezeichnet, wenn unabhängig von der Position der Quelle (bei konstanter Entfernung) die Energie des reproduzierten bzw. synthetisierten Schallfeldes konstant ist. Darüber hinaus darf es auch zu keinen Klangfärbungen in Abhängigkeit von der Quellposition kommen. Daraus folgt, dass keine Richtung bevorzugt behandelt wird.

Im Folgenden werden *transaurale* und *holophone* Beschallungssysteme vorgestellt. Unter transauralen Systemen versteht man Verfahren, die versuchen das gewünschte

Schalldrucksignal nur an den Stellen der beiden Ohren zu reproduzieren. Bei holophonen Systemen handelt es sich hingegen um einen Versuch das gesamte Schallfeld in einem definierten Bereich zu reproduzieren. Die Ortung der Schallquellen erfolgt dann wie unter natürlichen Gegebenheiten. Rein transaurale und rein holophone Systemen stellen jeweils die beiden Extreme dar. Die meisten vorgestellten Systeme lassen sich nicht eindeutig unter einen dieser Begriffe einordnen, da sie mehr oder weniger von beiden Ansätze abgeleitet wurden.

Eine exakte Kopie des originalen Schallfeldes ist bedingt durch die begrenzte Wahrnehmung des Gehörs gar nicht notwendig (siehe u.a. [Blauert, 1997], [Zwicker, 1982], [Fastl, 1982]). Die relevanten Größen des Schallfeldes können in zeitliche und räumliche Eigenschaften unterteilt werden. Zu den zeitlichen Eigenschaften zählt neben der linearen Übertragung, die notwendige Bandbreite und Dynamik. Die korrekte Ortung (Richtung und Entfernung → siehe *Grundlagen des räumlichen Hörens*), die Räumlichkeit¹ (*Spaciousness*), Quellenausdehnung (*apparent source width ASW*) sowie geeignete Diffusität (Einhüllung², *envelopment*) zählen zu den räumlichen Eigenschaften.

1.2.1 Stereofone bzw. zweikanalige Systeme

Diese Systeme folgen mehr dem transauralen Ansatz. Zweikanal-Stereofonie stellt den minimalen Aufwand für eine räumliche Schallreproduktion dar. Bei der Reproduktion unterscheidet man zwischen dem Versuch der realen und der funktionalen Nachbildung eines Schallfeldes. Im ersten Fall wird das originale Schallfeld durch zwei Mikrofone aufgenommen und ohne Signalverarbeitung über Lautsprecher bzw. Kopfhörer wiedergegeben.

Bei der Wiedergabe mit Lautsprechern eignet sich die Intensitätsstereofonie (Aufnahme mit zwei koinzidenten, gerichteten Mikrofonen³, deren Signal direkt den beiden Lautsprecher zugeordnet wird) besser als die Laufzeitstereofonie. Die gängige Anordnung der Lautsprecher und des idealen Abhörplatzes entspricht einem gleichseitigen Dreieck. Die Nachhallzeit im Wiedergaberaum sollte kürzer als jene des Aufnahmeraums sein, weiters sollten starke diskrete Reflexionen - verursacht durch die Raumform, Absorption und Lautsprecherposition - vermieden werden.

¹ Das Attribut ‚Räumlichkeit‘ tritt dann auf, wenn der Zuhörer den Eindruck hat, dass die Hörereignis aus einer größeren räumlichen Breite kommt, als sie der visuelle Eindruck liefert. Sie wird hauptsächlich durch frühe seitliche Reflexionen (Versatzzeit zum Direktschall kleiner 80ms) verursacht [Griesinger, 1997].

² Die Empfindung der Einhüllung wird bei entsprechend diffus verteilter Schallenergie wahrgenommen und durch späte seitliche Reflexionen (80 bis 150ms) verursacht [Griesinger, 1997].

³ Bei der Intensitätsstereofonie werden meist zwei koinzidente Druckgradientempfänger mit Achtercharakteristik (vgl. [Blumlein, 1931]) oder auch mit Nierencharakteristik verwendet. Die Achtercharakteristik ermöglicht eine nahezu lineare Abbildung der Schallquellen unter einem Blickwinkel von $\pm 45^\circ$ zwischen den beiden Lautsprecher. Bedingt durch die gegenphasige Empfindlichkeit der Mikrofonkapseln werden auch evt. nicht gewollte, abgewandte Quellen aufgenommen und entsprechend verkehrt abgebildet. Durch den Einsatz von Kapseln mit Nierencharakteristik werden diese Quellen zwar weitgehend unterdrückt, jedoch kommt es bei der Wiedergabe zu verzerrten Abbildungswinkeln.

Durch die quellrichtungsabhängigen Intensitätsunterschiede zwischen beiden Mikrofonsignalen bzw. Lautsprechersignalen entstehen bei der Wiedergabe Phantomschallquellen zwischen den beiden Lautsprechern. In der Nähe des optimalen Abhörpunktes kann (unter Freifeldbedingungen) das resultierende Schallfeld bei genügend großem Lautsprecherabstand als Superposition zweier ebenen Wellen beschrieben werden. Befinden sich die beiden Lautsprecher bezüglich der Referenzrichtung (entspricht der Blickrichtung, y-Achse) in einem Winkel von $\pm \vartheta_0$, so ergibt sich in Abhängigkeit von der Frequenz ω mit dem rechten Lautsprechersignal R und dem linken Lautsprechersignal L der Schalldruck P auf der Ohrachse (x-Achse):

$$P(x, \omega) = R \cdot e^{j k_x x} + L \cdot e^{-j k_x x} \quad (1.3)$$

wobei $k_x = k \cdot \sin \vartheta_0$ die Wellenzahl $k = \frac{\omega}{c}$ normiert auf die Einfallrichtung ϑ_0 ist.

Die Richtung der resultierenden Welle wird durch die Phasenfunktion $\varphi(x, \omega)$ beschrieben:

$$\varphi(x, \omega) = \arctan \frac{\text{Im}\{P\}}{\text{Re}\{P\}} \quad (1.4)$$

Für tiefe Frequenzen mit $k \cdot x \ll 1$ kann für $\arctan(kx)$ die Approximation $\arctan(kx) \approx kx$ verwendet werden, und man erhält somit:

$$\varphi(x, \omega) \approx \frac{R - L}{R + L} \cdot kx \cdot \sin \vartheta_0 \quad (1.5)$$

Dadurch ergibt sich für den Abbildungswinkel ψ der resultierenden Welle :

$$kx \cdot \sin \psi = \varphi(x, \omega) \quad (1.6a)$$

$$\sin \psi = \frac{R - L}{R + L} \sin \vartheta_0 \quad (1.6b)$$

Die so vorhergesagte Abbildungsrichtung der Phantomquelle stimmt relativ gut mit den Ergebnissen von subjektiven Hörtests [Blauert, 1997] überein. Da die Lautsprechersignale in der Intensitätsstereofonie phasensynchron sind, ergeben sich in Abhängigkeit vom Abhörposition entlang der x-Achse Kammfiltereffekte¹. Berücksichtigt man die Phasendifferenz an zwei Aufpunkten im Ohrabstand, so ergibt sich damit eine frequenzabhängige Richtungsordnung. Am Referenzabhörplatz zeigt diese jedoch einen eindeutigen Trend auf, für davon abweichende Orte ergeben sich jedoch mehrdeutige Richtungsinformationen. In Abbildung 1.7 sind Schallwellenfelder für unterschiedliche Pegeldifferenzen dargestellt.

¹ Diese Kammfiltereffekte sind unter Freifeldbedingungen wesentlich stärker ausgeprägt als unter gewöhnlichen Gegebenheiten. Durch die Raumreflexionen werden diese Unterschiede vermindert.

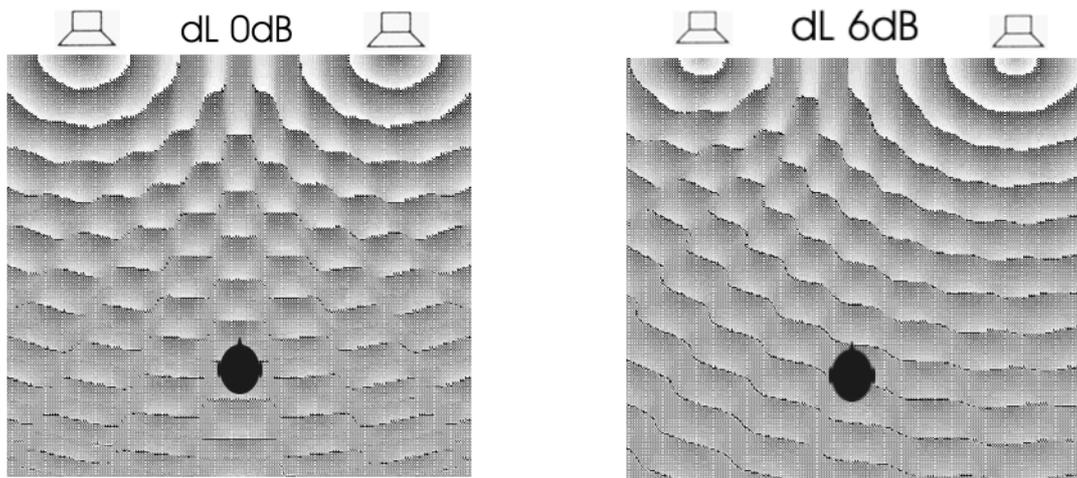


Abb. 1.7: (links) 0dB Pegeldifferenz bei 400Hz, (rechts) 6dB Pegeldifferenz bei 400Hz

Bei der Wiedergabe über Kopfhörer kann man von keiner natürlichen Wahrnehmung sprechen. Die Schallquellen werden im Kopf entlang einer Linie zwischen den beiden Ohren geortet (*Im-Kopf-Lokalisation*). Aufgrund der koinzidenten Mikrofonkapseln werden nur monaurale Rauminformationen aufgezeichnet.

Bei der Laufzeitstereofonie werden die Mikrofonkapseln an zwei entfernten Raumpunkten angebracht. In Abhängigkeit von der Quellposition entstehen zwischen den beiden Mikrofonen Laufzeitunterschiede¹. Beim natürlichen Hören treten die größten Laufzeiten für Quellen aus der $\pm 90^\circ$ Richtung auf und liegen in der Größenordnung von 0.7ms. Bei der Wiedergabe über Lautsprecher sind jedoch größere Laufzeitunterschiede (bis zu 1.5ms vgl. [Blauert, 1997]) notwendig. Diese Umstände können durch die Summenlokalisierung (*summing localization*) [Blauert, 1997] bzw. durch das Assoziationsmodell von Theile [Theile, 1980 und 1981], die auch das Übersprechen der Lautsprechersignale auf beide Ohren beschreiben, erklärt werden. Die wahrgenommene Richtung der Phantomschallquelle ist zusätzlich vom Signal abhängig. So bewirken schmalbandige Signale, abhängig vom Frequenzbereich, für gleiche Laufzeitdifferenzen unterschiedliche Richtungen². In Abbildung 1.8 sind Schallwellenfelder für unterschiedliche Laufzeitdifferenzen dargestellt.

Die mit Laufzeitstereofonie reproduzierten Schallquellen bei Lautsprecherwiedergabe stimmen nicht mit den Positionen der natürlichen Quellen überein. Die wahrgenommene Richtung variiert von Hörer zu Hörer und ist überdies relativ stark von kleinen Kopfbewegungen und vom Signal abhängig. Bei der Wiedergabe mit Kopfhörer ist hingegen eine Wahrnehmung außerhalb des Kopfes möglich, wodurch eine durchaus naturgetreue Abbildung erreicht wird.

¹ Da hier Druckempfänger (Kugelcharakteristik) verwendet werden, können die minimalen Intensitätsunterschiede vernachlässigt werden.

² Bedingt durch die geometrische Ausdehnung des Kopfes und durch den Ohrabstand ergeben sich nur für einen begrenzten Frequenzbereich eindeutige Richtungsinformationen aus den Laufzeitdifferenzen.

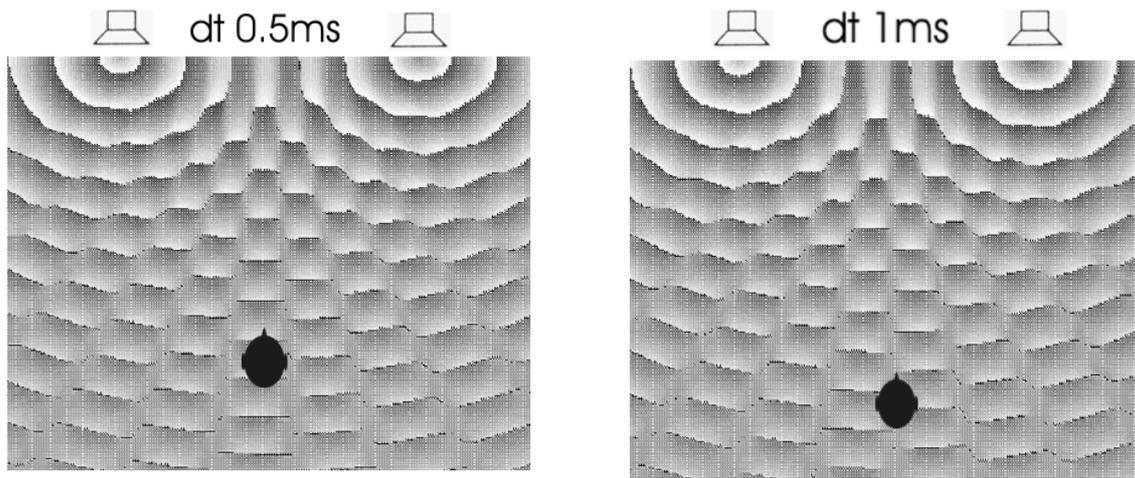


Abb. 1.8: (links) 0.5ms Laufzeitdifferenz bei 400Hz, (rechts) 1ms Laufzeitdifferenz bei 400Hz

Zusammenfassend lässt sich sagen: Die Intensitätsstereofonie eignet sich besser für die Lautsprecherwiedergabe. Sie bietet eine gut lokalisierbare und stabile Quellenabbildung. Es treten keine Phasendifferenzen zwischen den beiden Kanälen auf, dadurch ist auch die Monokompatibilität gewährleistet. Bedingt durch die monauralen Rauminformationen (koinzidente Mikrofonkapseln) werden die räumlichen Eigenschaften nur geringfügig abgebildet. Die reine Laufzeitstereofonie liefert bei der Kopfhörerwiedergabe einen weitgehend natürlichen Eindruck. Durch im Abstand d aufgestellten Mikrofone werden auch manche räumliche Eigenschaften wie Diffusität, Einhüllung aufgrund der reduzierten Signalkohärenz besser abgebildet. Bei der Wiedergabe mit Lautsprechern treten zusätzliche Laufzeitdifferenzen durch das Übersprechen des jeweils vom Ohr abgewendeten Lautsprechers (*Crosstalk* vgl. [Atal, Schroeder, 1966]) auf. Bedingt durch die Überlagerung der beiden Signale am Ohreingang, entsteht eine kammfilterartige Wirkung, diese führt vor allem im hochfrequenten Bereich zu Klangfärbungen. Werden die beiden Signale elektrisch addiert ergeben sich ebenfalls Überhöhungen und Auslöschungen, bedingt durch die Informations- Richtungskodierung in der Phase. Dadurch ist eine Monokompatibilität nicht gewährleistet.

Bei der Aufnahmetechnik gibt es einige Verfahren, die versuchen die Vorteile der reinen Intensitätsstereofonie und der reinen Laufzeitstereofonie zu vereinen. Dazu zählen das ORTF-Verfahren¹, die Trennkörper- oder auch Äquivalenzstereofonie (z.B. OSS, Jecklin-Scheibe²) oder aber Kunstkopftechnik. Die Wiedergabe von Kunstkopfsignalen erfolgt sinnvoller Weise mit Kopfhörer. Bei der Lautsprecherwiedergabe ohne Signalverarbeitung entsteht ein unerwünschtes Übersprechen (*cross talk*) vom linken Kanal zum rechten Ohr bzw. vom rechten Kanal zum linken Ohr. In der Literatur finden sich einige Ansätze, die dieses Übersprechen durch sogenannte *cross talk canceller* verhindern. Dazu zählen: 'Apparent sound source translator' [Atal, Schröder, 1966], 'Stereo Dipole' [Kirkeby, et. al. 1996a], 'Optimal Source Distribution for Virtual Acoustic Imaging' [Takeuchi, 2000a, 2000b]. Darüber hinaus sind die Außenohrübertragungsfunktionen des Kunstkopfes und die des

¹ ORTF wurde 1974 vom franz. Rundfunk als ein Stereostandardverfahren entwickelt.

² OSS Optimales Stereo Signal (vgl. [Jecklin, 1980]).

Hörers überlagert. Diese Überlagerung kann durch entsprechende Vorentzerrung (Filterung) kompensiert werden. Bei der Erzeugung von synthetischen Schallfeldern für Kopfhörerwiedergabe werden die virtuellen Schallquellen mit den entsprechenden HRTFs gefiltert (siehe Kapitel 4).

Über die beschriebenen, teils puristischen Ansätze hinaus gehend werden bei der Stereoreproduktion durch die Verwendung von Effekten wie Hall, Verzögerungen (z.B. bei der Polymikrofonie¹) und Filterung physikalische Eigenschaften des natürlichen Schallfeldes durch funktionale Pendanten ersetzt. Dadurch ergeben sich natürlich wirkende, konsistente Abbildungsbreiten und -tiefen sowie ein brauchbarer Raumeindruck.

1.2.2. Matrizierung und Panning bei Lautsprecherwiedergabe

Damit eine Schallquelle als Phantomschallquelle zwischen den Lautsprechern beliebig wandern und entsprechend geortet werden kann, müssen durch die Reproduktion möglichst viele im natürlichen Hören auftretende Lokalisationsmerkmale konsistent erzeugt werden [Gerzon, 1992a]. Erst dann ergibt sich für den Zuhörenden eine glaubwürdige und eindeutige (natürliche) Richtungswahrnehmung. Zusätzlich sollte die Quellrichtung möglichst stabil, also die Wellenfront möglichst konsistent sein, d.h. bei Bewegungen des Hörers darf sich diese Richtung nicht bzw. nur konsistent verändern, eine Forderung, die von klassischen Transauralsystemen nicht erfüllt wird. Dies lässt sich in die folgenden zwei Bedingungen an das reproduzierte Schallfeld fassen:

- 1) Die Energie des reproduzierten Schallfeldes muss unabhängig von der Position der virtuellen Quelle konstant sein. (*Energetische Betrachtung*)

Die Summenleistung der beiden Lautsprecher ist unabhängig für jeden Abbildungswinkel ϕ konstant.

$$L(\phi)^2 + R(\phi)^2 = 1 \quad (1.7)$$

(normiert auf die Signalleistung)

Dadurch ergibt sich die Kodierung des linken und rechten Kanals zu:

$$L(\phi) = S \cdot \sin\left(\frac{\pi \cdot (\phi_L + \phi)}{4 \cdot \phi_L}\right) \quad \text{und} \quad R(\phi) = S \cdot \cos\left(\frac{\pi \cdot (\phi_L + \phi)}{4 \cdot \phi_L}\right) \quad (1.8)$$

mit S = Signal, ϕ = Winkel der Schallquelle, ϕ_L = Winkel des linken Lautsprechers bzgl. der Blickrichtung.

¹ Unter Polymikrofonie versteht sich die Verwendung vieler verteilter, meist in Quellnähe platzierter Mikrofone.

- 2) Gleichmäßige Stabilität des Klangfeldes d.h.: die Lokalisationsschärfe soll über den ganzen Bereich gleich sein. (*Amplituden- bzw. Phasenbetrachtung*)

Wie oben gezeigt wird durch die phasensynchrone Lautsprechersignale L und R an der Stelle des Zuhörers eine resultierende ebene Wellenfront aus der Richtung ϕ erzeugt mit (tieffrequente Approximation):

$$\frac{\sin \phi}{\sin \phi_L} = \frac{L - R}{L + R} \quad (1.9)$$

Wählt man für ein Aufnahme/Wiedergabesystem den Aufnahmewinkel gleich dem Wiedergabewinkel¹, so ergibt sich bei der Verwendung dieses Phasengesetzes folgende Kodiervorschrift bzw. Richtcharakteristik der Mikrophone, um eine Quelle S aus dem Winkel ϕ wiedergeben zu können:

$$L(\phi) = \left(\frac{1}{2} + \frac{1}{2} \cdot \frac{\sin \phi}{\sin \phi_L} \right) \cdot S \quad \text{und} \quad R(\phi) = \left(\frac{1}{2} - \frac{1}{2} \cdot \frac{\sin \phi}{\sin \phi_L} \right) \cdot S \quad (1.10)$$

Die Lokalisation wird bei tiefen Frequenzen durch die Amplitude der Lautsprechersignale gesteuert. Bei höheren Frequenzen (>700 Hz), wo die Größenordnung des Kopfes in den Bereich der Wellenlänge kommt, treten Beugungserscheinungen auf. Die Lokalisation wird in diesem Frequenzbereich immer mehr durch diese Abschattungseffekte beeinflusst, was die energetische Betrachtung nahe legt.

Es zeigt sich, dass es bei der Wiedergabe mit zwei Lautsprechern nicht möglich beide Bedingungen gleichzeitig zu erfüllen (vgl. [Poletti, 1996a]). Befindet man sich zusätzlich nicht im optimalen Abhörpunkt, so zerfällt der Phantomquelle und die wahrgenommene Richtung wandert automatisch zum näher liegenden Lautsprecher ("*sweet spot*" Problem). Bei der Wiedergabe über mehr als zwei Lautsprecher können Phantomquellen erzeugt werden, die beiden genannten Bedingungen genügen. Einen allgemeinen theoretischen Rahmen für die Herleitung verschiedener Schallfeldkodierungen bietet Micheal Gerzon's

Theorie des Energie- und Schnellektors

Gerzon hat in seiner „*General Metatheory of Auditory Localisation*“ [Gerzon, 1992a] die Lokalisation als Verbindung dreier Einflussfaktoren in einem hierarchischen Modell zusammengefasst. Diese sind die Quellsignal selbst, die Richtcharakteristik der Quelle und der Einfluss der Kopforientierung (starr versus Kopfbewegung). In der Modellhierarchie der Lokalisation kann der auditive Wahrnehmungsapparat als ein System aufgefasst werden, welches auf eine aus dem Quellsignal abgeleitete Funktion aus einer bestimmten Richtung reagiert. Diese Funktion kann das lineare Signal der Schallwelle (lineare Größe, Theorie 1.Ordnung), die Intensität (quadratische Größe, Theorie 2.Ordnung) oder eine komplexere Größe sein. Als weitere Hierarchiedimension führt Gerzon den Grad der

¹ Dadurch werden jene Quellen, die sich im Bereich des Aufnahmewinkels befinden, ohne Winkelverzerrung eindeutig abgebildet, Quellen außerhalb des Aufnahmewinkels werden falsch abgebildet.

Richtungsabhängigkeit gemäß dem höchsten zur Richtungsmodellierung verwendeten Grad der sphärischen Harmonischen ein. Kernpunkt der Metatheorie ist, dass ein Aufnahme/Wiedergabe-System umso besser funktioniert, je mehr Einzelmodelle gleichzeitig in konsistenter Weise zusammengeführt werden.

Nach Gerzon kann eine Theorie 1.Ordnung durch das Schnelle-Modell (*Velocity Model*) beschrieben werden, in dem die wahrgenommene Richtung einer Phantomquelle durch den aus der Überlagerung aller Lautsprecher signale resultierenden Schnellevektor bestimmt ist. Die Kenntnis bzw. Reproduktion des Schalldrucks und der Schallschnelle im idealen Abhörplatz (Koordinatenursprung des Bezugssystems) ist ausreichend, um das Schallfeld auch in der Umgebung um den Ursprung und damit verbunden an den beiden Ohren zu bestimmen. Die nach dem Schnelle-Modell bestimmte Lokalisation wird auch *Makita-Lokalisation* genannt und ist für Frequenzen unter 700Hz von zentraler Bedeutung. Abschattungserscheinungen können in diesem Frequenzbereich vernachlässigt werden, die ITD dominieren in diesem Bereich grundsätzlich die Richtungswahrnehmung.

Die Theorie 2.Ordnung kann durch ein Energie-Modell (*Energy Model*) beschrieben werden. Die Richtung einer über mehrere Lautsprecher reproduzierten Schallquelle wird durch den resultierenden Energievektor (eigentlich die resultierende Intensität) berechnet und als "*Energy Vector Localisation*" bezeichnet. Die Modell ist gültig im Frequenzbereich von 500 Hz bis 5000 Hz.

Bei der Schallfeldreproduktion erfolgt die Berechnung des resultierenden Schnelle- und Energievektors mit den N Lautsprecherpositionen und deren Verstärkungsfaktoren. Die Berechnung erfolgt für den Hörer im Koordinatenursprung, der sinnvoller Weise in den ungefähren Mittelpunkt der Lautsprecheranordnung gelegt wird. In der Matrix \mathbf{S} sind die Positionen der N Lautsprecher $\mathbf{s}_i = (x_i, y_i, z_i)^T$ beschrieben.

$$\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N] \quad (1.11)$$

Zur Vereinfachung der allgemeinen Beschreibung werden alle Lautsprecher auf eine (virtuelle) Einheitskugel projiziert mit:

$$\tilde{\mathbf{s}}_i = \frac{\mathbf{s}_i}{\|\mathbf{s}_i\|} \quad (1.12)$$

$$\tilde{\mathbf{S}} = [\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \dots, \tilde{\mathbf{s}}_N] \quad (1.13)$$

Dazu müssen die Laufzeiten und die Grundverstärkungen korrigiert werden (siehe Gl. 1.14a und b), wobei c die Schallausbreitungsgeschwindigkeit ist.

$$gain_i = \frac{\|\mathbf{s}_i\|}{\max(\|\mathbf{s}_1 \dots \mathbf{s}_N\|)} \quad (1.14a)$$

$$delay_i = \frac{\max(\|\mathbf{s}_1 \dots \mathbf{s}_N\|) - \|\mathbf{s}_i\|}{c} \quad (1.14b)$$

Alle Lautsprecher befinden sich nun virtuell „gleich weit“ vom Hörer entfernt.

Die Annahme des tieffrequenten Lokalisationskriteriums, dass die wahrgenommene Richtung einer Schallquelle normal zur resultierenden Wellenfront liegt, führt zur Berechnung des resultierenden Schnellektors (*velocity vector*) des Reproduktionssystems.

$$r_v \cdot \mathbf{s}_v = \frac{\tilde{\mathbf{S}} \cdot \mathbf{g}}{\sum_{i=1}^N g_i} \quad (1.15)$$

Die Matrix $\tilde{\mathbf{S}}$ beschreibt das Lautsprecherlayout, und der Vektor \mathbf{g} beinhaltet als „lineare Gewichte“ die Verstärkungsfaktoren der einzelnen Lautsprecher. Die Summe der Gewichte dient zur Normierung des resultierenden Schnellektors. Der Einheitsvektor \mathbf{s}_v gibt die Richtung des resultierenden Schnellektors an, r_v stellt seinen Betrag dar. Für eine korrekte Richtungsabbildung muss \mathbf{s}_v mit der Richtung der abzubildenden Quelle \mathbf{s} übereinstimmen.

$$\mathbf{s}_v = \mathbf{s} \quad (1.16)$$

Eine weitere Bedingung ist, dass dessen Betrag gleich eins sein sollte wie bei einer realen Quelle im Freifeld.

$$r_v \equiv 1 \quad (1.17)$$

Wenn nur positive¹ Lautsprechergewichte g_i zulässig sind, gilt: $r_v \leq 1$, das Gleichheitszeichen gilt, wenn das Quellsignal nur zu einem Lautsprecher geführt wird.

Damit eine Quelle unabhängig von der Abbildungsrichtung immer „gleichlaut“ wahrgenommen wird, d.h. unabhängig davon, ob sie von einem oder mehreren Lautsprecher wiedergegeben wird, sollte die Summe der Lautsprechergewichte auf 1 normiert werden (*Amplitudennormierung*).

$$\sum_{i=1}^N g_i = 1 \quad (1.18)$$

Die Betrachtung des Energievektors ergibt sich aus dem Lokalisationskriterium für höhere Frequenzen. Hier reicht die Kenntnis von Druck und Schnelle im idealen Abhörplatz nicht mehr aus, da Abschattungen des Kopfes wirksam werden. Vielmehr ist hier räumliche Leistungsfluss im Abhörort zu verwenden, wobei der Schalldruck des Schnelle-Modells durch die Gesamtleistung und die Schnelle durch die „gerichtete Energie“, eigentlich die Intensität² ersetzt werden. In formaler Übereinstimmung mit 1.15 lässt sich der resultierende Energievektor ansetzen mit:

¹ Durch negative bzw. komplexe Lautsprechergewichte können die Bedingungen im idealen Abhörplatz immer erfüllt werden, allerdings führen sie bei Lautsprecheranordnungen für große Auditorien an entfernten Plätzen zu Artefakten wie „Phasigkeit“ oder einer falschen Richtungsabbildung.

² Gerzon verwendet in seinen Arbeiten konsequent den Terminus „Energy Vector“, obwohl es sich eigentlich um eine „gerichtete Leistung“, also im Intensität, handelt. Aus Konsistenzgründen – auch mit der auf Gerzon aufbauenden Literatur – wird in der Folge ebenfalls der Ausdruck „Energievektor“ verwendet.

$$r_e \cdot \mathbf{s}_e = \frac{\tilde{\mathbf{S}} \cdot \mathbf{q}}{\sum_{i=1}^N q_i} \quad (1.19)$$

Die wahrgenommene Richtung wird durch Summation der Energievektoren der einzelnen Lautsprecher gebildet. Der Vektor $\mathbf{q} = [g_1^2, g_2^2, \dots, g_N^2]$ beinhaltet die Leistungsgewichte. Die Summe der einzelnen Leistungsgewichte dient zur Normierung. Der Einheitsvektor \mathbf{s}_e gibt die Richtung des resultierenden Energievektors an, r_e seinen Betrag. Wiederum sollte gelten:

$$r_e \equiv 1 \quad (1.20)$$

Da aber alle Elemente von \mathbf{q} nicht negativ sind, ist r_e immer kleiner als eins, wenn das Quellensignal auf mehrer Lautsprecher verteilt ist. Gerzon schlägt daher als zusätzliche Anforderung an ein Wiedergabesystem die Maximierung dieses Wertes vor.

Die resultierende Leistung am Abhörplatz sollte für jede Quellposition konstant sein, was eine Normierung auf die Summe der Leistungsgewichte nach sich zieht (*Leistungsnormierung*).

$$\sum_{i=1}^N q_i = 1 \quad (1.21)$$

In Freifeld einer realen Quelle stimmt die Richtung der Schallschnelle und die des Energieflusses überein, beide Vektoren haben die Länge 1.

Pairwise Mixing – Diskretes Amplituden- oder Intensitätspanning

Für *pairwise mixing* wird das Quellensignal wie bei 2-Kanal-Stereo entweder nach Leistungs- oder Phasengesetz ausschließlich auf die der virtuellen Quelle direkt benachbarten Lautsprecher aufgeteilt. Die Vorgangsweise führt nur bedingt zu zufriedenstellenden Ergebnissen (es ist nicht möglich seitliche Phantomschallquellen z.B. zwischen links vorne und links hinten zu erzeugen, vgl. *Quadrophonie*). Es handelt sich hierbei um diskrete Panpot-Gesetze, d.h. wenn die Richtung der virtuellen Quelle mit der eines Lautsprechers übereinstimmt, wird nur dieser bespielt. Dadurch ergeben sich bei der Wahrnehmung von bewegten Quellen Unstetigkeitsstellen. Entweder erfolgt die Bewegung nicht kontinuierlich, d.h. eine konstante Winkelgeschwindigkeit entlang einer Kreisbahn ist nicht realisierbar, oder es treten aufgrund der sich ändernden Leistung sogenannte „Löcher“ auf. Zu diesen Verfahren zählen das *Vector Based Amplitude Panning* (VBAP)¹, basierend auf der lokalen Nachbildung des Schnellektors und das *Vector Based Intensity Panning* (VBIP), basierend auf der lokalen Nachbildung des Energievektors. Diese Verfahren führen bei der direkten Implementierung rasch an die genannten Grenzen bzw. zu Artefakten. Zur Minimierung dieser Artefakte müssen benachbarte Lautsprecher entweder nahe beieinander liegen oder die Lokalisierungsschärfe wird durch eine modifizierte Abbildungsvorschrift reduziert (vgl. [Pulkki,1997], [Dickins et al.,1999], [Jot et al.,1999], etc.).

¹ Analog dazu gibt es *Angle Based Amplitude Panning* (ABAP) und *Angle Based Intensity Panning* (ABIP).

Eine Verbesserung der Abbildungsqualität kann durch zusätzliche Parameter wie die Berechnung von Erstreflexionen nach dem Außenmauer-Modell bzw. „Raum-im-Raum“ Modell (siehe [Moore, 1982]), die Berücksichtigung der Luftabsorption, sowie die entfernungsabhängige Pegelveränderung sowie den Dopplereffekt bei bewegten Quellen erreicht werden (vgl. [Feldbauer, 1998]).

Nicht-diskrete Panpot-Gesetze

Das Quellensignal wird in Abhängigkeit von seiner Lage immer auf alle Lautsprecher nach einer bestimmten Vorschrift aufgeteilt. Gerzon in [Gerzon, 1992b-d] schlägt optimale Panpot-Gesetze vor. In Abbildung 1.9 ist der Verlauf der frequenzunabhängigen Verstärkungsfaktoren für die Abbildung durch 3 Lautsprecher gegeben. Da im Gegensatz zu den diskreten Panpot-Gesetzen hier die Signale, unabhängig vom Abbildungswinkel, immer auf alle Lautsprecher mit unterschiedlichen Anteilen aufgeteilt werden, werden die Bewegungen der Schallquellen als gleichmäßig und homogen empfunden.

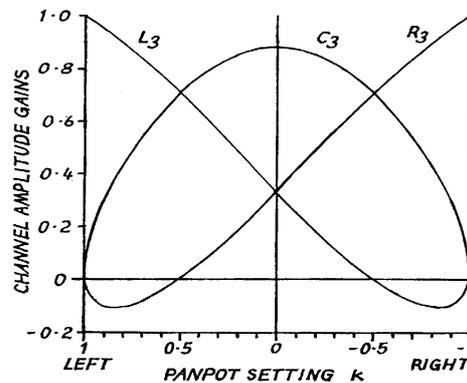


Abb. 1.9: „Optimum Panpot Law“ für 3 Lautsprecher (vgl. [Gerzon, 1992b]).

Eine weitere Verbesserung kann erreicht werden, wenn ein frequenzabhängiges Panpot-Gesetz verwendet wird, das der Energie- bzw. Phasenbedingung in den jeweiligen (überlappenden) Frequenzbereichen genügt:

- Phasentreue bis 700Hz
- Energieerhaltung ab 400-500Hz

Vor allem im überlappenden Bereich müssen konsistente Ergebnisse angestrebt werden. Energie- und Schnellektor einer Quelle müssen dabei die gleiche Richtung und unabhängig von der Richtung möglichst konstanten Betrag (gleich 1, normiert auf die Signalleistung) besitzen. Für eine Realisierung werden 2 Dekodiermatrizen verwendet, die jeweils nach der Energie- und Phasenbedingung unabhängig voneinander für die beiden Frequenzbereiche optimiert sind. Das Eingangssignal wird durch phasenkompensierte Shelving Filter in die zwei entsprechenden Frequenzbereiche aufgeteilt und den beiden Decoder zugeführt. Die Ausgangssignale der beiden Decoder für den jeweiligen Lautsprecher aufaddiert.

Somit wird es auch möglich, dass die beiden Signale L und R auf drei oder mehr Lautsprecher gewinnbringend matriziert werden können. Die Matrizierung von Stereosignale auf mehrere Lautsprecher, durch eine sogenannte „up-conversion matrix“ [Gerzon, 1992c,d]), vergrößert zwar den Abhörbereich (stabile Richtungswahrnehmung über einen größeren Bereich), jedoch verschlechtert sich die Lokalisierungsschärfe. Diese Phänomen wird durch gegenphasige Beaufschlagung der Lautsprecher hervorgerufen. Dennoch ist die Abweichung von der gewünschten richtungsunabhängigen Leistungskonstanz und der Phasentreue minimal. Ähnliche Ansätze finden sich bei [Miles, 1996], der die Wiedergabematrizen für unterschiedliche Lautsprecherlayouts auf reiner Vektoraddition aufbaut.

Poletti stellt in [Poletti, 1996] einen allgemeinen Ansatz für die Berechnung optimierter Kodierungsfunktionen für die Aufnahme und Wiedergabe mit verschiedenen Lautsprecher-Layouts vor. Für zwei Lautsprecher ergibt sich die dem Phasengesetz entsprechenden Stereokodierung (siehe Gl. 1.10). Für eine beliebige Anzahl von symmetrisch entlang eines Kreises angeordnete Lautsprecher (360° 2D-Surround-Anordnung) ergeben sich die „angular sinc functions“ (ASINC). Durch eine Modifikation ist auch eine Berechnung für nicht symmetrische Anordnungen sowie für reduzierte Abbildungswinkel (<360°) möglich. Eine wesentliche Eigenschaft der im folgenden definierten ASINC ist, dass bei der Übereinstimmung einer Quellposition mit einer Lautsprecherposition nur dieser Lautsprecher bespielt wird (diskretes Panpot-Gesetz). Stimmt die Quellposition nicht mit einer Lautsprecherposition überein, tragen alle Lautsprecher zur Erzeugung dieser Richtung bei (siehe Abbildung 1.10). Die Summe der Lautsprechergewichte ist unabhängig vom Abbildungswinkel θ immer konstant 1.

Die ASINC Funktionen leiten sich ab aus der Definition der zirkulären Sinc-Funktion (CSINC) für eine Lautsprecheranzahl N in Abhängigkeit vom Abbildungswinkel θ :

$$CSINC_N(\theta) = \frac{\sin(N \cdot \frac{\theta}{2})}{\sin(\frac{\theta}{2})} \quad (1.22)$$

Für geradzahlige N erhält man eine periodische Funktion mit der Periodenlänge 4π , für ungeradzahlige N ist die Periodenlänge 2π . Diese CSINC Funktionen lassen sich als Summe von überlagerten Kosinusfunktionen darstellen.

$$CSINC_N(\theta) = \sum_{n=-(N-1)/2}^{(N-1)/2} e^{jn\theta} = 1 + 2 \cdot \sum_{n=1}^{(N-1)/2} \cos(n\theta) \quad \text{N...ungerade} \quad (1.23a)$$

$$CSINC_N(\theta) = \sum_{n=-(N-1)/2}^{(N-1)/2} e^{jn\theta} = 2 \cdot \sum_{n=1/2}^{(N-1)/2} \cos(n\theta) \quad \text{N...gerade} \quad (1.23b)$$

Das Set von CSINC Funktionen für gerade bzw. ungerade N führt zur Definition der allgemeinen ASINC Funktionen mit:

$$ASINC_N(\theta) = \begin{cases} \frac{\sin(N \cdot \frac{\theta}{2})}{\sin(\frac{\theta}{2})} = CSINC_N(\theta) & N \dots \text{gerade} \\ \frac{\sin([N-1] \cdot \frac{\theta}{2})}{\sin(\frac{\theta}{2})} + \cos(\frac{N \cdot \theta}{2}) = CSINC_{N-1}(\theta) + \cos(\frac{N \cdot \theta}{2}) & N \dots \text{ungerade} \end{cases} \quad (1.24)$$

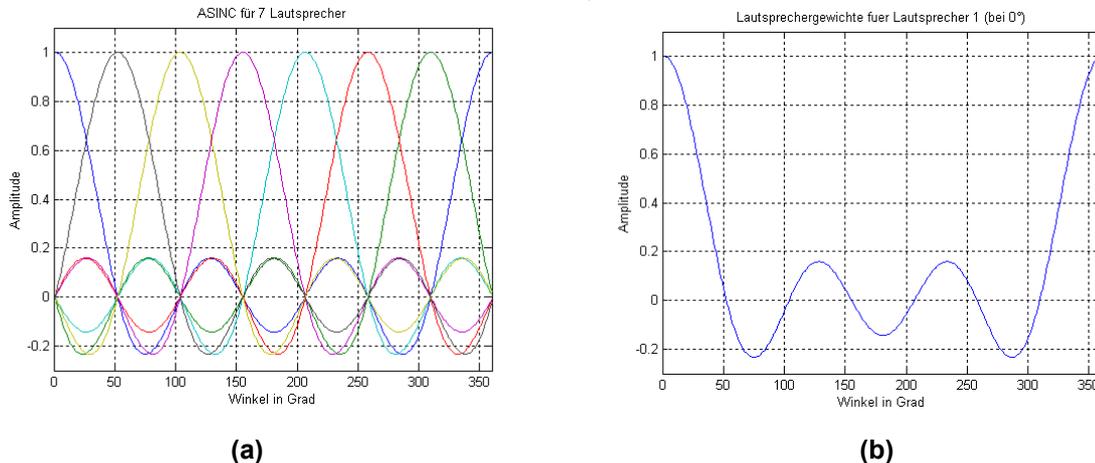


Abb. 1.10: Optimales Panning mit ASINC Funktionen (für $N=7$, symmetrische Anordnung).
(a) alle Lautsprechergewichte und (b) für Lautsprecher bei 0° .

1.2.3. Diskrete Surround Systeme [3/2, (3/4) bzw. 5.1, (7.1)]

Moderne Surround Systeme haben sich aus dem DOLBY SURROUND (ein Trademark und Format zugleich, siehe auch <http://www.dolby.com>) System heraus entwickelt. Die Anordnung der Lautsprecher sowie deren Funktionalität wurden zum Teil übernommen. Die Basiskonfiguration (3/2 bzw. 5.1) besteht aus 5 Lautsprecher. Zwei Lautsprecher befinden sich wie bei 2-Kanal-Stereo vorne in der linken und rechten Position. Es gibt einen Zentrallautsprecher, der im Kinobetrieb ursprünglich als Dialogquelle verwendet wurde und für eine verbesserte Lokalisationsstabilität von vorne sorgt. Des weiteren befinden sich rückseitig zwei Lautsprecher, die für den räumlichen Eindruck (Einhüllung) und für Spezialeffekte verwendet werden können. Für die Übertragung der Dolby Surround – Kanäle wurden diese in ein stereo-kompatibles 2-Kanal-Format kodiert, das sich durch schlechte Kanaltrennung und damit verbunden durch ungenügende räumliche Abbildung ausgezeichnet hat (vgl. [Hull, 1992]). Eine Verbesserung dazu wurde in [Link, Mandell, 1993] vorgestellt.

Bei neueren Surround Systemen, deren Kanäle im Gegensatz zum matrizierten Dolby Surround diskret übertragen werden, soll durch das Zusammensetzen von Schallfeldsektoren ein „vollständiges Schallfeld“ erzeugt werden, welches die Lokalisierungseigenschaften, Stabilitätsanforderungen und die gewünschte Räumlichkeit erfüllt. Diese Vorgangsweise entspricht einer funktionalen Schallfeldsynthese, wobei dennoch von einem sehr kleinen Abhörort - eher dem transauralen Ansatz entsprechend - ausgegangen wird. Durch den Einsatz von 5 diskreten Kanälen kann eine räumliche Abbildung präziser erfolgen (vgl.

[Theile, 1996, 2000], [Damaske, 1967]). Durch die Kanaltrennung kann vor allem die Richtung der einzelner Reflexionen exakt gesteuert werden und somit der gewünschte räumliche Eindruck und Einhüllung erzeugt werden [Griesinger, 1991, 1993, 1995, 1997]. Surround Systeme sind durch Empfehlungen und Richtlinien spezifiziert (vgl. [ITU-R BS. 775-1] und [SSF-01-SV-E1]). Dazu gehört die Anordnung und Spezifikation der Lautsprecher, die Bedingungen an den Abhörraum und das Einmessen des gesamten Wiedergabesystems.

1.2.4. Holophone Systeme

Unter einem holophonische Ansatz versteht man den Versuch der physikalischen Rekonstruktion eines Schallfeldes. In der Literatur finden sich verschiedene Systeme, die sich zum Teil stark durch ihre Komplexität unterscheiden.

- Holophonie (*Bounded Surface Control, BSC*)
- Ambisonic
- Wellenfeldsynthese (*Wave Field Synthesis, WFS*)

Bei BSC [Ise, 1999], [Takane et. al., 1999] handelt es sich um einen regelungstechnischen Ansatz. Es wird versucht das Schallfeld eines Ausschnitt des Aufnahme Raumes in einem anderen Raum zu reproduzieren, in dem die Schallfeldeigenschaften (Schalldruck bzw. Schnelle) auf einer den Primärraumausschnitt umschließenden Fläche den entsprechenden Raumpunkten des Wiedergaberaumes aufgeprägt werden. Dieser Ansatz leitet sich vom Kirchhoff-Helmholtz-Integral (siehe Kapitel 2) ab, welches besagt, dass die Zustände in einem abgeschlossenen quellenfreien Volumen durch die Kenntnis der zustandsbeschreibenden Größen entlang der begrenzenden Volumenoberfläche eindeutig beschreiben werden können.

Die klassischen Vertreter von holophonen Systemen, im wesentlichen Ambisonic und WFS, werden detailliert in Kapitel 2 behandelt, da sie die Grundlage für die im Rahmen dieser Arbeit entwickelten Applikationen bilden. Bei Ambisonic handelt es sich um eine lokale Wellenfrontrekonstruktion, wobei das Schallfeld durch ein Set von vektorialen Komponenten beschrieben wird. WFS hingegen versucht eine globale Wellenfrontrekonstruktion. Dazu werden die Ausbreitungspfade einer Schallwelle, ausgehend vom Huygens'schen Prinzip modelliert und synthetisiert.

Kapitel 2

Theoretische Ansätze

Holophonie

Ambisonic

Wellenfeld Synthese

Verknüpfung von Ambisonic und Wellenfeld Synthese

"...Gegeben sei eine primäre Quelle S_p , die ein Feld F_p in einem Volumen V ausstrahlt; weiters eine Oberfläche Z_0 , die S_p umgibt und V in zwei Bereiche teilt: V_1 innerhalb von Z_0 und V_2 außerhalb von Z_0 (Abbildung 2.1a). Es existiert nun ein Satz von sekundären Quellen S_H , die über Z_0 verteilt sind und ein Feld F_H ausstrahlen, welches Null innerhalb V_1 und gleich F_p innerhalb V_2 ist (Abbildung 2.1b). Dieser Satz von sekundären Quellen S_H kann die primäre Quelle zur Beschreibung des ausgestrahlten Feldes in V_2 ersetzen..." (sinngemäß nach [Huygens, 1690]).

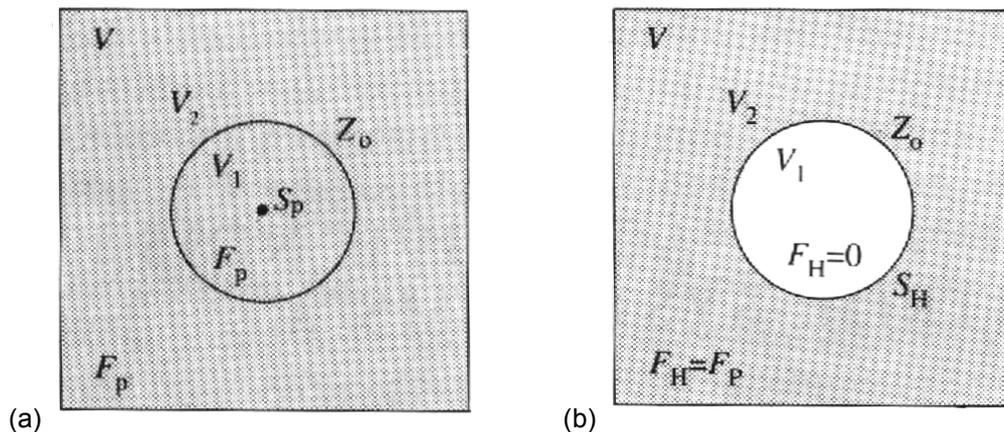


Abb. 2.1: Grafische Darstellung des Huygens'schen Prinzip.

Die Wellenfront einer sich ausbreitenden Welle kann durch räumlich verteilte sekundäre Quellen nachgebildet werden. Daraus folgt, dass die verteilten sekundären Quellen eine bzw. mehrere primären Quelle ersetzen können. Mit diesem Ansatz wird versucht ein primäres (=originales) Schallfeld in einem 2- bzw. 3-dimensionalen Bereich um einen Hörer zu reproduzieren. Dieser Systemansatz wird in Anlehnung an den Begriff Holographie mit „Holophonie“ bezeichnet (vgl. Kap.1.2).

Im Folgenden werden die Grundlagen der "holophonen Systeme" behandelt. Im Vergleich zu transauralen Beschallungsansätzen schaffen sie eine deutlich größere Hörumgebung, wodurch sich die Bewegungsfreiheit eines einzelnen Hörers vergrößert bzw. optimale Wiedergabe für mehrere Hörer erreicht werden kann.

2.1 Holophonie - Bounded Surface Control, BSC

Den Ausgangspunkt für holophone Verfahren stellt der genannte Satz von Huygens dar. Das Kirchhoff-Helmholtz Integral (siehe Gl. 2.1) beschreibt diesen Zusammenhang in mathematischer Form [Verheijen, 1997].

$$p(\vec{r}) = \iint_{\partial\Omega} \left[\vec{\nabla} p_0 \cdot \vec{n} - \frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4\pi R} dS_0 \quad \forall \vec{r} \in \Omega \quad (2.1)$$

Diese Gleichung besagt, dass der Schalldruck p innerhalb eines quellenfreien Volumen Ω durch ein Integral über die das Volumen begrenzende Oberfläche $\partial\Omega$ beschrieben werden kann, dessen Integrand eine Funktion des Drucks p_0 entlang $\partial\Omega$ entspricht. In Gleichung (2.1) ist k die Wellenkennzahl, und \vec{n} ist der normierte nach außen zeigende Flächennormalvektor. Der Vektor \vec{R} beschreibt den Ausbreitungspfad zwischen der sekundären Quelle an der Stelle \vec{r}_0 und dem Hörpunkt (Aufpunkt) \vec{r} (siehe Abbildung 2.2).

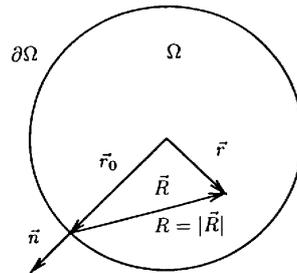


Abb. 2.2: Grafische Darstellung und Bezeichnungen

Im Kirchhoff-Helmholtz-Integral wird die Verteilung der sekundären Quellen durch Monopol- und Dipolquellen beschrieben. Die Amplituden dieser Quellen entsprechen dem Druckgradienten $\vec{\nabla} p_0$ sowie dem Druck p_0 des Schallfeldes entlang der begrenzenden Oberfläche $\partial\Omega$. Um ein Schallfeld innerhalb eines Volumens zu reproduzieren, muss daher die Aufnahme von Druckgradient und Druck entlang der begrenzenden Oberfläche mittels Schnelle- und Druckmikrofonen erfolgen. Die Wiedergabe erfolgt durch Lautsprecher, die eine Monopol- bzw. Dipolquellencharakteristik besitzen.

Beim BSC-Ansatz handelt es sich um einen regelungstechnischen Ansatz. Es wird versucht das Schallfeld eines beliebigen Ausschnitt des Primärraumes in einem anderen Raum zu reproduzieren, d.h. die Eigenschaften in jedem Raumpunkt des Aufnahmebereichs werden den Raumpunkten des Wiedergaberaumes aufgeprägt. Erfolgt die Wiedergabe – wie meist – nicht in einem total absorbierenden (schalltoten) Raum, so müssen die störenden Einflüsse der Reflexionen des Wiedergaberaumes kompensiert werden. Dazu wird das reproduzierte Schallfeld mit Mikrofonen räumlich abgetastet und mit den korrespondierenden Aufnahmesignalen verglichen. Die Lautsprechersignale werden in Abhängigkeit von der Differenz zwischen Aufnahme- und Sensorsignalen so gesteuert, dass die Differenz möglichst verschwindet. Dadurch kann eine eindeutige Nachbildung erreicht werden. In [Ise, 1999] und [Takane, et al. 1999] finden sich Systemansätze und Simulationen. Da es sich bei der Aufnahme und Wiedergabe um eine räumliche Abtastung handelt, kann räumliches Aliasing nur verhindert werden, wenn der Abstand zwischen den Aufnahme- bzw. Sensormikrofonen klein gewählt wird (vgl. Kap.3.1) oder der übertragene Frequenzbereich begrenzt wird. Der technische Aufwand bei der praktischen Umsetzung dieser Methode ist jedoch enorm und

steigt mit der Größe des Wiedergabebereiches und der angestrebten Genauigkeit und Grenzfrequenz der Reproduktion.

Zusammenfassung

Zu den Vorteilen des Holophonieverfahrens (BSC) zählt die relativ einfache Aufnahme mittels Schnelle- und Druckmikrofonen. Für eine exakte Abbildung ist jedoch in Abhängigkeit von der Größe des abzubildenden Volumens ein entsprechend hoher Aufwand notwendig, um räumliches Aliasing zu vermeiden. Einen weiteren Nachteil des Verfahrens bildet die praktische Realisierung der Monopol- bzw. Dipolquellen. Durch die regelungstechnisch unterstützte Wiedergabe können Fehler der Richtcharakteristik weitgehend kompensiert werden, allerdings müssen Aufnahme- und Sensorsignale nicht nur an der Oberfläche, sondern auch innerhalb des Volumens vorgesehen werden.

2.2 Ambisonic

Ambisonic bietet die Möglichkeit zur Aufnahme und Wiedergabe eines periphones¹ Schallfeldes. Es handelt sich dabei um ein skalierbares mehrkanaliges Aufnahme- und Wiedergabesystem, dessen Kodierungs- und Dekodierungsvorschrift bereits 1970 von Gerzon (vgl. Kap. 1.2) hergeleitet wurde. Für die Aufnahme realer Schallfelder wurde ein spezielles Mikrofon² verwendet, dessen Aufbau durch eine geeignet kombinierte, koinzidente Anordnung von Druckmikrofonen beschrieben werden kann. Die Wiedergabe erfolgte über ein Lautsprecherarray, welches um den Wiedergabebereich herum entlang einer Kreislinie (im 2 dimensionalen Fall) oder entlang einer Kugeloberfläche (im 3 dimensionalen Fall) aufgestellt wurde.

Die ersten Entwürfe und Implementierungen von Coder und Decoder verwendeten nur für Ambisonic 1. Ordnung. Dieses Aufnahme- und Wiedergabesystem kodiert das Schallfeld in 4 Übertragungskanäle besteht und wurde im Studiobereich unter dem Namen *B-Format* bekannt [Gerzon, 1992f]. Der visionäre Ansatz war seiner vom Stereoformat dominierten Zeit weit voraus, konnte sich jedoch am Markt aufgrund des höheren Speicheraufwands, der größeren Bandbreite und des fehlenden mehrkanaligen Tonträger nicht durchsetzen.

Im folgenden wird der allgemeine Ambisonic-Ansatz dargestellt und die Vorschriften für die Kodierung und Dekodierung für Systeme beliebiger Ordnung zusammengefasst.

¹ Der Begriff „Periphony“ wurde durch Gerzon geprägt und bedeutet eine 3-dimensionale Aufnahme bzw. Wiedergabe eines realen oder virtuellen Schallfeldes.

² Das Ambisonic-Mikrofon (MKV) wird von der Firma SoundField (<http://www.soundfield.com>) hergestellt und wurde in Zusammenarbeit mit Micheal Gerzon entwickelt. Dieses Mikrofon eignet sich allerdings nur zur Aufnahme der Ambisonicsignale bis zur 1. Ordnung.

2.2.1 Kodierungs- und Dekodierungsvorschriften für Ambisonic

Die Idee hinter dem Ambisonic-Modell basiert auf dem Vergleich einer aufgenommenen Referenzwelle mit der reproduzierten Welle des Systems [Bamford, 1995]. Dabei werden folgende Annahmen vereinbart:

Die aufgenommene Wellenfront (Referenzwelle) ist eine ebene Welle.

Diese Annahme ist gültig, solange die Schallquelle weit genug vom Hörer entfernt ist. Diese Annahme stellt keine Einschränkung der Allgemeinheit dar, da jede beliebige Wellenfront als Summe ebener Wellen dargestellt werden kann.

Die reproduzierte Welle eines Lautsprechers ist eine ebene Welle.

Auch diese Annahme ist gültig, solange die Lautsprecher weit genug vom Zuhörer entfernt sind.

Die sich Gesamtwelle des Ambisonic-Systems ergibt sich durch Superposition der abgestrahlten Teilwellen der einzelnen Lautsprecher. Sie ist wiederum eine ebene Welle. Der Ambisonic-Ansatz beschränkt sich auf die Analyse des Schallfeldes, das durch den Direktschall entsteht, d.h. es wird ausschließlich die Freifeldsituation betrachtet, evt. vorhandene Reflexionen werden vernachlässigt.

Auf Basis der Gleichsetzung von Referenz- und Systemwelle wird eine Kodierungsvorschrift für die Aufnahme und eine Dekodierungsvorschrift für die Wiedergabe abgeleitet. Die Aufnahme realer Schallfeldern erfolgt durch Mikrofone mit speziellen Richtcharakteristiken, die jedoch nur begrenzt zur Verfügung stehen. Virtuelle Schallfelder werden durch geeignete Signalverarbeitung kodiert. Die Wiedergabe erfolgt über ein Lautsprecherarray, das idealerweise symmetrisch auf der genannten Kugeloberfläche (für den 3D Fall) bzw. der Kreislinie (für den 2D Fall) verteilt ist.

In Abbildung 2.3 sind die relevanten geometrischen Größen für die Beschreibung der Kodierungs- bzw. Dekodierungsvorschriften dargestellt. Der normierte Vektor $\hat{\mathbf{y}} = [\sin \vartheta \cdot \cos \varphi, \sin \vartheta \cdot \sin \varphi, \cos \vartheta]^T$ beschreibt durch die Winkel (φ, ϑ) die Einfallrichtung der Referenzwelle $S_{Referenz}$, die gemäß Annahme als ebene Welle aufgefasst wird (siehe Gleichung 2.2). Mit dem Vektor $\mathbf{x} = x \cdot [\sin \theta \cdot \cos \phi, \sin \theta \cdot \sin \phi, \cos \theta]^T$ mit $\hat{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|$ wird ein beliebiger Beobachtungspunkt des Schallfeldes beschrieben.

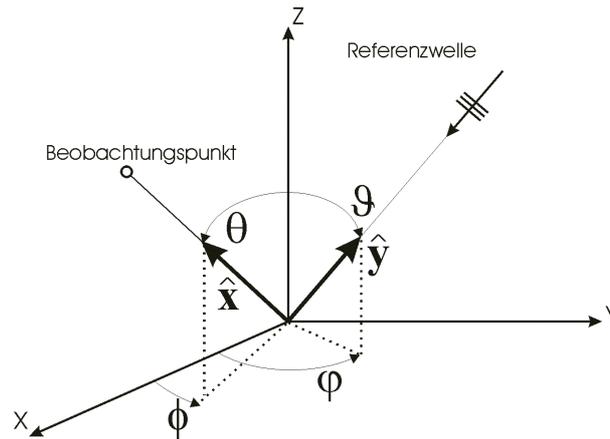


Abb. 2.3: Definition der Winkelbezeichnungen und beschreibende Größen in der 3D Darstellung des Ambisonic Systems

Erzeugt die Referenzschallquelle den Schalldruck $p_{\vartheta,\varphi}$, so kann unter Vernachlässigung der zeitlichen Abhängigkeit der Schalldruck im Punkt \mathbf{x} beschrieben werden mit:

$$S_{\text{Referenz}}(\mathbf{x}; k) = p_{\vartheta,\varphi} \cdot e^{ik\mathbf{x}(\hat{\mathbf{y}}^T \cdot \hat{\mathbf{x}})} \quad (2.2)$$

mit $i = \sqrt{-1}$ und der Wellenzahl $k = \frac{2\pi f}{c}$, die abhängig von der betrachteten Frequenz f und der Schallausbreitungsgeschwindigkeit c ist. Diese ebene Welle kann durch Kugelflächenfunktionen (*spherical harmonics*) in eine Reihe entwickelt werden, die sich bei Vernachlässigung der radialen Abhängigkeit als allgemeine Lösung der Wellengleichung (Gl. 2.3) für den 3-dimensionalen Raum in Kugelkoordinaten ergeben.

In Gleichung 2.3 ist die lineare, verlustlose, homogene Wellengleichung dargestellt. Sie beschreibt die Freifeldausbreitung in einem quellenfreien Volumen:

$$\Delta F(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} F(\mathbf{r}, t) = 0 \quad (2.3)$$

mit dem Laplace-Operator in Kugelkoordinaten

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \left[\frac{\partial^2}{\partial \theta^2} + \cot \theta \frac{\partial}{\partial \theta} \right] + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \quad (2.4)$$

Wird für die Funktion $F(\mathbf{r})$ (mit $\mathbf{r} = \mathbf{r}(r, \theta, \phi)$) eine zeitlich harmonische Größe der Form $F(\mathbf{r}, t) = F_f(\mathbf{r}) \cdot e^{j2\pi f t}$ gewählt, wobei das Subscript f die Abhängigkeit von der Frequenz repräsentiert, erhält man durch Einsetzen in Gleichung 2.3 die zeitunabhängige verlustlose Helmholtz Gleichung.

$$\Delta F(\mathbf{r}) + k^2 F(\mathbf{r}) = 0 \quad (2.5)$$

Diese Gleichung lässt sich durch die Variablenseparation mit

$$F(\mathbf{r}) = R(r) \cdot \Theta(\theta) \cdot \Phi(\phi) \quad (2.6)$$

lösen (vgl. [Ziomek, 1995], S.269 ff). Dadurch erhält man für die r-abhängige Funktion $R(r)$ folgende Lösung:

$$R(r) = A_m \cdot j_m(kr) + B_m \cdot y_m(kr) \quad m = 0, 1, 2, \dots \quad (2.7)$$

wobei j_m die sphärischen Besselfunktionen und y_m die sphärischen Neumannfunktionen m -ter Ordnung sind. Die sphärischen Besselfunktionen können wie folgt aus den Besselfunktionen J_n abgeleitet werden:

$$j_m(kr) = \sqrt{\frac{\pi}{2kr}} \cdot J_{m+1/2}(kr) \quad (2.8)$$

$$y_m(kr) = \frac{J_m(kr) \cos(m\pi) - J_{-m}(kr)}{\sin(m\pi)} \quad (2.9)$$

Da die sphärische Neumannfunktion eine Singularität bei $kr = 0$ aufweist und die Lösung endlich sein sollte, werden die Koeffizienten B_m gleich Null gesetzt. Somit ergibt sich die Lösung der radialen Variablen zu:

$$R(kr) = i^n \cdot j_n(kr) = i^n \cdot \sqrt{\frac{\pi}{2kr}} \cdot J_{n+1/2}(kr) \quad (2.10)$$

Die Lösung für die Elevation ergibt:

$$\Theta(\theta) = C_{mn} P_m^n(\cos\theta) + D_{mn} Q_m^n(\cos\theta) \quad m, n = 0, 1, 2, \dots \quad (2.11)$$

wobei $P_m^n(\cos\theta)$ die Legendre-Polynome 1. Art und $Q_m^n(\cos\theta)$ jene der 2. Art repräsentieren. Auch hier weist der zweite Term Singularitäten bei $\theta = 0$ bzw. $\theta = \pi$ auf, die Koeffizienten D_{mn} werden daher gleich Null gesetzt.

Die Lösung für den Azimut ergibt:

$$\Phi(\phi) = E_n \cos(n\phi) + F_n \sin(n\phi) \quad n = 0, 1, 2, \dots \quad (2.12)$$

Da beim Ambisonic-Ansatz die Modellierung durch ebene Welle erfolgt (siehe Annahmen) und nur die Situation im Koordinatenursprung betrachtet wird, ist die Lösung von der radialen Komponente unabhängig und kann somit vernachlässigt werden. Zieht man die verbleibenden Lösung von Elevation und Azimut zusammen, ergeben sich die Kugelflächenfunktionen $Y_{nm}(\mathbf{r})$.

$$F(\mathbf{r}) \cong \Theta(\theta) \cdot \Phi(\phi) = Y_{nm}(\mathbf{r}) \quad m, n = 0, 1, 2, \dots \quad (2.13)$$

Die Kugelflächenfunktionen $Y_{nm}(\mathbf{r})$ bilden ein Set von orthogonalen Basisvektoren. Durch die entsprechende Wahl von A_{nm} wird diese Orthogonalbasis normiert.

$$Y_{nm}(\mathbf{r}) = A_{nm} \cdot P_m^n(\cos \theta) \cdot e^{i \cdot n \phi} \quad (2.14)$$

$$\text{mit } A_{nm} = \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}}$$

Das Subskript n beim Legendre-Polynom 1. Art bezieht sich auf die Ordnung der Kugelflächenfunktion, das Subskript m zeigt den Mode an. Bei $m = 0$ reduziert sich das Legendre-Polynom auf die Legendre Funktion. Für jede Ordnung n existieren $2n+1$ Moden. Die entstehende Orthonormalbasis genügt der Eigenschaft (Orthogonalitätsrelation):

$$\oint Y_{nm}^*(\hat{\mathbf{r}}) \cdot Y_{pq}(\hat{\mathbf{r}}) \cdot d\hat{\mathbf{r}} = \delta_{np} \cdot \delta_{mq} \quad (2.15)$$

$$\text{mit der Kronecker delta Funktion } \delta_{np} = \begin{cases} 1 & \text{für } n = p \\ 0 & \text{für } n \neq p \end{cases}$$

Mit Hilfe der Kugelflächenfunktionen kann jede beliebige Funktion auf einer Kugeloberfläche in eine Reihe entwickelt werden. Damit wird die Referenzwelle aus Gleichung 2.2 wie folgt entwickelt:

$$S_{\text{Referenz}}(\mathbf{x}; k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n G_{nm}(x; k) \cdot Y_{nm}(\hat{\mathbf{x}}) \quad (2.16)$$

Durch diese Schallfeldrepräsentation ist bereits die Vorschrift für die Kodierung und für die Dekodierung implizit gegeben. Die Koeffizienten G_{nm} sind die Gewichte der einzelnen Kugelflächenfunktionen; sie sind vom Einfallswinkel unabhängig. Die Kugelflächenfunktionen beinhalten die Richtungsinformationen der abzubildenden (Referenz-)Quelle.

Die Kugelflächenfunktionen beschreiben die notwendigen Richtcharakteristiken für die Aufnahmemikrofone und werden im folgenden als Ambisonic-Signale bezeichnet. In Tabelle 2.1 sind diese richtungsabhängigen Terme bis zur 4. Ordnung zusammengefasst. Die Bezeichnung der Kugelflächenfunktionen (Ambisonic-Signale) wird durch das Superskript $\sigma \in [-1, +1]$ erweitert und symbolisiert folgende Lösungsschar (vgl. Gl. 2.14):

$$Y_{nm}^{\sigma}(\mathbf{r}) = \begin{cases} A_{nm} \cdot P_m^n(\cos \theta) \cdot \cos m \phi & \text{für } \sigma = 1 \\ A_{nm} \cdot P_m^n(\cos \theta) \cdot \sin m \phi & \text{für } \sigma = -1 \end{cases} \quad (2.17)$$

M	n	m	$Y_{n,m}^{+1}$	$Y_{n,m}^{-1}$
0	0	0	1	0
1	1	0	$\cos \theta$	0
	1	1	$\sin \theta \cos \phi$	$\sin \theta \sin \phi$
2	2	0	$\frac{1}{2}(3 \cos^2 \theta - 1)$	0
	2	1	$\sqrt{3}/2 \cdot \sin 2\theta \cos \phi$	$\sqrt{3}/2 \cdot \sin 2\theta \sin \phi$
	2	2	$\sqrt{3}/2 \cdot \sin^2 \theta \cos 2\phi$	$\sqrt{3}/2 \cdot \sin^2 \theta \sin 2\phi$
3	3	0	$\frac{1}{2}(5 \cos^3 \theta - 3 \cos \theta)$	0
	3	1	$\sqrt{6}/4 \cdot \sin \theta (5 \cos^2 \theta - 1) \cos \phi$	$\sqrt{6}/4 \cdot \sin \theta (5 \cos^2 \theta - 1) \sin \phi$
	3	2	$\sqrt{15}/2 \cdot \sin^2 \theta \cos \theta \cos 2\phi$	$\sqrt{15}/2 \cdot \sin^2 \theta \cos \theta \sin 2\phi$
	3	3	$\sqrt{10}/4 \cdot \sin^3 \theta \cos 3\phi$	$\sqrt{10}/4 \cdot \sin^3 \theta \sin 3\phi$
4	4	0	$\frac{1}{8}(35 \cos^4 \theta - 30 \cos^2 \theta + 3)$	0
	4	1	$\sqrt{10}/4 \cdot \sin \theta (7 \cos^3 \theta - 3 \cos \theta) \cos \phi$	$\sqrt{10}/4 \cdot \sin \theta (7 \cos^3 \theta - 3 \cos \theta) \sin \phi$
	4	2	$\sqrt{5}/4 \cdot \sin^2 \theta (7 \cos^2 \theta - 1) \cos 2\phi$	$\sqrt{5}/4 \cdot \sin^2 \theta (7 \cos^2 \theta - 1) \sin 2\phi$
	4	3	$\sqrt{70}/4 \cdot \sin^3 \theta \cos \theta \cos 3\phi$	$\sqrt{70}/4 \cdot \sin^3 \theta \cos \theta \sin 3\phi$
	4	4	$\sqrt{35}/8 \cdot \sin^4 \theta \cos 4\phi$	$\sqrt{35}/8 \cdot \sin^4 \theta \sin 4\phi$
usw.				

Tabelle 2.1: Halbnormalisierte Ambisonic-Koeffizienten (Kugelflächenfunktionen) bis zur 4. Ordnung

Die Richtcharakteristiken sind für Systeme bis zur 3. Ordnung in der Abb. 2.4 dargestellt.

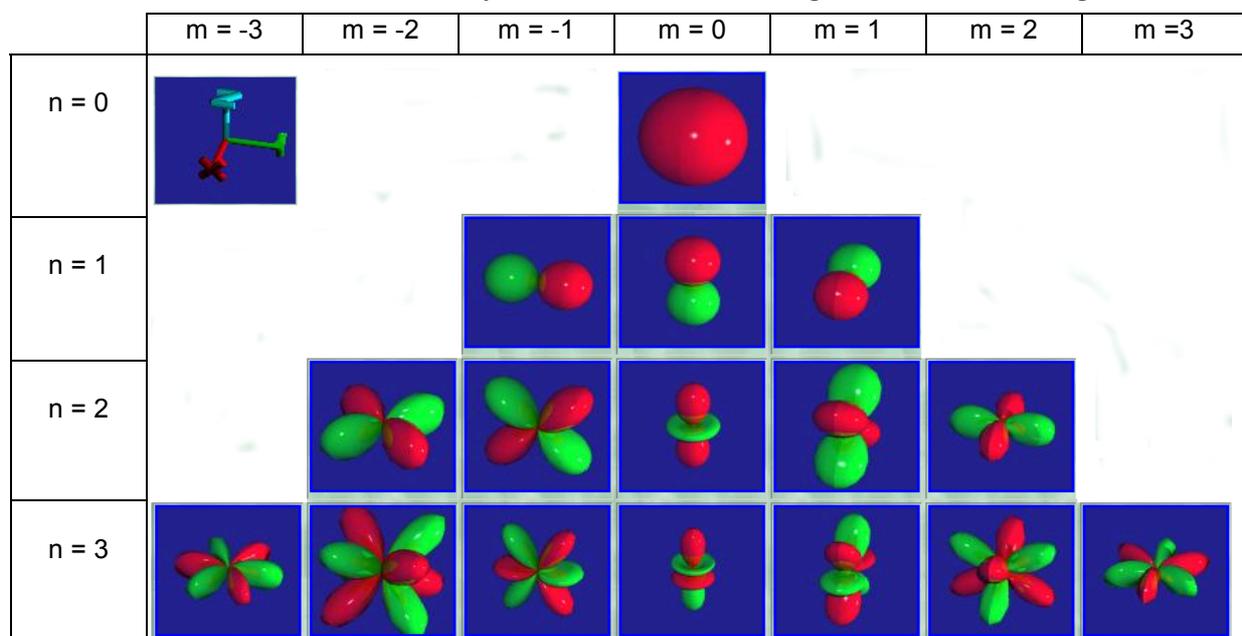


Abb. 2.4 Richtcharakteristiken $Y_{nm}(\cos \theta)$ für die Aufnahmemikrofone
 [vgl. <http://www.uniovi.es/~quimica.fisica/qcg/harmonics/harmonics.html>].

Unter Verwendung der Eigenschaften von Gl. 2.15 kann man die Koeffizienten G_{nm} wie folgt bestimmen

$$G_{nm}(x; k) = \oint Y_{nm}^*(\mathbf{y}) \cdot S_{\text{Referenz}}(\mathbf{x}; k) d\hat{\mathbf{x}} \quad (2.18)$$

Das Referenzfeld kann im Beobachtungspunkt \mathbf{x} wie folgt beschrieben werden

$$S_{\text{Referenz}}(\mathbf{x}; k) = p_{\vartheta, \varphi} \cdot \sum_{n=0}^{\infty} \sum_{m=-n}^n 4\pi \cdot R(kx) \cdot Y_{nm}^*(\hat{\mathbf{y}}_l) \cdot Y_{nm}(\hat{\mathbf{x}}) \quad (2.19)$$

und ergibt sich im Ursprung ($\mathbf{x} = \mathbf{0}$) als Beobachtungspunkt zu

$$S_{\text{Referenz}} = p_{\vartheta, \varphi} \cdot \sum_{n=0}^{\infty} \sum_{m=-n}^n Y_{nm}^*(\hat{\mathbf{y}}_l) \quad (2.20)$$

Somit kann eine ebene Welle aus einer Richtung $[\Phi, \Theta]$ durch unendlich viele eindimensionale Ambisonic-Signale vollständig repräsentiert und synthetisiert werden.

KODIERUNGSVORSCHRIFT

$$B_{m,n}^{\sigma} = Y_{mn}^{\sigma}(\Phi, \Theta) \cdot p_{\Phi, \Theta} \quad (2.21)$$

Die einzelnen Ambisonic-Signale (Gl. 2.21) lassen sich durch die Multiplikation eines Mono-Signals ($p_{\vartheta, \varphi}$) mit dem Funktionswert der Kugelflächenfunktion ($Y_{mn}^{\sigma}(\Phi, \Theta)$) für die Richtung $[\Phi, \Theta]$, aus der die Quelle die ebene Welle abstrahlt, darstellen.

Da die Übertragung von unendlich vielen Kanälen nicht möglich ist, wird die Reihenentwicklung nach einem bestimmten Glied abgebrochen, und es werden nur N Übertragungskanäle verwendet. Die Ordnung M der Kugelflächenfunktionen, nach der abgebrochen wird, bezeichnet man daher auch als Ordnung des Ambisonic-Systems.

Da die Kugelflächenfunktionen $Y_{m,0}^{-1} \equiv 0 \quad \forall m$ sind, ergibt sich die Anzahl N der benötigten Ambisonic-Signale für ein System der Ordnung M:

Für den dreidimensionalen Fall

$$N^{3D} = (M + 1)^2 \quad (2.22)$$

und für den zweidimensionalen Fall (d.h. $\theta = 0$)

$$N^{2D} = 2M + 1 \quad (2.23)$$

Die Bezeichnung der einzelnen Ambisonic-Signale erfolgt traditioneller Weise in Buchstaben des Alphabetes (siehe Gl. 2.24) bzw. nach entsprechend eigenen definierten Konventionen (vgl. Kap. 4).

$$\mathbf{B} = \begin{pmatrix} B_{0,0} \\ B_{1,1}^1 \\ B_{1,1}^{-1} \\ B_{1,0}^1 \\ \vdots \end{pmatrix} = \begin{pmatrix} W \\ X \\ Y \\ Z \\ \vdots \end{pmatrix} \quad (2.24)$$

Die Realisierung bis zur ersten Ordnung hat den Begriff Ambisonic geprägt. Dieses Set von Ambisonic-Signalen trägt die Bezeichnung B-Format. Die in der Literatur angeführten Kodierungsvorschriften [Gerzon, 1992f], [Bamford, 1995], [Jot et al., 1999], [Furse, 1999]¹, [Daniel, 2000] etc. unterscheiden sich durch verschiedene multiplikative Faktoren für jeden Kanal. Der W Kanal ($B_{0,0}$) wird meist nicht mit 1 sondern aus Gründen der ausgewogeneren Leistungsverteilung über die einzelnen Kanäle mit $1/\sqrt{2}$ gewichtet.

Das Syntheseschallfeld, das sich durch die Überlagerung der einzelnen Teilschallfelder der L Lautsprecher an den Positionen $\hat{\mathbf{y}}_l = [\sin \vartheta_l \cdot \cos \varphi_l, \sin \vartheta_l \cdot \sin \varphi_l, \cos \vartheta_l]^T$ ergibt, kann wie folgt beschrieben werden

$$S_{\text{Synthese}}(\mathbf{x}; k) = \sum_{l=1}^L p_l \cdot e^{ik\mathbf{x} \cdot \hat{\mathbf{y}}_l^T} \quad (2.25)$$

Entwickelt man das Syntheseschallfeld ebenfalls in Reihe von Kugelflächenfunktionen, erhält man

$$S_{\text{Synthese}}(\mathbf{x}; k) = \sum_{l=1}^L p_l \sum_{n=0}^{\infty} \sum_{m=-n}^n 4\pi \cdot R(kx) \cdot Y_{nm}^*(\hat{\mathbf{y}}_l) \cdot Y_{nm}(\hat{\mathbf{x}}) \quad (2.26)$$

Im Betrachtungspunkt ($\mathbf{x}=\mathbf{0}$) soll nun das Schallfeld der Referenzwelle (Gl. 2.20) durch die Superposition der Teilschallfelder der Lautsprecher (Gl. 2.26) resynthetisiert werden.

$$S_{\text{Referenz}} \equiv S_{\text{Synthese}} \quad (2.27)$$

Identität zwischen den beiden Schallfelder kann durch einen Koeffizientenvergleich der beiden Reihenentwicklungen erreicht werden. Dieser Koeffizientenvergleich führt zu den sogenannten „*matching conditions*“ in kompakter Notation.

$$p_{\vartheta, \varphi} \cdot Y_{mn}^{\sigma}(\hat{\mathbf{y}}) = \sum_{l=1}^L p_l \cdot Y_{mn}^{\sigma}(\hat{\mathbf{y}}_l) \quad (2.28a)$$

¹ Eine Auflistung der Kodierungsvorschriften bis zur 3. Ordnung, kombinierter Ordnungen sowie Decoder für unterschiedliche Lautsprecherlayouts finden sich unter : <http://www.muse.demon.co.uk/ref/speakers.html> .

und in ausgeschriebener Form:

$$\begin{aligned}
 p_{\vartheta,\varphi} &= \sum_{l=1}^L p_l \\
 p_{\vartheta,\varphi} \cdot P_{mn}(\cos \vartheta) \cdot \cos(n\varphi) &= \sum_{l=1}^L p_l \cdot P_{mn}(\cos \vartheta_l) \cdot \cos(n\varphi_l) \\
 p_{\vartheta,\varphi} \cdot P_{mn}(\cos \vartheta) \cdot \sin(n\varphi) &= \sum_{l=1}^L p_l \cdot P_{mn}(\cos \vartheta_l) \cdot \sin(n\varphi_l)
 \end{aligned} \tag{2.28b}$$

wobei für $m = 1, \dots, \infty$ und $n = 1, \dots, m$ gilt.

Da bei der Aufnahme nur eine endliche Anzahl N von Ambisonic-Signalen zur Verfügung stehen, reduzieren sich die ‘‘matching conditions’’ auf ein endliches Gleichungssystem:

$$\mathbf{B} = \mathbf{C} \cdot \mathbf{p} \tag{2.29}$$

wobei $\mathbf{p} = [p_1, p_2, \dots, p_L]^T$

$$\text{und } \mathbf{C} = \begin{bmatrix} Y_{0,0}^1(\hat{\mathbf{y}}_1) & Y_{0,0}^1(\hat{\mathbf{y}}_2) & \dots & Y_{0,0}^1(\hat{\mathbf{y}}_L) \\ Y_{1,0}^1(\hat{\mathbf{y}}_1) & Y_{1,0}^1(\hat{\mathbf{y}}_2) & \dots & Y_{1,0}^1(\hat{\mathbf{y}}_L) \\ \vdots & \vdots & \ddots & \vdots \\ Y_{M,M}^{-1}(\hat{\mathbf{y}}_1) & Y_{M,M}^{-1}(\hat{\mathbf{y}}_2) & \dots & Y_{M,M}^{-1}(\hat{\mathbf{y}}_L) \end{bmatrix}$$

Die Lautsprechersignale p_l ergeben sich durch Berechnung der Decodermatrix \mathbf{M}_d wie folgt:

$$\mathbf{p} = \mathbf{M}_d \cdot \mathbf{B} = \mathbf{C}^{-1} \cdot \mathbf{B} \tag{2.30}$$

Da im Allgemeinen $N \neq L$ ist, kann die Dekodiermatrix \mathbf{M}_d nur mit Hilfe der Pseudoinversen der Enkodiermatrix \mathbf{C} berechnet werden.

DEKODIERUNGSVORSCHRIFT

Bei der Dekodierung erfolgt eine Verteilung der N Ambisonic-Signale auf L Lautsprecher. Die Gewichtung der einzelnen Signale bei der Dekodierung wird nur von der Aufstellung der Lautsprecher und der Anzahl N (bzw. der Ordnung M des Ambisonic-Systems) bestimmt: Die Dekodierung ist somit von der Kodierung entkoppelt.

Der Vorschrift für den ‘‘Basic’’-Dekoder¹ ergibt sich aus der Lautsprecheranordnung wie folgt:

$$\mathbf{M}_{dBasic} = \text{pinv}(\mathbf{C}) = \mathbf{C}^T \cdot (\mathbf{C} \cdot \mathbf{C}^T)^{-1} \tag{2.31}$$

¹ Für andere Decoder-Vorschriften siehe Kap. 2.2.2

Um keine in den Ambisonic-Kanälen enthaltene (Richtungs-)Information bei der Dekodierung zu verlieren, muss Gl. (2.30) überbestimmt sein. Daraus folgt:

$$L \geq N \quad (2.32)$$

wobei L die Anzahl der Lautsprecher und N die Anzahl der Ambisonic-Kanäle repräsentiert.

Eine optimale Dekodierung ergibt sich, wenn N Ambisonic-Signale auf die $L=N$ Lautsprecher dekodiert werden. Beweise dafür finden sich bei [Gerzon, 1992a], bzw. [Poletti, 1996] und [Daniel, 2000]. Erfolgt jedoch eine frequenzabhängige Dekodierung, so gibt Daniel in [Daniel, 2000] bei einer Trennung in einen tieffrequenten ($<400\text{Hz}$) und einen hochfrequenten ($>700\text{Hz}$) Decoder (Werte für die Frequenzbereiche vgl. [Gerzon,1992h]) für den optimalen hochfrequenten Decoder $L \geq 2M + 2$ an.

Die Dekodierung wird nur von der Wahl der Lautsprecheraufstellung bestimmt. Eine Einschränkung bildet jedoch die eindeutige Lösbarkeit der Pseudoinversen für die gewählte Lautsprecheraufstellung. Die Pseudoinverse der Matrix \mathbf{C} liefert nur dann praktisch brauchbare Werte, wenn die Matrix gutartig konditioniert (kleine Konditionszahl)¹ ist.

Reguläre Lautsprecheranordnung

Für eine zwei dimensionale reguläre Lautsprecher Anordnung ($\varphi_n = 2\pi n/N$ mit $n = 0, \dots, N-1$) gilt:

$$\mathbf{C} \cdot \mathbf{C}^T = \mathbf{D} \quad (2.33)$$

wobei die Matrix \mathbf{D} eine Diagonalmatrix ist, die wie folgt für ein Ambisonic System M . Ordnung aufgebaut ist:

$$\begin{aligned} \mathbf{D} &= \text{diag}(d_0, d_1, \dots, d_{2M}) \\ &= \text{diag}\left(N, \sum_{n=0}^{N-1} \cos^2 \varphi_n, \sum_{n=0}^{N-1} \sin^2 \varphi_n, \dots, \sum_{n=0}^{N-1} \cos^2 (M\varphi_n), \sum_{n=0}^{N-1} \sin^2 (M\varphi_n)\right) \end{aligned} \quad (2.34)$$

Eine äquidistante Aufstellung der Lautsprecher entlang eines Kreises (regelmäßiges Polygon) oder einer Kugel (reguläre Polyeder) vereinfacht somit die Lösung der Dekodiergleichung zu:

$$\mathbf{M}_{\mathbf{d} \text{ Polyeder}} = \frac{1}{N} \cdot \mathbf{C}^T \quad (2.35)$$

¹ Die Konditionszahl einer Matrix berechnet sich aus dem Verhältnis des maximalen zum minimalen Eigenwert

Somit ergibt sich das Lautsprecher-signal p_i in Abhängigkeit von der Quelleneinfallrichtung $(\varphi, 90^\circ)$ für einen Lautsprecher an der Position $(\varphi_i, 90^\circ)$ für eine regelmäßige 2D Anordnung bei Verwendung eines Ambisonic Systems M. Ordnung:

$$\begin{aligned}
 p_i(\varphi, 90^\circ) &= \mathbf{M}_{\text{dPolyeder}} \cdot \mathbf{B}(\varphi, 90^\circ) = \\
 &= \frac{1}{N} \cdot [B_{0,0} + B_{1,1}^1(\varphi) \cdot \cos(\varphi_i) + B_{1,1}^{-1}(\varphi) \cdot \sin(\varphi_i) + B_{2,2}^1(\varphi) \cdot \cos(2\varphi_i) + B_{2,2}^{-1}(\varphi) \cdot \sin(2\varphi_i) + \dots] \\
 &= \frac{1}{N} \cdot \left[1 + 2 \cdot \sum_{m=1}^M \cos[m \cdot (\varphi - \varphi_i)] \right] \quad (2.36)
 \end{aligned}$$

Bei periphonen Systemen ist das Auffinden von optimalen Lösungen ungleich schwieriger. Verwendet man als Anordnung eine regulären Polyeder, so ist die Anzahl der Lautsprecher vorgegeben. Sollte die Anzahl jedoch frei wählbar sein, so muss eine Anordnung (Matrix) mit möglichst kleiner Konditionszahl gefunden werden. Die Varianz der Eigenwerte und damit verbunden die Konditionszahl kann durch Minimierung der maximalen Abstände bzw. der Raumwinkel zwischen den Lautsprechern reduziert werden¹.

Irreguläre Lautsprecheranordnung

Die Anwendung der „Basic“- Dekodiermatrix auf ein unregelmäßiges Lautsprecherlayout wird *Projektion* der Ambisonic-Kanäle auf die jeweiligen Lautsprecher genannt. Das Auffinden einer geeigneten Decodermatrix stellt auch für den zweidimensionalen Fall eine nicht triviale Aufgabe dar. Optimierte Decodervorschriften für einige standardisierte irreguläre Lautsprecherlayouts, wie z.B. für das 5.1 Format, gibt es bei [Gerzon, 1992f] und [Trebuchet, 1997] zu finden.

2.2.2 Optimierungskriterien für Dekodierungsvorschriften

Im folgenden werden Dekodierungsvorschriften für unterschiedliche Optimierungskriterien dargestellt. Eine explizite Untersuchung der Abbildungseigenschaften der ersten vier Vorschriften findet sich bei [Daniel, 2000].

- Basic (Max r_V)
- Max r_E
- In-Phase
- Super-Minimal
- Window-Applied

¹ Unter <http://www.research.att.com/~njas/packings/> finden sich Lösungen optimaler Verteilungen für unterschiedliche Punkte- (Lautsprecher-) Anzahlen.

Der „Basic“-Dekoder wurde bereits oben exemplarisch vorgestellt. Er eignet sich am besten für tiefe Frequenzen ($f < 700\text{Hz}$) und zentrale Hörpositionen. Durch die Anwendung dieser Dekodervorschrift wird sowohl für Systeme 1. Ordnung als auch für höhere Ordnungen mit regelmäßigen Lautsprecheranordnungen der Schnellevektor r_v (vgl. Kap. 1.2.2) der synthetisierten Schallquelle maximiert (daher auch die Bezeichnung $\text{Max } r_v$). Das Optimierungskriterium basiert auf der psychoakustischen Annahme, dass die Lokalisation für tiefe Frequenzen aufgrund der interauralen Phasendifferenz (infolge ITD vgl. Kap. 1) erfolgt. Es lässt sich zeigen [Makita, 1962], [Gerzon, 1992d,g], [Daniel et al., 1998], dass durch $|\vec{r}_v| \neq 1$ die Ausbreitungsgeschwindigkeit c' in Richtung \vec{r}_v von der allgemeinen Ausbreitungsgeschwindigkeit c verschieden ist. Dadurch kommt es zu einer zusätzlichen Phasenverschiebung (*phase lag*) und damit verbunden zur einer verfälschten ITD, die zu einer Fehllokalisierung der abgebildeten Quelle führt (Abbildungswinkel \neq Wahrnehmungswinkel). Ist der Betrag des Schnellektors unabhängig vom Winkel gleich 1, so stimmen Abbildungs- und Wahrnehmungswinkel überein und die Abbildungsrichtung bleibt auch bei Drehung des Kopfes stabil. Bei Werten kleiner eins nähert sich die wahrgenommene Richtung unter Kopfdrehung der Medianebene. Diese Änderung kann auch als Zunahme der Elevation interpretiert werden. Für Werte größer eins wandert die wahrgenommene Richtung in Richtung der Frontalebene (stärkere Lateralisation).

Die Herleitung der Dekodierungsvorschriften für den „Max r_E “- und den „In-Phase“-Decoder findet sich bei [Daniel et al., 1998], [Daniel, 2000] bzw. bei [Furse, 1999]. Beide Decoder werden durch eine Modifikation der Basis Decodermatrix D_{Basic} (vgl. Gl. 2.31) durch Korrekturgewichte realisiert.

Die Modifikation des Basic Decoders zu einem $\text{Max } r_E$ Decoder ergibt sich durch die Berechnung der Korrekturterme g_m wie folgt:

$$\mathbf{M}_{d_{\text{Max } r_E}}(\{g_m\}) = \mathbf{M}_{d_{Basic}} \cdot \text{diag}(g_0, g_1, g_1, \dots, g_M, g_M) \quad (2.37)$$

Als Beispiel für ein System 2. Ordnung schreibt sich Gl. 2.30 wie folgt nun:

$$\begin{pmatrix} P_1 \\ \vdots \\ P_N \end{pmatrix} = \mathbf{M}_{d_{Basic}} \cdot \begin{pmatrix} g_0 \cdot W \\ g_1 \cdot X \\ g_1 \cdot Y \\ g_2 \cdot U \\ g_2 \cdot V \end{pmatrix} \quad (2.38)$$

Der Betrag des Energievektors berechnet sich für ein 2D Ambisonic System M. Ordnung zu:

$$r_E = \frac{2 \sum_{m=1}^M g_m \cdot g_{m-1}}{g_0^2 + 2 \sum_{m=1}^M g_m^2} \quad (2.39)$$

Die Korrekturgewichte g_m können durch partielles ableiten der Gl.2.39 bestimmt werden. Für eine reguläre Anordnung (siehe Kap. 2.2.1) ist der maximal erreichbare Betrag des Energievektors begrenzt mit:

$$r_E^{\max} = \cos \frac{\pi}{2M + 2} \quad (2.40)$$

Somit ergeben sich die optimalen Gewichte zu:

$$g_m = \Gamma \cdot \cos \frac{m\pi}{2M + 2} \quad m = 0, 1, \dots, M \quad (2.41)$$

wobei $\Gamma = g_0$ ein positiver Wert ist, der zur Normierung der Decodervorschrift verwendet wird. Mit der Wahl $\Gamma = 1$ wird auf die Signalamplitude bzw. dem Druck normiert. Durch die Wahl $\Gamma = \sqrt{N/(M+1)}$ wird auf die Gesamtenergie normiert.

Der „Max r_E “-Decoder versucht die Länge des reproduzierten Energievektor (vgl. auch Kap.1.2.2) zu maximieren und ist daher am besten für hohe Frequenzen bei zentraler Hörposition geeignet. Der Betrag des Energievektors $|\vec{r}_E|$ beschreibt den Energieanteil aus der Richtung der abzubildenden Quelle und ist immer kleiner (gleich¹) eins. Ähnlich wie beim Schnellektor ergibt sich bei zu kleinen Werten eine zu geringe Lateralisation bzw. es treten Elevationswahrnehmungen auf. Im allgemeinen führt ein zu kleiner Betrag des Energievektors auch zu einer Zunahme der Lokalisationsunschärfe (*lokalisierung blur*, vgl. [Blauert, 1974] bzw. Kap1.1.1).

In großen Auditorien, wo sich viele Hörer außerhalb des Referenzabhörorts befinden und daher die einzelnen Lautsprecher in sehr unterschiedlicher Entfernung vom Hörer stehen, würde die Verwendung des „Basic“-Decoders zu Fehllokalisierungen auf Grund des Precedence-Effekts kommen, da auch Lautsprecher, die nicht in Quellrichtung liegen, zum Teil erhebliche, zueinander gegenphasige Signale abstrahlen. In solchen Fällen eignet sich der „In-Phase“-Decoder, bei dem gegenphasige Anteile aus den nicht in Quellrichtung liegenden

¹ Eine Gleichheit liegt nur dann vor, wenn ausschließlich ein Lautsprecher das gesamte Quellsignal wiedergibt. Der Maximalwert, der bei einem Ambisonic System erreicht werden kann, hängt von der realisierten Ordnung des Systems ab siehe Gl.2.39.

Lautsprechern unterdrückt werden. Damit sind die L Lautsprechersignale p_i für alle Quellenrichtungen immer positiv (inkl. Null).

$$p_i(\varphi) = g_0 + 2 \sum_{m=1}^M g_m \cos(m\varphi) \quad \forall i, \varphi \geq 0 \quad (2.42)$$

Die Berechnung der Korrekturgewichte g_m für den *In phase* Decoder ergibt sich durch die Bedingung der Gl. 2.42 und wird erfüllt wenn gilt:

$$\left. \frac{d^{2n} p_i}{d\varphi^{2n}} \right|_{\varphi = \pi} = 0 \quad \text{für } n = 0, \dots, M-1 \quad (2.43)$$

$$\frac{d^{2n} p_i}{d\varphi^{2n}}(\varphi) = 2 \sum_{m=1}^M g_m (-1)^n m^{2n} \cos(m\varphi) \quad \text{wenn } n > 1 \quad (2.44)$$

Die Korrekturgewichte $g'_m = g_m/g_0$ berechnen sich im 2D Fall zu :

$$g'_m{}^{2D} = \frac{(M!)^2}{(M+m)!(M-m)!} \quad (2.45)$$

mit $g_0 = N/\sqrt{M+1}$, und im 3D Fall zu:

$$g'_m{}^{3D} = \frac{M!(M+1)!}{(M+m+1)!(M-m)!} \quad (2.46)$$

mit $g_0 = \sqrt{N(2M+1)/(M+1)^2}$ erfolgt die Normierung auf die Gesamtenergie.

Sollte ein Decoder sowohl für den Schnellektor als auch für den Energievektor optimal ausgelegt werden, so müssen die Gewichtungsfaktoren der Lautsprechersignale frequenzabhängig sein. In der Praxis kann dies durch phasenkompensierte Shelving-Filter realisiert werden. Der Übergang zwischen den beiden Decodertypen sollte möglichst glatt erfolgen. Ein Richtwert für die Übergangsfrequenz ist 700 Hz [Gerzon, 1992f-h].

In [Poletti, 1996a] erfolgt die Berechnung der optimalen komplexen Gewichte durch Minimierung der mittleren quadratischen Fehler (LMS – Methode, *least mean squares methode*) für den 2-dimensionalen Fall. Dabei wird das Schallfeld einer ebenen Welle mit dem synthetisierten Schallfeld an N Punkten entlang einer Kreislinie mit dem Radius r um den Ursprung herum verglichen. Die Minimierung der Fehler an diesen N Punkten führt zu den optimalen Gewichten, die abhängig vom untersuchten Radius sind. Der „*Super-Minimal*“-Decoder liefert als Ergebnisse die in Kap. 1 vorgestellten ASINC-Funktionen (vgl. [Poletti, 1996a] bzw. [Daniel, 2000]).

Window-Applied Decoder

Der „*Window-Applied*“-Decoder fußt auf dem „*Basic*“-Decoder, wobei eine Optimierung der Lautsprechergewichte auf Basis einer signaltheoretischen Betrachtung – hier für den 2-dimensionalen Fall - erfolgt. Die Kugelflächenfunktionen weisen eine Periodizität von 2π auf, d.h. die Richtcharakteristiken bei Kodierung und Wiedergabe können in eine Fourier-Reihe entwickelt werden, deren Koeffizienten die einzelnen Ambisonic-Signale sind. Durch unterschiedliche Gewichtung der Ambisonic-Signale kann die Abhängigkeit der Lautsprechersignale von der Schalleinfallrichtung gezielt eingestellt werden. Diese Gewichtung entspricht der Fensterung im FIR-Filterdesign, wodurch hier ebenfalls gezielt „Haupt- und Nebenkeulen“ entworfen werden können. Unter dem Begriff der Hauptkeule versteht man hier jener Bereich des (Aufnahme-)Schallfeldes, der durch einen Lautsprecher bei der Wiedergabe repräsentiert wird. Der Bereich der Nebenkeulen umfasst die vom jeweiligen Lautsprecher entfernten Bereiche des Schallfeldes (außerhalb seines erwünschten Wirkungsbereich). Je schmaler die Hauptkeule ist, desto exakter kann eine Quelle lokalisiert werden. Je geringer die Dämpfung der Nebenkeulen ist, desto wahrscheinlicher ist eine Fehlortung, vor allem außerhalb des Referenzabhörorts. Um eine möglichst homogene Quellbewegung zu realisieren zu können, muss die Amplitudensumme (evt. auch (oder Leistungssumme) der Lautsprechersignale über den gesamten Schallwiedergabebereich (360°) konstant sein. Die Verwendung des „*Window-Applied*“-Decoder wird in Kapitel 4 anhand eines Beispiels diskutiert.

2.2.3 Eigenschaften der Ambisonic-Kodierung und Alternativen

Für ein Mono-Signal kann durch Angabe einer Richtungsinformation $[\Phi, \Theta]$ mit Hilfe der Kugelflächenfunktionen durch eine einfache lineare Operation (Gewichtung) eine Ambisonic-Repräsentation einer ebenen Welle gefunden werden. Da die Kodierungsvorschrift linear ist, kann das kodierte Schallfeld durch folgende Operationen manipuliert werden.

- Verstärkung

Durch Multiplikation des gesamten Schallfeldes mit einem Faktor g ergibt sich eine Lautstärkenänderung.

$$\tilde{\mathbf{B}} = g \cdot \mathbf{B} = g \cdot \mathbf{Y} \cdot p \quad (2.47)$$

- Superposition

Werden die Ambisonic-Signale mehrere Schallfelder addiert, so ergibt das resultierende Ambisonic-Schallfeld die Überlagerung der einzelnen virtuellen Quellen. Ein Ambisonic-

Schallfeld kann somit beliebig viele virtuelle Schallquellen bei gleich bleibender Anzahl von Ambisonic-Kanälen besitzen.

$$\tilde{\mathbf{B}} = \sum_i \mathbf{B}_i \quad (2.48)$$

- Rotation

Da Ambisonic auf den Kugelflächenfunktionen basiert, können Schallfeldrotationen um alle Achsen durch eine Matrizenmanipulation durchgeführt werden.

$$\tilde{\mathbf{B}} = \mathbf{R} \cdot \mathbf{B} = \mathbf{R}_x(\Psi) \cdot \mathbf{R}_y(\Theta) \cdot \mathbf{R}_z(\Phi) \cdot \mathbf{B} \quad (2.49)$$

Bei der Rotation eines Schallfeldes unterscheidet man die Drehungen um die x-Achse (\mathbf{R}_x „Tilt“, „Rollen“), um die y-Achse (\mathbf{R}_y „Tumble“, „Nicken“) und um die z-Achse (\mathbf{R}_z „Rotation“, „Drehen“). Eine Zusammenstellung dieser Rotationsmatrizen findet sich bei [Zmölnig, 2002].

Kombinierte Systemordnungen - Alternativen

Um die zu Verfügung stehende Bandbreite von Speichermedien (z.B. Mehrspurbänder) und gängige technische Infrastruktur (z.B. Soundkarten) optimal zu nutzen, können Systemordnungen auch gezielt kombiniert werden. Ein typisches Beispiel für eine kombinierte Kodierungsvorschrift stellt ein periphones System mit horizontaler Kodierung 3. Ordnung (7 Kanäle) und vertikaler Kodierung 1. Ordnung (1 Kanal) dar. Die gegenüber der Elevation deutlich verbesserte Auflösung in der Horizontalebene steht im Einklang mit der Lokalisationsfähigkeit des auditiven Systems. Ein periphones System in vollständiger 2. Ordnung verfügt über eine geringere Auflösung in der Horizontalebene, benötigt jedoch 9 Kanäle. Bei Systemen mit kombinierten Ordnungen kommt es zu Einschränkungen bei der Rotation des Schallfeldes. Für das genannte Beispiel wäre nur eine Rotation in der Horizontalebene (um die z-Achse) möglich.

2.2.4 Untersuchung der Fehler bei idealen Wiedergabebedingungen

Der *Integrated Wavefront Error* (vgl. [Bamford, 1995]) kann zur Evaluierung der Güte eines resynthetisierten Schallfeldes herangezogen werden. Bei diesem Fehlermaß wird die Differenz zwischen Referenz- und Syntheseschallfeld über einen Kreisumfang (2D, Gl. 2.50) bzw. über eine Kugeloberfläche (3D, Gl. 2.51) mit Radius r integriert und normiert.

$$D^{2D} = \frac{1}{|2\pi \cdot p_\Phi|} \int_0^{2\pi} |S_{Referenz,\Phi} - S_{Synthese}| \cdot d\phi \quad (2.50)$$

$$D^{3D} = \frac{1}{4\pi \cdot p_{\Phi, \Theta}^2} \int_0^{2\pi} \int_{-\pi/2}^{+\pi/2} |S_{Referenz, \Phi, \Theta} - S_{Synthese}| \cdot d\theta \cdot d\phi \quad (2.51)$$

Durch die Normierung ist bei einem fehlenden Synthesefeld ($S_{Synthese} = 0$) der Fehler $D=1$. Vor allem bei größeren Auditorien ist der Verlauf des Fehlers als Funktion von kr , also in Abhängigkeit von der Frequenz bzw. der Entfernung vom Mittelpunkt von Bedeutung. Im Folgenden ist der D-Error für ein Ambisonic System 2. Ordnung (2D), reguläres Lautsprecherlayout mit 5 Lautsprecher in 5m Entfernung, für einen Basic-Decoder („blau“) und für einen In-Phase Decoder („grün“) dargestellt.

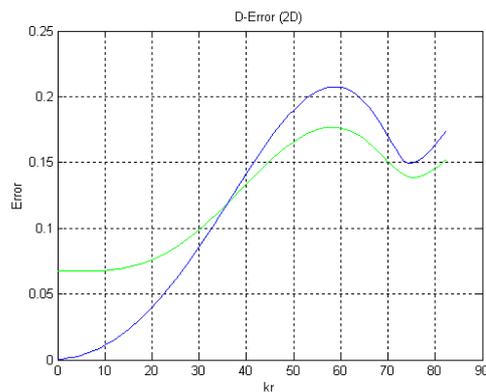


Abb. 2.5a: D-Error für eine Quelle bei 0°.

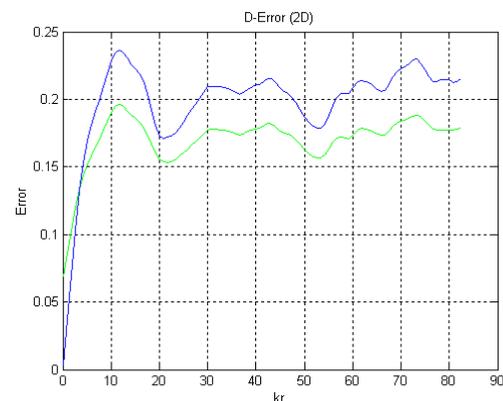


Abb. 2.5c: D-Error für eine Quelle bei 18°.

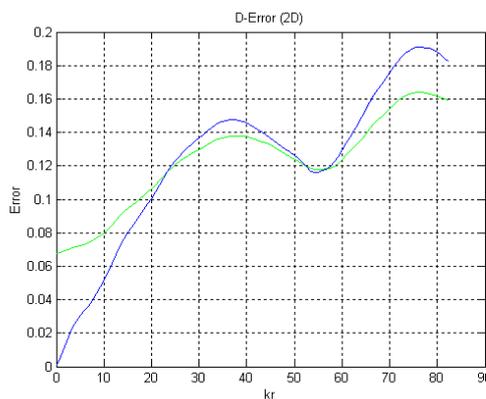


Abb. 2.5b: D-Error für eine Quelle bei 3°.

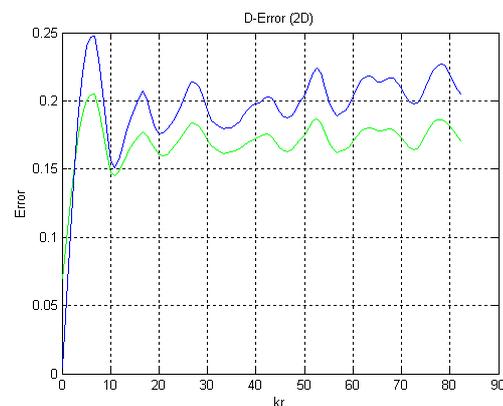


Abb. 2.5d: D-Error für eine Quelle bei 36°.

Aus der Abbildungsreihe Abb.2.5a-d geht hervor, dass der D-Error beim Basic-Decoder stark vom Abbildungswinkel abhängt. Er steigt mit zunehmender Entfernung bzw. zunehmender Frequenz, wie zu erwarten war, an. Stimmt der Abbildungswinkel mit einer Lautsprecherposition überein so wird er für tiefe Frequenzen bzw. kleine Entfernungen (vgl. Annahme für Ambisonic-Ansatz) vernachlässigbar klein. Bezieht man für diesen Fall das Systemschallfeld auf eine Referenzquelle die sich ebenfalls im Abstand der Lautsprecher

befindet so wird der D-Error unabhängig von kr gleich Null. Für Positionen zwischen zwei Lautsprecher wird der Fehler maximal.

Bei der Verwendung eines In-Phase-Decoders kann dagegen ein beinahe ausgewogener Wiedergabebereich unabhängig vom Abbildungswinkel erwartet werden. Der Fehler liegt für kleine kr meist über den Basic-Decoder bleibt aber für höhere Frequenzen bzw. größere Entfernungen zum Ursprung unter dem Basic-Decoder. Wählt man die Systemordnung hoch genug, so ist für größere Auditorien der In-Phase Decoder zu empfehlen.

In [Ward, Abhayapala, 2001] wird das Fehlermaß durch den normierten Rundungsfehler (*normalized truncation error*), der durch Abbruch der Reihenentwicklung beim M . Reihenglied entsteht, in Abhängigkeit von der realisierten Ordnung M definiert (Gl. 2.52). Die Integration erfolgt über die Einheitskugel.

$$\varepsilon_M(kx) = \frac{\int |S_{Referenz}(\mathbf{x}; k) - S_{Synthese}(\mathbf{x}; k)|^2 \cdot d\hat{\mathbf{x}}}{\int |S_{Referenz}(\mathbf{x}; k)|^2 \cdot d\hat{\mathbf{x}}} \quad (2.52)$$

Um den Vergleich unterschiedlicher Systeme zu vereinfachen, wird für die Wiedergabe infinitesimale Winkelauflösung (unendliche Lautsprecheranzahl) angenommen. Für ein vollständiges periphones System ergibt dieses Fehlermaß (siehe Appendix in [Ward, Abhayapala, 2001]):

$$\varepsilon_M(kx) = 1 - \sum_{m=0}^M (2m+1) \cdot (j_m(kx))^2 \quad (2.53)$$

Daraus folgt:

- Der normalisierte Rundungsfehler ist unabhängig von der Richtung der Quelle $\hat{\mathbf{y}}$.
- Der normalisierte Rundungsfehler ist nur vom Produkt der Wellenzahl k mit dem Radius x abhängig.

D.h. bei gegebener Ordnung M schrumpft für steigende Wiedergabefrequenzen f mit $k = 2\pi \cdot f/c$ bei einem zulässigen maximalen Fehler ε_{\max} der Wiedergabebereich $r \leq x$ (vgl. Abb. 2.6).

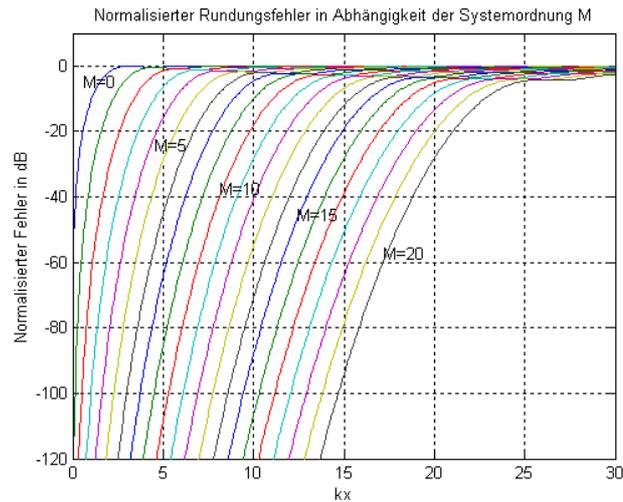


Abb. 2.6 Normalisierter Rundungs- (bzw. Abbildungs-) Fehler (vgl. [Ward, Abhayapala, 2001]).

Bei der Fehlerbetrachtung stellt sich die Frage, ob der Fehler für Punkte innerhalb einer untersuchten Kugeloberfläche den Fehler an der Oberfläche nicht überschreiten. Aus der Abbildung 2.5 ist ersichtlich, dass für eine definierte Ordnung M die Fehlerfunktion monoton steigend mit $k \cdot x$ ist. Wird daher der maximal zulässige Fehler ϵ_{\max} an einer Oberfläche $x = x_0$ für die größte abzubildende Frequenz f festgelegt, so ist der Fehler für alle Punkte $x < x_0$ für den betrachteten Frequenzbereich kleiner.

Infolge einer exakten Reproduktion des Schalldrucks entlang einer Kugeloberfläche, mit Hilfe der Kugelflächenfunktionen, wird auch der Druckgradient exakt reproduziert. Durch Gültigkeit des Kirchhoff-Helmholtz Theorems wird dadurch sichergestellt, dass innerhalb der betrachteten Kugeloberfläche das Schallfeld eindeutig beschrieben ist (vgl. Einleitung Kap. 2) und somit auch exakt reproduziert wird.

Der Druck entlang einer allgemeinen Kugeloberfläche definiert durch den Radius $r = |\mathbf{x}|$ ist mit Gl. 2.19 beschrieben. Der Druckgradient normal zur begrenzenden Oberfläche wird daher nur durch den Term $R(kx)$ beschrieben. Aus $R(kx) = i^n \cdot j_n(kr)$ (vgl. Gl. 2.10) und deren Ableitung in radialer Richtung folgt:

$$\frac{d}{d\zeta} j_n(\zeta) = j_{n-1}(\zeta) - \frac{n+1}{\zeta} j_n(\zeta) \quad (2.54)$$

Die beiden Terme, in obiger Gleichung, würden bei einem Koeffizientenvergleich des Referenz- und des Systemschallfeldes (vgl. Gl. 2.28a) auf beiden Seiten auftreten und somit keine weitere Auswirkung auf die Koeffizienten der Kugelflächenfunktionen haben.

Für einen normierten Fehler von 4% (-14dB), der eine ausreichende Genauigkeit für die meisten praktischen Anwendungen darstellt¹ (vgl. [Ward, Abhayapala, 2001]), kann der Zusammenhang für die notwendige Systemordnung M bei gegebener Wiedergabeanforderungen $k \cdot x$ (d.h. Bereichsgröße und Bandbreite) wie folgt bestimmt werden:

$$M = \lceil kx_0 \rceil$$

wobei $\lceil \cdot \rceil$ eine Aufrundung zur nächsten ganzen Zahl bedeutet.

Dieser Zusammenhang ist aus Abbildung 2.5 ersichtlich und in [Ward, Abhayapala, 2001] bestätigt.

Zusammenfassung

Der Vorteil von Ambisonic ist die relativ einfache Kodierung und Dekodierung. Ein kodiertes Schallfeld kann wie ein grafisches Objekt beliebig mit Rotationsmatrizen gedreht, gekippt oder geneigt werden. Je höher die Systemordnung gewählt wird, umso genauer bzw. umso größer wird der Wiedergabebereich, in dem ein definierter Fehler nicht überschritten wird. Da es sich bei Ambisonic um ein koinzidentes Aufnahmeverfahren handelt, Mikrofone aber eine gewisse physikalische Ausdehnung besitzen, und komplexe Mikrofoncharakteristika (d.h. $\cos(2\phi)$, $\sin(2\phi)$, etc.) nur begrenzt existieren, ist dieses Verfahren für Aufnahmen von realen Schallfelder Ordnungen größer 1 zur Zeit nicht realisierbar. Hingegen ergibt sich bei synthetisch erzeugten Schallfelder nur eine Grenze durch die Rechenleistung und die Übertragungsbandbreite. Die Anzahl der notwendigen Übertragungskanäle leitet sich aus der realisierten Systemordnung ab. Bei der Verwendung von Kugelflächenfunktionen wird nicht nur der Druck, sondern auch der Druckgradient entlang einer Kugeloberfläche bzw. Kreislinie akkurat reproduziert, und somit gilt dies auch für den eingeschlossenen Bereich (vgl. *Kirchhoff-Helmholtz-Integral*). Der Zusammenhang zwischen Größe des Wiedergabebereiches bzw. Bandbreite und der notwendigen Systemordnung lässt sich für einen normierten Fehler von 4% mit $M = \lceil kx_0 \rceil$ beschreiben. Der Wiedergabefehler kann durch eine geeignete Wahl der Lautsprecherpositionen und -gewichtung bzgl. Raum und Frequenz unterschiedlich verteilt – optimiert werden.

¹ Die psychoakustisch auffälligsten Fehler ergeben sich durch Signalanteile aus nicht in Quellrichtung liegenden Lautsprechern. Wenn diese um mind. 14 dB unter den Signalen aus Quellrichtung liegen, so werden sie für tiefe und mittlere Frequenzbereiche auch dann von den „Hauptkeulen“-Signalen verdeckt, wenn man unter ungünstiger Kopforientierung ein bestimmte Dämpfung der „Hauptkeulen“-Signale zum kontralateralen Ohr gegenüber ungedämpften „Nebenkeulen“-Signalen berücksichtigt (siehe ILDs von typischen HRFTs).

2.3 Wellenfeldsynthese (Wave Field Synthesis)

Die Wellenfeldsynthese (WFS, *Wave Field Synthesis*) zählt ebenso wie der Ambisonic-Ansatz zu den holophonen Verfahren. Während Ambisonic auf einer lokalen Schallfeldrekonstruktion im/um den Koordinatenursprung basiert, wird bei WFS eine globale Rekonstruktion angestrebt. Der Ansatz der Wellenfeldsynthese leitet sich direkt aus dem Huygens-Fresnel'schen Prinzip ab, das im folgenden dargestellt und neu formuliert wird.

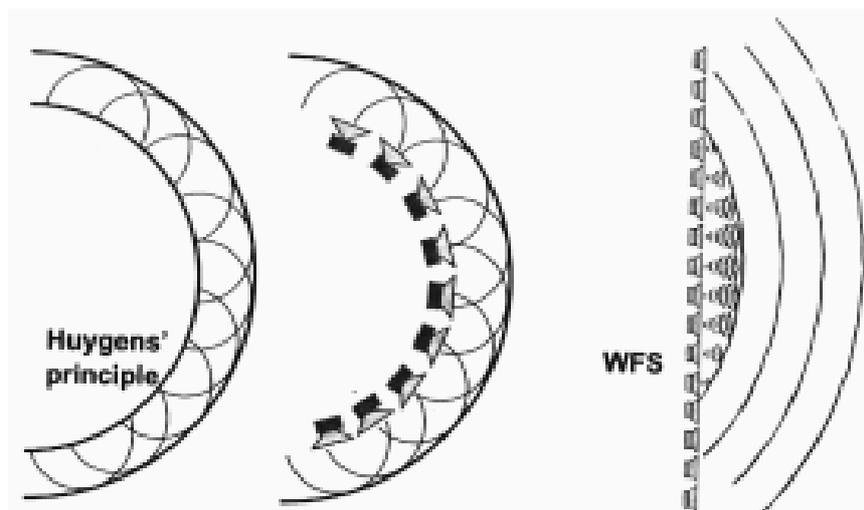


Abb. 2.7: Grafische Ableitung der Wellenfeldsynthese aus dem Huygens-Fresnel'schen Prinzip.

Durch geeignete Verteilung von sekundären Quellen kann das Schallfeld einer primären Quelle nachbildet werden (vgl. [Boone, 2001]).

Eine primäre Quelle sendet eine Welle aus, die sich mit der Geschwindigkeit c im Raum ausbreitet. Nach der Zeit t hat sich am Ort $c \cdot t$ eine primäre Wellenfront ausgebildet. Jeder Punkt dieser Wellenfront kann als Ausgangspunkt einer neuen Elementarwelle aufgefasst werden. Durch Überlagerung dieser Elementarwellenfelder zum Zeitpunkt $(t + \Delta t)$ entsteht eine neue Wellenfront am Ort $c \cdot (t + \Delta t)$. Diese neue Wellenfront entspricht der natürlichen Ausbreitung der primären Wellenfront. Das heißt: durch eine bestimmte Anordnung von sekundären Quellen ist es möglich, das Schallfeld einer realen oder virtuellen Primärquelle nachzubilden (siehe Abb. 2.7). Basierend auf den Kirchhoff- und Rayleigh-Theoremen führte Berkhout in [Berkhout, 1988] das Konzept der Wellenfeldsynthese in der Akustik ein. Durch eine geeignete Ansteuerung von Monopol- und Dipolquellen (Lautsprecher), die auf einer ebenen Fläche verteilt sind, sollte es möglich sein das Schallfeld einer primären Quelle, das durch planar verteilte Druck- und Druckgradientenempfänger (Mikrofone) aufgenommen wurde, in einem Bereich um einen Hörer getreu wiederzugeben (siehe Abb. 2.8).

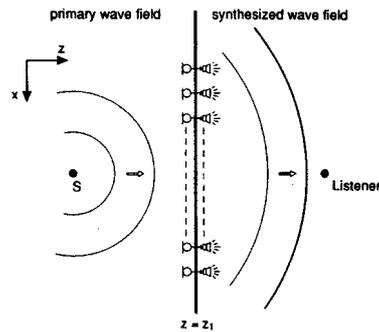


Abb. 2.8: Illustration des WFS-Prinzips (vgl. [Berkhout et al., 1993]).

Erfolgt die Wiedergabe nicht am Ort der Aufnahme, so ist eine Extrapolation der Schallausbreitung zwischen Aufnahmepunkt und Wiedergabepunkt notwendig (vgl. Abb. 2.9). Diese Extrapolation wird in der Literatur durch den Syntheseoperator \mathbf{W} bezeichnet.

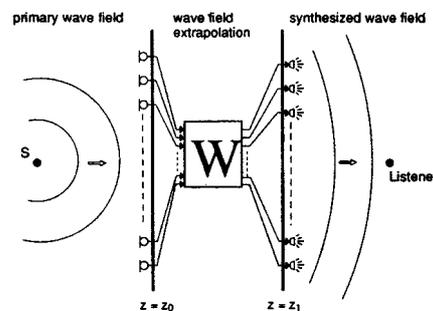


Abb. 2.9: Verallgemeinerung des WFS-Prinzips (vgl. [Berkhout et al., 1993]).

Aus [Berkhout et al., 1993] und [Start et al., 1999] geht hervor, dass für viele Anwendungen die Aufnahme durch quellennahe, nachgeführte Mikrofonierung¹ (*tracked close miking*) bessere Ergebnisse erzielt werden können als durch flächendeckende Mikrofonarrays (*remote miking*). Durch das WFS-Konzept können akustische Wellenfelder zeitlich und räumlich präzise gesteuert werden, wodurch naturgetreue Live-Verstärkung oder künstliche Raumadaptierungen ermöglicht werden. In [Start, 1997] wurde der Einsatz von WFS für reine Live-Verstärkung nach physikalischen und perceptiven Kriterien evaluiert. In [de Vries et al., 1994] und [Sonke et al., 1997] finden sich Ansätze zur künstlichen Verhallung auf Basis der WFS. Die Implementierung eines vollständigen WFS Systems (Live Verstärkung, Aufnahme und Erzeugung einer virtuellen Raumakustik, etc.), sowie eine objektive und subjektive Evaluierung dieses Systems wird in [Verheijen, 1997] ausführliche diskutiert.

Im folgenden werden die Signale der Wiedergabelautsprecher aus einem vereinfachten Kirchhoff-Helmholtz Integral (KHI) für praktische Applikationen hergeleitet.

¹ Das close-miking Prinzip kann für eine oder mehrere Klangquellen verwendet werden. Diese Aufnahmen werden als ‚notional sources‘ (fiktive Quellen) bezeichnet und entsprechend ihrer räumlichen Positionierung durch den Syntheseoperator \mathbf{W} behandelt.

2.3.1 Kirchhoff-Helmholtz Integral

Eine konkrete mathematische Beschreibung der feldtheoretischen Fakten für den Schalldruck P_A innerhalb der geschlossenen Oberfläche S eines quellenfreien Volumens V (Abb. 2.10) liefert das aus der Wellengleichung und dem zweiten Green'schen Theorem abgeleitete Kirchhoff-Helmholtz Integral (KHI) [Berkhout, 1982]:

$$P_A = \frac{1}{4\pi} \int_S (P \nabla G - G \nabla P) \bar{n} dS \quad (2.55)$$

$$G = \frac{e^{-jk \cdot r}}{|\mathbf{r}|} + F \quad (2.56)$$

Wählt man vorerst in der allgemeinen Form der Green'schen Lösungsfunktion G (Gl. 2.56) die Funktion $F=0$ und fügt dies zusammen mit der Bewegungsgleichung (Gl. 2.57) in Gl. 2.55 ein, erhält man das KHI für homogene Medien Gl. 2.58 (vgl. dazu auch Gl. 2.1).

$$\frac{\partial P}{\partial n} = -j\omega \rho_0 V_n \quad (2.57)$$

$$P_A = \frac{1}{4\pi} \int_S \left[\left(P \frac{1+jkr}{r} \cos \varphi \frac{e^{-jkr}}{r} \right) + \left(j\omega \rho_0 V_n \frac{e^{-jkr}}{r} \right) \right] dS \quad (2.58)$$

Aus der Interpretation des KHI resultieren zwei Aussagen:

1.) Bei Kenntnis des Drucks P und der Normalkomponente V_n der Teilchengeschwindigkeit an der geschlossenen Oberfläche S eines quellenfreien Volumens V kann der Druck in jedem Punkt A innerhalb von V errechnet werden (Abb. 2.10).

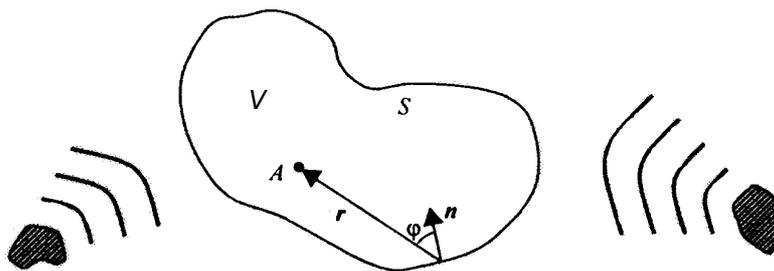


Abb. 2.10: Veranschaulichung des Kirchhoff-Helmholtz Integrals (KHI) (vgl. [Verheijen, 1997]).

2.) Jedes Schallfeld kann durch eine Verteilung von Monopol- und Dipolquellen entlang S synthetisiert werden. Die Stärke jedes Monopols ist durch die Normalkomponente der Teilchengeschwindigkeit V_n (zweiter Integrand) festgelegt, jene der Dipolquellen durch den Schalldruck P (erster Integrand) entlang der geschlossenen Oberfläche S (Abb. 2.11).

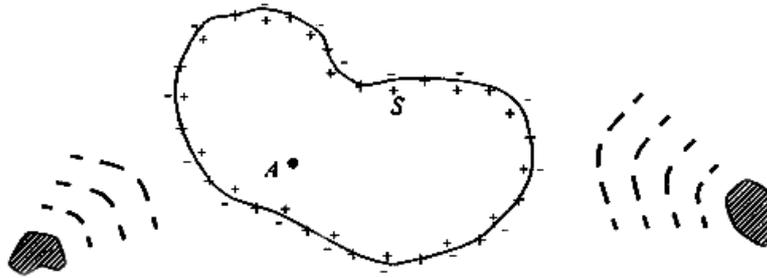


Abb. 2.11: Monopol- und Dipolquellenverteilung entlang der Hülle des Volumens V (vgl. [Verheijen, 1997]).

Die originale Quellenverteilung wird im Folgenden als *primäre Quellenverteilung* bezeichnet, die Verteilung der Monopol- und Dipolquellen als *sekundäre Quellenverteilung*.

2.3.2 Rayleigh Integrale

Durch spezielle Wahl der Hüllfläche von V und Einschränkung auf entweder Mono- oder Dipolquellen als Sekundärstrahler ergeben sich praktikable Näherungslösungen des KHI [Berkhout, 1982].

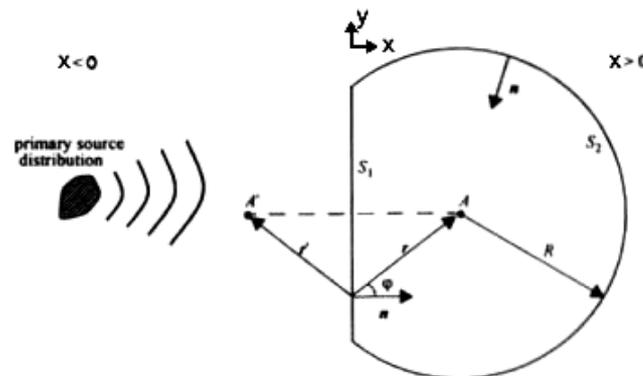


Abb. 2.12: Das Volumen V wird nun durch die Ebene S_1 bei $x=0$ und die Halbkugel S_2 im Raum $x>0$ begrenzt.

Wie in Abb.2.12 dargestellt, vereinfacht man die Integrationsoberfläche S auf eine Ebene S_1 bei $x=0$ und eine Halbkugelfläche S_2 im Halbraum $x>0$. Jetzt nimmt man an, das Schalldruckfeld P im rechten Halbraum ($x>0$) werde durch Primärquellen im linken Halbraum ($x<0$) hervorgerufen, der Raumbereich innerhalb S sei quellenfrei. Rückt man die Teilfläche S_2 in immer größere Entfernung, verschwindet der Beitrag dieser Quellen zum KHI für $R \rightarrow \infty$. Der Integrationsbereich darf auf die Ebene entlang $x=0$ reduziert werden.

Der Druck im Punkt A wird wiederum durch Substitution von Gleichung (2.56) in das Integral (in Gl. 2.59) berechnet:

$$P_A = \frac{1}{4\pi} \int_{S_1} \left[P \frac{\partial}{\partial n} \left(\frac{e^{-jkr}}{r} + F \right) - \left(\frac{e^{-jkr}}{r} + F \right) \frac{\partial P}{\partial n} \right] dS \quad (2.59)$$

Findet man nun geeignete Funktionen F , die, eingesetzt in die Green'sche Lösungsfunktion (Gl. 2.56) den ersten oder zweiten Integranden in Gl. 2.59 aufheben, sind nur Monopol- oder nur Dipolquellen zur Reproduktion des Primärfeldes erforderlich.

Rayleigh Integral I

Um den ersten Integranden aus Gl. 2.59 zu eliminieren, wählt man für die Funktion F

$$F = \frac{e^{-jkr'}}{r'} \quad (2.60)$$

und das Kirchhoff-Helmholtz Integral wird auf das Rayleigh Integral I vereinfacht:

$$P_A = \frac{1}{2\pi} \int_{S_1} j\omega\rho_0 V_n \frac{e^{-jkr}}{r} dS \quad (2.61)$$

Eine Verteilung von sekundären Monopolquellen entlang einer ebenen Fläche, deren Stärke nun das Zweifache der Teilchengeschwindigkeit normal zur Oberfläche beträgt, reicht somit aus, das primäre Wellenfeld im Halbraum $x > 0$ zu synthetisieren. Allerdings wird im Bereich $x < 0$ eine Spiegelwelle erzeugt, die wegen der Abwesenheit von Dipolquellen nicht kompensiert werden kann. Das hat jedoch keinen Einfluss auf das Gebiet $x > 0$.

Rayleigh Integral II

Ebenso ist die Reduktion der sekundären Quellen auf eine Dipolverteilung erreichbar. Die Elimination des zweiten Terms aus Gl. 2.59 gelingt durch eine geeignete Wahl der Funktion

$$F = -\frac{e^{-jkr'}}{r'} \quad (2.62)$$

Das so erhaltene Rayleigh Integral II beschreibt das Feld einer Dipolverteilung entlang S_1 , angetrieben mit der doppelten Stärke des an dieser Fläche herrschenden Schalldrucks. Für die im Bereich $x < 0$ erzeugte Spiegelwelle gilt dasselbe wie für das Rayleigh Integral I.

$$P_A = \frac{1}{2\pi} \int_{S_1} P \frac{1+jkr}{r} \cos\varphi \frac{e^{-jkr}}{r} dS \quad (2.63)$$

2.3.3 Syntheseoperatoren

Funktionen zur Gewichtung der Lautsprechersignale sollen gefunden werden. Abhängig von der Reproduktionsanordnung existieren verschiedene Syntheseoperatoren zur Berechnung der komplexen Lautsprechergewichte.

2.3.3.1 Der 3D Syntheseoperator

In Übereinstimmung mit den 3D Rayleigh Integralen (Gl. 2.61 und Gl. 2.63) erfolgt die Rekonstruktion beliebiger Primärsignale $S(\omega)$ mit einer, zwischen dem primären Quellbereich und dem Auditorium liegenden, planaren Verteilung von Sekundärquellen (Abb. 2.12).

Für omnidirektional abstrahlende Lautsprecher sind die komplexen Gewichte proportional zur Normalkomponente V_n der Schallschnelle. Das Lautsprechersignal für Monopolquellen lautet

$$Q_m = S(\omega) j \omega \rho_0 V_n \quad (2.64)$$

Lautsprecher mit Dipolcharakteristik werden proportional dem an ihrer Position herrschenden Schalldruck des nachzubildenden Quellsignals gewichtet. Das Lautsprechersignal ist demnach

$$Q_d(\omega) = S(\omega) \frac{1 + jkr}{r} \cos \varphi \quad (2.65)$$

2.3.3.2 Der 2½D Syntheseoperator

Für Bildschirmmanwendungen (z.B. Audiodisplay vgl.[Strauß, 2002]) oder den Einsatz auf Bühnen, in Kinos und ähnlichen Wiedergabesituationen wäre eine flächige, sichtbehindernde Applikation problematisch. Zudem ist das räumliche Auflösungsvermögen der menschlichen Hörwahrnehmung in vertikaler Richtung geringer als in horizontaler Richtung [Blauert, 1997]. Eine eindimensionale Lautsprecheranordnung ist wesentlich leichter in die Szene zu integrieren und somit obligat. Damit reduziert sich der mögliche Quellenraum theoretisch auf zwei Dimensionen (Azimut und Distanz). Das Schallfeld wird als von der vertikalen Achse unabhängig angenommen, was zur Folge hat, dass statt dreidimensionaler Punktquellen nur noch vertikale Linienquellen existieren können. Für das 2D-Rayleigh Integral I ergibt sich im Fernfeld folgende Näherung [Berkhaut, 1982]:

$$P(\mathbf{r}_A, \omega) = \rho c \sqrt{\frac{jk}{2\pi}} \int_{x=0} V_n(y, \omega) \frac{e^{-jkr}}{\sqrt{r}} dy \quad \text{für } kr \gg 1 \quad (2.66)$$

Aus Gl. 2.66 ist ersichtlich, dass die Amplitude der Sekundärquellen nun proportional $1/\sqrt{r}$ gedämpft wird, was dem zylindrischen Feld einer Linienquelle entspricht. Die Anwendung des 2D-Operators nach Gl. 2.66 zur Reproduktion eines 3D-Schallfeldes mittels einer sekundären Linienverteilung kann folglich nicht das gewünschte Resultat liefern.

Eine praxisnahe Lösung muss direkt aus dem dreidimensionalen Feld einer virtuellen Quelle die Signale zum Betrieb einer zweidimensionalen Lautsprecheranordnung synthetisieren. Als Dämpfungsfaktor für die Amplitude der Sekundärquellen wird ein, der natürlichen Schallausbreitung einer Punktquelle entsprechendes, $1/r$ Verhalten gefordert.

Zur Veranschaulichung der Herleitung des $2\frac{1}{2}D$ -Operators dient die Abbildung 2.13. In der Horizontalebene sei eine Punktquelle S positioniert. Für deren Schalldruck am Ort \mathbf{r} schreibt man, mit $S(\omega)$, dem Primärsignal und $G(\varphi, \vartheta, \omega)$, der Richtcharakteristik der Quelle,

$$P(\mathbf{r}, \omega) = S(\omega) G(\varphi, \vartheta, \omega) \frac{e^{-jkr}}{r} \quad (2.67)$$

Eine flächige sekundäre Quellenverteilung von Monopolquellen ($G(\varphi, \vartheta, \omega) = 1$) in der Frontalebene bei $x=0$ synthetisiert im Abhörpunkt R (vgl. Gl. 2.61) den Schalldruck

$$P_{\text{synth}} = \frac{1}{2\pi} \int_{xz\text{-Ebene}} j\omega\rho_0 V_n(\mathbf{r}, \omega) \left[\frac{e^{-jk\Delta r}}{\Delta r} \right] dx dz \quad (2.68)$$

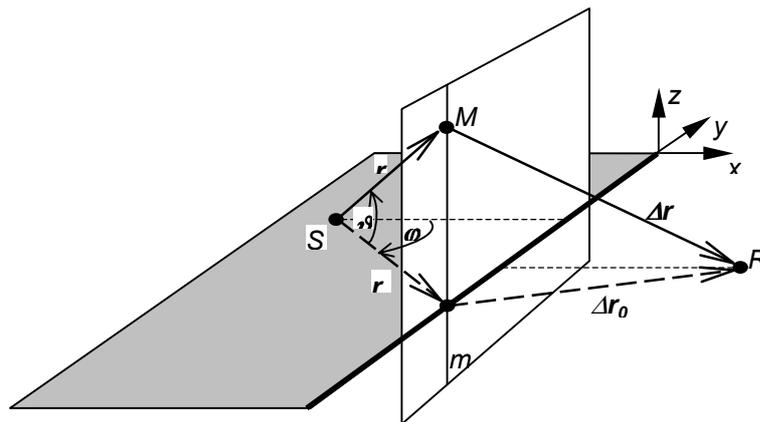


Abb. 2.13: Syntheseoperator nach Start [Start, 1997]. Primärquelle S und Empfänger R liegen in der xy -Ebene. Vektor r weist von S zur Sekundärquelle M auf der Linie m , Δr zeigt von der Sekundärquelle zum Empfänger. Projektionen in die xy -Ebene sind mit Index 0 versehen.

Die Auswertung des exakten 3D Rayleigh Integrals Gl. 2.68 in z -Richtung führt das Flächenintegral in ein Linienintegral entlang der y -Achse über. Um das zu erreichen, passt man die Bewegungsgleichung (Gl. 2.57) an die geometrischen Gegebenheiten (Abb. 2.13) an, wobei für $\varphi = \arctan(y/x)$ und $\vartheta = \arctan(z/\sqrt{x^2 + y^2})$ gilt, und erhält

$$j\omega\rho_0 V_n(\mathbf{r}, \omega) = S(\omega) \frac{e^{-jkr}}{r} \left\{ \frac{\sin \varphi}{r \cos \vartheta} \frac{\partial G}{\partial \varphi} + \frac{\cos \varphi \sin \vartheta}{r} \frac{\partial G}{\partial \vartheta} + \frac{1 + jkr}{r} G \cos \varphi \cos \vartheta \right\} \quad (2.69)$$

Nach Einsetzen dieses Ausdrucks in Gl. 2.68 erlangt das Rayleigh Integral die Form

$$I = \int_{-\infty}^{\infty} f(z) e^{jk\phi(z)} dz \quad (2.70)$$

Zur Lösung von Integralen dieses Typs bedient man sich der „Stationary Phase Method“ [Bleistein, 1984]. Diese Näherungsmethode beruht darauf, dass für eine vertikale Quellenverteilung entlang der Linie m ein Punkt der stationären Phase von ϕ (d.h. $\phi'=0$) unter der Bedingung $kr_0 \gg 1$ gefunden werden kann. Das ist der Schnittpunkt der Linie m mit der y -Achse (in Abb. 2.13 mit z_0 bezeichnet). Die Quelle in diesem Punkt liefert im Vergleich zu allen anderen Punktquellen M entlang m den größten Beitrag zur Synthetisierung des Schallfeldes, und fungiert in weiterer Folge als genäherter Repräsentant für alle Punktquellen entlang m . Wendet man das Verfahren auf alle sekundären Spaltenquellen in der yz -Ebene an, resultiert daraus eine sekundäre Linienverteilung entlang der y -Achse. Als Zwischenergebnis für den synthetisierten Schalldruck erhält man das 2½D Rayleigh Integral.

$$P_{synth} = S(\omega) \sqrt{\frac{jk}{2\pi}} \int_{-\infty}^{\infty} \sqrt{\frac{\Delta r_0}{r_0 + \Delta r_0}} G(\varphi, 0, \omega) \cos \varphi \frac{e^{-jkr_0}}{\sqrt{r_0}} \frac{e^{-jk\Delta r_0}}{\Delta r_0} dy \quad (2.71)$$

Der Amplitudenfaktor $\sqrt{\Delta r_0 / (r_0 + \Delta r_0)}$ in Gleichung (2.71) ist abhängig von Δr_0 , der Distanz Sekundärquelle-Abhörpunkt. Somit wäre die Gewichtung jedes Lautsprechers eine Funktion des jeweiligen Abhörpunktes R , was nicht realisierbar ist.

Eine praktikable Lösung wird erreicht, wenn man eine durch erneute Anwendung der „Stationary Phase Method“ errechnete Referenzlinie $x=\Delta x_0$ verwendet (Abb. 2.14). Das führt zu einem konstanten Amplitudenfaktor für alle Zuhörer entlang dieser Linie.

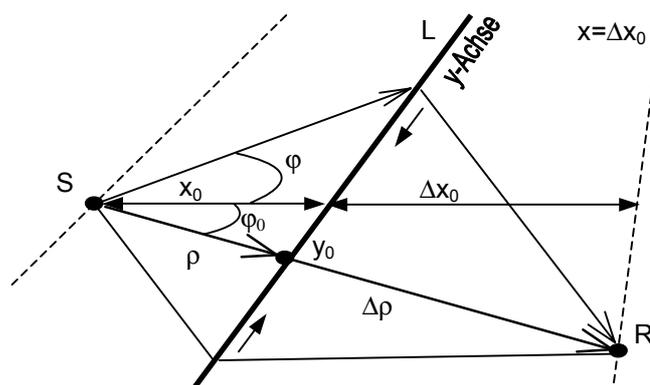


Abb. 2.14: Die Suche nach einem Punkt stationärer Phase entlang der Linie L , führt auf den Schnittpunkt y_0 der sekundären Linienquelle L mit der Verbindungslinie Primärquelle-Empfänger (ρ , $\Delta\rho$) (vgl. [Strauß, 2002]).

Für Positionen entlang der Linie $x=\Delta x_0$ ist der Amplitudenfaktor $\sqrt{\Delta x_0 / (x_0 + \Delta x_0)}$ nun konstant und liefert die korrekte Druckamplitude. In den Abhörpunkten abseits der

Referenzlinie ergeben sich allerdings Amplitudenfehler, die jedoch durch geeignete Adjustierung des Referenzabstandes Δx_0 minimiert werden können [Start, 1997].

Die Genauigkeit der Näherung durch den Referenzabstand hat den selben Grad wie die Näherung der Integration in z-Richtung, bedeutet also keine Qualitätsminderung des Resultats von Gl. 2.71. Der synthetisierte Schalldruck wird somit angeschrieben als

$$P_{synth} = S(\omega) \sqrt{\frac{jk}{2\pi}} \int_{-\infty}^{\infty} \sqrt{\frac{\Delta x_0}{x_0 + \Delta x_0}} G(\varphi, 0, \omega) \cos \varphi \frac{e^{-jkr}}{\sqrt{r}} \frac{e^{-jk\Delta r}}{\Delta r} dy \quad (2.72)$$

Aus der Definition des 2½D Syntheseoperators (Gl. 2.72) für eine Linienverteilung von Monopolquellen $W_{Punkt} = \frac{e^{-jk\Delta r}}{\Delta r}$ lässt sich als Funktion für das Antriebssignal Q_m der Lautsprecher extrahieren.

$$Q_m(y, \omega) = S(\omega) \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{\Delta x_0}{x_0 + \Delta x_0}} G(\varphi, 0, \omega) \cos \varphi \frac{e^{-jkr}}{\sqrt{r}} \quad (2.73)$$

Für das Antriebssignal Q_m einer Linienverteilung sekundärer Dipolquellen mit dem Syntheseoperator $W_{Dipol} = jk \cos \varphi \cdot \frac{e^{-jk\Delta r}}{\Delta r}$ ergibt aus der analogen Ableitung des Rayleigh Integral II (siehe Gl. 2.63) die Funktion

$$Q_d(y, \omega) = S(\omega) \sqrt{\frac{1}{2\pi jk}} \sqrt{\frac{\Delta x_0}{x_0 + \Delta x_0}} G(\varphi, 0, \omega) \frac{e^{-jkr}}{\sqrt{r}} \quad (2.74)$$

Die Gültigkeit des 2½D Operators aus Gl. 2.72 lässt sich im Fernfeld beweisen [Start, 1996]. Für weit von den Sekundärquellen entfernte Zuhörerpositionen gilt $\Delta r/r \gg 1$, und man erhält für das synthetisierte Schalldruckfeld:

$$P_{synth} = S(\omega) \sqrt{\frac{jk}{2\pi}} \int_{-\infty}^{\infty} \cos \varphi \frac{e^{-jkr}}{\sqrt{r}} \frac{e^{-jk\Delta r}}{\Delta r} dy_L \quad \text{für } \frac{\Delta r}{r} \gg 1 \quad (2.75)$$

Mit der Fernfeldnäherung der Normalkomponente der Schnelle V_x^{Linie} einer Linie von Monopolpunktquellen

$$V_x^{Linie}(y_L) = \frac{1}{\rho_0 c} \sqrt{\frac{2\pi}{jk}} S(\omega) \cos(\varphi) \frac{e^{-jkr}}{\sqrt{r}} \quad (2.76)$$

erhält der Ausdruck (Gl. 2.77) die selbe Form wie das 3D Rayleigh Integral (Gl. 2.66).

$$P(\mathbf{r}_R) = \frac{1}{2\pi} \int_{-\infty}^{\infty} j\omega \rho_0 V_x^{Linie}(y_L) \frac{e^{-jk\Delta r}}{\Delta r} dy_L \quad \text{für } \frac{\Delta r}{r} \gg 1 \quad (2.77)$$

Zur Rekonstruktion des Wellenfeldes einer primären Monopol-Punktquelle mit einer horizontalen Linienverteilung von Monopolquellen, muss die Quellstärke jeder sekundären

Punktquelle proportional zur Normalkomponente der Schnelle V_x^{Linie} einer durch den Ort der Primärquelle laufenden vertikalen Monopol-Linienquelle sein.

Die zu Beginn der Ableitung getroffene Annahme, dass alle Quellen und Zuhörer in der selben Ebene positioniert sein müssen, gilt nach wie vor. Vertikale Auslenkung der Sekundärquellen, die klein gegen die Distanz Primär- zu Sekundärquelle und klein gegen den Abstand Sekundärquelle zu Zuhörer sind, können jedoch vernachlässigt werden [Berkhout et al., 1993].

2.3.3.3 Syntheseoperatoren für gekrümmte Linienformen

Der 2½D Syntheseoperator kann auf beliebige sekundäre Linienverläufe erweitert werden, lediglich eine Anpassung der Schreibweise an die Geometrie (siehe Abb. 2.15) allgemeiner Kurvenformen ist notwendig [Start, 1996].

Als Funktion der Lautsprechersignale Q_m erhält man für sekundäre Monopolquellen¹

$$Q_m(\vec{r}, \omega) = S(\omega) \sqrt{\frac{jk}{2\pi}} \sqrt{\frac{\Delta r}{r + \Delta r}} \cos \varphi \frac{e^{-jkr}}{\sqrt{r}} \quad (2.77)$$

und für sekundäre Dipolquellen

$$Q_d(\vec{r}, \omega) = S(\omega) \sqrt{\frac{1}{2\pi jk}} \sqrt{\frac{\Delta r}{r + \Delta r}} \frac{e^{-jkr}}{\sqrt{r}} \quad (2.78)$$

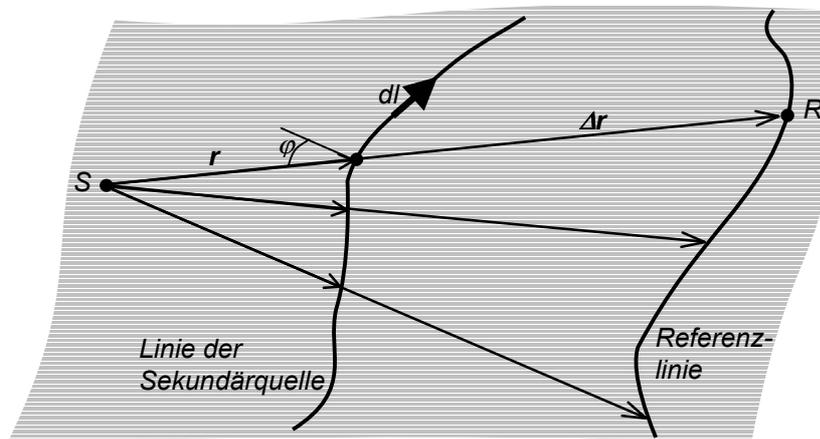


Abb. 2.15: Anpassung des Syntheseoperators an die geometrischen Verhältnisse allgemeiner Kurven (vgl. [Strauß, 2002])

¹ Die Gl. 2.77 bildet die Grundlage für die Berechnung der komplexen Lautsprechergewichte bei der Distanzkodierung in Kapitel 3

2.3.3.4 Einbindung der Abstrahlcharakteristik der Lautsprecher

Ein Aspekt bei der Planung von Wiedergabeeinrichtungen, die in Wellenfeldsynthese betrieben werden sollen, ist die Einbindung der Abstrahlcharakteristik G der verwendeten Lautsprecher. Unter Fernfeldbedingungen ($kr \gg 1$) ist die Annahme einer, von der Distanz unabhängigen Funktion $G(\phi, \omega)$ gültig.

Das Abstrahlverhalten der gedachten primären Linienquelle wurde im Ansatz zur Entwicklung des Syntheseoperators (Gl. 2.67) bereits berücksichtigt und ist invers verknüpft mit der Abstrahlcharakteristik der sekundären Lautsprecher [de Vries, 1996].

$$G_{Linie}(\phi, \omega) = \sqrt{\frac{jk}{2\pi}} \frac{\cos \phi}{G_{LS}(\phi, \omega)} \quad (2.79)$$

Besitzen die Arraylautsprecher eine Monopolcharakteristik, dann werden die Linienquellen zu Dipole:

$$G_{LS}(\phi, \omega) = 1 \Rightarrow G_{Linie}(\phi, \omega) = \sqrt{\frac{jk}{2\pi}} \cos \phi \quad (2.80)$$

Besitzen die Arraylautsprecher eine Dipolcharakteristik, dann werden die Linienquellen zu Monopole:

$$G_{LS}(\phi, \omega) = jk \cos(\phi) \Rightarrow G_{Linie}(\phi, \omega) = \sqrt{\frac{1}{2\pi jk}} \quad (2.81)$$

Adaptiert man die Lautsprechersignale Q_m nach Gl. 2.79, transferiert man dadurch die Abstrahlcharakteristik der Lautsprecher virtuell auf jene der gedachten primären Quelle. Nichtideale Abstrahlcharakteristika einzelner Lautsprecher können auf diese Weise kompensiert werden. Im Prinzip kann jeder Typ von Lautsprecher für Anwendungen der Wellenfeldsynthese benutzt werden. Wichtig ist, dass bei lautsprecherbasierten Implementierungen alle verwendeten Lautsprecher einheitliche Charakteristik aufweisen [de Vries, 1996].

2.3.4 Artefakte

2.3.4.1 Durch endliche Abmessungen

In der Praxis limitiert man die seitliche Ausdehnung des Arrays. Der Reproduktionsbereich (Abb. 2.16) reduziert sich dadurch auf jene Positionen, für die entlang der sekundären Quellenverteilung ein stationärer Phasenpunkt gefunden werden kann.

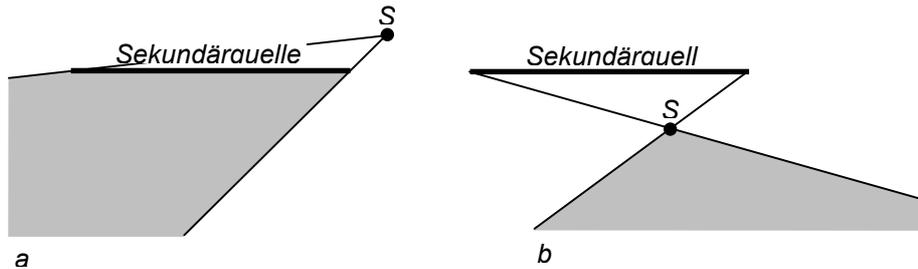


Abb. 2.16: Reproduktionsareal für Primärquellen (vgl. [Strauß, 2002]),

- a) die hinter dem Array positioniert sind.
b) die (virtuell) vor dem Array fokussiert werden.

Durch endliche Integrationsgrenzen fallen nun die Beiträge der Quellen außerhalb des aktiven Bereichs des Arrays zur Rekonstruktion weg. Der so entstehende Fehlerterm ε (Gl. 2.82) ruft Artefakte hervor. Im synthetisierten Wellenfeld treten Schattenwellen auf, deren Ausbreitungszentren in den Unstetigkeitsstellen des Arrays liegen (Abb. 2.17a).

$$\varepsilon(\omega) = \sqrt{\frac{jk}{2\pi}} \sum_{m=-\infty}^{-(M+1)} P(y_m, \omega) \cos(\varphi_m) \frac{e^{-jk\Delta r_m}}{\sqrt{\Delta r}} \Delta y + \sqrt{\frac{jk}{2\pi}} \sum_{m=M+1}^{\infty} P(y_m, \omega) \cos(\varphi_m) \frac{e^{-jk\Delta r_m}}{\sqrt{\Delta r}} \Delta y \quad (2.82)$$

Diese Erscheinung verursacht Schwierigkeiten bei der Lokalisation. Wenn das Zeitintervall zwischen der erwünschten und der Schattenwellenfront etwa 50ms übersteigt, werden separate Echos hörbar. Für kürzere Zeitintervalle tritt spektrale Signalverfärbung (Kammfilter Effekt) auf, welche Einfluss auf die Auswertung binauraler Faktoren in der Hörwahrnehmung nimmt [Blauert, 1997],[de Vries, 1994].

Kompensationsquellen

Ein Weg zur Reduktion dieser Artefakte ist das Hinzufügen von Kompensationsquellen (Abb. 2.17c), die mit einem dem Fehlerterm ε angenäherten Fehlersignal beaufschlagt werden.

$$Q_m(\vec{r}_{edge}, \omega) = S(\omega) \sqrt{\frac{1}{2\pi jk}} \sqrt{\frac{\Delta x_0}{x_0 + \Delta x_0}} \frac{\cos \varphi_{edge}}{\sin \varphi_{edge} - \sin \beta_0} \frac{e^{-jkr_{edge}}}{\sqrt{r_{edge}}} \quad (2.83)$$

und so die Schattenwellen weitgehend eliminieren [de Vries, 1994]. Das Resultat sind deutlich reduzierte Störungen, deren Wellenfront aber immer noch klar erkennbar ist (Abb. 2.17d). Außerhalb des Reproduktionsbereichs vergrößert sich die Störung.

Fensterungsmethode (Tapering Window)

Die andere Methode beruht darauf, die Energie der Schattenwelle zeitlich und räumlich zu „verschmieren“ [Vogel, 1993]. Dazu blendet man die Amplituden der Lautsprecher signale an den Enden des Arrays z.B. mit einer Cosinusfunktion aus. Die Fehlerenergie wird dadurch zwar nicht kompensiert, die zeitliche Aufweitung verhindert aber ein hörbares Echo. Außerhalb des Reproduktionsbereichs werden die Artefakte ebenso abgeschwächt (Abb. 2.17b).

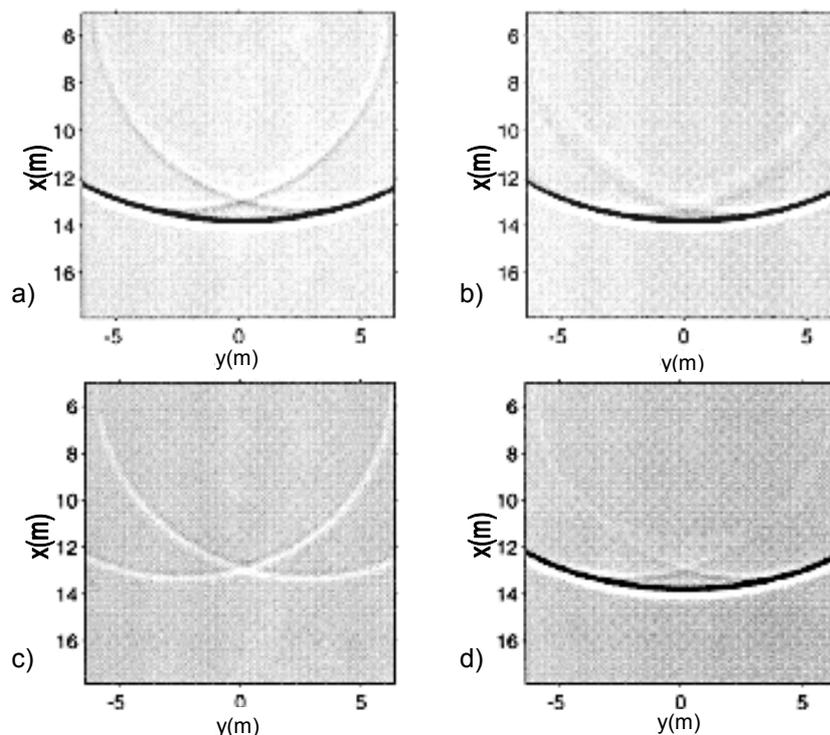


Abb. 2.17:

- a) Durch die endliche Länge des Arrays werden deutlich sichtbare Schattenwellen hervorgerufen.
- b) Die Anwendung der Fensterungsmethode (Tapering Window) „verschmiert“ die Artefakte.
- c) Das von den Kompensationsquellen produzierte Signal ε
- d) Ergebnis der Kompensationsmethode.

Zusammenfassend kann gesagt werden, dass die Störung durch endliche Abmessungen der Anordnung abnimmt, je kleiner der Beitrag der äußersten Lautsprecher zur Rekonstruktion des Schallfeldes ist, also je näher zum Zentrum des Arrays die virtuelle Quelle verschoben wird. Wenn eine Anordnung von mehreren Arrays Ecken aufweist, so werden dadurch im Vergleich zu den Fehlern durch die endliche Ausdehnung nur vernachlässigbare Störungen hervorgerufen [Verheijen, 1997].

2.3.4.2 Durch die Diskretisierung (Räumliches Aliasing)

Reale Anwendungen der Wellenfeldsynthese erfordern eine Näherung der infinitesimal fein angeordneten Sekundärquellen durch diskret positionierte Lautsprecher. Die bisher in diesem Kapitel entwickelten Integrale zur Berechnung der Antriebssignale, werden durch Summen

$$\int F(x) dx \rightarrow \sum_n F_n \Delta x \quad (2.84)$$

ersetzt. Analog zur Abtastung im Zeitbereich ergeben sich dadurch Einschränkungen bezüglich des Wiedergabefrequenzbereichs. Das räumliche Sampling Theorem

$$f_{\max} = \frac{c}{2\Delta x} \leq f_{\text{Nyq}} = \frac{c}{2\Delta x \sin \theta} \quad (2.85)$$

muss erfüllt sein, wobei Δx der Abstand zwischen den Lautsprechern ist.

Eine Begrenzung des Frequenzbandes der Primärquellen auf die Nyquistrate (siehe Gl. 2.85) würde zwar Aliasing verhindern, klangliche Einbußen sind aber nicht im Sinne einer qualitativ hochwertigen Audiowiedergabe [Start, 1997].

Reduktion der räumlichen Bandbreite

In Abbildung 2.18a ist der Fall eines Arrays, dessen Lautsprecher in zu großen Abständen Δx angeordnet sind. Es entsteht eine räumliche Überlagerung der Komponenten (Aliasing). Als praktikable Lösung bietet sich die räumliche Bandbreitenreduktion [Verheijen, 1997] an.

Der Vorgang gliedert sich in 3 Schritte:

- Reduktion der räumlichen Bandbreite durch Überabtastung des Eingangssignals der Lautsprecher und ein räumliches Anti-Aliasing-Filter. Das einfallende Schallfeld wird mit sehr feiner Auflösung abgetastet ($\Delta \ell = \pi/k_{x,N}^{\Delta \ell}$, Abb. 2.18b) und anschließend einer räumlichen Tiefpassfilterung ($k_x^{\infty} = k_{x,N}^{\Delta x} = \pi/\Delta x < \pi/\Delta \ell = k_{x,N}^{\Delta \ell}$, Abb. 2.18c) unterzogen.
- Abtasten (Resampling) dieses Wellenfeldes an den Positionen der Lautsprecher (A/D-Wandler). Die Periodizität im k_x - k -Diagramm (Abb. 2.18d) entspricht jener aus Abb. 2.18a, jedoch diesmal ohne räumliches Aliasing.
- Rekonstruktion des Wellenfeldes aus den erhaltenen Abtastwerten durch Ansteuerung von gerichteten Lautsprechern mit diesem adaptierten Eingangssignal (D/A-Wandler). Die Form des Rekonstruktionsfilters wird durch die Abstrahlcharakteristik der Lautsprecher bestimmt (Abb. 2.18e) und lässt sich durch spezielles Design der Abstrahlfläche gestalten [Kyriakakis, 1998], [Heil et al., 1992].

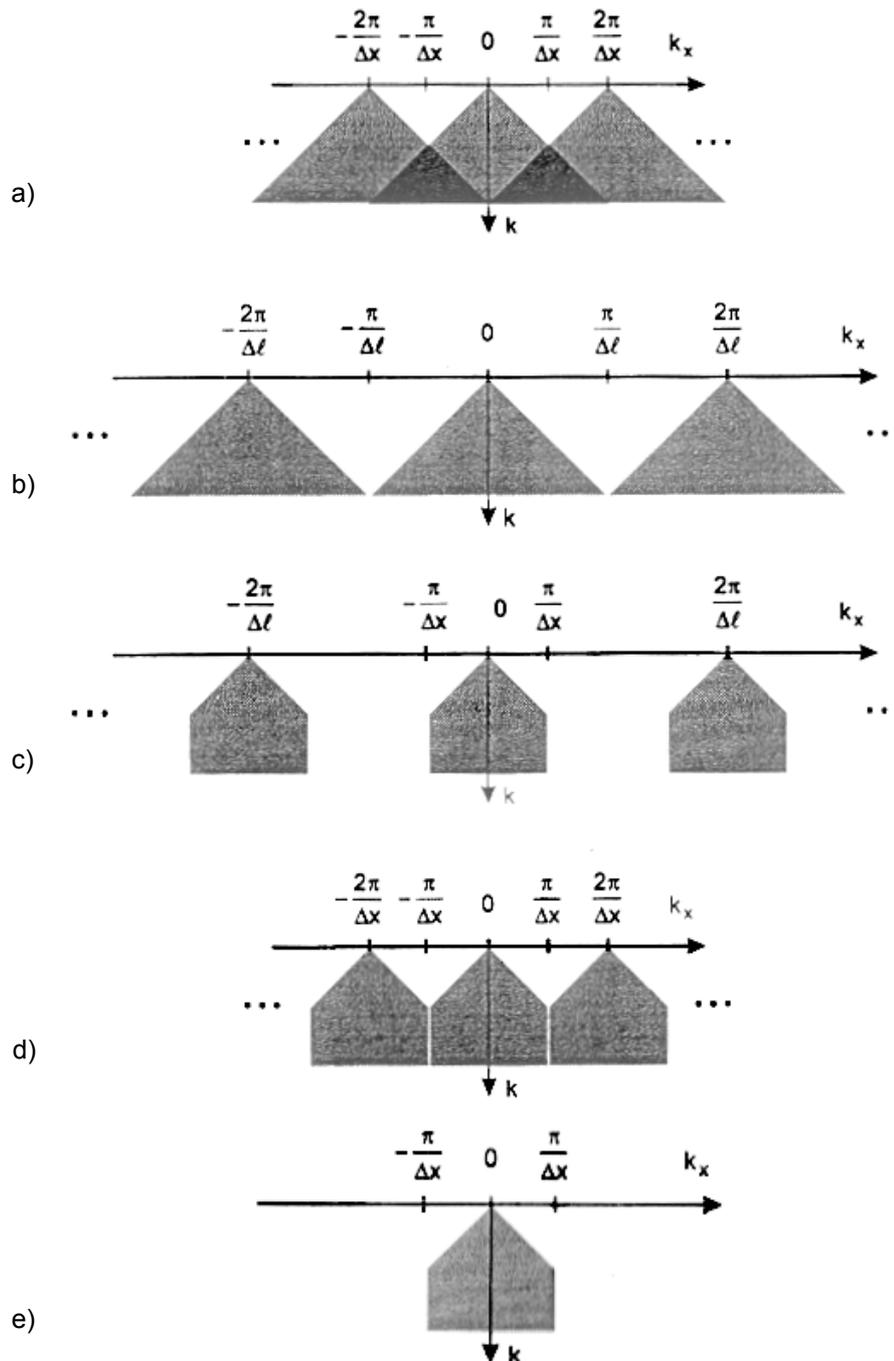


Abb. 2.18: Die Methode der räumlichen Bandbreitenreduktion [Start et al., 1995]

- a) Abtastung mit dem Intervall Δx , Aliasing entsteht.
- b) Überabtastung mit dem Intervall $\Delta \ell$, kein Aliasing
- c) Räumliche Tiefpass-Filterung, cut-off Frequenz $\pi/\Delta x$
- d) Resampling mit Δx
- e) Räumliches Rekonstruktionsfilter

2.4 Verknüpfung von Ambisonic und Wellenfeldsynthese

Die Wellenfeldsynthese bzw. allgemeiner der holophone Ansatz kann in den Ambisonic-Ansatz übergeführt werden. Bei der Holophonie wird das Schallfeld innerhalb eines abgeschlossenen quellenfreien Volumens durch die Kenntnis von Druck und Schnelle (partielle Normalkomponente) entlang der begrenzenden Volumenoberfläche beschrieben. Bei Ambisonic hingegen wird das Schallfeld durch die Kenntnis sämtlicher Richtungskomponenten (bzw. n-ten Ableitungen des Drucks) in einem zentralen Punkt beschrieben, wobei die Quellen unendlich weit entfernt sind und ebene Wellen abstrahlen. In beiden Fällen wird das Schallfeld (durch endlich großen Aufwand) vollständig beschrieben.

Im Folgenden wird gezeigt, dass die allgemeine Beschreibung eines Schallfeldes (Quellenverteilung) durch den holophonen Ansatz in die Beschreibung des Ambisonic Ansatz übergeführt werden kann. Dieser Sachverhalt wird exemplarisch für den allgemeinen 2-dimensionalen Fall nachgewiesen. Dabei werden folgende drei Annahmen vereinbart (vgl. [Nicol, Emerit, 1999]).

- *Das originale Schallfeld ist eine ebene Welle. Das stellt keine Einschränkung der Allgemeinheit dar, da jede Welle als Superposition von ebenen Wellen dargestellt werden kann.*
- *Die sekundären Quellen sind entlang einer kreisförmigen Linie angeordnet.*
- *Die sekundären Quellen sind vom "Abhörgebiet" weit genug entfernt, sodass ihre abgestrahlten akustischen Felder wiederum durch ebene Wellen beschrieben werden können.*

Der Übergang von der 3-dimensionalen Problemstellung, die durch das Kirchhoff-Helmholtz Integral (siehe Gl. (2.1)) beschrieben wird, zum 2-dimensionalen Fall (Wellenzustände in einer Ebene), bei dem die sekundären Quellen entlang des Kreisumfangs verteilt sind, erfolgt durch die 'Stationary Phase Method' (vgl. Kapitel 2.3.3.2). Als Integrationsfläche wird ein unendlich ausgedehnter Zylinder mit Radius r_0 um die z-Achse angenommen (zweite Annahme), der wiederum als Überlagerung von entlang z verlaufenden Linienquellen gedacht wird.

$$p(\vec{r}) = \int_0^{2\pi} \int_{-\infty}^{\infty} \left[\vec{\nabla} p_0 \cdot \vec{n} - \frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4R\pi} r_0 \cdot dz_0 \cdot d\phi_0 \quad \forall \vec{r} \in \Omega \quad (2.86)$$

Will man nun die Feldzustände innerhalb einer bestimmten Ebene z.B. in $z=z_0$ beschreiben, so kann man das unendliche Integral unter Voraussetzung der Gültigkeit der 'Stationary Phase Method' wie folgt berechnen:

$$I = \int_{-\infty}^{+\infty} \left[\vec{\nabla} p_0 \cdot \vec{n} - \frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4R\pi} dz_0 \cong \frac{1}{2\sqrt{jk2\pi}} \cdot \left[\vec{\nabla} p_0 \cdot \vec{n} - \frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{\sqrt{R}} \Big|_{z_0=0}$$

Wird für das allgemeine (jedoch zeitfrei) zu beschreibende Schallfeld p_{org} eine horizontale ebene Welle (Amplitudenbetrag a , Wellenzahl k und Einfallswinkel θ) angenommen (*erste Annahme*), die für jeden beliebigen Punkt (r_0, ϕ_0) in einer Ebene durch $p_{\text{org}}(r_0, \phi_0) = a \cdot e^{jk r_0 \cos(\phi_0 - \theta)}$ beschrieben wird, so ergibt sich für den Druck innerhalb des Kreises mit dem Radius r_0 :

$$p(\vec{r}) = p(r, \phi) = \frac{r_0}{2\sqrt{jk2\pi}} \int_0^{2\pi} \left[\frac{\partial p_0}{\partial r_0} - \cos \alpha (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{\sqrt{R}} d\phi_0 \quad \text{für } r < r_0 \quad (2.87)$$

$$\text{mit } R = \sqrt{r^2 + r_0^2 - 2rr_0 \cos(\phi - \phi_0)} \quad \text{und } \cos \alpha = \frac{r \cos(\phi - \phi_0) - r_0}{R}$$

Der Vektor \vec{R} beschreibt den Ausbreitungspfad zwischen der Sekundärquelle an der Stelle \vec{r}_0 und dem Aufpunkt \vec{r} (vgl. Abbildung 2.2). Die holophone Beschreibung des Schallfeldes in Gl. (2.86) beschreibt das allgemeine Schallfeld durch die Ausbreitungsterme von Monopol-

$$\left(\frac{e^{-jkR}}{4R\pi} \right) \text{ als auch von Dipolquellen } \left(\frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \cdot \frac{e^{-jkR}}{4R^2\pi} \right).$$

Befinden sich nun die sekundären Quellen weit genug vom Hörer (Aufpunkt) entfernt (*dritte Annahme*) d. h. dies ist der Fall, wenn die Entfernung größer als die betrachtete maximale Wellenlänge ist: $|\vec{r}_0| > \lambda_{\text{max}}$, dann gilt der Übergang $\lim_{r_0 \rightarrow \infty}$ und somit ergeben sich folgende Näherungen:

$$(1) \quad \lim_{r_0 \rightarrow \infty} \frac{1 + jkR}{R} = jk$$

$$(2) \quad \lim_{r_0 \rightarrow \infty} \cos \alpha = -1$$

$$(3) \quad \lim_{r_0 \rightarrow \infty} \frac{e^{-jkR}}{\sqrt{R}} = \frac{e^{-jkr_0}}{\sqrt{r_0}} \cdot e^{jkr \cos(\phi - \phi_0)}$$

Unter Verwendung der oben genannten Näherungen ergibt sich für den Druck aus Gl. (2.87) nun folgende Vereinfachung:

$$p(\vec{r}) = p(r, \phi) = \frac{\sqrt{r_0}}{2\sqrt{jk2\pi}} \cdot e^{-jkr_0} \int_0^{2\pi} \left[\frac{\partial p_0}{\partial r_0} + jkp_0 \right] e^{jkr \cos(\phi - \phi_0)} d\phi_0 \quad \text{für } r < r_0 \quad (2.88)$$

Im 2-dimensionalen Fall kann eine ebene Welle, ähnlich wie Abschnitt 2.2.1, durch eine Reihenentwicklung in zylindrische Bessel- und Kosinusfunktionen dargestellt werden. Die Lösungsschar, die durch die Variablen-Separation für $\Theta(\theta)$ erhalten wurde entfällt hier. Für die Annahme $\lim_{r_0 \rightarrow \infty}$ kann die Besselfunktion durch ihre asymptotische Näherung ersetzt werden (vgl. [Bronstein et al., 1999] S. 506, Gl. 9.56a).

$$\lim_{r_0 \rightarrow \infty} J_n(kr_0) = \sqrt{\frac{2}{kr_0\pi}} \cdot \cos \left[kr_0 - \left(n + \frac{1}{2}\right) \cdot \frac{\pi}{2} \right]$$

Wendet man diese Eigenschaft auf den Term in den eckigen Klammern in Gl. 2.88 an, so ergibt sich folgende weitere Vereinfachung:

$$\lim_{r_0 \rightarrow \infty} \left[\frac{\partial p_0}{\partial r_0} + jkp_0 \right] = a \cdot \sqrt{\frac{2jk}{r_0\pi}} \cdot e^{jkr_0} \left[1 + 2 \sum_{m=1}^{+\infty} \cos[m(\phi_0 - \theta)] \right]$$

Durch einsetzen in Gl.2.88 erhält man:

$$p(\vec{r}) = p(r, \phi) = \frac{a}{2\pi} \int_0^{2\pi} \left[1 + 2 \sum_{m=1}^{+\infty} \cos[m(\phi_0 - \theta)] \right] e^{jkr \cos(\phi - \phi_0)} d\phi_0 \quad \text{für } r < r_0 \quad (2.89)$$

Diskretisiert man das Integral in Gl. 2.89 und somit auch die bisher kontinuierlich betrachtete Quellenverteilung entlang des Zylinderumfanges so Linie, so erhält man eine Summendarstellung und der Normierungsfaktor 2π wird durch die Anzahl N der diskreten Punkte ersetzt. Diese N Punkte können als sekundäre Quellen interpretiert werden, die entlang einer Kreislinie angeordnet sind. Somit schreibt sich die Gl. 2.89 nun neu wie folgt:

$$p(\vec{r}) = p(r, \phi) = \frac{a}{N} \sum_{i=0}^{N-1} \left[1 + 2 \sum_{m=1}^M \cos[m(\phi_i - \theta)] \right] \cdot e^{jkr \cos(\phi - \phi_i)} \quad (2.90)$$

mit N Anzahl der sek. Quellen, ϕ_i der Position der sekundären Quellen, θ dem Einfallswinkel der primären Welle und M dem Approximationsgrad der asymptotisch genäherten Besselreihe.

Das Ergebnis in Gl. 2.90 ist identisch mit der Beschreibung des Schallfeldes im Ambisonic Domain (vgl. Gl. 2.36 in Abschnitt 2.2.1). Diese Beschreibung des Schallfeldes wird in der Literatur auch als Quellenverteilung ('source distribution') bezeichnet.

Die Quellenverteilung für eine ebene Welle mit der komplexen Amplitude A kommend aus der Richtung θ_0 ist eine Deltafunktion an der Stelle $\theta = \theta_0$ und kann wie folgt beschrieben werden:

$$S_{0,\theta_0}(\theta) = 2\pi A \cdot \sum_{m=-\infty}^{m=\infty} \delta(\theta - \theta_0 - 2m\pi) = A \cdot \sum_{m=-\infty}^{m=\infty} e^{jm(\theta - \theta_0)} \quad (2.91)$$

Wird ein (2D) Ambisonic System M.Ordnung verwendet, so ergibt sich durch den Abbruch der Reihenentwicklung in Gl. 2.91 das Quellenspektrum zu

$$S_{0,\theta_0,M}(\theta) = A \cdot \sum_{m=-M}^{m=M} e^{jm(\theta - \theta_0)} = A \left\{ 1 + 2 \cdot \sum_{m=1}^M \cos[m(\theta - \theta_0)] \right\} \quad (2.92)$$

In der Arbeit wird diese Form, je nach Anwendung, auch als winkelabhängiges Lautsprecher-signal oder Richtcharakteristik bezeichnet (vgl. Kapitel 3 und Kapitel 4).

Ausgehend von der Eigenschaft, dass sich der holophone und der Ambisonic Ansatz verbinden lassen und die Beschreibung des Schallfeldes durch das Quellenspektrum eine kompakte Notation darstellt, sind damit die Grundlagen für eine Aufnahme von höheren Ambisonic Ordnungen bereits gelegt. In [Poletti, 2000] und [Nicol, Emerit, 1999] wird ein Aufnahmesystem mit kreisförmig angeordneten Mikrofonen vorgestellt. Durch die Verwendung dieser kompakten Notation wird in [Poletti, 2000] eine Herleitung einer geschlossenen Beschreibung der Aufnahmerichtcharakteristiken möglich. Auf deren Aufbau und die damit verbundenen Probleme wie Phasen- und Amplitudenkompensation wird hier nicht näher eingegangen.

Zusammenfassung

Es wurde gezeigt, dass die Beschreibung des Schallfeldes mit Hilfe des holophonen Ansatzes identisch (unter Verwendung allgemeiner Annahmen) mit jener des Ambisonic Ansatzes ist. Dadurch ist eine vollständige Beschreibung des Schallfeldes durch beide Ansätze möglich. Durch deren Äquivalenz lassen sich folgende Eigenschaften ableiten:

Aus der theoretischen Betrachtung ist der Ambisonic Ansatz gegenüber dem Holophonie Ansatz zu bevorzugen. Die räumliche Kodierung ist effizienter und die Dekodierung ermöglicht eine perfekte Wiedergabe. Begrenzt ist dieser Ansatz nur durch Aufnahme der Ambisonic Signale höherer Ordnungen. Beim Holophonie Ansatz kommen zusätzliche Fehler durch die ‚stationary phase approximation‘ und durch räumliches Aliasing bei der Wiedergabe hinzu.

In der Praxis ergeben sich jedoch bei der Aufnahme von Ambisonic Signale höherer Ordnung massive Probleme, sowie die Wiedergabe von ebenen Wellen stellt Einschränkungen dar. Daher erscheint der holophone Ansatz brauchbarer.

Welcher Ansatz im Einzelfall geeigneter ist, hängt vom Anwendungsbereich und von den Forderungen ab.

Kapitel 3

Lautsprecherapplikationen

Der IEM CUBE

Objektive Beurteilungsgrößen für Schallfelder

Kompensation von Raumreflexionen

3D Verhallungskonzepte

Distanz & Positionierung von virtuellen Quellen

Im folgenden Kapitel wird ein realisiertes System, der IEM CUBE¹, vorgestellt. Neben der Live-Verstärkung von natürlichen Quellen und Wiedergabe von gespeicherten Audiodaten in den üblichen standardisierten Wiedergabeformaten können auch die raumakustischen Parameter gezielt verändert werden.

Für die Beschreibung der Wiedergabeeigenschaften des Systems werden objektive Beurteilungsgrößen vorgestellt. Die Einführung des komplexen Schnellektors, sowie die des aktiven Intensitätsvektors ermöglicht eine repräsentative grafische Darstellung und objektive Beurteilung von holophonen Systemen. Darüber hinaus können diese Größen bei der Entwicklung und Optimierung von zusätzlichen Applikationen verwendet werden.

Zur Erzeugung einer variablen steuerbaren Akustik werden Ansätze zur Reduktion der raumcharakteristischen ersten diskreten Reflexionen sowie Grundlagen der dreidimensionalen Verhallung vorgestellt.

Durch die Verbindung der in Kapitel 2 vorgestellten Ansätze wird ein modulares, skalierbares akustisches Rendering Verfahren für „virtuelle „*Soundscapes*“ präsentiert.

3.1 Der IEM – CUBE

Der IEM CUBE dient als Versuchslabor für virtuelle Realitäten und ist gleichsam ein Musikinstrument für Komponisten und Musiker. In dieser Arbeit wird auf die Umsetzung von virtueller Raumakustik eingegangen. Ziel ist es die Hörwahrnehmung derart gezielt und reproduzierbar zu beeinflussen, dass technisch erzeugte Reize von ihren natürlichen Pendanten nicht mehr unterscheidbar sind (vgl. Kapitel 1.1).

3.1.1 Wahl des Renderingalgorithmus

Im IEM CUBE sollte eine periphone Wiedergabelösung realisiert werden. Mit holophonen Ansätzen sollten sich eine beliebige Anzahl von virtuellen Quellen an beliebigen Orten abbilden lassen. Des Weiteren sollten die gängigen standardisierten Wiedergabeformate wie Mono, Stereo, 3-4-5 Kanal Stereo, Quadrophonie, Surround (5.1 bzw. 3/2 Format) diskret wie auch virtuell realisierbar sein. Neben der Live-Verstärkung von natürlichen Quellen und Wiedergabe von gespeicherten Audiodaten sollten die raumakustischen Parameter gezielt verändert werden können.

Der eigentliche holophone Ansatz (vgl. Kapitel 2.1, BSC) eignet sich für solche Anforderungen nicht, da für den Adaptionsprozess echte Holophonie-Aufnahmen benötigt werden.

Die Wellenfeldsynthese ist ebenso relativ aufwendig. Vor allem bei der Reproduktion von dreidimensionalen (periphonen) Schallfeldern steigt der Aufwand immens an.

¹ Cube: Computer unterstützte Beschallungseinheit.

Der Ambisonic Ansatz ist nicht nur weniger komplex, sondern auch besser skalierbar, wodurch er sich für eine Vielzahl von Anwendungen und Umgebungen eignet. Aus diesem Grund wurde entschieden, die Beschallungsanlage auf ‚Ambisonic‘-Basis zu implementieren.

3.1.2 Anforderungen an das Beschallungssystem

Die Anforderungen an die Hardware wird bestimmt durch die Forderung nach hoher, gleich bleibender Klangqualität für alle Wiedergabekanäle. Folgende Kriterien sind die Grundlage bei der Auswahl der technischen Ausrüstung:

- Linearer Frequenzgang im hörbaren Bereich
- Ausreichende Dynamik
- Signal-Rauschabstand
- Latenzzeit
- Ausreichend große Anzahl von Kanälen
- Sicherheit

Die Betriebssicherheit des Gesamtsystems ist für Aufführungen von großer Bedeutung. Trotz großer Dynamik darf es zu keinen systembedingten gesundheitsschädigenden¹ Lautstärken kommen. Ein wichtiger Punkt stellt auch das Synchronisationskonzept dar. Bei Ausfall einer Komponente des Systems (u.U. Verlust der Steuerbarkeit) müssen die Ausgänge stumm geschaltet oder zumindest auf einen verträglichen Pegel begrenzt werden.

Bei mehrkanaligen Beschallungsanlagen und der Annahme von kohärenten Signalen kommt es im Hörraum zu Addition der Signalamplituden der K Lautsprecher (LS). Der maximal zulässige Pegel berechnet sich bei diesem Sonderfall für einen Dynamikumfang von 120dB_A wie folgt:

$$L_{\text{Gesamt}} = 20 \log \frac{K \cdot p_{\text{EinzelLS}}}{p_0} = 20 \log K + L_{\text{EinzelLS}} < 120\text{dB}_A \quad (3.1)$$

Das Signal eines einzelnen Lautsprechers sollte bei digitaler Vollaussteuerung einen Pegel von

$$L_{LS,\text{max}} = 120 - 20 \log K \quad (3.2)$$

nicht übersteigen. Somit ist der Einsatz von nichtlinearen Modulen (Limiter²) unabdingbar.

¹ Die Schmerzgrenze liegt über 120dB_A . Abhängig von Pegel und Wirkungsdauer kann eine derartige Schalleinwirkung auch zu irreversiblen Hörschäden führen.

² Limiter dienen nur zur Sicherheit und sollten den Wiedergabebereich so wenig wie möglich verändern. Kurzfristige Pegelspitzen sollten dabei geringer als lang anhaltende hohe Pegel bedämpft werden.

3.1.3 Module

Da die Signalverarbeitung rein digital erfolgt, kann die Beschallungseinheit formal in folgende drei Abschnitte unterteilt werden:

- **Eingangsstufe (analog/digital)**

Die vorliegenden Signale werden der Verarbeitungsstufe in digitaler Form angeboten. Jedes Signal kann statische (Die Zuweisung der Position erfolgt in der darauffolgenden Einheit.) oder dynamische Positionen (Die Zuweisung von Azimut, Elevation und Entfernung erfolgt durch zusätzliche MIDI Daten.) besitzen.

- **Verarbeitungsstufe (digital)**

Diskrete virtuelle Quellen (z.B. auch virtuelle Lautsprecher bei der Wiedergabe von herkömmlichen Stereoaufnahmen) werden in das Ambisonic-Format kodiert. Es erfolgt eine Summation der Ambisonic-Repräsentationen der einzelnen Quellen. Im Ambisonic-Bereich können die in Kapitel 2 erwähnten Modifikationen recheneffizient realisiert werden. Des Weiteren erfolgt in dieser Stufe das Einschleifen von Effekten sowie die Umsetzung der in den folgenden Abschnitten 3.3 bis 3.5 vorgestellten Applikationen. Abschließend wird die Ambisonic-Repräsentation auf die reale Lautsprecher-Konfiguration dekodiert.

- **Wiedergabestufe (analog)**

Die vom Ambisonic-Dekoder berechneten diskreten Lautsprechersignale werden den realen Lautsprechern zugeordnet.

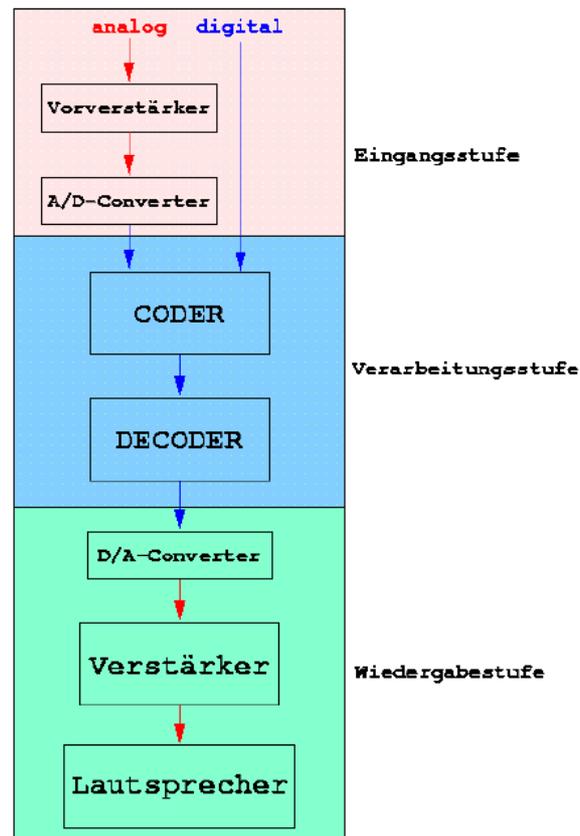


Abb. 3.1: Hardware Grundstruktur – 3 stufiges Konzept (vgl. [Zmölnig, 2002]).

Das Kernstück der Hardware ist die Kodier- und Dekodiereinheit. Die Implementierung erfolgt auf einer allgemeinen Hardwareplattform (PC). Dadurch sind im Gegensatz zu einer DSP-Lösung Hard- und Software nicht zu eng miteinander verbunden. Die Programmierung erfolgt in der grafischen Echtzeit-Programmiersprache PD¹. Dieser Code ist unabhängig von der speziellen Hardware, so dass Hard- und Softwarekomponenten voneinander unabhängig erweitert und erneuert werden.

3.1.4 Wahl der Lautsprecheranordnung

Die Lautsprecheranzahl wurde aus praktischen Kriterien - wie baulichen Gegebenheiten, zur Verfügung stehenden Übertragungskanäle und natürlich dem verfügbaren Budget - auf 24 festgelegt. Damit lässt sich allgemein ein 3-dimensionales Ambisonic-Schallfeld 3. Ordnung resynthetisieren (vgl. dazu Kapitel 2: $(3+1)^2 = 16$). Um eine möglichst gleichmäßige Ambisonic-Dekodierung zu ermöglichen - sollte wie bereits im Kapitel 2 beschrieben - eine Cluster-Bildung vermieden und die Lautsprecher möglichst in den Knoten eines regelmäßigen Polyeders angeordnet werden. Es wurde daher eine Lösung angestrebt, bei der die Lautsprecher auf einer Halbkugel über den Köpfen des Publikums angeordnet sind. Die Wiedergabe erfolgt somit für die obere Hemisphäre, d.h. Schallquellen von unten können prinzipiell nicht realisiert werden.

¹ PD steht für Pure Data. Weitere Informationen sowie Software Download siehe URL: <http://iem.kug.ac.at/pd/>.

Um die „Ebenen-Wellen“-Voraussetzung für die optimale Funktion des Ambisonic-Ansatzes (siehe Abschnitt 2.2.1) zu gewährleisten, ist eine maximale Entfernung der Lautsprecher zum Auditorium anzustreben. Die Lautsprecher entlang einer Kugeloberfläche und gleichzeitig in jeweils maximaler Entfernung vom zentralen Hörer zu montieren ist aus den vorgegebenen räumlichen Bedingungen nicht möglich. Die möglichen Lautsprecherpositionen beschreiben somit die Oberfläche eines Rotationsellipsoids. Die Abweichung von den idealen Positionen (Kugeloberfläche) kann durch eine Kalibrierung des Gesamtsystems (d.h. Verstärkungs- und Laufzeitkorrektur der einzelnen Lautsprecher, vgl. Kapitel 1.2) kompensiert werden. Liegen die Lautsprecher möglichst nahe an den begrenzenden Raumflächen, dann können zusätzliche Klangfärbungen aufgrund von Reflexionen vermieden werden. Störende Kammfiltereffekte treten somit erst bei höheren Frequenzen auf. Diese sind jedoch vernachlässigbar, da mit zunehmender Frequenz eine gerichtete Abstrahlung zu erwarten ist.

Die regelmäßige Verteilung der Lautsprecher über eine Hemisphäre stellt ein nicht-triviales Problem dar. Aus den folgenden Kriterien wurde ein vereinfachter Ansatz für die Anordnung gefunden:

- *Pro Flächenanteil sollten gleich viele Lautsprecher vorhanden sein.*
(Energetische Betrachtung)
- *Diskrete Realisierung von standardisierten Formaten.*
- *Ausreichende Winkelauflösung*

Da das menschliche Gehör eine bessere Lokalisationsfähigkeit in der Horizontalebene hat [Blauert, 1997], ist es sinnvoll in der Horizontalen relativ mehr Lautsprecher anzubringen als in der Höhe. Somit liegt die einfache Idee nahe die 24 Lautsprecher in drei (in sich symmetrischen) Ringen anzuordnen, wobei die Anzahl der Schallwandler pro Ring nach oben hin abnimmt (z.B. 12, 8 und 4).

Damit die Decodermatrix gut konditioniert ist (vgl. Abschnitt 2.2, Dekodierungsvorschrift), müssen die minimalen Abstände zwischen benachbarten Lautsprechern maximiert werden. Durch eine symmetrische Verteilung entlang eines Ringes und eine Verdrehung der Lautsprecherringe zueinander können diese Abstände auf relativ einfache Weise optimiert werden.

Bei der Anordnung von drei Lautsprecherringen mit je 12, 8 und 4 Lautsprechern ergibt sich nach dem obigen energetischen Kriterium für die Oberflächen eine Proportionsvorgabe von 10:6:2 (siehe Abbildung 3.2). Dabei liegen die Lautsprecherringe jeweils auf der Schwerelinie der einzelnen Teilflächen. Der erste Lautsprecherring ist in einer Höhe von 1.4m angebracht. Diese Höhe entspricht der unteren ‚Schallquellenkante‘ und ist ein geeigneter Mittelwert für sitzende und stehende Personen. Für die reale Lautsprecheranordnung, die durch einen Rotationsellipsoiden beschrieben werden kann, ergeben sich somit aus den mit der

Lautsprecheranzahl gewichteten Flächenverhältnissen V die relativen Höhen H_1 und H_2 der beiden oberen Lautsprecherringe (siehe Abbildung 3.2). Eine mathematische Herleitung dieser beiden Höhen findet sich im Anhang.

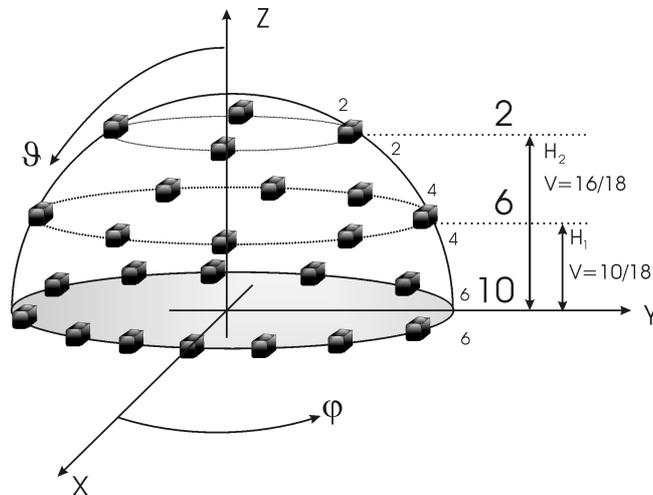


Abb. 3.2: Flächenproportionen für eine 3 Lautsprecherring Anordnung zu je 12, 8 und 4 Lautsprecher.

Damit die Lautsprecheranordnung auch der gewünschten Forderung der diskreten Realisierung der standardisierten Formate wie 5.1 nachkommt, musste in den beiden vorderen Quadranten des untersten Lautsprecherringes eine höhere Winkelauflösung vorgesehen werden. Da das menschliche Gehör vorne die geringste Lokalisationsunschärfe aufweist, hat diese Wahl auch eine zusätzliche Verbesserung für das Gesamtsystem zur Folge. Diese Anordnung wurde so optimiert, dass die idealen Abhörpunkte der unterschiedlichen diskreten Reproduktionsformate mit dem Referenzhörort des gesamten Systems bestmöglich übereinstimmen.

Eine ausreichende Winkelauflösung ist zum Einen dann erreicht, wenn durch den Panning-Algorithmus bei gegebener Lautsprecheranordnung eine virtuelle Schallquelle möglichst glatt¹ im Quellenraum² bewegt werden kann.

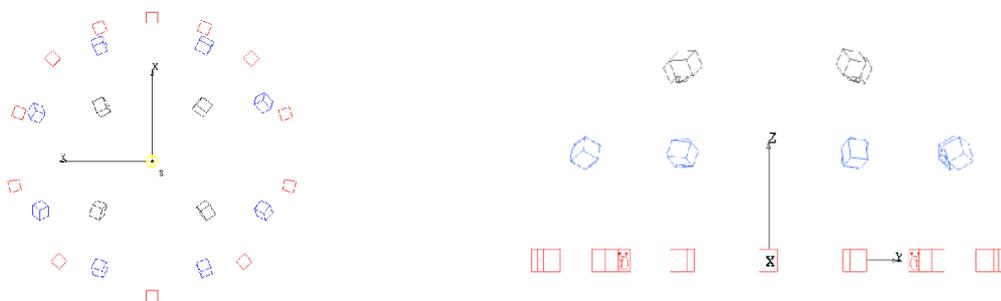


Abb. 3.3: (links) Grund- und (rechts) Seitenriss der Lautsprecheranordnung des IEM CUBE (vgl. [Zmölnig, 2002]).

¹ Der Begriff ‚glatt‘ beschreibt hier die Empfindung der wahrgenommenen Quellenaufweitung. Eine Bewegung wird als ‚glatt‘ empfunden, wenn sich die wahrgenommene Quellenaufweitung winkelunabhängig ist, also für Positionen zwischen Lautsprecher nicht von jenen an einer Lautsprecherposition unterscheidet.

² Als Quellenraum wird jener Bereich bezeichnet, in dem virtuelle Quellen abgebildet werden können.

Die Minimalvariante für dreidimensionales Ambisonic 3. Ordnung benötigt 16 Lautsprecher. Verglichen mit den theoretischen Betrachtungen sind somit zu viele Lautsprecher vorhanden. Ergänzt man darüber hinaus die hemisphärische Anordnung gedanklich auf eine vollständige sphärische Anordnung, so stehen 36 Lautsprecher¹ zur Verfügung, wobei natürlich virtuelle Quellen mit negativer Elevation ausgeschlossen sind. Daher ist ein vollständiges Ambisonic System 5. Ordnung (d.h. $(5+1)^2=36$ Übertragungskanäle, die entsprechend auf die 24 Lautsprecher dekodiert werden müssen) realisierbar. Um Übertragungskanäle zu sparen, wurde eine gemischte Kodierung (vgl. Kapitel 2.2.3 Kombinierte Systemordnungen) mit 5. Ordnung horizontal und 3. Ordnung vertikal, somit 18 Übertragungskanälen gewählt.

Ring	Lautsprechernummer											
I	1	2	3	4	5	6	7	8	9	10	11	12
<i>j</i>	0°	22°	45°	71.5°	102°	138.5°	180°	221°	258°	288°	315°	338°
<i>J</i>	90°	90°	90°	90°	90°	90°	90°	90°	90°	90°	90°	90°
II	13	14	15	16	17	18	19	20				
<i>j</i>	25°	69°	116°	155°	205°	244°	296°	335°				
<i>J</i>	62°	62°	62°	62°	62°	62°	62°	62°				
III	21	22	23	24								
<i>j</i>	45°	135°	225°	315°								
<i>J</i>	30.5°	33°	33°	30.5°								

Tabelle 3.1 Übersicht und Angabe der Lautsprecherpositionen

Eine detaillierte Beschreibung der verwendeten Hard- und Software für die Umsetzung des Beschallungskonzeptes findet sich in [Zmölnig, 2002].

3.2 Objektive Beurteilungsgrößen für Schallfelder

Zur Beschreibung der akustischen Eigenschaften eines realen Schallfeldes gibt es in der Raumakustik (vgl. [Beranek, 1996],[Griesinger, 1997]) bzw. in der Beschallungstechnik (vgl. [Ahnert, Steffen, 1993]) zahlreiche subjektive und objektive Beurteilungsgrößen. Diese Größen sind beim Entwurf und bei der Simulation von synthetisch erzeugten Schallfeldern jedoch nur bedingt verwertbar, da hier vor allem die korrekte Lokalisation der abgebildeten Schallquellen im Vordergrund steht. Der radiale Fehler (vgl. Kapitel 2, Abb. 2.5) ist für eine Fehlerabschätzung über einen größeren Bereich brauchbar, jedoch kann über die

¹ Der unterste Ring (I) liegt am Äquator der gedachten Kugeloberfläche. Somit gehören 6 Lautsprecher zur oberen Hemisphäre und die verbleibenden 6 zur unteren. Daher berechnet sich die Anzahl der Lautsprecher für die gesamte Kugeloberfläche zu: $(24-6)*2=36$.

wahrgenommene Richtung und die Qualität dieses Eindrucks mit herkömmlichen Beschreibungsgrößen keine Aussage getroffen werden. Daher werden für die qualitative und quantitative Beurteilung und Optimierung der hier zu untersuchenden Systemen Größen benötigt, welche mit der Perzeption des reproduzierten Schallfeldes im Mittelpunkt (bzw. im idealen Abhörpunkt) der Wiedergabeeinheit korrelieren. Dazu zählen der Schnelle- und der Energievektor (vgl. Kapitel 1). Gerzon hat in [Gerzon, 1992d] den Schnellevektor als brauchbare Beschreibungsgröße für die einfallende Richtung einer tieffrequenten ($f < 700\text{Hz}$) Schallwelle bezeichnet und den Energievektor für die Beschreibung von höherfrequenten Komponenten vorgeschlagen.

Da der menschliche Kopf eine endliche Ausdehnung besitzt, sind jene Größen, welche die Eigenschaften in einem größeren Bereich beschreiben meist hilfreicher. Alternative Ansätze finden sich in der Literatur, die den Druck am Ort der Ohren des Hörers durch Verwendung der HRTFs bestimmen und die daraus resultierenden interauralen Zeit- und Pegeldifferenzen berechnen. Daniel [Daniel, 2000] hat diesen Ansatz für die Untersuchung der Richtungsabbildung von 2-dimensionalen Ambisonic-Systemen verwendet.

Im Folgenden werden Feldgrößen vorgestellt, die eine Visualisierung des Schallfeldes ohne Einbeziehung des Hörers ermöglichen und ebenfalls ein Maß für die holophonen Eigenschaften des Systems darstellen. Die Herleitung erfolgt für den zweidimensionalen Fall. Es wird dabei zwischen Schnelle-, Intensitäts- und Energiefluss-Methoden unterschieden.

3.2.1 Schnelle-Methode

Der Schnellevektor entspricht der normalisierten Schnelle im Zentrum des Wiedergabebereiches (für $r=0$), wobei die Lautsprecher entlang einer Kreislinie an den Winkeln \mathbf{q}_n positioniert sind und eine ebene Welle mit der Amplitude a_n abstrahlen. Somit ergibt sich der Schnellevektor zu:

$$\mathbf{v} = \frac{\sum a_n \cos(\mathbf{q}_n)}{\sum a_n} \cdot \mathbf{i} + \frac{\sum a_n \sin(\mathbf{q}_n)}{\sum a_n} \cdot \mathbf{j} \quad (3.3)$$

wobei \mathbf{i} und \mathbf{j} die Einheitsvektoren entlang der x bzw. y Achse repräsentieren.

In [Daniel et al., 1999] wurde eine verallgemeinerte Form des Schnellevektors für beliebige Schallfelder vorgeschlagen, der im Punkt \mathbf{r} wie folgt beschrieben werden kann:

$$\mathbf{V}(\mathbf{r}) = \mathbf{r} \cdot c \cdot \frac{\mathbf{v}(\mathbf{r})}{p(\mathbf{r})} \quad (3.4)$$

wobei \mathbf{r} die (Normal-)Dichte der Luft, c die Schallausbreitungsgeschwindigkeit, \mathbf{v} den Schnellevektor und p den Schalldruck beschreibt. Diese Größe kann zur Beschreibung der tieffrequenten Richtungslokalisierung verwendet werden und ist zeitunabhängig.

Für komplexwertige monochromatische Felder mit der allgemeinen Form

$$p(\mathbf{r}, t) = |p(\mathbf{r}, t)| \cdot e^{j \cdot [\mathbf{w} \cdot t - f(\mathbf{r})]} = P(r, t) \cdot e^{j \cdot [\mathbf{w} \cdot t - f(\mathbf{r})]} \quad (3.5)$$

ergibt sich die Schnelle aus der Bewegungsgleichung¹ $-\nabla p = \mathbf{r} \cdot \partial_t \mathbf{v}$ bei linearer Schallausbreitung unter Freifeldbedingungen zu

$$\mathbf{v}(\mathbf{r}) = \frac{1}{k \cdot \mathbf{r} \cdot c} (\nabla f + j \frac{\nabla P}{P}) p \quad (3.6)$$

wobei die Positions- und Zeitabhängigkeit noch implizit vorhanden ist. Durch Einsetzen der Gl. 3.5 und 3.6 in Gl. 3.4 erhält man den verallgemeinerten Schnellevektor wie folgt

$$\mathbf{V} = \frac{1}{k} (\nabla f + j \frac{\nabla P}{P}) = \Omega + j\Phi \quad (3.7)$$

Dieser Schnellevektor ist definiert für jeden Punkt in einem Schallfeld und kann daher zur Untersuchung und Beschreibung von Systemen in größeren Bereichen verwendet werden. Der Realteil des Schnellevektors $\Omega = \nabla f/k$ beschreibt die wahrgenommene Richtung, da der Phasengradient in die Ausbreitungsrichtung zeigt. Der Imaginärteil $\Phi = 1/k (\nabla P/P)$ wird hingegen zur Beschreibung der Quellenaufweitung („image broadening“) bzw. der Phasigkeit („phasiness“) verwendet. Bei der Verwendung von mehreren Schallquellen entstehen durch die Überlagerung der Einzelfelder Bereiche mit geringen Druckamplituden (destruktive Interferenz). In diesen Bereichen wird der Imaginärteil des Schnellevektors sehr groß, d.h. der Begriff „Phasigkeit“ wird mit jenen Bereichen verbunden, wo weitgehende Schallauslöschung stattfindet.

3.2.2 Schallintensitäts-Methode

Für das oben beschriebene monochromatische Schallfeld ergibt sich die komplexwertige Momentan-Intensität nach [Heyser, 1986]:

$$\mathbf{I}_c = \frac{1}{2} \cdot [p(\mathbf{r}, t) + j \cdot \hat{p}(\mathbf{r}, t)] \cdot [\mathbf{v}(\mathbf{r}, t) - j \cdot \hat{\mathbf{v}}(\mathbf{r}, t)] = \frac{1}{2} \cdot p_a(\mathbf{r}, t) \cdot \mathbf{v}_a^*(\mathbf{r}, t) \quad (3.8)$$

wobei \hat{p} und $\hat{\mathbf{v}}$ jeweils die Hilberttransformierte der Komponenten p und \mathbf{v} sind. Das Subscript a steht für analytische Funktion und der Asterix steht für eine konjugiert Größe.

Die Momentan-Intensität ergibt sich unter Verwendung der Gleichung 3.5 und 3.6 zu

$$\mathbf{I}_c = \frac{P^2}{2k \cdot \mathbf{r} \cdot c} \left(\nabla f - j \cdot \frac{\nabla P}{P} \right) = \frac{P^2}{2 \cdot \mathbf{r} \cdot c} \cdot \mathbf{V}^* \quad (3.9)$$

¹ Schallschnelle und Schalldruck sind durch die skalare Potentialfunktion $\Psi(\mathbf{r}, t)$ miteinander verbunden, wobei $\mathbf{v} = \nabla \Psi$ und $p = -\mathbf{r} \cdot \partial_t \Psi$ gilt. Diese Potentialfunktion löst die homogene Wellengleichung $(\Delta + k^2)\Psi = 0$

Die Größe in Gl. 3.9 ist zeitinvariant, da für die Berechnung die konjugierte Schnelle verwendet wird.

In der Praxis handelt es sich bei der Beschreibung der Felder um reale Größen, darum wird nun zur Berechnung der reellwertigen Intensität der Realteil von Schalldruck (aus Gl. 3.5) und Schallschnelle (aus Gl. 3.6) verwendet. Diese ergibt sich wie folgt zu:

$$\mathbf{I}_R = \frac{P^2}{r \cdot w} \left\{ \nabla \mathbf{f} \cdot \cos^2(\mathbf{w} \cdot \mathbf{t} - \mathbf{f}) - \frac{\nabla P}{2 \cdot P} \cdot \sin[2(\mathbf{w} \cdot \mathbf{t} - \mathbf{f})] \right\} \quad (3.10)$$

Der erste Term in Gl. 3.10 stellt die aktive Intensität dar. Diese ist direkt proportional dem reellwertigen Teil der komplexwertigen Momentan-Intensität. Die aktive Intensität ist proportional dem Gradienten der Phase des Schallfeldes und steht somit normal zur Wellenfront und zeigt in die Richtung der Schallausbreitung. Schnelle und Druck sind in der aktiven Intensität in Phase und ihr zeitliches Mittel ist ein Maß für den mittleren Schallenergiefluss.

Der zweite Term in Gl. 3.10 stellt die reaktive Intensität dar. Ihre Einhüllende ist proportional dem Gradienten der Druckamplitude und dem Imaginärteil der komplexwertigen Momentan-Intensität. Sie steht normal zu den druckgleichen Flächen. Das zeitliche Mittel der reaktiven Intensität ist Null.

3.2.3 Die komplexe Energieflussgeschwindigkeit \mathbf{u}

Ein weiteres sinnvolles Maß ist die Geschwindigkeit des Energieflusses. Dazu wird die Energiedichte w in Gl. 3.11, die sich aus der kinetischen (1.Term) und potentiellen Energiedichte (2.Term) zusammensetzt, für ein reelles Schallfeld eingeführt.

$$w = \frac{r}{2} \cdot \left[\|\mathbf{v}\|^2 + \frac{|p|^2}{(rc)^2} \right] \quad (3.11)$$

Der normalisierte komplexe Intensitätsvektor wird durch Normierung des komplexen Intensitätsvektors (vgl. 3.2.2) auf die Energiedichte gebildet [Jacobsen, 1989] und stellt dadurch eine physikalisch brauchbare Größe dar. Die Momentangeschwindigkeit des Energieflusses ist definiert (vgl. [Schiffreer, Stanzial, 1994]) durch:

$$\mathbf{u}_E = \frac{\mathbf{I}}{w} \quad (3.12)$$

Das Verhältnis der zeitlich gemittelten¹ Energieflussgeschwindigkeit $u = \langle \mathbf{I} \rangle / \langle w \rangle$ und der Schallausbreitungsgeschwindigkeit c ist nach [Farina, Ugolotti, 1998] ein brauchbares alternatives Maß für die Beschreibung von Surround-Sound Schallfelder.

¹ Die zeitliche Mittelung wird durch das Symbol $\langle \rangle$ dargestellt.

Die Energiedichte eines komplexen monochromatischen Schallfeldes wird definiert, ähnlich wie in Gl. 3.8, mit einem zusätzlichen Faktor $\frac{1}{2}$ und ergibt durch Einsetzen für \mathbf{v} und p aus Gl. 3.5 und 3.6

$$w_c = \frac{\mathbf{r}}{4} \cdot \left[\|\mathbf{v}\|^2 + \frac{|p|^2}{(rc)^2} \right] = \frac{P^2}{2rc^2} \cdot \frac{1}{2} (1 + \|\mathbf{V}\|^2) \quad (3.13)$$

Die komplexe Form der Momentangeschwindigkeit des Energieflusses ergibt sich somit zu:

$$\mathbf{u}_c = \frac{\mathbf{I}_c}{w_c} = c \cdot \frac{\mathbf{V}^*}{\frac{1}{2} \cdot (1 + \|\mathbf{V}\|^2)} \quad (3.14)$$

Diese Größe ist zeitinvariant und durch die Normierung auf die Energiedichte werden Probleme, die durch Singularitäten des komplexen Schnellektors auftreten, vermieden. Der Betrag des komplexen Vektors liegt immer im Bereich zwischen Null und der Ausbreitungsgeschwindigkeit c .

Der Realteil der komplexen Momentangeschwindigkeit des Energieflusses (siehe Gl. 3.15) weist die selben Eigenschaften wie die aktive Intensität auf und wird daher in der englischsprachigen Literatur auch als active velocity¹ (vgl. [Poletti, 2000]) bezeichnet.

$$\mathbf{u}_{cR} = c \cdot \frac{(1/k)\nabla \mathbf{f}}{\frac{1}{2} \cdot (1 + \|\mathbf{V}\|^2)} \quad (3.15)$$

Das Ergebnis in Gl. 3.15 entspricht der zeitlich gemittelten Größe in Gl. 3.12 und somit der mittleren Fortpflanzungsgeschwindigkeit der Wirkleistung. Die active velocity \mathbf{u}_{cR} repräsentiert jedoch nicht die Ausbreitungsgeschwindigkeit der Wellenfronten. Die Geschwindigkeit der Wellenfront wird wie folgt bestimmt:

$$c_p = c \frac{k}{|\nabla \mathbf{f}|} = \frac{c}{|\operatorname{Re}\{\mathbf{V}\}|} \quad (3.16)$$

und ist für allgemeine Darstellungen weniger brauchbar, da c_p unendliche Werte annehmen kann, wenn der Phasengradient gegen Null geht.

Der Imaginärteil der komplexen Momentangeschwindigkeit des Energieflusses (Gl. 3.17) weist die selben Eigenschaften wie die reaktive Intensität auf und wird daher in der Literatur auch als reactive velocity bezeichnet. Da die zeitlich gemittelte reaktive Intensität Null ist, kann der Imaginärteil der Gl. 3.14 nicht mit dem Transport der Schallenergie in Beziehung gesetzt werden.

¹ Der Begriff velocity wird in der Akustik üblicherweise mit der Schnelle in Verbindung gebracht. Hier handelt es sich um die mittlere Fortpflanzungsgeschwindigkeit der Wirkleistung. Aus Konsistenzgründen zur englischsprachigen Literatur wird auch hier der Begriff active velocity verwendet.

$$\mathbf{u}_{cl} = c \cdot \frac{(1/k) \frac{\nabla P}{P}}{\frac{1}{2} \cdot (1 + \|\mathbf{V}\|^2)} \quad (3.17)$$

3.2.4 Grafische Repräsentationsbeispiele

Die in Kapitel 3.2.1-3 eingeführten Schallfeldindikatoren werden nun anhand des IEM CUBE (eingeschränkt auf den zweidimensionalen Fall) angewendet und einander gegenübergestellt. Als Lautsprecherlayout dient für die Simulation die Lautsprecheranordnung des I. Ringes (siehe Tabelle 3.1). Die Kodierung erfolgt für ein 2D Ambisonic-System 5. Ordnung. Die Entfernung der realen Lautsprecher und der virtuellen Quellen sind mit 5m vorgegeben. Die Abbildungen zeigen, falls nicht anders angegeben, ein monochromatisches Schallfeld mit 1kHz. Für die Dekodierung wurde die Vorschrift des Basic-Decoders gewählt.

Die Position der abzubildenden Quelle wird zwischen zwei realen Lautsprecherpositionen gewählt, da für diese Position die auftretenden Fehler maximal werden. Liegt die abzubildende Quelle an der Position eines Lautsprechers, so sind die Fehler in Abhängigkeit von Lautsprecheranordnung und Wahl der Dekodierungsvorschrift meist vernachlässigbar klein (vgl. Kapitel 2.2.4).

Für die folgenden Abbildungen wurde die Position zwischen dem 2. und 3. Lautsprecher (I. Ring, siehe Tabelle 3.1) auf 33.5° festgelegt. Die entsprechenden Lautsprechergewichte sind in Abbildung 3.4 dargestellt. Die Gewichte der Lautsprecher 2 und 3 sind nicht identisch (Differenz ca. 1dB), da die Anordnung der übrigen Lautsprecher im I. Ring nicht äquidistant ist. Der Beitrag der benachbarten Lautsprecher maximiert den Betrag des Schnellektors (siehe Kapitel 2.2.2). Aus der Abbildung der normierten Lautsprechergewichte (normiert auf den größten Beitrag) ist jedoch zu erwarten, dass diese Beiträge (alle kleiner -13dB) für einen ideal sitzenden Hörer beinahe irrelevant sind.

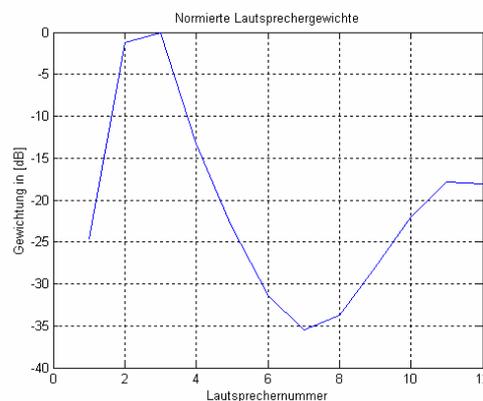


Abb. 3.4: Normierte Lautsprechergewichte in [dB].

In der folgenden Gegenüberstellung (Abb. 3.5 a und b) sind die Schalldruckfelder des Referenzfeldes (linke Seite) und des Ambisonic-Systemfeldes (rechte Seite) einer virtuellen Punktquelle in 5m Abstand vom Referenzabhörort (das entspricht dem mittleren Lautsprecherabstandes) dargestellt.

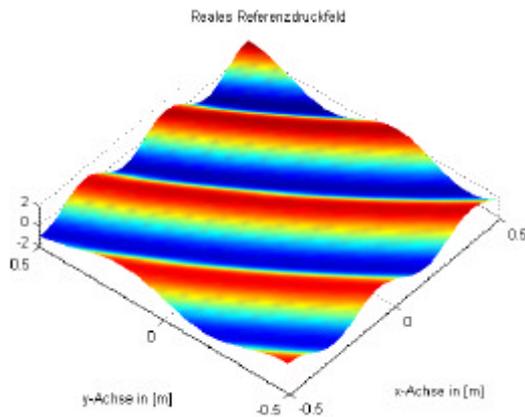


Abb. 3.5a: Realteil des Schalldruckfeldes der Referenzquelle. x und y-Achse in [m].

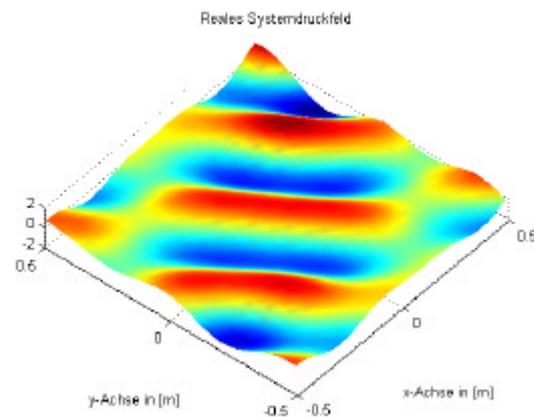


Abb. 3.5b: Realteil des Schalldruckfeldes des Systemfeldes. x und y-Achse in [m].

Der Fehler, der sich durch die Ambisonic-Approximation ergibt, kann wie in Abbildung 3.6 (a) als quadrierter absoluter Schalldruckfehler bzw. in (b) als D-Error (siehe Kapitel 2.2.4) dargestellt werden. Anhand dieser Abbildungen ist jedoch nur eine energetische Fehlerbetrachtung über ein definiertes Areal möglich.

Für einzelne Frequenzen ergibt sich für den Fehler eine ‚hügelartige Landschaft‘, d.h. es gibt weiter vom Mittelpunkt entfernte Orte, an denen der Fehler geringer ist als in mittelpunktsnahen Bereichen. Durch geeignetes Decoder-Design kann dieser Fehler besser verteilt werden.

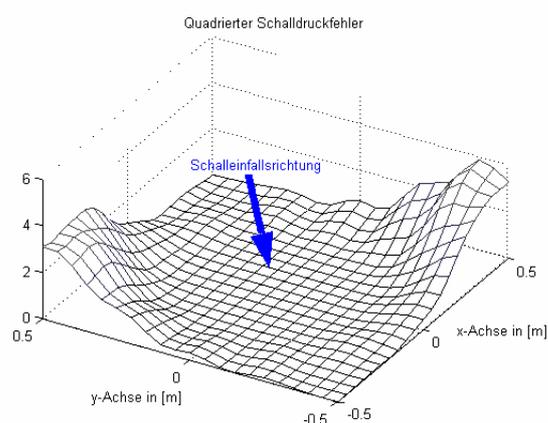


Abb. 3.6a: Quadrierter Schalldruckfehler.

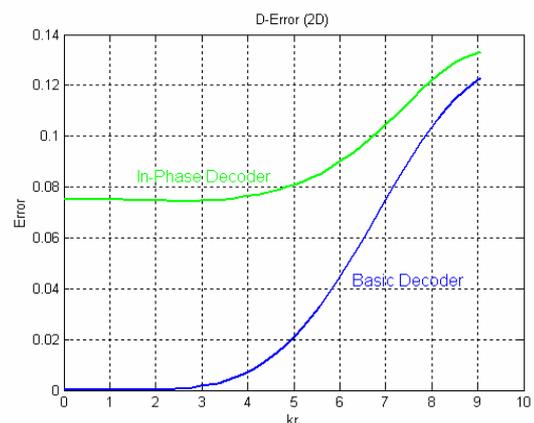


Abb. 3.6b: D-Error für Basic & In-Phase Decoder.

Durch die Darstellung der zeitlich gemittelten aktiven Intensität (,active intensity'-Vektor) ist auch eine Aussage über die Richtung des Energieflusses möglich (siehe Abbildung 3.7a für das Referenzschallfeld und 3.7b für das reproduzierte Schallfeld).

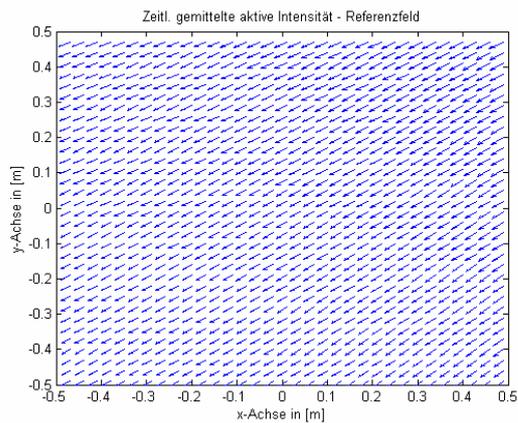


Abb. 3.7a: Zeitlich gemittelte aktive Intensität des Referenzschallfeldes.

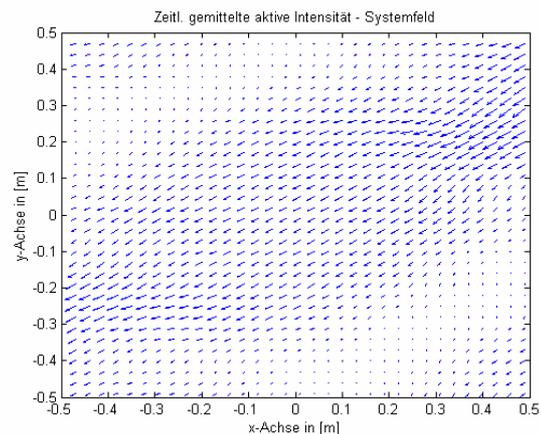


Abb. 3.7b: Zeitlich gemittelte aktive Intensität des Systemschallfeldes.

Im Systemfeld ergeben sich parallel zur Hauptflussrichtung der Energie Orte, an denen keine eindeutige Energieflussrichtung vorliegt. Es kommt zur Bildung von „Energieflusswirbel“, wobei der Druck in diesen Bereichen durch destruktive Interferenz gegen Null geht. Da der ‚active intensity‘-Vektor direkt proportional zum quadrierten Schalldruck ist, verschwindet auch dessen Betrag in diesen Bereichen. Der Höreindruck ergibt hier einen „undifferenzierbaren Richtungseindruck“ bzw. eine Form von „Im-Kopf-Lokalisation“. Diese Wirbelpunkte (-bereiche) sind somit durch geringen Druck und große „Phasigkeit“ gekennzeichnet. Eine ausführliche Untersuchung dieser speziellen Orte in einem Schallfeld findet sich bei [Mann et al., 1987].

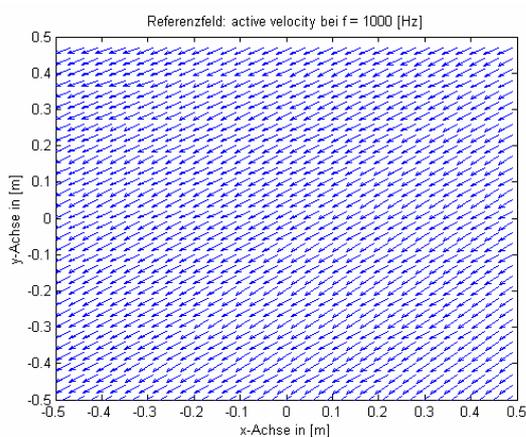


Abb. 3.8a: active velocity des Referenzschallfeldes.

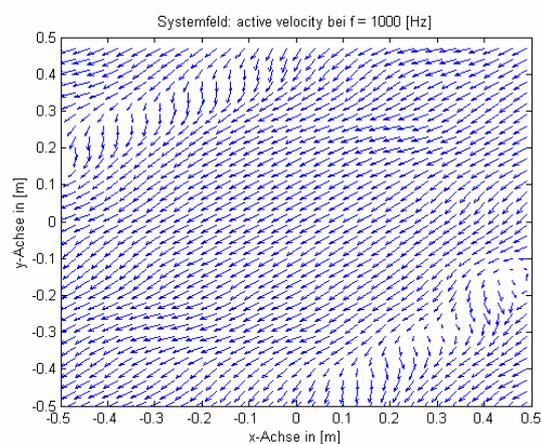


Abb. 3.8b: active velocity des Systemschallfeldes.

Der ‚active intensity‘-Vektor und der ‚active velocity‘-Vektor weisen die selbe Richtung auf, jedoch variiert der Betrag des ‚active velocity‘-Vektors weniger stark bedingt durch die Normierung auf die Energiedichte (vgl. Abbildung 3.8a,b und 3.7a,b).

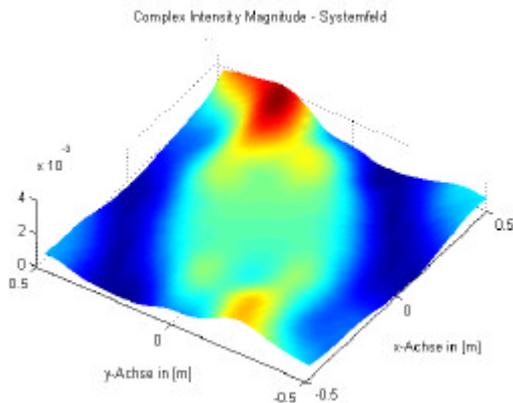


Abb. 3.9: Betrag des ‚active intensity‘ Vektors des Systemfeldes.

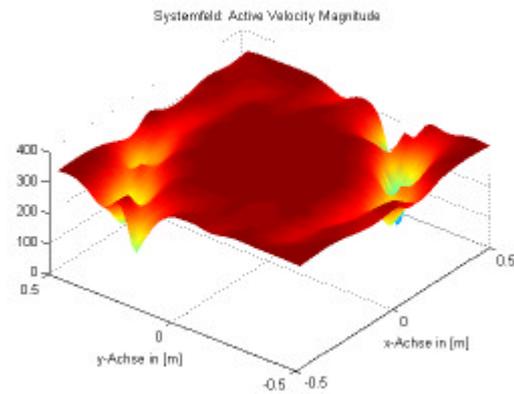


Abb.3.10: Betrag des ‚active velocity‘ Vektors des Systemfeldes.

Der Betrag des ‚active intensity‘-Vektors weist zwei Maxima in der maximalen Energieflussrichtung auf, während der Betrag des ‚active velocity‘ Vektors einen flachen breiten Bereich um den Ursprung in der maximalen Energieflussrichtung aufweist. Der Betrag des ‚active velocity‘ Vektors ist um den Ursprung gleich der Schallausbreitungsgeschwindigkeit (vgl. $c=340[m/s]$ bei 20°).

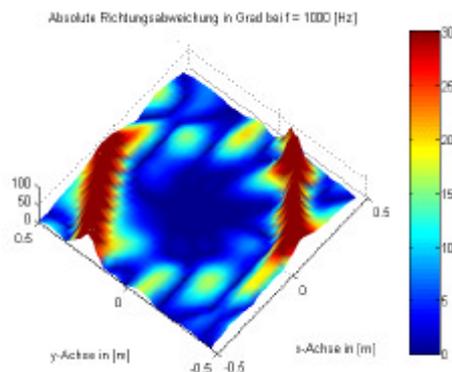


Abb. 3.11: Richtungsabweichung des System- vom Referenzfeldes in Grad, ohne Korrektur der Abweichung von der Ausbreitungsgeschwindigkeit.

Aus der Richtung des Schnellektors kann im Vergleich zwischen System- und Referenzfeld die wahrnehmbare Richtungsabweichung bestimmt werden (siehe Abb.3.11). Jene Bereiche in Abbildung 3.10, die von der Ausbreitungsgeschwindigkeit ($c = 340[m/s]$) abweichen, ergeben eine zusätzliche Veränderung der wahrgenommenen Richtung (vgl. Kapitel 1).

Da Abbildungen von einzelnen Frequenzen wenig Auskunft über das gesamte Erscheinungsbild und Wiedergabeverhalten von Breitbandsignalen geben, sind im Folgenden die Druckverteilung in den Referenz- und Systemschallfeldern bei Impulsanregung dargestellt. Die Systemparameter wurden nicht geändert. In Abbildung 3.12a und b ist jeweils eine hanningförmige Impulsanregung mit einer Bandbreite bis 1.5kHz dargestellt. Die Nachbildung ist für den innersten Bereich der Darstellung (Durchmesser bis ca. 60cm) nahezu perfekt – „exakte Wellenfeldsynthese“.

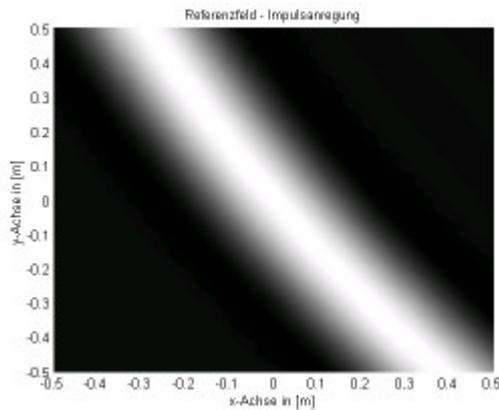


Abb. 3.12a: Referenzschallfeld bei Impulsanregung mit $\Delta f = 100\text{Hz}-1500\text{Hz}$.

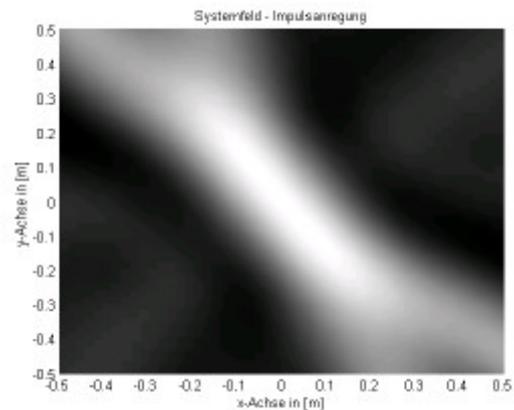


Abb. 3.12b: Systemschallfeld bei Impulsanregung mit $\Delta f = 100\text{Hz}-1500\text{Hz}$.

In Abbildung 3.13a,b ist jeweils eine Impulsanregung mit einer Bandbreite bis 8kHz dargestellt. Die Nachbildung weicht hier schon sehr deutlich von ihrer Vorlage ab. Der exakte Wiedergabebereich ist sehr klein.

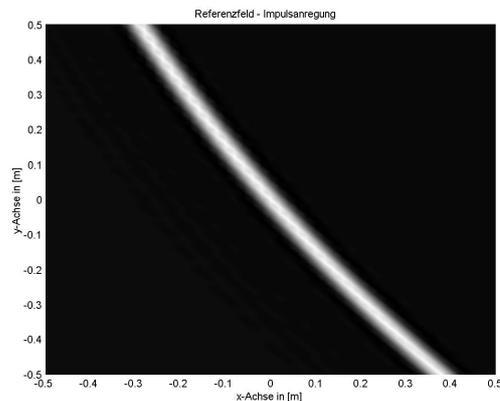


Abb. 3.13a: Referenzschallfeld bei Impulsanregung mit $\Delta f = 400\text{Hz}-8\text{kHz}$.

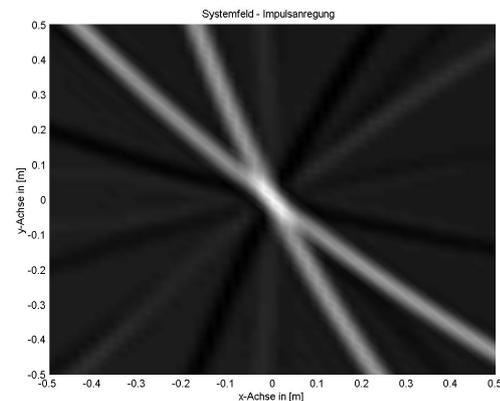


Abb. 3.13b: Systemschallfeld bei Impulsanregung mit $\Delta f = 400\text{Hz}-8\text{kHz}$.

Zusammenfassung

Durch die in Kapitel 3.2 eingeführten Größen ist es möglich die allgemeinen Eigenschaften der Artefakte, den Einfluss unterschiedlicher Dekodierungsvorschriften (vor allem die des Window-Applied Decoders) sowie die Systemperformance bei der Wiedergabe über größere Areale zu beschreiben und zu beurteilen.

Bedingt durch die geringere Schwankung des Betrags des ‚active velocity‘ Vektors ist die Energieflussrichtung in jeden Punkt deutlich erkennbar, während die Bildung von Energieflusswirbeln bei anderen Fehlermaßen nicht erkennbar sind. Somit scheint diese Größe zur Beurteilung von Systemen zur Schallreproduktion bzw. bei der Parameterauswahl sehr geeignet zu sein.

Ein Ambisonic-System kann für niedrige Ordnungen (und somit auch entsprechend wenigen Lautsprechern) als eine tieffrequente Approximation der Wellenfeldsynthese aufgefasst werden.

Gegenüber der stereophonen Wiedergabe erfüllt die Ambisonic-kodierte Wiedergabe die geforderte Homogenität und Kohärenz (vgl. Kapitel 1.2) und stellt dahingehend eine wesentliche Verbesserung dar. Bei einem Vergleich mit der Stereodipoltechnik (vgl. [Gardner, 1998]) bzw. allgemein mit transauralen Systemen (vgl. Kapitel 1.2) ist der größere Wiedergabebereich - gleichzeitiges Hörerlebnis für mehrere Hörer – als wesentlicher Unterschied und Vorteil zu nennen. Zu den Nachteilen der transauralen Systeme gegenüber dem Ambisonic-Verfahren zählt auch die Klangfärbung, die durch die Richtungskodierung und Unterdrückung des Kanalübersprechens (cross-talk cancellation) entsteht.

Die Surroundwiedergabe im 5.1 bzw. 3/2 Format stellt einen funktionalen Ansatz zur Schallfeldreproduktion dar. Sie ist jedoch im Vergleich zu Ambisonic auf die Wiedergabe in einer Ebene beschränkt. Bei der Quellenpositionierung wurde hierbei ursprünglich die Panoramatechnik der Stereowiedergabe übernommen. Untersuchungen (vgl. dazu [Bosun, 2001]) bestätigen die Sinnhaftigkeit der globalen Interpolationstechnik (Ambisonic) für die Quellenpositionierung auch für Surround Anwendungen.

In [Pulkki, 2001] wurde bereits eine Vergleichsuntersuchung zwischen Ambisonic und VBAP auf Basis eines binauralen auditorischen Modells durchgeführt. Die Ergebnisse (z.B. gute Richtungsabbildung bis ca.1kHz bei Verwendung Ambisonic 2.Ordnung) bestätigen zum Teil die Kenntnisse der theoretischen Grundlagen in Kapitel 2.2 (vgl. dazu Abschnitt 2.2.4). Bei der Beurteilung höherer Ordnungen bzw. Hinzunahme von mehreren Lautsprechern entstehen größere Abweichungen der ITD und ILD von den Referenzrichtungen. Diese wurden auch bei der Entwicklung des objektiven Fehlermaß (vgl. Kapitel 5) beobachtet. Bei der Untersuchung zur periphonen Anordnung [Pulkki, 2001] wurde ein irreguläres Lautsprecherlayout verwendet, dessen Konditionszahl sehr groß ist (vgl. Kapitel 2.2). Die Berechnung der Lautsprecher-signale wurde explizit nur für den 2-dimensionalen Fall angegeben. Dabei wurden suboptimale Dekodierer verwendet. Die publizierten Ergebnisse für die tatsächlichen Wiedergabequalität von Ambisonic sind daher nur bedingt aussagekräftig.

Für einen Vergleich Ambisonic - Wellenfeldsynthese können qualitativ keine wesentlichen Vorteile genannt werden. In [Nicol, Emerit, 1999] ist die Schallfeldreproduktion anhand beider Systeme im 2D Fall eingehend untersucht und verglichen worden. Bei entsprechend hohem Aufwand auf beiden Seiten sind beide Verfahren äquivalent. Jedoch überzeugt grundsätzlich die Effizienz und Skalierbarkeit von Ambisonic.

3.3 Kompensation von Raumreflexionen

Bei der Wiedergabe aufgenommener Signale kommt es zur Überlagerung des Aufnahme- und des Wiedergaberaumes. Bei der Raumkompensation wird der meist störende Einfluss des Wiedergaberaumes auf das zu reproduzierende Schallfeld durch eine spezielle Filterung der Lautsprecher-signale reduziert. Es wird für einen begrenzten Raumausschnitt eine Freifeldwiedergabe angestrebt. Das Ziel ist eine bestmögliche Unterdrückung früher diskreter Raumreflexionen (Wand, Decken und Bodenreflexionen bis ca. 100ms, vgl. [Griesinger, 1997], [Beranek, 1996]), da diese für die Wahrnehmung der raumcharakteristischen Eigenschaften von zentraler Bedeutung sind. Durch Ausblendung der ersten Reflexionen des physikalischen Raumes kann daher jede beliebige gewünschte Akustik einem Raum (-ausschnitt) aufgeprägt werden.

In den folgenden Abschnitten werden Ansätze zur Kompensation von Raumreflexionen vorgestellt.

3.3.1 durch Messungen im Ambisonic-Bereich

Vorausgesetzt sei eine Aufnahme in einem Ambisonic-Format. Bei der Wiedergabe im freien Schallfeld treten bei der Schallfeldreproduktion keine zusätzlichen Fehler¹ auf. Bei der Wiedergabe in einem geschlossenen Raum treten jedoch die Reflexionen des Wiedergaberaumes zusätzlich in Erscheinung.

Mit einem Ambisonic-System kann für eine "virtuelle Schallquelle" (siehe Kapitel 2) ein natürliches Schallfeld rekonstruiert werden. Jede Reflexion kann als eine "Schallquelle" interpretiert und damit aus Ambisonic-Komponenten so synthetisiert werden, dass diese die ursprüngliche Raumreflexion durch destruktive Interferenz kompensieren. Die notwendigen Komponenten können durch Messung der richtungsabhängigen Raumimpulsantworten quasi in einem Punkt bestimmt werden, wie im folgenden Abschnitt anhand eines Ambisonic Systems 1. Ordnung für den 2-dimensionalen Fall im Punkt (0,0) gezeigt wird.

3.3.1.1 Herleitung für ein 2D-System mit Ambisonic 1.Ordnung

Die Wiedergabe erfolgt mit einem kreisförmigen Lautsprecherarray bestehend aus L Lautsprechern. Durch die begrenzenden Flächen des Wiedergaberaumes kommt es zu Reflexionen. Diese Reflexionen bringen das Originalsignal gemäß dem Spiegelquellenmodell

¹ Wenn man von der Modellannahme ausgeht, dass die abgestrahlten Wellen eines Lautsprechers in einer entsprechenden Entfernung als ebene Welle betrachtet werden können und die Ordnung des Systems hoch genug gewählt ist.

mit entsprechender Verzögerung, Abschwächung und Klangfarbenveränderung aus einer anderen Richtung wieder.

Für eine Freifeldwiedergabe berechnen sich die Lautsprecher-Signale L_i aus den Ambisonic-Signalen W_{org} , X_{org} und Y_{org} wie folgt:

$$\begin{pmatrix} L_1 \\ \vdots \\ L_L \end{pmatrix} = \begin{pmatrix} \text{Decoder} \end{pmatrix} \cdot \begin{pmatrix} W_{org} \\ X_{org} \\ Y_{org} \end{pmatrix} \quad (3.18)$$

Der Decoder gewichtet die originalen Drucksignale W_{org} , X_{org} und Y_{org} in Abhängigkeit von der Lautsprecherposition und bildet durch Überlagerung die Lautsprecher-Signale (vgl. Kapitel 2).

Die k -te diskrete Raumreflexion wird durch die Parameter Amplitude a_k , Verzögerung t_k und Richtung \mathbf{q}_k modelliert. Die Raumimpulsantwort des i -ten Lautsprechers h_i aus der Richtung \mathbf{f}_i kann wie folgt formuliert werden:

$$h_i(t) = \mathbf{d}_{\mathbf{f}_i}(t - t_0) + \sum_{k=1}^{+\infty} a_k \cdot \mathbf{d}_{\mathbf{q}_{i,k}}(t - t_i) \quad (3.19)$$

Die direktionalen Raumimpulsantworten können mit einem *Soundfield microphone*, mit der Methode von Farina¹ (siehe [Farina, Ugonotti, 1998], [Farina, Tronchin, 1998]), oder mit einem Mikrofonarray samt Nachverarbeitung (vgl. Kap 2.4 bzw. [Poletti, 2000], [Nicol, Emerit, 1999]) gemessen werden und liegen somit im Ambisonic-Format (zumindest 1. Ordnung) vor. Die Ambisonic-Signale der Impulsantwort des i -ten Lautsprechers können wie folgt dargestellt werden:

$$W_{h_i}(t) = \mathbf{d}(t - t_0) + \sum_{k=1}^{+\infty} a_k \cdot \mathbf{d}(t - t_{i,k}) \quad (3.20a)$$

$$X_{h_i}(t) = \mathbf{d}(t - t_0) \cdot \cos(\mathbf{f}_i) + \sum_{k=1}^{+\infty} a_k \cdot \mathbf{d}(t - t_{i,k}) \cdot \cos(\mathbf{q}_{i,k}) \quad (3.20b)$$

$$Y_{h_i}(t) = \mathbf{d}(t - t_0) \cdot \sin(\mathbf{f}_i) + \sum_{k=1}^{+\infty} a_k \cdot \mathbf{d}(t - t_{i,k}) \cdot \sin(\mathbf{q}_{i,k}) \quad (3.20c)$$

Die ersten Terme der Gl. (3.20a bis c) beschreiben die direkte Schallkomponente und werden für die folgenden Betrachtungen entfernt. Für den Anteil der Reflexionen unter der Berücksichtigung, dass diese zur Eliminierung der Raumreflexionen verwendet werden (negatives Vorzeichen) ergeben sich mit dem Übergang von $t \rightarrow n$ folgende Kompensationsfilter (negative Vorzeichen zur destruktiven Interferenz):

¹ Farina verwendet dabei eine orthogonale Anordnung mit sieben ungerichteten Mikrofonen. Die Richtungsinformation (Ambisonic-Signale) wird durch die Differenz der entsprechenden Mikrofon-Signale bestimmt (Approximation der Differentiation).

$$W_{h_{i,r}}(n) = -\sum_{k=1}^{+\infty} a_k \cdot \mathbf{d}(n - m_{i,k}) \quad (3.21a)$$

$$X_{h_{i,r}}(n) = -\sum_{k=1}^{+\infty} a_k \cdot \mathbf{d}(n - m_{i,k}) \cdot \cos(\mathbf{q}_{i,k}) \quad (3.21b)$$

$$Y_{h_{i,r}}(n) = -\sum_{k=1}^{+\infty} a_k \cdot \mathbf{d}(n - m_{i,k}) \cdot \sin(\mathbf{q}_{i,k}) \quad (3.21c)$$

Das Lautsprechersignal des i-ten Lautsprechers für das freie Schallfeld lautet (vgl. Gl. (3.18)):

$$L_i(n) = a_i \cdot W(n)_{org} + b_i \cdot X(n)_{org} + c_i \cdot Y(n)_{org} \quad (3.22)$$

Die Koeffizienten a_i , b_i und c_i in Gl. (3.22) sind die Elemente der i. Zeile der Ambisonic-Decodermatrix. Die Kompensationssignale für die Reflexionen werden durch Faltung¹ der Lautsprechersignale mit den dazugehörigen Ambisonic-Raumimpulsantworten berechnet.

$$W_{L_{i,r,comp}}(n) = L_i(n) \otimes W_{h_{i,r}}(n) \quad (3.23a)$$

$$X_{L_{i,r,comp}}(n) = L_i(n) \otimes X_{h_{i,r}}(n) \quad (3.23b)$$

$$Y_{L_{i,r,comp}}(n) = L_i(n) \otimes Y_{h_{i,r}}(n) \quad (3.23c)$$

Somit ergeben sich bei Kompensation der vom Lautsprecher L_i erzeugten Reflexionen die Lautsprechersignale L^* aus den Ambisonic-Signalen:

$$\begin{aligned} \begin{pmatrix} L_1^* \\ \vdots \\ L_L^* \end{pmatrix} &= \begin{pmatrix} \text{Decoder} \\ \vdots \\ \text{Decoder} \end{pmatrix} \cdot \begin{pmatrix} W_{org} \\ X_{org} \\ Y_{org} \end{pmatrix} + \begin{pmatrix} \text{Decoder} \\ \vdots \\ \text{Decoder} \end{pmatrix} \cdot \begin{pmatrix} W_{L_{i,r,comp}} \\ X_{L_{i,r,comp}} \\ Y_{L_{i,r,comp}} \end{pmatrix} \\ &= \begin{pmatrix} \text{Decoder} \\ \vdots \\ \text{Decoder} \end{pmatrix} \cdot \begin{pmatrix} W_{org} \\ X_{org} \\ Y_{org} \end{pmatrix} + \begin{pmatrix} \text{Decoder} \\ \vdots \\ \text{Decoder} \end{pmatrix} \cdot \begin{pmatrix} W_{h_{i,r}}(n) \otimes (a_i W_{org} + b_i X_{org} + c_i Y_{org}) \\ X_{h_{i,r}}(n) \otimes (a_i W_{org} + b_i X_{org} + c_i Y_{org}) \\ Y_{h_{i,r}}(n) \otimes (a_i W_{org} + b_i X_{org} + c_i Y_{org}) \end{pmatrix} \\ &= \begin{pmatrix} \text{Decoder} \\ \vdots \\ \text{Decoder} \end{pmatrix} \cdot \left[\mathbf{d}(n) \cdot I \cdot \begin{pmatrix} W_{org} \\ X_{org} \\ Y_{org} \end{pmatrix} + \begin{pmatrix} W_{h_{i,r}}(n) \\ X_{h_{i,r}}(n) \\ Y_{h_{i,r}}(n) \end{pmatrix} \otimes \begin{pmatrix} a_i W_{org} + b_i X_{org} + c_i Y_{org} \\ a_i W_{org} + b_i X_{org} + c_i Y_{org} \\ a_i W_{org} + b_i X_{org} + c_i Y_{org} \end{pmatrix} \right] \end{aligned} \quad (3.24a)$$

¹ Für die Faltungsoperation wird in den folgenden Gleichungen das Symbol \otimes verwendet.

$$= \begin{pmatrix} \\ \\ \\ \end{pmatrix} \cdot \begin{pmatrix} \mathbf{d}(n) & 0 & 0 \\ 0 & \mathbf{d}(n) & 0 \\ 0 & 0 & \mathbf{d}(n) \end{pmatrix} + \begin{pmatrix} a_i W_{h_{i,r}}(n) + b_i W_{h_{i,r}}(n) + c_i W_{h_{i,r}}(n) \\ a_i X_{h_{i,r}}(n) + b_i X_{h_{i,r}}(n) + c_i X_{h_{i,r}}(n) \\ a_i Y_{h_{i,r}}(n) + b_i Y_{h_{i,r}}(n) + c_i Y_{h_{i,r}}(n) \end{pmatrix} \otimes \begin{pmatrix} W_{org} \\ X_{org} \\ Y_{org} \end{pmatrix}$$

...und man erhält folgendes Ergebnis:

$$\begin{pmatrix} L_1^* \\ \vdots \\ L_L^* \end{pmatrix} = \begin{pmatrix} \\ \\ \\ \end{pmatrix} \cdot \begin{pmatrix} \mathbf{d}(n) + a_i W_{h_{i,r}}(n) & b_i W_{h_{i,r}}(n) & c_i W_{h_{i,r}}(n) \\ a_i X_{h_{i,r}}(n) & \mathbf{d}(n) + b_i X_{h_{i,r}}(n) & c_i X_{h_{i,r}}(n) \\ a_i Y_{h_{i,r}}(n) & b_i Y_{h_{i,r}}(n) & \mathbf{d}(n) + c_i Y_{h_{i,r}}(n) \end{pmatrix} \otimes \begin{pmatrix} W_{org} \\ X_{org} \\ Y_{org} \end{pmatrix} \quad (3.24b)$$

Für eine 2-dimensionale Anordnung mit Ambisonic 1.Ordnung erhält man bei der Berücksichtigung der Reflexionen, die durch die L Lautsprecher hervorgerufen wurden die kompensierten Lautsprechersignale L_{comp} :

$$\begin{pmatrix} L_{1,comp} \\ \vdots \\ L_{L,comp} \end{pmatrix} = \begin{pmatrix} \\ \\ \\ \end{pmatrix} \cdot \begin{pmatrix} \mathbf{d}(n) + \sum_{i=1}^L a_i W_{h_{i,r}} & \sum_{i=1}^L b_i W_{h_{i,r}}(n) & \sum_{i=1}^L c_i W_{h_{i,r}}(n) \\ \sum_{i=1}^L a_i X_{h_{i,r}}(n) & \mathbf{d}(n) + \sum_{i=1}^L b_i X_{h_{i,r}}(n) & \sum_{i=1}^L c_i X_{h_{i,r}}(n) \\ \sum_{i=1}^L a_i Y_{h_{i,r}}(n) & \sum_{i=1}^L b_i Y_{h_{i,r}}(n) & \mathbf{d}(n) + \sum_{i=1}^L c_i Y_{h_{i,r}}(n) \end{pmatrix} \otimes \begin{pmatrix} W_{org} \\ X_{org} \\ Y_{org} \end{pmatrix} \quad (3.25)$$

Zusammenfassung

Durch diese Vorgangsweise benötigt man abhängig von der Anzahl N der Übertragungskanäle bei Ambisonic-Ordnung M insgesamt N^2 (2 dimensional $(2M+1)^2$ bzw. 3 dimensional $(M+1)^4$) Filter zur Eliminierung der Raumreflexionen des Wiedergaberaumes. Die Anzahl der Filter ist somit unabhängig von der Anzahl der Lautsprecher.

In der praktischen Umsetzung lässt sich die Raumkompensation nur für tiefe Frequenzen (bis ca. 500 Hz) stabil einsetzen, da sich für höhere Frequenzen durch Messfehler und die nicht reproduzierbaren Ausbreitungsbedingungen (Einfluss von Publikum, Temperatur, Luftfeuchtigkeit, etc.) die gewünschte destruktive Interferenz in eine konstruktive verwandelt und somit den störenden Raumeinfluss eher noch verstärkt.

3.3.2 durch Messungen an einer Oberfläche, zeitinvariant.

Bei der Raumkompensation durch zeitinvariante Impulsantworten wird - angelehnt an das Kirchhoff-Helmholtz-Theorem (vgl. Kap.2) - durch eine Mikrofonanordnung (diese werden im Folgenden als Sensoren bezeichnet) ein Raumbereich definiert, in dessen Inneren eine Kompensation der Reflexionen angestrebt wird. Die Übertragungsfunktionen von jedem Lautsprecher zu jeder Sensorposition werden gemessen und mit der gewünschten Freifeldübertragung in Beziehung gesetzt. Die Minimierung eines geeigneten Optimalitätskriteriums (meist der mittlere quadratische Fehler) führt zu Filtern für die

kompensierenden Lautsprecher, die – auf das Signal des Primärlautsprechers angewendet – die Reflexionen an den Sensorpunkten reduzieren (siehe Abb.3.14). Der Vorteil dieses Ansatzes liegt darin, dass das Verhältnis von Quell- zu Kompensationssignal variabel einstellbar ist und damit Rechenaufwand, Kompensation und Artefakte in gewissen Grenzen optimiert werden können.

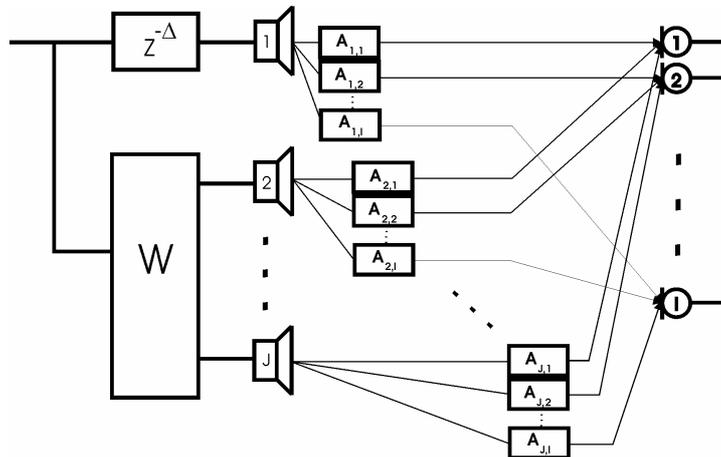


Abb. 3.14: Strukturschaltbild für die Bestimmung der zeitinvarianten Kompensationsfilter W von frühen diskreten Raumreflexionen.

3.3.2.1 Sensoranordnung

Bei der Wahl der Sensorenanordnung ist die Größe des zu optimierenden Wiedergabebereiches und der angestrebte Frequenzbereich von wesentlicher Bedeutung. Für die Ausdehnung des Arrays ist die untere Grenzfrequenz von Bedeutung. Ist die Ausdehnung zu gering, kann eine eindeutige Richtung für entsprechend langwelligem Schall nicht detektiert werden. Bei hohen Frequenzen sind geringe Abstände zwischen den Sensorpositionen anzustreben, um 'spatial aliasing' (vgl. Kapitel 2.3.4.2) bei der räumlichen Abtastung im gewünschten Frequenzbereich zu vermeiden. Die obere Grenzfrequenz einer Anordnung wird durch f_{\max} wie folgt festgelegt:

$$f_{\max} = \frac{c}{2 \cdot d \cdot \sin \alpha_{\max}} \quad (3.26)$$

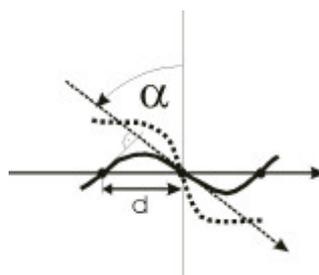


Abb. 3.15: Räumliche Abtastung einer Schallwelle an drei Raumpunkten.

Um die verwendbare Bandbreite bei gegebener Sensoranzahl zu erhöhen, sollten äquidistante Anordnungen vermieden werden. Vielmehr sollten die Teilungsverhältnisse der Sensordistanzen ein größtmögliches gemeinsames Vielfaches besitzen.

Mit Hilfe eines raumakustischen Simulationsprogramms (vgl. [Krejci, 2001]), das auf dem Spiegelquellenmodell basiert, können verschiedene Mikrofonanordnungen ohne aufwendige Messung rasch getestet und bzgl. Anordnung und Sensorenanzahl optimiert werden (vgl. Abbildung 3.16).

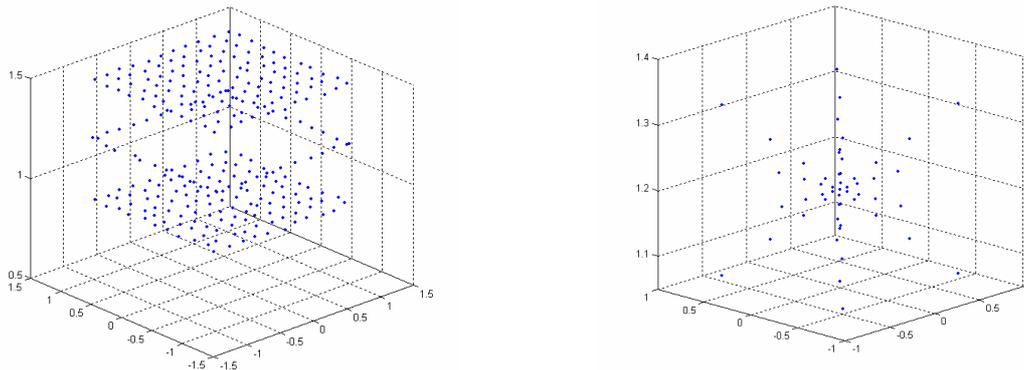


Abb. 3.16: Räumliche Abtastung durch unterschiedliche Sensorarrays
Links: nicht optimiert. Rechts: optimiert.
Die Punkte stellen jeweils Sensorpunkte dar.

3.3.2.2 Bestimmung der Kompensationsfilter

Die Kompensationsfilter sollen - wie bereits eingangs erwähnt - vor allem die tieffrequenten Anteile der frühen Reflexionen eliminieren. Hochfrequente Anteile werden unter natürlichen Umständen durch die Ausbreitungsdämpfung der Luft¹ (vgl. [Cremer, Müller, 1978]) und durch die tendenziell zu hohen Frequenzen hin zunehmende Absorption raumbegrenzender Flächen relativ früh stark abgeschwächt.

Zuerst erfolgt die Messung der Übertragungsfunktionen zwischen jedem Lautsprecher und jeder Sensorposition. Diese Übertragungsfunktionen werden mathematisch durch eine Übertragungsmatrix beschrieben. Aus den gesuchten Direktsignalanteilen wird ein Zielvektor gebildet.

Zielvektor, Übertragungsmatrix und gesuchte Filter bilden ein Normalgleichungssystem, das durch Minimierung des mittleren quadratischen Fehlers (Least Mean Squares, LMS) gelöst werden kann. Die Sensoranzahl I wird größer als die Lautsprecheranzahl J gewählt. Dadurch erhält man ein überbestimmtes Gleichungssystem, dessen Lösung stets eindeutig ist.

¹ Eine Schallwelle mit 10 kHz erfährt bedingt durch die Luftabsorption bereits nach 100ms (ca. 34m) eine Dämpfung von 5dB !

Die Ausgangsgleichung zur Bestimmung der gesuchten Filter lautet somit:

$$\underline{\mathbf{Y}} = \underline{\mathbf{A}} \cdot \underline{\mathbf{X}} \quad (3.27)$$

wobei $\underline{\mathbf{X}}$...die gesuchte Filter im Frequenzbereich,

$\underline{\mathbf{A}}$...die Übertragungsfunktionen im Frequenzbereich und

$\underline{\mathbf{Y}}$...die Zielfunktionen (Sensorsignale) im Frequenzbereich darstellen.

Mit $\underline{\mathbf{Y}} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_J]^T$ und $\mathbf{Y}_1 = [y_{1,1}(f), y_{1,2}(f), \dots, y_{1,I}(f)]^T$ etc.

sowie $\underline{\mathbf{X}} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J]^T$ und $\mathbf{X}_1 = [x_{1,1}(f), x_{1,2}(f), \dots, x_{1,J}(f)]^T$ etc.

$$\underline{\underline{\mathbf{A}}} = \begin{bmatrix} \mathbf{A} & 0 & \dots & 0 \\ 0 & \mathbf{A} & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{A} & 0 \\ 0 & 0 & \dots & 0 & \mathbf{A} \end{bmatrix}$$

Die Matrix \mathbf{A} beinhaltet sämtliche Übertragungsfunktionen zwischen den I Sensoren und J Lautsprecher.

Bevor die Messung der Übertragungsfunktionen erfolgt, wird die Beschallungsanlage im Ursprung der Sensoranordnung kalibriert¹, d.h. für jeden Lautsprecher erfolgt ein Laufzeit- und Verstärkungsabgleich.

Aus praktischen Gründen bieten sich für die Messung der Impulsantworten Maximallängenfolgen (Maximum Length Sequences, MLS)² an. Zu den Vorteilen gegenüber der periodischen Impulstestung zählt die Verbesserung des Signal-Rausch-Abstandes durch höheren Energieeintrag und Mittelung über mehrere Messperioden sowie die weitgehende Immunität gegen kurzzeitige akustische Störungen während der Messung (vgl. [Rife, Vanderkooy, 1989]). Die Periodendauer der gewählten Folge muss länger als die zu messende Raumimpulsantwort sein. Die Berechnung der Impulsantworten erfolgt offline.

Aus den ML-Folgen und den gemessenen Systemantworten berechnet sich die Impulsantwort des j-ten Lautsprechers auf den i-ten Sensor wie folgt³:

$$h_{i,j}(t) = \text{DFT}^{-1} \left\{ \frac{S_{i,j}(f) \cdot \text{MLS}(f)^*}{|\text{MLS}(f)|^2} \right\} = \text{DFT}^{-1} \{ S_{i,j}(f) \cdot \text{MLS}(f)^* \} \quad (3.28)$$

¹ Dadurch wird ein definierter Zustand gewährleistet. Hinsichtlich des Ambisonic-Ansatzes (vgl. Kapitel 2) befinden sich nach Berücksichtigung der Laufzeiten und Verstärkungsfaktoren alle Lautsprecher auf einer virtuellen Kugeloberfläche.

² Alternativ auch logarithmische Sweeps.

³ Für die praktischen Berechnungen muss die Latenzzeit der digitalen Soundkarte (Messkarte) und der AD bzw. DA Konverter abgezogen werden. Die verbleibenden Samples bis zum Direktschall entsprechen dann der „akustischen Flugzeit“ vom Lautsprecher zum Sensor.

wobei $MLS(f)$ die diskrete Fourier-Transformierte der gewählten ML-Folge und $S_{i,j}(f)$ die Transformierte des i -ten Sensorsignals bei Anregung des j -ten Lautsprechers darstellt.

Da nur die frühen Reflexionen von Interesse sind, werden diese Impulsantworten auf ca. 93 ms, hier $N = 4096$ Punkte bei einer Abtastrate von 44.1 kHz, durch Fensterung mit $w(n)$ gekürzt. Dadurch werden im Fall des IEM-CUBE, laut einer Raumsimulation basierend auf der Spiegelquellenmethode, Reflexionen bis zur 3. Ordnung vollständig berücksichtigt.

$$\tilde{h}_{i,j}(n) = h_{i,j}(m) \cdot w(n) \quad (3.29)$$

$$\text{mit } w(n) = \begin{cases} 1 & 0 \leq n < \frac{3N}{4} \\ \frac{4}{N-4}(N-1-n) & \text{für } \frac{3N}{4} \leq n < N \\ 0 & \text{sonst} \end{cases}$$

3.3.2.3 Lösung im Frequenzbereich

Aus den Impulsantworten zwischen j -ten Lautsprecher und i -ten Sensor (siehe Gl. 3.29) berechnet sich mit Hilfe der diskreten Fourier-Transformation die Übertragungsfunktionen $A_{i,j}(f)$ wie folgt¹:

$$A_{i,j}(f) = \text{DFT}_{2N} \{ \tilde{h}_{i,j}(n) \} \quad (3.30)$$

$$\mathbf{A}(f) = \begin{bmatrix} \text{diag}(A_{1,1}(f)) & \text{diag}(A_{2,1}(f)) & \cdots & \text{diag}(A_{J,1}(f)) \\ \text{diag}(A_{1,2}(f)) & & & \vdots \\ \vdots & & & \\ \text{diag}(A_{1,I}(f)) & \cdots & & \text{diag}(A_{J,I}(f)) \end{bmatrix} \quad (3.31)$$

wobei gilt:

$$\text{diag}(A_{i,j}) = \begin{bmatrix} A_{i,j}(f_1) & 0 & 0 & \cdots & 0 \\ 0 & A_{i,j}(f_2) & 0 & & \\ 0 & 0 & A_{i,j}(f_3) & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & A_{i,j}(f_{2N}) \end{bmatrix}$$

Im folgenden wird die Frequenzabhängigkeit der einzelnen Größen nicht weiter explizit gekennzeichnet, außer es besteht eine dringende Notwendigkeit.

Der Vektor \mathbf{X} repräsentiert die gesuchten Filtersets zur Kompensation der einzelnen Lautsprecher. Mit dem Vektor \mathbf{Y} können die gewünschten Direktschallanteile an den I

¹ Der Subskript bedeutet, dass eine 2N Punkte DFT durchgeführt wird. Dadurch wird in der folgenden Filterung eine zirkuläre Faltung verhindert.

Sensoren für alle J Lautsprecher vorgegeben werden. Diese werden im folgenden als Zielfunktionen bezeichnet.

Für die Zielfunktionen \mathbf{Y} werden anstelle von Dirac-Impulsen, die eine vollbandige Lösung liefern würden, Hanning-Impulse verwendet, durch deren Länge N_h die Bandbreite der Lösung gezielt gesteuert werden. Je größer die Länge N_h gewählt wird, umso stärker werden die hohen Frequenzen abgeschwächt (siehe als Beispiel Abbildung 17). Somit handelt es sich nunmehr um eine tieffrequente Approximation.

$$w_h(n) = \begin{cases} 0.5 \cdot [1 - \cos(\frac{2\pi \cdot n}{N_h})] & 0 \leq n < N_h \\ 0 & \text{sonst} \end{cases} \quad (3.32)$$

$$W_h(f) = \text{DFT}_{2N}\{w_h(n)\}$$

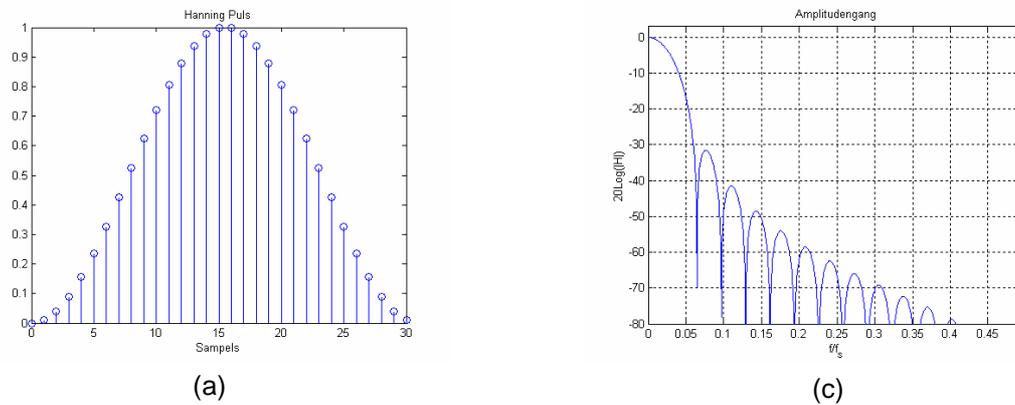


Abb. 3.17: (a) Pulsform in Samples, (b) Amplitudengang für einen Hanning Puls mit $N_h=31$.

Die Positionierung dieser gefilterten Impulse und deren Amplitude ergibt sich durch die Berechnung der normierten Kreuzkorrelation zwischen den gemessenen Impulsantworten und den Hanning-Impulsen. Das Maximum der Korrelationsfunktion ergibt die Position p und die Amplitude m für die Zielfunktion.

$$KKF_{i,j}(n) = \text{DFT}^{-1} \left\{ \frac{A_{i,j}(f) \cdot W_h(f)^*}{W_h(f) \cdot W_h(f)^*} \right\} \quad (3.33)$$

$$p_{i,j} = \arg[\max_n(KKF_{i,j}(n))] \quad (3.34)$$

$$m_{i,j} = KKF_{i,j}(p_{i,j}) \quad (3.35)$$

Die Zielfunktionen $\mathbf{Y}(f)$ ergeben sich somit durch die entsprechende Positionierung¹ $p_{i,j}$ und Gewichtung $m_{i,j}$ der künstlichen, gewünschten „Direktschall Anteile“ an den Sensoren $S_{i,j}$.

¹ Sollte sich das Maximum nicht genau an der Position eines Samples befinden, so kann durch eine geeignete Interpolation eine effiziente Approximation für tiefe Frequenzen realisiert werden (vgl. [Kirkeby et al., 1996b]).

Für den j -ten Lautsprecher ergeben sich die Zielfunktionen zu:

$$\mathbf{Y}_j(f) = \begin{bmatrix} m_{1,j} \cdot W_h(f) \cdot e^{-j \cdot \left(\frac{2\mathbf{p} \cdot f \cdot (\mathbf{p}_{1,j} + t)}{f_s} \right)} \\ m_{2,j} \cdot W_h(f) \cdot e^{-j \cdot \left(\frac{2\mathbf{p} \cdot f \cdot (\mathbf{p}_{2,j} + t)}{f_s} \right)} \\ \vdots \\ m_{l,j} \cdot W_h(f) \cdot e^{-j \cdot \left(\frac{2\mathbf{p} \cdot f \cdot (\mathbf{p}_{l,j} + t)}{f_s} \right)} \end{bmatrix} \quad (3.36)$$

wobei t für ein ‚processing delay‘ steht. Gemeinsam mit dem Verzögerungsglied für die Ausbreitung des Direktschallanteils bildet es das sogenannte ‚modeling delay‘.

Die gesuchten Filter ergeben sich z.T. durch die Invertierung der Übertragungsfunktionen zwischen den Kompensationslautsprecher und den Sensoren. Dadurch ergeben sich sämtliche Poltypen, auch akausale instabile Pole. Das ‚modeling delay‘ ist notwendig um die Kausalität der gesuchten Filter zu gewährleisten. Für das ‚processing delay‘ wird im Allgemeinen die Dauer der halben Filterlänge des gesuchten Kompensationsfilter gewählt. Der exakte Wert ist nicht kritisch und üblicherweise ist der Bereich von möglichen Werten groß, wobei die Ergebnisse nur geringfügig divergieren (vgl. [Widrow, Stearns, 1985]). Fällt die Energie der gesuchten Filter zu Beginn und Ende ihrer Impulsantwort ab, so ist das ein Indikator für die richtige Wahl des ‚processing-delays‘.

Im Folgenden wird die Bestimmung der Filter bzw. die Lösung des Gleichungssystems anhand des Lautsprechers 1 exemplarisch dargestellt.

$$\mathbf{Y}_1 = \mathbf{A} \cdot \mathbf{X}_1 \quad (3.37)$$

Die Lösung der Normalgleichung erhält man durch Minimierung des quadratischen Fehlers. Da die Matrix \mathbf{A} nicht quadratisch ist, muss ihre Pseudoinverse¹ gebildet werden.

$$\mathbf{X}_1 = (\mathbf{A}^H \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^H \cdot \mathbf{Y}_1 \quad (3.38)$$

Bei dieser Lösung erhält man für alle Lautsprecher ein Filterset, d.h. auch Lautsprecher 1 (L1) wird zur Kompensation herangezogen. Das approximierte Freifeldsignal ergibt sich an den Sensorpunkten und damit im definierten Raumausschnitt durch akustische Überlagerung.

Da sich die Kompensationssignale gefilterte Versionen des L1-Signals darstellen, ist in der Praxis eine Klangverfärbung zu erwarten. Eine Verbesserung wird erreicht, wenn der selektierte Lautsprecher nicht zu seiner eigenen Kompensation beiträgt. Für diesen Lösungsansatz müssen die Zielfunktionen durch Partitionierung von Gl. 3.37 entsprechend modifiziert werden .

¹ H steht für den hermitesch transponierten Operator (transponiert und anschließend konjugiert).

$$\mathbf{Y}_1 - \mathbf{A}_1 \cdot \mathbf{x}_1 = \mathbf{A}_2 \cdot \mathbf{x}_2 \quad (3.39)$$

$$\text{mit}^1 \mathbf{x}_1 = \mathbf{x}_{1,1}(f) = \text{DFT}_{2N} \{ \text{HP}(w_h(n) \otimes \mathbf{d}(n - N)) \}$$

wobei HP – Hochpassfunktion, zur Berücksichtigung der AC-Kopplung des Systems.

$$\text{und } \mathbf{x}_2 = [\mathbf{x}_{1,2}(f), \mathbf{x}_{1,3}(f), \dots, \mathbf{x}_{1,J}(f)]^T$$

$$\mathbf{A}_1 = [\mathbf{A}_{1,1}, \mathbf{A}_{1,2}, \dots, \mathbf{A}_{1,l}]^T$$

$$\mathbf{A}_2 = \begin{bmatrix} \mathbf{A}_{2,1} & \mathbf{A}_{3,1} & \dots & \mathbf{A}_{J,1} \\ \mathbf{A}_{2,2} & \mathbf{A}_{3,2} & & \mathbf{A}_{J,2} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_{2,l} & \mathbf{A}_{3,l} & \dots & \mathbf{A}_{J,l} \end{bmatrix}$$

Dadurch wird der Anteil der Kompensation auf einfache Weise regelbar und somit auch der Grad der dabei entstehenden Klangverfärbung. Da nur eine tieffrequente Lösung angestrebt wird, werden für die Bestimmung der modifizierten Zielfunktionen sowohl das Anregungssignal \mathbf{x}_1 des Lautsprecher 1 als auch die Zielfunktionen \mathbf{Y}_1 durch Hanning Impulse modelliert. Dadurch werden die Zielfunktionen für das Sensorarray modifiziert und beschreiben die inversen Reflexionen ohne Direktsignal.

$$\tilde{\mathbf{Y}}_1 = \mathbf{Y}_1 - \mathbf{A}_1 \cdot \mathbf{x}_1 \quad (3.40)$$

Die vor allem bei hohen Frequenzen schlechte Konditionierung der Matrix A wird durch einen frequenzabhängigen Regularisierungsparameter $\mathbf{a}(f)$ berücksichtigt, und es ergibt sich für die Lösung des Systems das Filterset \mathbf{x}_2 zur Kompensation der frühen Reflexionen, die durch den Lautsprechers 1 verursacht werden zu:

$$\mathbf{x}_2 = (\mathbf{A}_2^H \cdot \mathbf{A}_2 + \mathbf{a}(f) \cdot \mathbf{E})^{-1} \cdot \mathbf{A}_2^H \cdot \tilde{\mathbf{Y}}_1 \quad (3.41)$$

wobei E die Einheitsmatrix bezeichnet.

Eine Daumenregel (vgl. [Kirkeby et al., 1996b]) für eine geeignete Wahl von \mathbf{a} bzw. $\mathbf{a}(f)$ besagt, dass die Konditionszahl der zu invertierenden Matrix maximal zwischen 1000 und 5000 liegen soll.

Für die praktische Berechnung ist eine Generierung eines expliziten Direktschallanteils überflüssig, da die Zielfunktionen hier den invertierten Reflexionen entsprechen. Es genügt also das Ende des Direktschalls zu detektieren. Die gemessenen Impulsantworten im Falle des IEM CUBE weisen allerdings starke Reflexionen innerhalb der ersten 2-3ms auf. Es handelt sich dabei um Bodenreflexionen, die nicht kompensiert werden können, da für diese Richtung

¹ Das Symbol \otimes steht für eine zirkuläre Faltung.

keine Kompensationslautsprecher vorhanden sind. Daher beschränkt man sich auf die Kompensation der Reflexionen, die nach dem Direktschall und nach der ersten Reflexion (Bodenreflexion) erfolgen.

Abhängig von der Wahl des Sensorarrays ergibt sich eine obere und untere Grenzfrequenz. Innerhalb dieses Bereiches können durch die Raumkompensation in etwa Freifeldbedingungen erreicht werden. Für die Berechnung der Filter und bei deren Implementierung ist daher eine zusätzliche Abtastratenreduktion zu empfehlen. Eine Machbarkeitsstudie wird in [Krejci, 2001] beschrieben. Dabei wurden bei einer Abtastratenreduktion um einen Faktor 8 ($f_{s,red}@5512.5\text{Hz}$) Kompensationsfilter mit einer Länge von 1024 Punkten (ca.185ms) durchgeführt. Die Energie der Reflexionen konnte dabei durch die Filterung durchschnittlich um 12dB reduziert werden, wodurch manche Reflexionen unter die zeitliche Mithörschwelle sanken, für andere die Lautheit mehr als halbiert wurde. Das Ergebnis einer Lautsprecher-Sensor-Kombination ist in Abbildung 18a,b dargestellt.

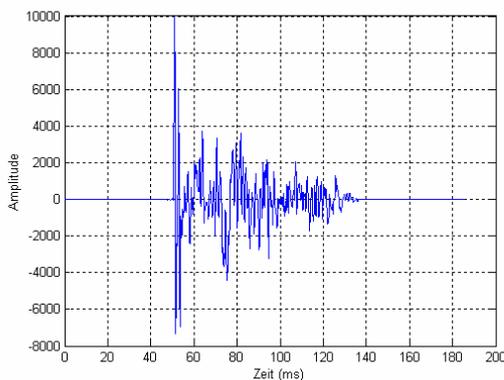


Abb. 3.18a: Ungefilterte Impulsantwort von Lautsprecher 1 an der Sensorposition 1.

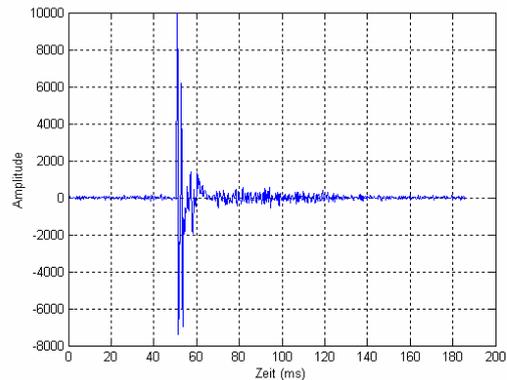


Abb. 3.18b: Simulierte gefilterte Impulsantwort von Lautsprecher 1 an der Sensorposition 1.

Es gibt keine Garantie, dass die gefundene optimale Lösung im Frequenzbereich auch tatsächlich eine realisierbare Lösung im Zeitbereich darstellt (vgl. [Kirkeby et al., 1996b]).

3.3.2.4 Lösung im Zeitbereich

Die Lösung im Zeitbereich hat den Vorteil, dass sie realisierbar ist und man die Filterlänge der gesuchten Filter relativ frei wählen kann.

Im Folgenden werden die Impulsantworten aller Lautsprecher auf alle Sensoren mit $\underline{\mathbf{A}}$ bezeichnet, die gesuchten Filterimpulsantworten werden durch den Vektor \mathbf{X} repräsentiert. Auch hier handelt es sich um ein Normalgleichungssystem.

Bei der Gradientenmethode (vgl. [Widrow, Stearns, 1985]) wird für jedes neue Eingangssample der Gradient einer Kostenfunktion, im speziellen einer quadrierten Fehlerfunktion zwischen Ziel- und Sensorsignalen die in Abhängigkeit von den Filterkoeffizienten berechnet. Die Koeffizienten der Kompensationsfilter werden entgegen Gradientenrichtung adjustiert, sodass sich der Gesamtfehler in Richtung eines lokalen Minimums bewegt. Da die Fehlerfunktion quadratisch von den Filterkoeffizienten abhängt, ist dieses Minimum auch global. Die Struktur des verwendeten Algorithmus zeigt Abbildung 3.19.

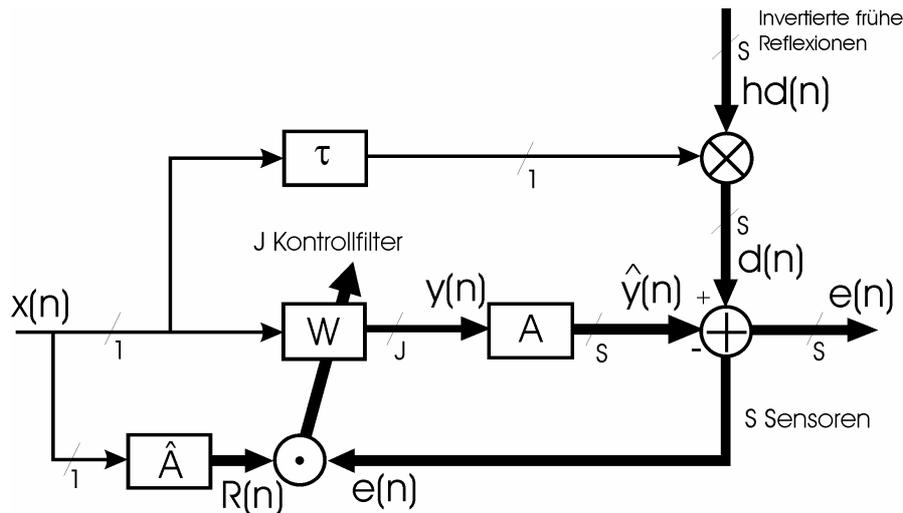


Abb. 3.19 Adaptiver Regelkreis zur Bestimmung der Filterkoeffizienten der Kompensationsfilter W .

Das Symbol \otimes steht für den Faltungsoperator.

Das System besteht aus J Kompensationslautsprechern mit den in der Matrix A zusammengefassten Übertragungspfaden zu den S Sensoren¹. Die Lautsprechersignale werden durch die Filter W bestimmt, mit deren Hilfe an den Sensoren jene Signale erzeugt werden sollen, die dem verzögerten Eingangssignal gefaltet mit den invertierten frühen Reflexionen entsprechen. Das Verzögerungsglied t (modeling delay) berücksichtigt die Verarbeitungszeit im primären Pfad und die Verzögerungszeit für die Ausbreitung zwischen Lautsprecher und Sensor. A bzw. \hat{A} beinhalten die Impulsantworten aller J Lautsprecher auf alle S Sensoren.

Im Folgenden wird auch hier nur die Kompensation eines Lautsprechers betrachtet. Die Berechnungsschritte folgen dem Adaptionsprozess für einen ‚Multiple Error FXLMS‘ - Algorithmus nach [Elliott, Nelson, 1993].

Als Eingangssignal $x(n)$ wird weißes Rauschen verwendet. Zu Beginn des adaptiven Prozesses weichen die Kompensationssignale $\hat{y}(n)$ von den Zielfunktionen $d(n)$ ab und verursachen dadurch an den S Sensoren Fehlersignale, deren Ziel ist eine energetische

¹ Die Indizes bzgl. der Sensoren werden im folgenden zur besseren Unterscheidbarkeit mit s bezeichnet.

Minimierung Ziel des Verfahrens ist. Das Zielsignal am Sensor s berechnet sich aus den invertierten frühen Reflexionen $hd_s(n)$, die mit dem zeitlich um t verzögerten Eingangssignal gefaltet werden.

$$d_s(n) = -\sum_{k=0}^K hd_s(k) \cdot x(n-k-t) \quad (3.45)$$

K gibt die Länge der invertierten frühen Reflexionen an.

Das Eingangssignal für den j -ten Kompensationslautsprecher berechnet sich aus der Faltung des Eingangssignals mit dem dazugehörigen Kompensationsfilter

$$y_j(n) = \sum_{i=0}^{I-1} w_{j,i} \cdot x(n-i) \quad (3.46)$$

I steht für die Anzahl der Koeffizienten der Kompensationsfilter w .

Das Subscript j bezeichnet den Lautsprecher, Subscript i gibt den Koeffizienten im Filter an.

Das Lautsprechersignale $y_j(n)$ werden mit den Übertragungspfaden gefiltert und führen für jeden Sensor zu einem Differenzsignal:

$$e_s(n) = d_s(n) - \sum_{j=1}^J \sum_{m=0}^{M-1} A_{s,j,m} \cdot y_j(n-m) \quad (3.47)$$

M steht für die Anzahl der Koeffizienten der Impulsantworten in A .

Durch Einsetzen für $y_j(n)$ aus Gl. 3.46 in Gl. 3.47 erhält man für den Fehler

$$e_s(n) = d_s(n) - \sum_{j=1}^J \sum_{i=0}^{I-1} A_{s,j,m} \cdot w_{j,i} \cdot x(n-i-m) \quad (3.48)$$

und weiter

$$e_s(n) = d_s(n) - \sum_{j=1}^J \sum_{i=0}^{I-1} R_{s,j}(n-i) \cdot w_{j,i} \quad (3.49)$$

$$\text{mit } R_{s,j}(n) = \sum_{m=0}^{M-1} A_{s,j,m} \cdot x(n-m) \quad (3.50)$$

$R(n)$ wird durch die Filterung des Eingangssignals mit der Impulsantwort vom j -ten Kompensationslautsprecher zum s -ten Sensor bestimmt.

Die Koeffizienten der Kompensationsfilter werden derart aktualisiert, dass die momentane Kostenfunktion $J(n)$ reduziert wird. Diese berechnet sich aus Summe der quadrierten Fehlersignale.

$$J(n) = \sum_{s=1}^S e_s^2(n) \quad (3.51)$$

Ein Minimum liegt dann vor, wenn die Ableitung der Kostenfunktion nach allen Kompensationsfiltern gleich Null ist.

$$\frac{\partial J(n)}{\partial w_{j,i}} = 2 \sum_{s=1}^S e_s(n) \cdot \frac{\partial e_s(n)}{\partial w_{j,i}} = 2 \sum_{s=1}^S e_s(n) \cdot R_{s,j}(n-i) \quad (3.52)$$

Jeder Filterkoeffizient der Kompensationsfilter wird nun durch einen Betrag proportional zu $-\frac{\partial J(n)}{\partial w_{j,i}}$ aktualisiert. Somit ergibt sich für den Adaptionprozess folgende Vorschrift für das

Kompensationsfilterset zum Zeitpunkt $n+1$:

$$w_{j,k}(n+1) = w_{j,k}(n) - \mathbf{a} \sum_{s=1}^S e_s(n) \cdot R_{s,j}(n-i) \quad (3.53)$$

mit \mathbf{a} als Konvergenzkoeffizient bzw. Adaptionsschrittgröße.

Der maximale Konvergenzkoeffizient kann wie folgt bestimmt werden

$$\mathbf{a}_{\max} = \frac{1}{R^2 \cdot (I + t)} \quad (3.54)$$

und ist damit abhängig von der mittleren Leistung der gefilterten Referenzsignale, der Anzahl der Filterkoeffizienten und der gesamten Verzögerungszeit des sekundären Pfades in Samples. Die Wahl des ‚modeling delays‘ bestimmt somit die zeitliche Dynamik des Gesamtsystems und reduziert den maximalen Konvergenzkoeffizient. Da es sich hier um eine ‚off-line‘ Berechnung handelt, ist die Wahl dieses Parameters nicht kritisch.

Zusammenfassung

Es wurden Lösungswege für die zeitinvariante Kompensation von frühen Reflexionen im Zeit- und Frequenzbereich vorgestellt.

Bei der Lösung im Frequenzbereich ist die Filterlänge der Kompensationsfilter für eine gefundene Lösung nicht beliebig veränderbar. Die Lösung garantiert keine ‚gutmütigen‘ (‚well-behaved‘) Filter im Zeitbereich.

Bei der Lösung im Zeitbereich kann die Filterlänge flexibler variiert werden. Dieser Ansatz führt jedoch zu einem großen Berechnungs- und Speicheraufwand, weshalb bei Verwendung einer direkten Lösung der Frequenzbereich vorzuziehen ist. Abhilfe schafft die adaptiv Berechnung der Filter. Modifikationen und Verbesserungen des adaptiven Prozesses hinsichtlich Konvergenzverhalten, Berechnungsaufwand und Speicheraufwand finden sich z.B. in [Douglas, 1997], [González, López, 2001] und [Bouchard, Quednau, 2000].

Allgemein gilt es festzuhalten, dass sich durch zeitinvariante Kompensationsansätze nur für stark höhenbedämpfte Räume zufrieden stellende Ergebnisse erzielen lassen.

3.3.3 durch Messungen an einer Oberfläche, zeitvariant.

Wie bereits in Kapitel 2 erwähnt, ist eine exakte Schallfeldreproduktion mit Hilfe von Ambisonic bzw. Holophonie nur dann realisierbar, wenn die akustischen Eigenschaften des Wiedergaberaumes vernachlässigt werden können. Da im Allgemeinen der Wiedergaberaum nicht vernachlässigt werden kann, wird im Folgenden ein neuer Ansatz zur Schallfeldreproduktion vorgestellt, der mit Hilfe von adaptiven Filtern diese Aufgabe erfüllt.



Abb. 3.20: (a) Aufnahme- und (b) Wiedergabeschema mit gerichteten/ungerichteten Mikrofonen und adaptiv angesteuerten Lautsprechern.

Zu Beginn steht eine Aufnahme mit einem Mikrofonarray, bestehend aus M Druck- und Schnellemikrofonen. Die Mikrofonsignale werden über eine Wiedergabematrix \mathbf{W} auf ein Lautsprecherlayout mit L Lautsprechern aufgeteilt und von diesen abgestrahlt. Gleichzeitig wird das Schallfeld mit S ($=M$) Sensoren abgetastet. Die Positionierung der Sensoren erfolgt wie bei der Aufnahme. Mittels adaptiver Filter, die durch die Sensorsignale an den räumlichen Abtastpunkten und die Aufnahme gesteuert werden, wird die Wiedergabematrix \mathbf{W} modifiziert. Dadurch stellen sich nach Konvergenz der Filter an den Sensoren (theoretisch) dieselben Signale wie bei der Aufnahme ein.

In Abbildung 3.21 ist das Strukturbild für den adaptiven Prozess dargestellt (vgl. [Kuo, Morgan, 1999],[Stearns, 1988]).

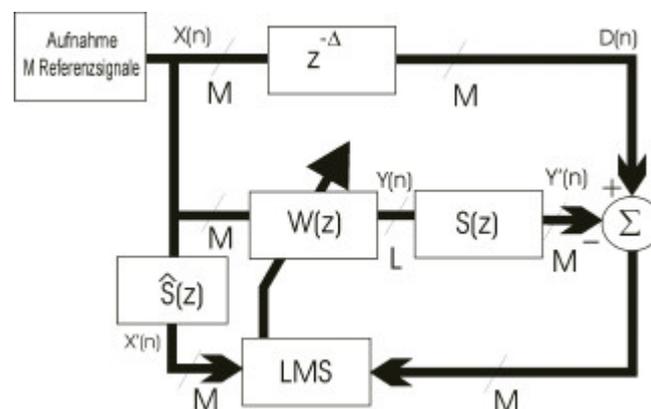


Abb. 3.21: Blockschaltbild eines adaptiven Mehrkanal -"Feedforward"-Systems zur inversen Modellierung.

Die Berechnung der Filtermatrix \mathbf{W} erfolgt analog dem Verfahren aus dem vorigen Abschnitt 3.3.2.5. Da der adaptive Prozess im Gegensatz zur oben beschriebenen Anwendung jetzt als Echtzeitsystem arbeitet, ist eine Modifizierung der Kostenfunktion ratsam, die eine Regularisierung der Filterkoeffizienten beinhaltet und damit eine Beschränkung der abgegebenen Lautsprecherleistung nach sich zieht. Dadurch werden übergroße Leistungen nahe Singularitäten vermieden. Die neue Kostenfunktion lautet

$$\mathbf{x}(n) = \sum_{s=1}^M e_s^2(n) + \mathbf{g} \cdot \sum_{l=1}^L \sum_{m=1}^M \mathbf{w}_{l,m}^T(n) \cdot \mathbf{w}_{l,m}(n) \quad (3.55)$$

wobei $\mathbf{w}_{l,m}$ der Filtervektor (der Länge I) vom m -ten Aufnahmekanal zum l -ten Wiedergabelautsprecher darstellt. Aus dieser Kostenfunktion ergibt sich der sogenannte "Leaky FXLMS"-Algorithmus. Die Filterkoeffizienten berechnen sich mit:

$$\mathbf{w}_{l,m}(n+1) = \mathbf{n} \cdot \mathbf{w}_{l,m}(n) + \mathbf{a} \sum_{s=1}^S \mathbf{x}'_{l,m,s}(n) \cdot e_s(n) \quad (3.56)$$

Der maximale Wert für \mathbf{a} ergibt sich wie aus Gl. (3.54). Der Gewichtungsfaktor \mathbf{n} ("leakage factor") liegt im Intervall $[0 < \mathbf{n} < 1]$ und wird bestimmt durch $\mathbf{n} = 1 - \mathbf{a} \cdot \mathbf{g}$. Er hat eine stabilisierende Wirkung auf den adaptiven Algorithmus (vgl. den Regularisierungsparameter in 3.3.2.3). Die konvergente Lösung ist jedoch mit einem systematischen Fehler behaftet. Der Faktor γ stellt also einen Kompromiss zwischen dem systematischen Fehler der konvergierten Gewichte und der möglichen Leistungskontrolle dar.

Zusammenfassung

Die aufgezeichneten Mikrofonensignale werden über eine Wiedergabematrix auf beliebig positionierte Lautsprecher aufgeteilt. Die Wiedergabematrix wird durch adaptive Filter realisiert, welche die invertierten Übertragungsfunktionen der sekundären Pfade, d.h. die Eigenschaften des Wiedergaberaumes, nachbilden.

Die Wiedergabe mittels adaptiver Filter ermöglicht es - zumindest theoretisch - für geeignete Lautsprecheranordnung an den „Wiedergabe“-Mikrofonen (Sensoren) den gleichen Signalverlauf wie bei der Aufnahme zu erhalten und damit bei geeigneter Positionierung der Aufnahme- bzw. Wiedergabemikrofone entlang einer volumsbegrenzenden Oberfläche innerhalb dieses Volumens das Aufnahmeschallfeld zu reproduzieren. In der Praxis würde jedoch entweder das Wiedergabevolumen sehr klein sein oder die Anzahl der Aufnahmekanäle prohibitiv hoch sein.

Darüber hinaus werden bei der Wiedergabe $L \cdot M$ ($=L \cdot S$) Filter benötigt, d.h. die Anzahl der Filter ist abhängig von der Anzahl der Aufnahme- (Sensor-) Mikrofone und der verwendeten Lautsprecher. Dieser Ansatz ist somit im Vergleich mit der „Lösung im Ambisonic-Bereich“ (vgl. Abschnitt 3.3.1) deutlich rechenaufwendiger und stellt daher keine Verbesserung dar.

Bestehende und in der Literatur bereits weitgehend untersuchte Verfahren wurden dargestellt und in das Ambisonic Gesamtsystem eingebunden. Die Kompensation von unerwünschten Raumreflexionen im Wiedergaberaum sind auch direkt im Ambisonic Bereich möglich. Durch richtungsabhängige Filterungen früher Reflexionen können die raumcharakteristischen Eigenschaften ausgeblendet werden. Der vorgestellte theoretische Ansatz stellt keine unmittelbare Performance-Verbesserung dar, sondern zeigt vor allem die Vielseitigkeit von Ambisonic. Im Vergleich zu herkömmlichen Verfahren ist hier zum Teil eine Verbesserung hinsichtlich der Recheneffizienz erkennbar. Ebenso wie bei den bekannten Verfahren ist der sinnvoll nutzbare Wirkungsbereich örtlich sowie spektral stark begrenzt.

3.4 Dreidimensionaler Nachhall

Neben der Positionierung der virtuellen Quellen, die durch Multiplikation der Quellsignale mit den Sphärischen Harmonischen $Y_{m,n}^s(\mathbf{j}_k, \mathbf{J}_k)$ (vgl. Kapitel 2.2), Superposition der so gewonnenen Ambisonic-Kanäle und anschließender Dekodierung auf das vorhandene Lautsprecherarray realisiert wird, ist die Gestaltung einer künstlichen variablen Raumakustik von großem Interesse. Im vorher gehenden Abschnitt wurden Ansätze präsentiert, die es ermöglichen den bestehenden physikalischen Raum auszublenden. Dies stellt die Basis für die nun folgenden Vorschläge zur Erzeugung einer virtuellen Raumakustik dar.

Die Verhallungskonzepte basieren wiederum auf dem Ambisonic-Ansatz. Ziel ist ein skalierbares System, dessen perzeptiv relevante Parameter den Raumeindruck gestalten lassen (z. B. richtungsabhängiger Hall). Im Folgenden werden theoretische Ansätze und Ideen vorgestellt, die in der Zukunft als Ausgangspunkt für praktische Versuche dienen sollen.

3.4.1 Modellierung von Raumimpulsantworten

Für die Simulation von beliebigen Raumsituationen sind Verhallungsalgorithmen notwendig, die auf gemessenen oder modellierten Raumimpulsantworten basieren. Eine Raumimpulsantwort kann durch das Direktsignal, frühe diskrete Reflexionen und den diffusen Nachhall modelliert werden. Grundsätzlich wäre eine separate Behandlung jeder virtuellen Quelle durch eine eigene direktionale Raumimpulsantwort notwendig. Bedingt durch die dafür notwendige hohe Rechenleistung stellt dieser Ansatz derzeit keine gangbare Lösung dar.

Eine Alternative ist die eine gemeinsame Verhallung der Ambisonic-Repräsentation. Dieses Verfahren erinnert an die Vorgangsweise bei herkömmlichen 2-Kanalproduktionen. Dadurch wäre die notwendige Rechenleistung unabhängig von der Anzahl der virtuellen Schallquellen. Allerdings können im Ambisonic-Bereich die einzelnen Schallquellen nicht mehr getrennt

manipuliert werden. Die Richtungen der ersten Reflexionen, die eng mit der Position der jeweiligen virtuellen Quelle im Raum verbunden sind, würden dadurch nur bedingt korrekt simuliert werden.

Algorithmen zur Raumsimulation können in zwei große Gruppen unterteilt werden (vgl. [Sarti, Tubaro, 2001]).

Wellentheoretische Methoden

Dabei handelt es sich um numerische Verfahren, die aus wellentheoretischen Überlegungen abgeleitet sind.

- FEM (*finite elements method*)
- BEM (*boundary elements method*)
- FDM (*finite differences method*)

Siehe dazu [Kuttruff, 1995], [Botteldooren, 1995], [Savioja et al., 1995].

Geometrische Methoden

- Spiegelquellenmethode (*image source method*)
- Strahlenmodelle (*path tracing, ray tracing, beam tracing*)
- Schallteilchenmodelle (Strahlungsaustauschmethode, *radiosity method*)

Siehe dazu [Allen, Berkley, 1979], [Borish, 1984], [Krokstad et al., 1968], [Kulowski, 1985]. Geometrische Methoden berücksichtigen jedoch keine Beugungseffekte bei tiefen Frequenzen. Die *Uniform Theory of Diffraction (UTD)* ([Mc Namara et al., 1990], [Tsingos et al., 2001]) stellt eine mögliche Methode zur Berechnung dieser Beugungseffekte dar.

Für Echtzeitanwendungen fordert die begrenzt verfügbare Rechenleistung Vereinfachungen. Wellentheoretische Modelle scheiden daher aus. Meist wird der direkte Schall und die ersten diskreten Reflexionen, die einen Eindruck über die Abmessungen des Raumes sowie die Entfernung und Ausdehnung der Quellen geben (vgl. [Griesinger, 1997]), individuell durch geometrische Methoden berechnet und die späten Reflexionen als diffuser Nachhall durch rekursive Filter modelliert. Eine übersichtliche Zusammenfassung dieses Verfahrens sowie ein praktischer Realisierungsansatz findet sich in [Savioja et al., 1999].

Die ersten Reflexionen können sehr recheneffizient nahezu korrekt mit der Hilfskonstruktion von Spiegelquellen simuliert werden.

Befindet sich der Zuhörer angenommen im Zentrum eines kugelförmigen Raumes, so ergeben sich unabhängig von der Dimension des Wiedergabesystems pro Schallquelle eine Spiegelquelle 1. Ordnung. Die Schallereignisrichtung der Spiegelquelle ist identisch mit jener der Schallquelle. Dieser Ansatz ist auch als Außenmauermodell [Moore, 1982] bekannt. Er ist sehr recheneffizient, jedoch ist der damit erreichbare Räumlichkeitseindruck noch nicht sehr

beeindruckend. Im Fall eines einfachen Rechteckraumes ergeben sich für jede virtuelle Schallquelle zusätzlich vier (2D-Fall, Rechteck siehe Abbildung 3.22) bzw. sechs (3D-Fall, Quader) Spiegelquellen. Im Vergleich zum Außenmauermodell rechtfertigt der dadurch erreichte Raumeindruck den gesteigerten Rechenaufwand.

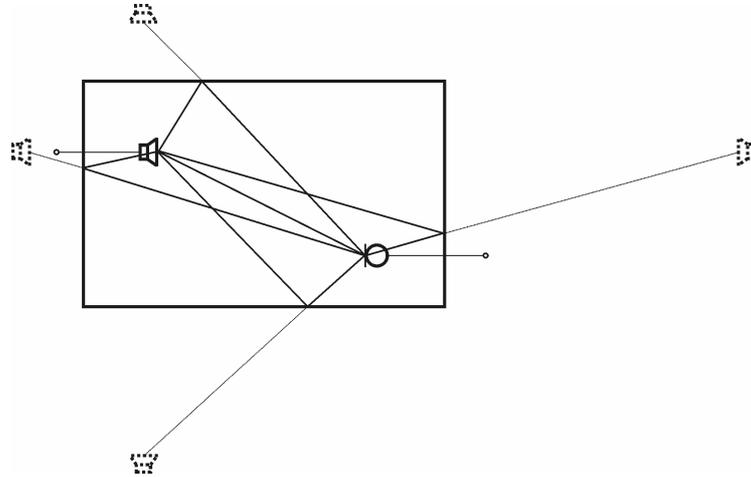


Abb. 3.22: Simulation der ersten Reflexionen basierend auf dem Spiegelquellenmodell für einen einfachen rechteckigen Raum (Spiegelquellenmodell 1.Ordnung).

Werden höhere Ordnungen von Spiegelquellen inkludiert und zusätzlich bei den Reflexionen Absorptionsgrade der Wände und Änderungen des Phasenganges berücksichtigt (d.h. ein eigener Filter für jede Reflexion), kann der Natürlichkeit des Raumeindrucks wesentlich gesteigert werden [Sontacchi et al., 2002b].

Bei Verwendung des Spiegelquellenprinzips steigt die Anzahl der virtuellen Spiegelquellen N_{vir} exponential mit der Ordnung e des Systems an¹.

$$N_{vir} = \frac{n_w}{n_w - 2} [(n_w - 1)^e - 1] \quad (3.57)$$

wobei n_w die Anzahl der reflektierenden Wände darstellt.

In [Sarti, Tubaro, 2001] wird ein recheneffizienter Algorithmus vorgestellt, der einen physikalisch plausiblen Übergang von Erstreflexionen zu Diffushall verwendet (siehe Abb.3.23). Die Erstreflexionen werden mit Hilfe eines Ringbuffers (*tapped delay*), die diffusen Reflexionen durch ein *Waveguide Digital Network* (WDN) implementiert. Dabei erfolgt eine Zerlegung der reflektierenden Wände in Knotenpunkte (*wall nodes*). Für jeden Knotenpunkt wird eine Streumatrix berechnet, welche die akustischen Eigenschaften der Reflexion (Absorptionskoeffizienten, usw.) berücksichtigt. Die Streumatrix beschreibt somit das Verhältnis der einfallenden Schallwelle zur reflektierten Schallwelle. Die Beschickung

¹ Diese Anzahl muss noch um die Anzahl der nicht sichtbaren virtuellen Spiegelquellen, abhängig von der Raumform, reduziert werden

des WDN erfolgt durch Zuordnung der Erstreflexionen zu den jeweiligen Knotenpunkten. Die Erzeugung des Diffusfelds ist mittels rekursiver Filtern implementiert, auf deren Entwurf hier nicht näher eingegangen wird.

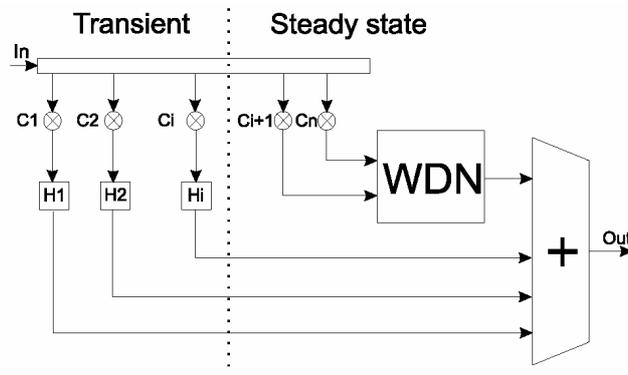


Abb. 3.23: Raumsimulation mittels Ringbuffer für die frühen Reflexionen und einem „Waveguide Digital Network“ für den Diffusschall [Sarti, Tubaro, 2001].

Die Koeffizienten C_i in der Abbildung 3.23 bewirken eine distanzabhängige Dämpfung der Reflexionssignale, die Filter H_i eine distanzabhängige Klangfärbung.

Die Beschaffenheit des diffusen Schallfeldes kann als unabhängig von der Quellposition modelliert werden. Daher kann bei Ambisonic-Kodierung des Schallfeldes eine einfache Realisierung durch Verhallung des omnidirektionalen W-Signals bewerkstelligt werden. Bei komplexeren Raumformen oder bei akustischer Koppelung von Nebenräumen mit einem Hauptraum (z. B. in Kirchen mit Haupt-, Seiten- und Querschiffen sowie mit zusätzlichen Kuppeln) wäre auch die Speisung der Diffushall-Erzeugung richtungsabhängig vorzusehen.

3.4.2 Einfaches funktionales Modell für 3D-Hall

Ein einfaches Konzept strebt nicht nach exakter physikalischer Nachbildung, sondern stellt eine funktionale Approximation einer 3D-Diffusverhallung dar.

Das Ambisonic-kodierte Schallfeld wird auf vier an den Ecken eines Tetraeders angeordneten virtuelle Lautsprecher dekodiert (siehe Abbildung 3.24).

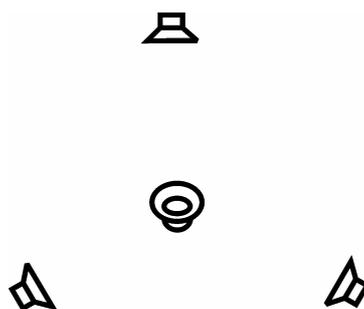


Abb. 3.24: Schallfelddekodierung auf vier Raumrichtungen, schematisiert.

Die Signale dieser virtuellen Lautsprecher werden durch unterschiedliche Hallalgorithmen verhallt. Dadurch ist eine Steuerung von unterschiedlichen richtungsabhängigen Diffusschallanteilen möglich (siehe Abb 3.25).

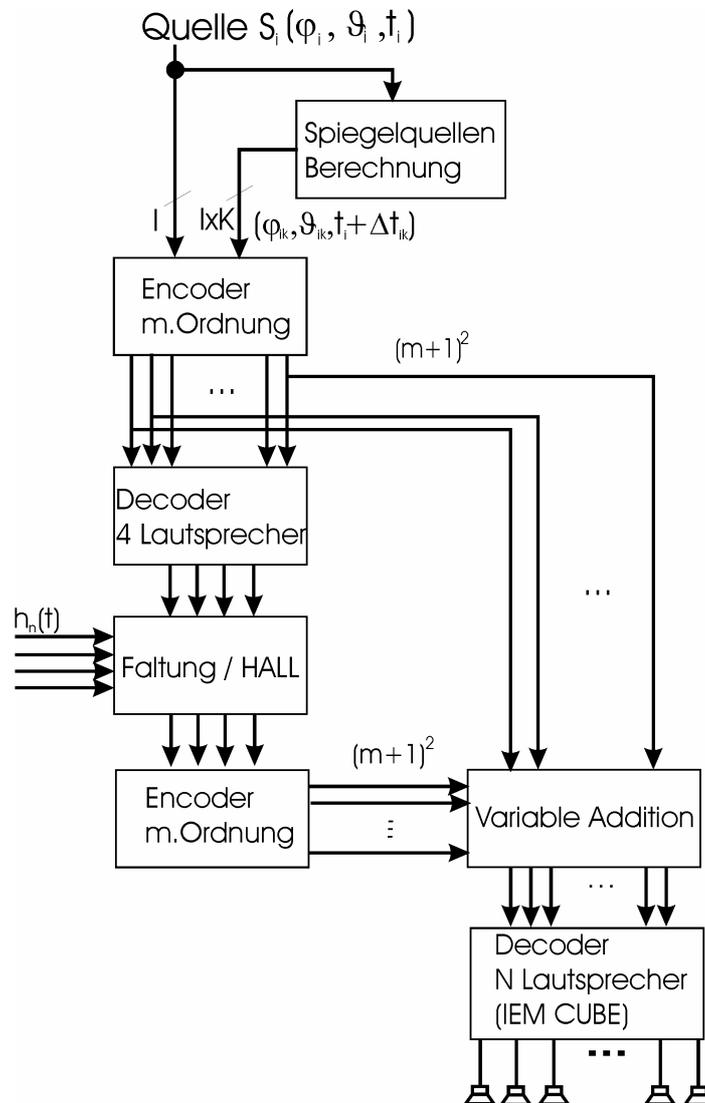


Abb. 3.25: 3D Verhallungskonzept basierend auf dem Ambisonic Ansatz 1. Ordnung.

3.4.3 Physikalisch modellierter 3D-Hall

Für die Erzeugung von physikalisch modelliertem 3D-Hall müssen die Raumimpulsantworten realer Räume dreidimensional erfasst werden, d.h. neben der zeitlichen Detektierung werden auch die Richtungen der auftretenden Reflexionen festgehalten. Die entweder mit einem Soundfieldmikrofon (für Ambisonic 1. Ordnung) bzw. mit Mikrofonarrays (vgl. [Poletti, 2000], für höhere Ordnungen) gemessen werden. Alternativ dazu können für beliebige Räume, die geometrisch erfasst und modelliert werden, dreidimensionale Raumimpulsantworten durch raumakustische Simulationsprogramme (vgl. [Krejci, 2001]) berechnet werden.

Die so erhaltenen direktionalen Raumimpulsantworten werden in den Ambisonic-Bereich m . Ordnung kodiert.

Liegt das zu verhallende Schallfeld nun ebenfalls in Ambisonic vor, so erfolgt die Verhallung durch entsprechende Faltung der einzelnen Repräsentationssignale mit den kodierten Raumimpulsantworten. Das Prinzip ist in Abbildung 3.26 als Strukturschaltbild dargestellt.

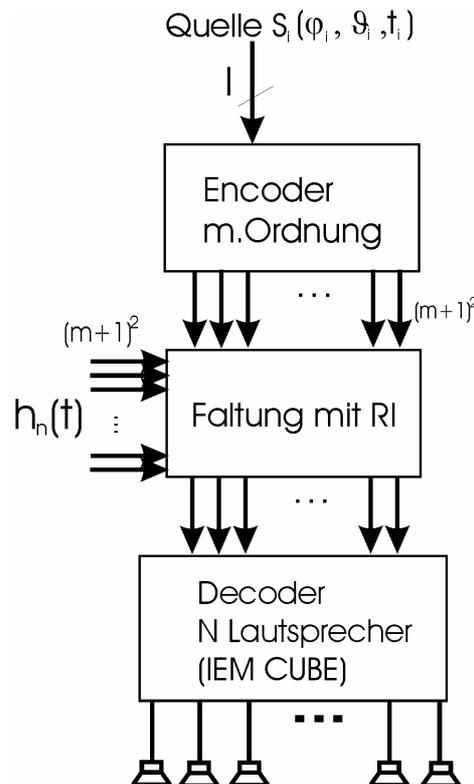


Abb. 3.26: Verhallung mit gemessenen oder berechneten 3D Raumimpulsantworten.

3.4.4. Gestaltbare Raumakustik

Eine Veränderung der Akustik eines vorhandenen physikalischen Raumes kann durch Eingriff in dessen Nachhallzeit erfolgen. Durch eine Kombination eines Wiedergabesystems mit einem Aufnahmesystem kann die Nachhallzeit bei kontrollierter Rückkopplung gezielt verlängert werden. Grundsätzlich lassen sich diese Systeme in die zwei Kategorien „in-line“ Systeme und „non in-line“ Systeme unterteilen.

Ziel eines „in-line“ Konzepts ist die Aufnahme des Direktschalls, der künstlich verhallt und anschließend an unterschiedlichen Raumpunkten wiedergegeben wird. Deshalb müssen die Mikrofone möglichst nahe bei den Schallquellen und entfernt von den Wiedergabelautsprechern positioniert werden. Sie können lokale Eigenschaften der Raumakustik wie Klarheit, wahrgenommene Quellweite (apparent source width, ASW) und frühe Abfallzeit (early decay time - EDT) für Quellen auf der Bühne leicht erhöhen.

Bei „non in-line“ Systemen werden die Mikrofone hingegen außerhalb des Hallradius platziert und nehmen hauptsächlich das diffuse Hallfeld auf. Dieses wird zusätzlich verstärkt und - wenn nötig - künstlich verändert, bevor es mit Lautsprecher an anderen Punkten wiedergegeben wird. Die Mikrofone sind hier mindestens außerhalb des Hallradius des zunächst stehenden aktiven Wiedergabelautsprechers anzuordnen. Diese Systeme eignen sich gut, um globale raumakustischen Eigenschaften wie Lautheit, Nachhallzeit und das Gefühl der Schalleinhüllung (envelopment) zu erhöhen.

Neuere Entwicklung gehen in die Richtung einer Kombination der beiden Systeme, um die Nachteile der Einzelsysteme zu kompensieren und die Eingriffsmöglichkeiten auf die raumakustischen Eigenschaften zu erweitern. Diese Systeme werden Hybrid Systeme genannt (vgl. LARES [Griesinger, 1991] , VRAS [Poletti, 1993 u. 1996b]).

Das wesentlichste Kriterium bei der Dimensionierung solcher Systeme stellt die Wahl der Schleifenverstärkung dar. Zu kleine Verstärkungsfaktoren bewirken nur bedingt eine wahrnehmbare Vergrößerung des Raumes, d. h. die damit erzielbare Verlängerung der Nachhallzeit ist zu gering¹. Bei zu großen Verstärkungsfaktoren treten ohne zusätzliche Vorkehrungen Artefakte wie vor allem Klangfärbungen und Mitkopplungsneigung auf. Bei [Ahnert, Reichard, 1981] findet sich eine Beschreibung verschiedener Ansätze zur Erhöhung der maximalen Schleifenverstärkung. Durch den Übergang zu zeitvarianten Übertragungsfunktionen zwischen Aufnahme- und Wiedergabepunkten kann eine Mitkopplung verhindert werden. Verwendet werden Wobbelung (vgl. [Griesinger, 1991], [Lokki, Hiipakka, 2001]), Frequenzverschiebung ([Schröder, 1959], [Schröder, 1962], [Burkhard, 1963]) sowie Kanalschaltung (vgl. [Ahnert, 1975]), die für den Ambisonic-Ansatz besonders geeignet zu sein scheint. Durch zeitvariante Rotationsmatrizen kann die Kanalschaltung relativ einfach und recheneffizient realisiert werden.

Zusammenfassung

Die Implementierung von 3 dimensional Hallalgorithmen ermöglicht eine im Vergleich zu traditionellen Surroundsystemen weitgehende Berücksichtigung der Erkenntnisse der Psychoakustik im Bereich der räumlichen Wahrnehmung und kann im vorgestellten Ambisonic System recheneffizient umgesetzt werden.

¹ $T_m = \frac{T_0}{1 - v^n}$ nach [Curtis, 1969], wobei T_m Nachhallzeit mit elektroakust. Anlage, T_0 natürl. Nachhallzeit, v Schleifenverstärkung und $n=1$ bei Druckaddition (Nahfeld) und $n=2$ bei Energieaddition (Fernfeld).

3.5 Distanzkodierung virtueller Quellen

Die Modellierung der Distanzkodierung basiert auf der Krümmung der einfallenden Direktschall-Wellenfront. Das Schallfeld einer nahen, punktförmigen Schallquelle besitzt eine relative starke Krümmung (Kugelwelle, Kugelstrahler), wohingegen das Schallfeld einer weit entfernten Quelle immer mehr einer ebenen Welle entspricht (siehe Abb. 3.27).

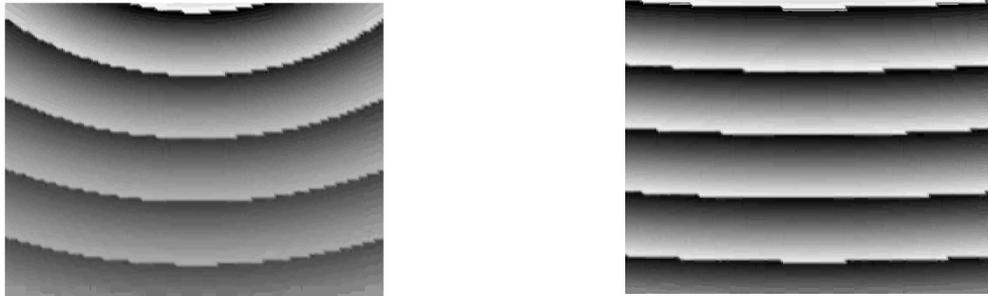


Abb. 3.27: Monochromatische Wellenfronten (800 Hz) für eine punktförmige Quelle. Links in 2m und rechts in 11m Entfernung, bezogen auf die jeweilige Mitte des Darstellungsbereiches. Größe des Darstellungsbereich 2x2m.

Im 1. Kapitel wurden bereits Phänomene zur Beurteilung der Quellenentfernung diskutiert. Diese sind zum Teil unbefriedigend und nicht ausreichend für eine realistische Reproduktion. Der in Kapitel 2.2 vorgestellte Ambisonic-Ansatz bietet die Möglichkeit, relativ recheneffizient und mit einer relativ geringen Lautsprecheranzahl auch für größere Auditorien zufriedenstellende Ergebnisse für variabel positionierbare virtuelle Schallquellen zu erreichen. Mit diesem Ansatz allein kann jedoch die wahrgenommene Entfernung zwischen Hörer und virtueller Quelle nicht verändert werden. Sie ist mehr oder weniger von der Entfernung der physikalischen Lautsprecher abhängig. Der Ansatz der Wellenfeldsynthese (WFS) in Kapitel 2.3 ermöglicht eine freie Distanzkodierung, jedoch zum Preis einer rechen- und hardwareintensiven Lösung.

Der hier vor zu stellende Ansatz setzt auf die physikalische Reproduktion der originalen Krümmung des Schallfeldes einer Referenzquelle. Es wird dabei eine Kompromisslösung zwischen Ambisonic und WFS angestrebt, die eine skalierbare und hardwareeffiziente Umsetzung ermöglicht. Bedingt durch den Wunsch nach geringer Kanalanzahl und die physikalische Ausdehnung der elektroakustischen Wandler sind derzeit realisierbare Systeme meist tieffrequente Approximationen.

Die Krümmung der synthetisierten Wellenfront kann durch die Verteilung des komplexen Schnellektors (bzw. des komplexen Intensitätsvektors) und somit durch die Energieflussrichtung eines Schallfeldes (vgl. 3.2) beschrieben werden. Dadurch kann die Distanzkodierung auch objektiv optimiert werden. Die subjektive Qualität der durch die Wellenfrontkrümmung wahrgenommenen Entfernungsempfindung ist bisher noch nicht ausreichend untersucht worden. Dieser Umstand ist vor allem dadurch zu begründen, dass die Krümmung nur bei wenigen Beschallungssystemen kontrolliert erzeugt werden kann.

3.5.1 Ansatz

Der folgende Ansatz zur Positions- und Distanzsteuerung von virtuellen Quellen ist ein 2-stufiges Konzept, welches die Vorteile der Wellenfeldsynthese (vgl. Kap 2.3) und des Ambisonic-Ansatz (Ambisonic unter Verwendung höherer Ordnung, im folgenden HOA 'higher order ambisonics') zu vereinen versucht (vgl. [Sontacchi, Höldrich, 2002]). Bei der Lautsprecherwiedergabe wird das Hauptaugenmerk auf die Anzahl der notwendigen Übertragungskanäle und auf die Unabhängigkeit der Kodierungsvorschrift von der Lautsprecheranordnung gelegt.

In der ersten Stufe erfolgt für jede abzubildende Quelle in Abhängigkeit von ihrer Entfernung zum idealen Abhörpunkt eine Transformation (Abbildung 3.28, Part 1) mit Hilfe des WFS-Ansatzes (vgl. Kap 2.4) auf ein virtuelles Lautsprecherarray (Anzahl und Anordnung siehe Abschnitt 3.5.1.1). Durch diese Transformation soll die Entfernung gesteuert werden. Die dadurch erhaltenen virtuellen Lautsprechersignale werden anschließend in der zweiten Stufe, in Abhängigkeit vom abzubildenden Einfallswinkel der Quelle, in den Ambisonic-Bereich kodiert. Die Ambisonic-Kanäle, deren Anzahl unabhängig von der Anzahl der Schallquellen ist, werden gespeichert und durch eine vom Wiedergabe-Layout abhängige Dekodierungsvorschrift über ein reales Lautsprecherarray wiedergegeben (Abbildung 3.28, Part 2).

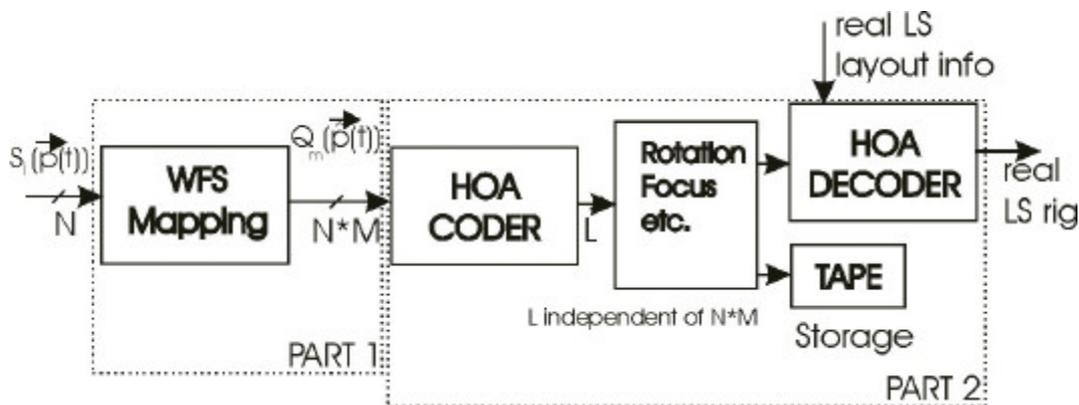


Abb. 3.28: Systemdesign Blockschaubild für die Wiedergabe mit Lautsprecher

Die folgenden Betrachtungen und Herleitungen erfolgen für den zwei-dimensionalen Fall.

3.5.2 Anzahl und Anordnung der virtuellen Lautsprecher

Die Anordnung und Anzahl der Lautsprecher ist, ähnlich wie bei der Sensoranordnung in Kapitel 3.3, ausschlaggebend für die Wiedergabebandbreite und die räumliche Auflösung. Sind die Lautsprecher zu weit voneinander entfernt, so können hochfrequente Richtungskomponenten nicht eindeutig abgebildet werden (räumliches Aliasing). Ist jedoch die Gesamtapertur der virtuellen Lautsprecheranordnung zu gering, so wird die Abbildung für

tieffrequente Richtungskomponenten durch relativ hohe Gewichte (zusätzliche Energie) erkauft und der optimale Wiedergabebereich um den Ursprung verkleinert. Artefakte werden in peripheren Abhörpunkten zusätzlich verstärkt.

Aus den minimalen Abständen der virtuellen Lautsprecher (Winkelauflösung $\Delta j_{L,\min}$) ergibt sich die minimal notwendige Ambisonic-Ordnung. Die maximale Winkelauflösung für ein 2D-Ambisonic-System ergibt sich aus der Systemordnung M .

$$j_{\text{Amb}} = \frac{2p}{2M+1} \quad (3.58)$$

Somit kann für ein virtuelles Lautsprecherlayout mit einem $\Delta j_{L,\min}$ die notwendige Ambisonic- Ordnung M berechnet werden mit

$$M_{\min} = n \cdot \left\lceil \frac{p}{j_{L,\min}} - \frac{1}{2} \right\rceil \quad (3.59)$$

$\lceil \cdot \rceil$ bedeutet, dass M_{\min} zum nächst größeren Integerwert gerundet wird.

Der Faktor n ergibt sich aus der vollständigen Trennung der räumlichen Hauptkeulen und sollte idealerweise zwischen 1 und 2 liegen damit $j_{\text{Amb}} \leq \Delta j_{L,\min}$ gilt.

Wird die Ordnung zu gering gewählt, können die virtuellen Lautsprecher nicht mehr getrennt voneinander aufgelöst werden. Die diskrete Verteilung der Quellen über den Winkel ('angular source distribution' über $2p$) wird zunehmend verschmiert. Ideal wären einzelne spatiale Impulse zur Beschreibung der kodierten virtuellen Quellen, jedoch bewirkt die Kodierung in den Ambisonic Domain eine (räumliche) Faltung mit der Richtcharakteristik des Ambisonicdecoder/decoder (siehe Abb 3.29).

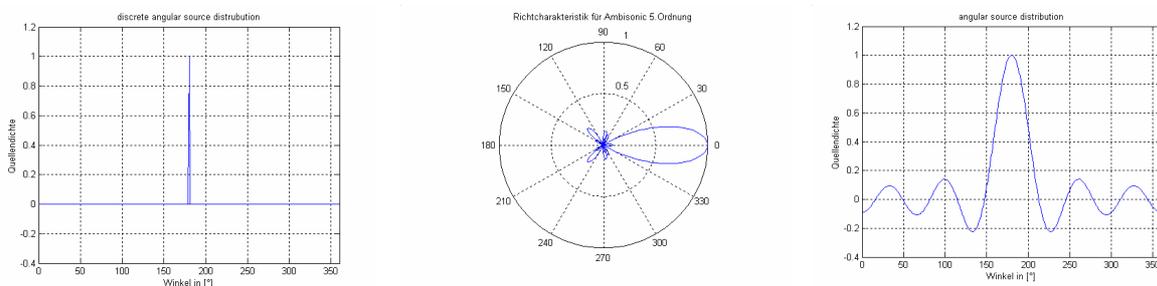


Abb. 3.29: Diskrete räumliche Quellenrepräsentation (links), Richtcharakteristik eines Ambisonic Coders/Decoders (mitte) und Quellenverteilung (angular source distribution) durch die endliche Kodierung (rechts).

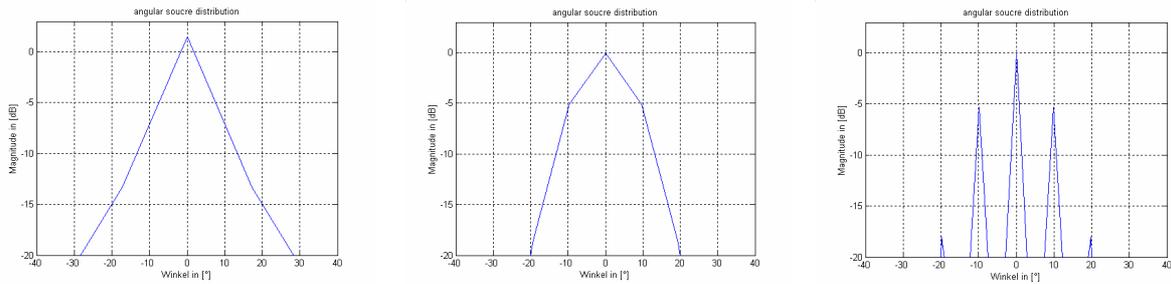


Abb. 3.30: Die 'angular source distribution' für fünf (virtuelle) Quellen bei Verwendung unterschiedlicher Ambisonic Ordnungen. Links $M=10$, mitte $M=18$ und rechts $M=36$.

In Abbildung 3.30 sind die normierten Beträge der komplexen Gewichte von fünf symmetrisch angeordneten virtuellen Quellen dargestellt. Die zu kodierenden Werte betragen 0dB für 0° , -5.3dB für $\pm 10^\circ$ und -18.3dB für $\pm 20^\circ$. Der Winkelabstand zwischen den Quellen beträgt jeweils 10° . Nach Gleichung 3.64 ergibt sich als minimale Ambisonic-Ordnung $M=18$. Auf der linken Seite in Abbildung 3.30 wurde eine Ambisonic-Kodierung 10. Ordnung verwendet. Aus dem Verlauf der kodierten Gewichte (angular source distribution) ist unschwer eine stark Verschmierung der wahren Gewichte erkennbar. In der Mitte der Abbildungsserie kann man bereits die wahren Werte ablesen, wobei die einzelnen Quellen noch nicht isoliert erkennbar sind ($n=1$). Auf der rechten Seite der Abbildungsfolge erfolgt mit $M=36$ eine perfekte Reproduktion der Gewichte ($n=2$).

Die Anordnung der virtuellen Lautsprecher erfolgt entlang eines beschränkten Bereiches einer Kreislinie (2D) bzw. auf einem Kugelflächensegment (3D). Jeder virtuelle Lautsprecher ist vom Ursprung (idealer, zentraler Abhörpunkt) gleich weit entfernt und symmetrisch um die Blickrichtung (x-Achse) positioniert (siehe Abb 3.31).

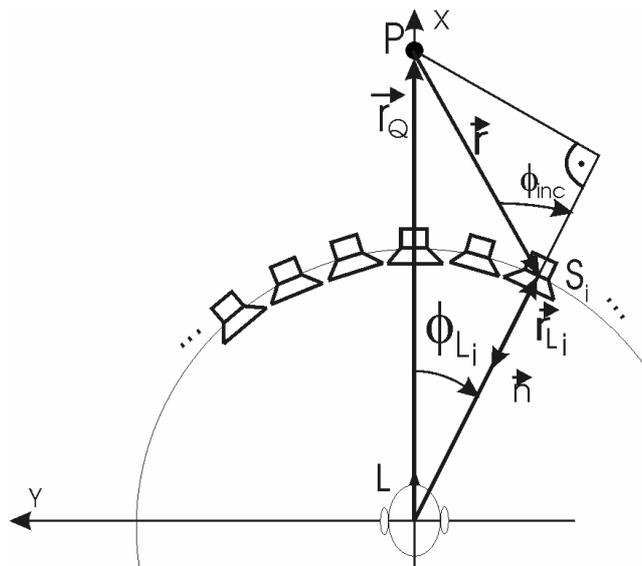


Abb. 3.31: Die virtuellen Lautsprecherquellen befinden sich auf einer Kreislinie in den beiden vorderen Quadranten, wobei sie um die Blickrichtung (x-Achse) zentriert sind.

3.5.3 Berechnung der Lautsprecher-signale Q_m mit Hilfe der WFS

Die Berechnung der virtuellen Lautsprecher-signale erfolgt analog zur Herleitung in Kapitel 2.4. Die Geometrie des virtuellen Lautsprecherarrays sowie deren Bezeichnungen ist in Abbildung 3.31 gegeben.

Die abzubildende primäre Quelle P an der Position \vec{r}_Q besitzt das Quellspektrum $S(\mathbf{w})$. Die virtuellen Lautsprecher (sekundäre Quellen S_i) an den Positionen \vec{r}_{L_i} müssen mit den Lautsprecher-signalen $Q_{m_i}(\mathbf{w})$ angesteuert werden, um in einer Umgebung um den idealen Abhörpunkt L (Koordinatenursprung in Abbildung 3.31) das Feld der Primärquelle zu reproduzieren.

$$Q_{m_i}(\mathbf{w}) = Q_m(\vec{r}_{L_i}, \mathbf{w}) = D(\vec{r}_i, \mathbf{w}) \cdot S(\mathbf{w}) \quad (3.60)$$

$$P_{prim}(\vec{r}, \mathbf{w}) = S(\mathbf{w}) \cdot A_{prim}(\vec{r}_{Q,r}, \mathbf{w}) \quad (3.61)$$

$$\text{mit } \vec{r}_{Q,r} = \vec{r}_Q - \vec{r}_r$$

$$P_{syn}(\vec{r}, \mathbf{w}) = \sum_{i=1}^I Q_m(\vec{r}_{L_i}, \mathbf{w}) \cdot A_{S_i}(\vec{r}_{L,r}, \mathbf{w}) \quad (3.62)$$

$$\text{mit } \vec{r}_{L,r} = \vec{r}_L - \vec{r}_r$$

Das Lautsprecher-signal ist abhängig von der Anordnung der virtuellen Lautsprecher, deren Entfernung vom Ursprung sowie von der Entfernung der abzubildenden Quelle. $A_{S_i}(\vec{r}_{L,r}, \mathbf{w})$ beschreibt die Ausbreitung des Schallfeldes von der jeweiligen sekundären Quelle und wird durch die Ausbreitungsfunktion einer Punktquelle modelliert.

Die „Driving Function“ $D(\vec{r}_i, \mathbf{w})$ beschreibt die Übertragungsfunktionen zwischen der Primärquelle und den Sekundärquellen (virtuelle Lautsprecher). Aus Gl. 2.77 in Kapitel 2.4 ergibt sich bei sekundären Monopolquellen entlang gekrümmter Linienformen und unter Verwendung der geometrischen Relationen in Abbildung 3.31 für die „Driving Functions“

$$D(\vec{r}_i, \mathbf{w}) = \sqrt{\frac{jk}{2p}} \cdot \sqrt{\frac{|\vec{r}_{L_i}|}{|\vec{r}_i| + |\vec{r}_{L_i}|}} \cdot \cos \mathbf{f}_{inc_i} \cdot \frac{e^{-jk|\vec{r}_{L_i}|}}{\sqrt{|\vec{r}_{L_i}|}} \quad (3.63)$$

Nach einigen Umformungen und unter der Verwendung der geometrischen Beziehungen $r_0 = |\vec{r}_{L_i}|$ und $r_Q = |\vec{r}_Q|$ erhält man:

$$r_i = |\vec{r}_i| = \sqrt{r_0^2 + r_Q^2 - 2 \cdot r_0 \cdot r_Q \cdot \cos \mathbf{f}_{L_i}} \quad (3.64)$$

$$\cos \mathbf{f}_{inc_i} = \frac{r_Q \cdot \cos \mathbf{f}_{L_i} - r_0}{r_i} \quad (3.65)$$

Für $r_Q > r_0$, ergeben sich die „Driving functions“ zu:

$$D(\vec{r}_i, \mathbf{w}) = \sqrt{\frac{jk}{2\mathbf{p}}} \cdot \sqrt{\frac{r_0}{r_0 + r_i}} \cdot \frac{r_Q \cdot \cos \mathbf{f}_{L_i} - r_0}{r_i} \cdot \frac{e^{-jk r_i}}{\sqrt{r_i}} \quad (3.66)$$

Für $r_Q < r_0$ ergeben sich die „Driving functions“ zu:

$$D(\vec{r}_i, \mathbf{w}) = \sqrt{\frac{k}{j2\mathbf{p}}} \cdot \sqrt{\frac{r_0}{r_0 - r_i}} \cdot \frac{r_Q \cdot \cos \mathbf{f}_{L_i} - r_0}{r_i} \cdot \frac{e^{-jk r_i}}{\sqrt{r_i}} \quad (3.67)$$

und für $r_Q = r_0$ ergibt sich die „Driving functions“ zu:

$$D(\vec{r}_i, \mathbf{w}) = \begin{cases} 1 & \text{für } \cos \mathbf{f}_{inc} = 1 \\ 0 & \text{sonst} \end{cases} \quad (3.68)$$

Für den Fall $r_Q = r_0$ kann die Quelle durch einen einzigen (virtuellen bzw. realen) Lautsprecher dargestellt werden (vgl. Gleichung 3.72). Im allgemeinen Fall werden die Ausbreitungsterme $1/\sqrt{r_i}$, die sich aus den „Driving functions“ ergeben, vernachlässigt. Dies führt zu Amplitudenfehlern, die jedoch bei Beschränkung auf einen kleinen Wiedergabebereich tolerierbar sind. Die Distanzkodierung kann zusammengefasst durch ein Verstärkungs- und Verzögerungsglied (abhängig vom Distanzverhältnis r_Q/r_0) sowie einen $\sqrt{\mathbf{w}}$ -proportionalen Filter realisiert werden. Wie bereits im Kapitel 2.4 gezeigt wurde, müssen bei endlichen Arrays zusätzlich noch die Randeffekte kompensiert werden. Eine praktische Realisierung wurde in [Strauß, 2002] bzw. [Sontacchi et al., 2002a] vorgestellt.

Die Anzahl der virtuellen Lautsprecher sowie deren Abstände zueinander sind die Designparameter für dieses Distanzkodierungsmodell. Diese können nach unterschiedlichen Kriterien optimiert werden. Die Array-Apertur bestimmt die Ausdehnung des optimalen Wiedergabebereiches. Abhängig von der Entfernung der primären Quelle zum Array ergibt sich für die Krümmung des Schallfeldes bei hohen Frequenzen eine Grenzfrequenz durch die Lautsprecherabstände. Eine Vergrößerung der Lautsprecherabstände bei gegebener Anzahl vergrößert somit den Wiedergabebereich auf Kosten von verstärktem Aliasing, d.h. die Grenzfrequenz für eindeutige räumliche Wiedergabe sinkt. Für hohe Frequenzen müssten die Abstände entsprechend kleiner gewählt werden, um eine eindeutige Rekonstruktion zu ermöglichen. Werden für tiefe Frequenzen ebenfalls kleine Abstände verwendet, so hängt das Ergebnis stark vom Berechnungsweg der Lautsprechersignale ab (siehe 3.5.1.3).

3.5.4 Optimierung der Lautsprecher signale mit Hilfe des LMS Ansatzes

Die „Driving Functions“ (Filter) können auch durch einen LMS Ansatz bestimmt werden (vgl. [Kirkeby, Nelson, 1993]).

Die Berechnung der komplexen Gewichte \mathbf{a} (Volumbeschleunigung, zeitliche Ableitung der Quellenstärke) für die virtuellen Lautsprecher erfolgt über die Minimierung eines mittleren quadratischen Fehlers. Dazu wird ein Gitter um den Ursprung aufgespannt, für das jeweils das Referenzdruckfeld \mathbf{p} der abzubildenden Quelle und das synthetisierte Feld $\hat{\mathbf{p}}$ bestimmt wird.

Die Kostenfunktion J ergibt sich aus der Summe der quadrierten Druckdifferenzen zwischen den beiden Feldern an den R Referenzpunkten.

$$J = \sum_{r=1}^R |e_r|^2 = \mathbf{e}^H \cdot \mathbf{e} \quad (3.69)$$

$$J = (\mathbf{p} - \hat{\mathbf{p}})^H \cdot (\mathbf{p} - \hat{\mathbf{p}}) = (\mathbf{p} - \mathbf{H}\mathbf{a})^H \cdot (\mathbf{p} - \mathbf{H}\mathbf{a}) \quad (3.70)$$

Die Matrix \mathbf{H} in Gleichung 3.70 (auch modifizierte Impedanzmatrix genannt [Kirkeby, Nelson, 1993]), beschreibt dabei die Geometrie bzw. die Ausbreitungsterme der einzelnen virtuellen Lautsprecher zu den jeweiligen Referenzpunkten.

$$J = \mathbf{p}^H \cdot \mathbf{p} - \mathbf{a}^H \cdot (\mathbf{H}^H \cdot \mathbf{p}) - (\mathbf{p}^H \cdot \mathbf{H}) \cdot \mathbf{a} + \mathbf{a}^H \cdot (\mathbf{H}^H \cdot \mathbf{H}) \cdot \mathbf{a} \quad (3.71)$$

Der optimale Gewichtsvektor \mathbf{a}_0 ergibt sich durch partielles ableiten der Kostenfunktion J nach jedem Gewicht a_i . In kompakter Schreibweise ergibt sich als Lösung die Pseudoinverse (pinv) der Matrix \mathbf{H}

$$\mathbf{a}_0 = (\mathbf{H}^H \cdot \mathbf{H})^{-1} \cdot \mathbf{H}^H \cdot \mathbf{p} = \mathbf{H}_{\text{pinv}} \cdot \mathbf{p} \quad (3.72)$$

Auch hier sollte die Matrix $\mathbf{H}^H \cdot \mathbf{H}$ eine kleine Konditionszahl besitzen (vgl. [Nelson, Elliott, 1992]). Aus Untersuchungen (vgl. auch [Kirkeby, Nelson, 1993]) geht somit hervor, dass für tiefe Frequenzen ein größerer Abstand zwischen den virtuellen Quellen geeigneter ist (siehe Abbildung 3.32).

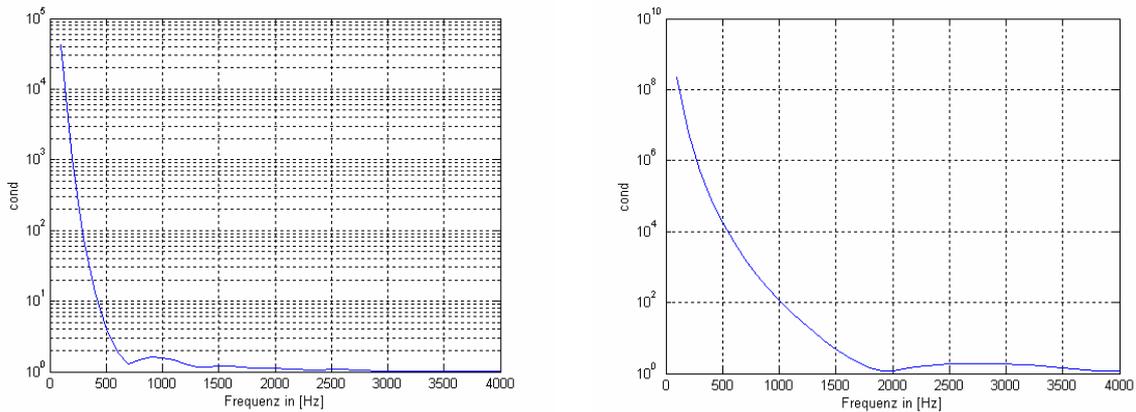


Abb. 3.32: Konditionszahl cond der Matrix $\mathbf{H}^H \cdot \mathbf{H}$ für zwei Layouts.

Links: Winkel zwischen den virtuellen Quellen 30° . Rechts: 10° .

Die Berechnung erfolgt für einen Referenzausschnitt mit 1x1 Meter.

Im Referenzausschnitt sind 51x51 Referenzpunkte äquidistant angeordnet.

Die Eigenfrequenz des Systems liegt bei 17kHz und somit weit außerhalb des verwendeten Messbereiches.

Die Konditionszahl wird nur durch die Anordnung der Quellen und Referenzpunkte bestimmt und ist völlig unabhängig von der untersuchten Distanzkodierung. Für ein optimiertes Modell sollten daher für hohe und tiefe Frequenzen unterschiedliche Layouts verwendet werden (siehe [Takeuchi, Nelson, 2000b], [Täschl, 2000]) was für die gesuchte tieffrequente Approximation dem Wunsch nach einer niedrigen Ambisonic-Ordnung entgegen kommt.

Eine vereinfachte Modellierung der Filter für variable Distanzen unter Verwendung des LMS Ansatzes ist nicht durch frequenzunabhängige Delays und Verstärkungsfaktoren realisierbar (vgl. Abbildung 3.33). Im Gegensatz zu den Filtern, die sich durch den WFS-Ansatz ergeben, besitzen diese relativ große Gewichte für tiefe Frequenzen. Im Vergleich zum WFS-Ansatz kann die notwendige virtuelle Quellenanzahl reduziert werden. Durch die Verwendung eines symmetrischen Layouts reduziert sich zusätzlich die Anzahl der benötigten Filter.

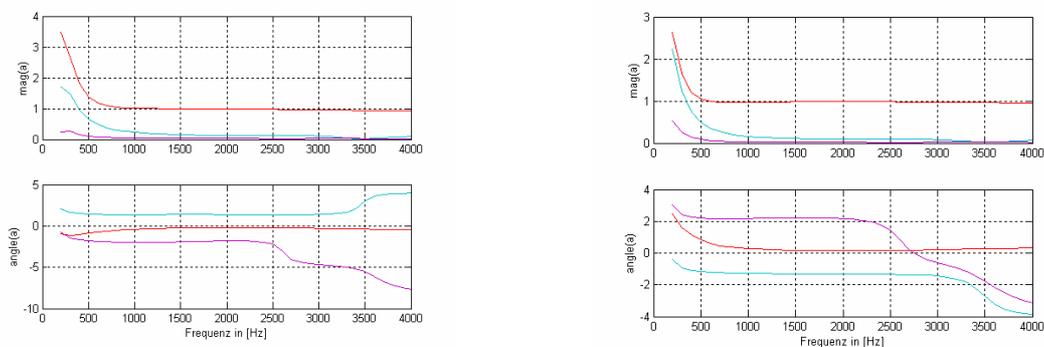


Abb. 3.33: Verlauf der komplexen Gewichte (Amplituden- und Phasengang) für ein Layout bestehend aus 5 virtuellen Lautsprechern in 5m Entfernung mit konst. Versatzwinkel von 10° .

Links: Für eine Quelle in 3m Entfernung. Rechts: Für eine Quelle in 10m Entfernung.

3.5.5 Objektive Analyse der erzeugten Schallfelder

Anhand des oben verwendeten Lautsprecherlayouts (5 virtuelle LS in 5 m Entfernung, Versatzwinkel von 10°) wird das synthetisierte monochromatische Schallfeld, bei einer Frequenz von 1kHz, für eine kodierte Quelle aus 2m Entfernung analysiert. Die Signale virtuellen Lautsprecher werden mit einem Ambisonic-System 18. Ordnung kodiert und mit Hilfe des basic-Decoders (vgl. Kapitel 2.2) auf ein symmetrisches Lautsprecherlayout dekodiert.

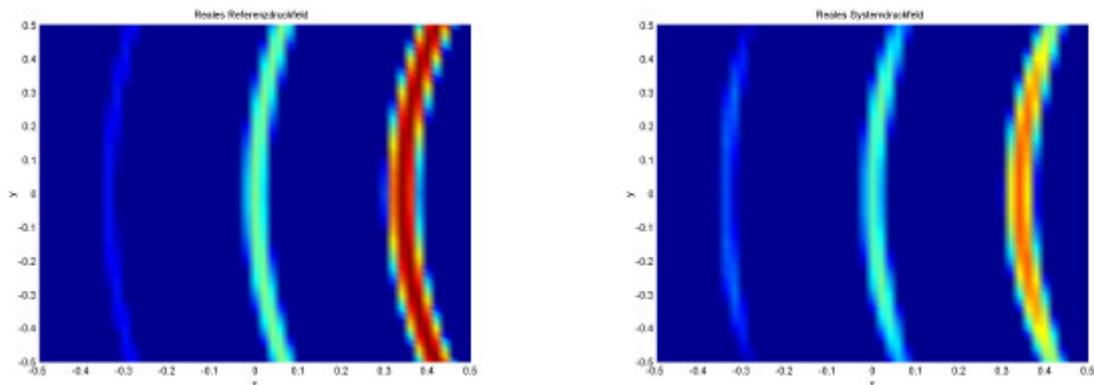


Abb. 3.34: Links: Reales Schalldruckfeld einer monochromatischen Schallquelle mit 1kHz in 2.m Entfernung (Referenzfeld).

Rechts: Synthetisiertes Feld mit 5 virtuellen Quellen und Ambisonic 18.Ordnung.

Zur verbesserten Hervorhebung des Verlaufs der Wellenfronten wurde in Abbildung 3.34 die Amplitudenwerte nicht-linear verzerrt dargestellt. Betrachtet man jeweils jene Wellenfronten die sich in einer Umgebung von $x=0$ ergeben, so stimmen die Verläufe von Referenz- und Synthesefeld relativ gut überein.

In Abbildung 3.35 sind die dazugehörigen relativen Amplitudenfehler in dB sowie die Richtungsabweichungen in Grad dargestellt.

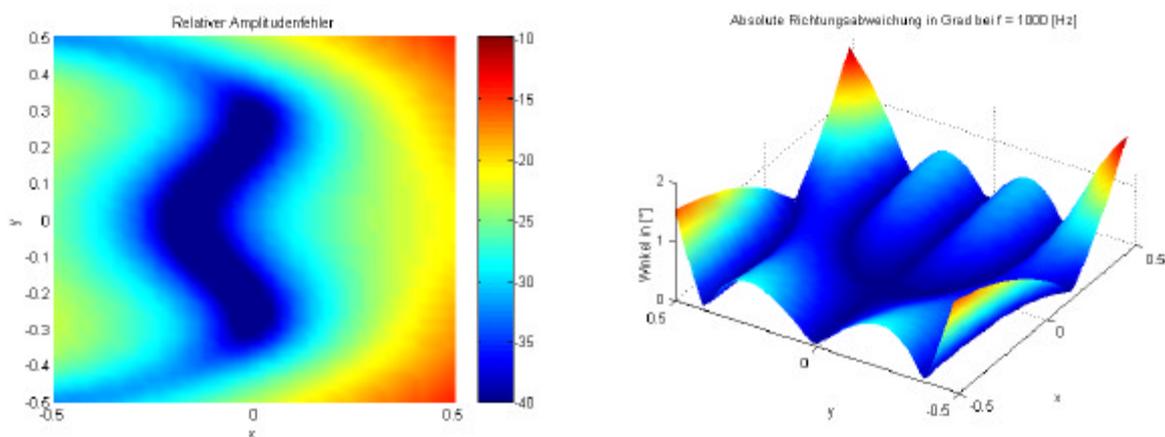


Abb. 3.35: Links: Relativer Amplitudenfehler in dB. Rechts: Richtungsabweichung in Grad.

Mit Hilfe des Schallschnellektors \mathbf{u} (zeitlich gemittelte Geschwindigkeit der Schallflussenergie, vgl. Kapitel 3.2) kann die Richtung des Energieflusses und somit die Krümmung der erzeugten Wellenfront ermittelt werden (siehe Abb.3.36). Der Schnellektor steht normal zur Wellenfront und kann für jeden Raumpunkt, in dem die Ausbreitung erfolgt, bestimmt werden (vgl. Kapitel 3.2).

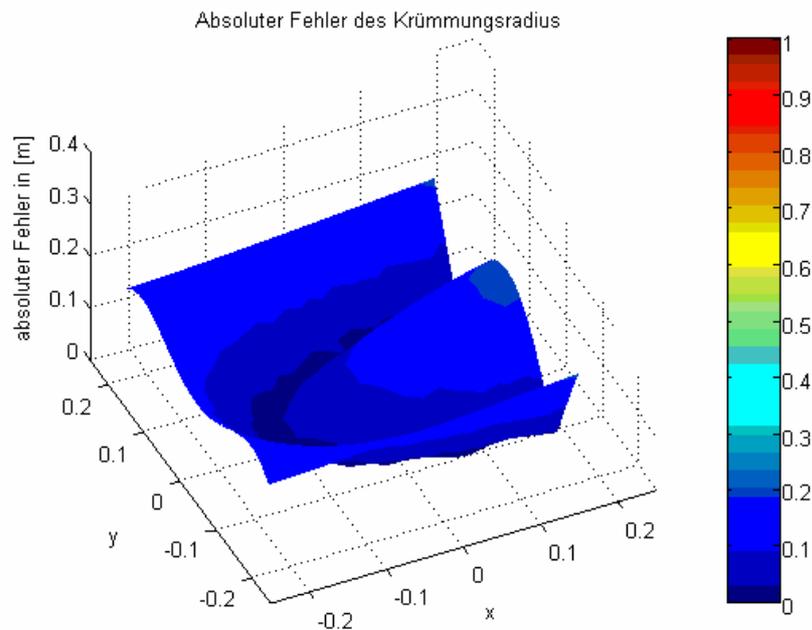


Abb. 3.36: Betrag der absoluten Abweichung des Krümmungsradius im synthetisierten Schallfeld.

Bisher wurde das synthetisierte Feld unter Verwendung eines basic Decoders für Übereinstimmung der Quellrichtung mit der Richtung eines realen Wiedergabelautsprechers dargestellt. Betrachtet man verschiedene Winkelverhältnisse, so verschlechtert sich der mittlere Abbildungsfehler E_{LS} mit

$$E_{LS} = \left(\frac{J}{R} \right)^{1/2} = \sqrt{\frac{\sum_{r=1}^R |e_r|^2}{R}} \quad (3.73)$$

beinahe um den Faktor 2 (vgl. Abbildung 3.37). Durch eine optimierte Decodierung (z.B. in-phase bzw. window-applied Decoder) verteilt sich der Abbildungsfehler gleichmäßiger und wird unabhängig vom Abbildungswinkel.

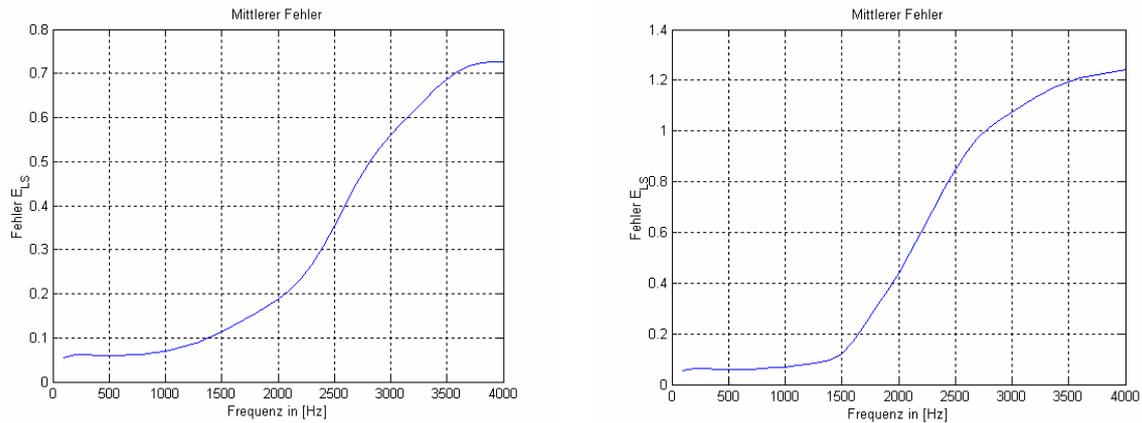


Abb. 3.37: Gegenüberstellung des mittleren Fehlers für den optimalen Abbildungswinkel (identisch mit einer Lautsprecherposition) und dem $\frac{1}{2}$ Versatzwinkel (zwischen zwei benachbarten Lautsprecher).

Zusammenfassung und Ausblick

Bei der Distanzkodierung können kaum Vergleiche mit anderen bisherigen Entwicklungen angestellt werden, da diese meist auf völlig anderen Modellierungskonzepten beruhen. Ansätze von Gerzon in [Gerzon, 1992i] basieren zum Beispiel auf der psychoakustischen Modellierung der zeitlichen und räumlichen Reflexionsmuster. Da die Distanzwahrnehmung erst auf einer höheren Ebene durch eine komplexe Auswertung von mehreren psychoakustischen Faktoren erfolgt (vgl. Kapitel 1.1), wurde hier ein Lösungsansatz angestrebt, der durch physikalischen Reproduktion des Wellenfeldes einer entfernten Quelle nahe kommt. Anders als bei Richtungswahrnehmung ist die Entfernungswahrnehmung zum Teil relativ ungenau und beschränkt (geringer akustischer Horizont). Aufbauend auf den evaluierten Ergebnissen bezüglich der Distanzwahrnehmung, die mit der Wellenfeldsynthese erreicht werden [Start, 1997], wurde eine WFS basierte Distanzkodierung in den Ambisonic Ansatz integriert. Start zeigt in seinen Untersuchungen, dass die synthetisch erzeugten Entfernungen mit Hilfe der Wellenfeldsynthese ähnliche Ergebnisse wie reale Quellen liefern. Die Ergebnisse (7 Versuchspersonen, 165 Versuche d.h. 5 Wiederholungen von 33 unterschiedlichen Quellpositionen) bei der Bewertung von drei Entfernungsebenen (alle innerhalb 30m, Abstand zueinander jeweils ca. 2m) liegen mit 49% Erkennungsrate für reale und 40% für synthetisch reproduzierte Quelle über der Zufallswahrscheinlichkeit von 33%. Die relative geringe Erkennungsrate bei der Entfernungsschätzung bei realen wie auch bei synthetisch erzeugten Quellen zeigt die Grenzen der Entfernungswahrnehmung.

In dieser Arbeit wurde basierend auf den Ergebnissen von Start die Qualität der synthetischen Schallfelder nur theoretisch untersucht. Eine Hörversuch ist für die Verifikation noch ausständig.

Dieses Modell ist auch für die Distanzkodierung bei Kopfhörerwiedergabe, die im nächsten Kapitel vorgestellt wird, denkbar, wird dort aber nicht mehr explizit behandelt.

Kapitel 4

Kopfhörerapplikationen

Berechnung eines binauralen Kopfhörersignals

Implementierung eines 2D Systems 3.Ordnung

Binauraler Kopfhörer und Erweiterungen

Bei der Kopfhörerwiedergabe erfordert die originalgetreue Reproduktion eines aufgezeichneten Schallfeldes sowie die Spatialisierung von Monoquellen eine Filterung der virtuellen Quellen mittels Außenohrübertragungsfunktionen (HRTFs), welche den Unterschied der Signale an den beiden Trommelfellen in Abhängigkeit vom Einfallswinkel der Schallwelle beschreiben (siehe Kapitel 1, Grundlagen des räumlichen Hörens, HRTF). Diese Filteroperationen können durch FIR- oder auch IIR-Filter modelliert werden (siehe [Brandenstein, Unbehauen, 1992], [Jot et al., 1995], [Fellner, 1998]), auf die hier nicht näher eingegangen wird.

Bei den HRTFs kann man zwischen individualisierten, standardisierten und funktionalen Kurven unterscheiden. Der Vorteil der individualisierten Übertragungsfunktionen besteht in einer verbesserten Lokalisation. Versuchspersonen, die über eine gute natürliche Lokalisation¹ verfügen, können mit HRTFs von Versuchspersonen mit schlechten natürlichen Lokalisationsfähigkeiten virtuelle Quellen schlechter orten als mit ihren eigenen. Hingegen verschlechtert sich die Ortung jener Probanden mit verminderten natürlichen Lokalisationseigenschaften bei Verwendung von Übertragungsfunktionen guter Lokalisierer im Wesentlichen nicht. Møller [Møller et. al., 1995] führte zahlreiche Messungen zur Bestimmung von individuellen Unterschieden und allgemeinen Strukturen durch. Dabei zeigte er auch die Problematik der Reproduzierbarkeit von individuellen Übertragungskurven auf, die selbst durch geringe Streuung der Messgeräte (Mikrofone, Messsonden und Lautsprecher) oder Fehlpositionierung der Versuchspersonen bzw. Messsonden auch unter annähernd „idealen Bedingungen“ entstehen können. Darüber hinaus lassen objektive Unterschiede (im Zeit- oder Frequenzbereich) zwischen HRTFs nicht direkt auf perzeptive Unterschiede schließen. Dadurch eignen sich für kommerzielle Produkte eher standardisierte Übertragungsfunktionen, die mit einem Kunstkopf vermessen werden (z.B. KEMAR-HRTF Kurven). In [Cheng, Wakefield, 2001] findet sich eine Gegenüberstellung der gängigen Messverfahren (wie Impulsmessung, Weißes gausverteilter Rauschen, MLS Folgen, Golay (complementary) codes). Ein weiterer Nachteil von individualisierten und standardisierten Übertragungsfunktionen stellt im allgemeinen der große Speicheraufwand (Übertragungsfunktionen für „alle Richtungen“) und die Rechenleistung für die Filterung dar. Eine Reduktion der Filterlängen kann durch die Modellierung der HRTFs mit Hilfe der *Hauptkomponenten-Methode* (Principal Components Analysis, Karhunen-Loève Expansion siehe [Kistler, Wightman, 1992], [Martens, 1987], [Chen et al., 1995]) oder auch durch die *Balanced Model Truncation* (siehe [Mackenzie et al., 1997]) erreicht werden, auf die hier im folgenden jedoch nicht näher eingegangen wird.

Funktionale Übertragungsfunktionen können von schematisierten bzw. zusammengesetzten mathematisch beschreibbaren und physikalisch messbaren Modellen abgeleitet werden, und

¹ „Gute Richter“: Personen, die bei natürlichen Schallquellen, z.B. räumlich verteilte Lautsprecher, nur geringe Abweichung zwischen tatsächlicher und wahrgenommener Richtung sowie hohe Reproduzierbarkeit erzeugen.

benötigen neben einem reduzierten Speicherbedarf meist auch eine geringere Rechenleistung. Ziel ist es durch Änderung weniger Parameter, die wenn möglich linear unabhängig sind, die Übertragungsfunktionen an den Hörer optimal anzupassen (vgl. [Brown, Duda, 1998]).

Die menschliche Hörwahrnehmung erlaubt eine Verbesserung der Lokalisation von Schallereignissen mittels kleiner Kopfdrehungen (siehe Kapitel 1, Grundlagen des räumlichen Hörens, Peilbewegungen). Um dieses Phänomen in *Virtual Reality* (VR)-Applikationen zugänglich zu machen ist ergänzend die Aufnahme der Kopforientierung durch einen Headtracker notwendig. Dadurch erhält man ein dynamisches System, das die Problematik der zeitvarianten Interpolation zwischen den verschiedenen HRTF-Kurven ergibt. Da diese nur für diskrete Winkel zur Verfügung stehen, wird für davon abweichende Positionen zwischen Filterkoeffizienten der benachbarten Richtungen (meist linear) interpoliert. Für geringe Winkelinkremente ist diese Vorgangsweise zulässig, jedoch steigt mit zunehmender Genauigkeit - d.h. für verbesserte Klangqualität und Abbildungstreue - auch der damit benötigte Speicherplatz für die Filterkoeffizienten. Eine Lösungsmöglichkeit bietet die Verwendung eines virtuellen Ambisonic-Systems an, das die Kopfdrehung mittels einer zeitvarianten Rotationsmatrix und zeitinvarianter Filter berücksichtigt.

Im Kapitel 2 wurde das Ambisonic-Verfahren behandelt, das die dreidimensionale Reproduktion eines Schallfeldes über Lautsprecher ermöglicht. Im Folgenden wird die Berechnung eines binauralen Stereosignals durch die Anwendung der Ambisonic-Technik vorgestellt. In einer Zwischenstufe des Algorithmus werden virtuelle Lautsprecher verwendet, die eigentliche Wiedergabe erfolgt jedoch mit Kopfhörer. Die virtuellen Lautsprecher befinden sich an zeitlich unveränderlichen Punkten. Durch die Filterung dieser Lautsprechersignale, in Abhängigkeit von ihrer Position, mit den entsprechenden Außenohrübertragungsfunktionen wird ein binaurales Kopfhörersignal berechnet.

4.1 Berechnung eines binauralen Kopfhörersignals

Um ein binaurales Kopfhörersignal zu erhalten ist es notwendig, dass eine aus einer bestimmten Raumrichtung einfallende Schallwelle mit der entsprechenden Außenohrübertragungsfunktion gefiltert wird. Als einfallende Schallwellen werden in diesem Fall die Signale der virtuellen Lautsprecher aus dem Ambisonic-System betrachtet. Als HRTFs stehen die KEMAR-Kunstkopfmessungen mit 710 Impulsantworten, davon 72 für die Horizontalebene (Elevation 0°) in einer Auflösung von 5° , zur Verfügung. Die Lautsprechersignale werden mit den entsprechenden Impulsantworten gefiltert und zu einem binauralen Kopfhörersignal superponiert (siehe Abbildung 4.1).

Zuerst erfolgt eine allgemeine Herleitung für den 2 dimensionalen Fall. Im Abschnitt 4.2 wird auf eine praktische Implementierung eines 2D-Systems 3. bzw. 4. Ordnung eingegangen (vgl. [Sontacchi et al., 2002b]).

Das Kopfhörersignal L für den linken und R für den rechten Kanal ergibt sich durch Filterung der Lautsprechersignale p_1 bis p_N mit den Außenohrimpulsantworten $h_{i,L}(n)$ für das linke und $h_{i,R}(n)$ für das rechte Ohr:

$$\begin{pmatrix} L \\ R \end{pmatrix} = \begin{pmatrix} h_{1,L}(n) & h_{2,L}(n) & \cdots & h_{N,L}(n) \\ h_{1,R}(n) & h_{2,R}(n) & \cdots & h_{N,R}(n) \end{pmatrix} \otimes \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{pmatrix} \quad (4.1)$$

wobei das Zeichen \otimes für die Faltung steht.

Im Folgenden werden alle Größen im Frequenzbereich betrachtet. Dazu werden alle Größen mit Hilfe der diskreten Fourier Transformation (DFT) in den Frequenzbereich gebracht. Die Länge der DFT wird durch die doppelte Länge ($2M$, M Punkte zero padding) der verwendeten Impulsantworten bestimmt und steht im Subscript der Transformationsoperation. Unendlich lange Signale werden in Blöcke der Länge M unterteilt und in Blockfiltertechnik verarbeitet. Die Filterung selbst entspricht im Frequenzbereich einer Multiplikation. Die DFT der Außenohrimpulsantworten lautet:

$$H_{i,L} = DTF_{2M} \{h_{i,L}(n)\} \quad (4.2)$$

wobei $h_{i,L}(n) \neq 0$ für $0 \leq n \leq M - 1$

Die Lautsprechersignale $\mathbf{P} = [P_1, P_2, \dots, P_N]^T$ berechnen sich aus den Ambisonic-Signalen für ein 2D-System M . Ordnung wie folgt:

$$\mathbf{P} = \mathbf{Decoder} \cdot \begin{bmatrix} W \\ X1 \\ Y1 \\ X2 \\ Y2 \\ \vdots \\ XM \\ YM \end{bmatrix} \quad (4.3)$$

$$\mathbf{Decoder} = \frac{1}{N} \begin{pmatrix} \mathbf{a}_1 & 2\mathbf{b} \cos \mathbf{f}_1 & 2\mathbf{g} \sin \mathbf{f}_1 & 2\mathbf{d} \cos 2\mathbf{f}_1 & 2\mathbf{e} \sin 2\mathbf{f}_1 & \dots \\ \mathbf{a}_2 & 2\mathbf{b} \cos \mathbf{f}_2 & 2\mathbf{g} \sin \mathbf{f}_2 & 2\mathbf{d} \cos 2\mathbf{f}_2 & 2\mathbf{e} \sin 2\mathbf{f}_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\ \mathbf{a}_N & 2\mathbf{b} \cos \mathbf{f}_N & 2\mathbf{g} \sin \mathbf{f}_N & 2\mathbf{d} \cos 2\mathbf{f}_N & 2\mathbf{e} \sin 2\mathbf{f}_N & \dots \end{pmatrix} \quad (4.4)$$

Die Faktoren α , β , γ , δ und ϵ etc. sind für eine regelmäßige Lautsprecheranordnung gleich eins (siehe Kapitel 2). Das Schallfeld, das in den Ambisonic-Signalen $\mathbf{Amb} = [W, X1, Y1, X2, Y2, \dots]^T$ abgebildet ist, kann in beliebige Richtungen gedreht werden. Man erhält nach der Drehung die neuen Ambisonic-Signale $\mathbf{Amb}' = [W', X1', Y1', X2', Y2', \dots]^T$.

Die unrotierten Signale für eine Quelle, mit $P_{\mathbf{y}} = P_{\mathbf{y}}(t, f)$, aus der Richtung \mathbf{y} lauten:

$$W = \frac{P_{\mathbf{y}}}{\sqrt{2}} \quad \begin{array}{l} X1 = P_{\mathbf{y}} \cos(\mathbf{y}) \\ Y1 = P_{\mathbf{y}} \sin(\mathbf{y}) \end{array} \quad \begin{array}{l} X2 = P_{\mathbf{y}} \cos(2\mathbf{y}) \\ Y2 = P_{\mathbf{y}} \sin(2\mathbf{y}) \end{array} \quad \text{etc.} \quad (4.5)$$

Bei einer Rotation (Drehung um die z-Achse) um den Winkel \mathbf{r} gilt: $\cos(\mathbf{y}) \rightarrow \cos(\mathbf{y} + \mathbf{r})$, $\sin(\mathbf{y}) \rightarrow \sin(\mathbf{y} + \mathbf{r})$, $\cos(2\mathbf{y}) \rightarrow \cos(2(\mathbf{y} + \mathbf{r}))$, $\sin(2\mathbf{y}) \rightarrow \sin(2(\mathbf{y} + \mathbf{r}))$, etc. Mit den Sumsensätzen für die Sinus- und Kosinus-Funktionen folgt für die um einen Winkel ρ rotierten Signale:

$$\begin{pmatrix} W' \\ X1' \\ Y1' \\ X2' \\ Y2' \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & \cos \mathbf{r} & -\sin \mathbf{r} & 0 & 0 & \cdots \\ 0 & \sin \mathbf{r} & \cos \mathbf{r} & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cos 2\mathbf{r} & -\sin 2\mathbf{r} & \cdots \\ 0 & 0 & 0 & \sin 2\mathbf{r} & \cos 2\mathbf{r} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \cdot \begin{pmatrix} W \\ X1 \\ Y1 \\ X2 \\ Y2 \\ \vdots \end{pmatrix} \quad (4.6)$$

$$\text{bzw. } \mathbf{Amb}' = \mathbf{ROT} \cdot \mathbf{Amb}$$

Aus den obigen Gleichungen berechnet sich jetzt ein binaurales Stereosignal für ein rotiertes Schallfeld um den Winkel ρ aus den Ambisonic-Signalen:

$$\begin{pmatrix} L \\ R \end{pmatrix} = \mathbf{Filter} \cdot \mathbf{Decoder} \cdot \mathbf{ROT} \cdot \mathbf{Amb} \quad (4.7)$$

$$\mathbf{Filter} = \begin{pmatrix} H_{1,L} & H_{2,L} & \cdots & H_{N,L} \\ H_{1,R} & H_{2,R} & \cdots & H_{N,R} \end{pmatrix} \quad (4.8)$$

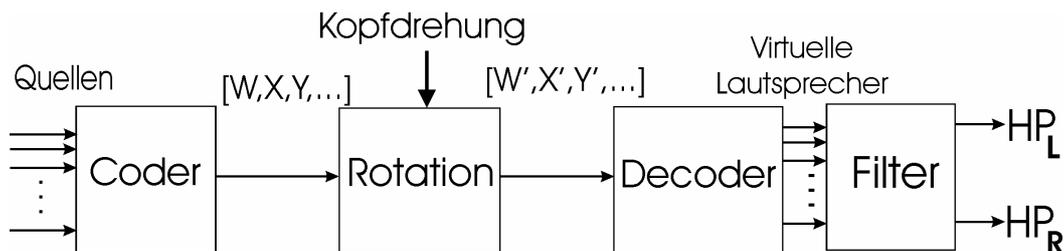


Abb. 4.1: Strukturschaltbild zur Berechnung eines binauralen Kopfhörersignals

4.1.1 Rotationsmatrizen für den 3-dimensionalen Fall erster Ordnung

Rotationsmatrix bei einer Drehung des Kopfes um die z-Achse (Drehwinkel ρ):

$$\begin{pmatrix} W' \\ X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \mathbf{r} & -\sin \mathbf{r} & 0 \\ 0 & \sin \mathbf{r} & \cos \mathbf{r} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} W \\ X \\ Y \\ Z \end{pmatrix} \quad (4.9)$$

Zusätzlich zur Rotation (Drehung des Kopfes um die z-Achse) können auch Bewegungen des Kopfes um die x-Achse (Neigen), und Bewegungen um die y-Achse (Nicken) berücksichtigt werden. Die Matrix bei einer Drehung um die x-Achse (Drehwinkel ξ) lautet

$$\begin{pmatrix} W' \\ X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \boldsymbol{x} & -\sin \boldsymbol{x} \\ 0 & 0 & \sin \boldsymbol{x} & \cos \boldsymbol{x} \end{pmatrix} \begin{pmatrix} W \\ X \\ Y \\ Z \end{pmatrix}. \quad (4.10)$$

Bei einer Drehung um die y-Achse (Nicken) lautet die Matrix (Drehwinkel ζ)

$$\begin{pmatrix} W' \\ X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \boldsymbol{V} & 0 & -\sin \boldsymbol{x} \\ 0 & 0 & 1 & 0 \\ 0 & \sin \boldsymbol{V} & 0 & \cos \boldsymbol{V} \end{pmatrix} \begin{pmatrix} W \\ X \\ Y \\ Z \end{pmatrix}. \quad (4.11)$$

Zusammengesetzte Bewegungen lassen sich durch Multiplikation der einzelnen Matrizen berücksichtigen. Es lassen sich auch für Systeme höherer Ordnung Rotationsmatrizen bestimmen¹.

4.2 Implementierung eines 2D-Systems, 3. Ordnung

Im Folgenden wird nun auf Optimierungsmöglichkeiten in der praktischen Implementierung² eingegangen. Dazu gehört die Anzahl der Übertragungskanäle und Lautsprecher sowie deren Positionierung.

Aus Gleichung 4.1 geht hervor, dass die Filteranzahl bei der direkten Umsetzung von der Anzahl der virtuellen Lautsprecher und damit von der Ambisonic-Ordnung abhängig ist. Untersuchungen in (vgl. [Gerzon, 1992a], [Poletti, 1996a], [Bamford, 1995] etc.) zeigen, dass bei einer Übereinstimmung von Lautsprecheranzahl mit der Anzahl von Kodierungskanäle die geringsten Abbildungsfehler erreichbar sind (siehe Abb 4.2). Je höher die gewählte Ordnung,

¹ Ausgangspunkt sind die Richtungskosinusse für die entsprechende Drehrichtung. Für höhere Kodierungsordnungen (ausgehend für die z-Richtung) werden im kartesischen System die Terme niedriger Ordnung bzw. richtungsunabhängige Terme mittels der Kugeloberflächenbeziehung $(x^2+y^2+z^2)=1$ entsprechend erweitert. Die Kugelflächenfunktionen werden durch einsetzen der jeweiligen 1. Ambisonic Ordnung gebildet. Dadurch erhält man zwei neue Sets von Kugelflächenfunktionen (bzgl. der yz-Ebene und x-Achse, und bzgl. zx-Ebene und y-Achse). Die Rotationsmatrizen ergeben sich, indem man in den einzelnen Kanalkodierungsgleichungen, des zur Rotation korrespondierenden Polarkoordinatensystems, j durch $j+r$ ersetzt (Ausnahme: System bzgl. y-Achse – hier wird j durch $j-r$ ersetzt), sie als Linearkombination aller Kodiergleichungen der jeweiligen Ambisonic-Ordnung gleichsetzt und dann eine Variablenseparation durchführt.

² Das System wurde am Institut für Elektronische Musik und Akustik, Graz, auf unterschiedlichen Plattformen realisiert. Implementierungen erfolgten jeweils mit *MAX (FTS)* auf SGI-Origin für 5 Eingangskanäle, mit *PD (pure data)* unter NT und Linux mit 8 Eingangskanäle, und für 2 Eingangskanäle auf einem DSP-Board TMS320C6711 von Texas Instruments.

umso größer wird der optimale Abhörbereich bzw. umso größer wird die Bandbreite für eine exakte Abbildung.

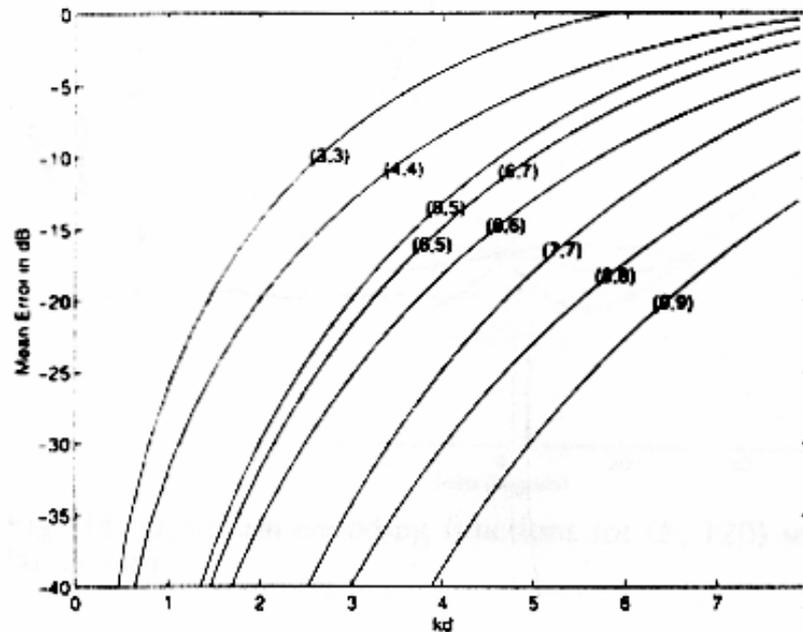


Abb. 4.2: Gemittelter Fehler über die Einfallsrichtungen für unterschiedliche Systemlösungen (N, M) mit N Lautsprecher und M Übertragungskanäle, (vgl. [Poletti, 1996a]).

4.2.1 Positionierung der Lautsprecher und Filterreduktion

Der Design des virtuellen Lautsprecherlayouts hat eine Auswirkung auf die Anzahl der benötigten Filter und der Abbildungseigenschaften. Aus [Poletti, 1996] geht hervor, dass grundsätzlich eine ungerade Anzahl von Lautsprecher eine bessere Lösung liefert, da im 2D-Fall die Anzahl der Kodierungskanäle M aus der Systemordnung L mit $M=2L+1$ hervorgeht. Wird ein Lautsprecherlayout spiegelsymmetrisch bezüglich der x -Achse (Blickrichtung nach vorne) gewählt, so lässt sich die Anzahl der benötigten Filter halbieren.

Um den Lokalisationseigenschaften des auditiven Systems Rechnung zu tragen, sollte die Anzahl der Lautsprecher vorne größer sein. Gleichzeitig sollte nicht ein einzelner Lautsprecher in der 0° -Richtung positioniert werden. Dadurch werden für kleine Peilbewegungen die Unterschiede für ipsi- und kontralaterales Ohr maximiert. Ist die Latenzzeit des Gesamtsystems kleiner 60ms (vgl. [Wenzel, 1997], [Mackensen, Reichenauer und Theile, 1998]), so kann bei entsprechender Genauigkeit des Headtrackers und geeigneter Wahl der Systemordnung eine exakte Lokalisation erfolgen.

Bei der Wahl eines Ambisonic-Systems 3. Ordnung bleibt der mittlere Abbildungsfehler (Mean Error) für den Bereich $k \cdot d \leq 4.8$ bei unter 20% (-14dB, siehe Abb. 4.2). Für einen mittleren Ohrabstand von 17cm ($d = 0.085$) ist daher eine gute Abbildung bis ca. 3kHz möglich. Die Anzahl der virtuellen Lautsprecher M ergibt sich nach obigen Überlegungen zu 7 ($M=2 \cdot 3 + 1 = 7$), die Positionierung ist in Abbildung 4.3 zu sehen.

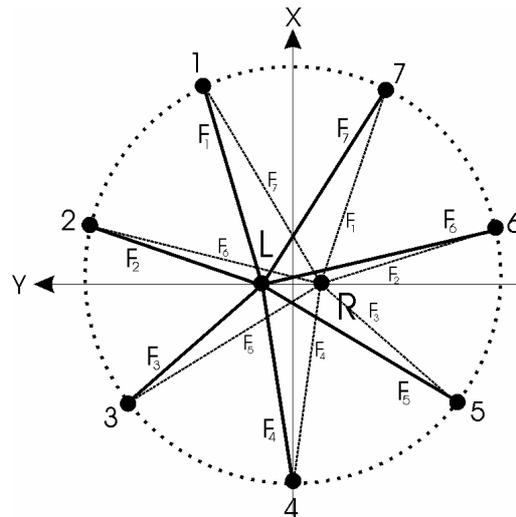


Abb. 4.3: Spiegelsymmetrisches Lautsprecherlayout mit Filter (Übertragungsfunktionen) F_i .

Die Übertragungsfunktionen bzw. die notwendigen Filter zwischen den Lautsprecherpositionen 1-7 und den beiden Ohren L und R sind mit F_i bezeichnet (dicke Linien für das linke und dünne Linien für das rechte Ohr). Werden Kopf, Pinna und Torso bzgl. linkem und rechtem Ohr als symmetrisch betrachtet, ergeben sich für übereinstimmende Übertragungspfade die gleichen Filter F_i . Mit der spiegelsymmetrischen Anordnung gemäß Abbildung 4.3 ergeben sich die Binauralsignale L_{Ohr} und R_{Ohr} aus den virtuellen Lautsprechersignalen wie folgt:

$$\begin{pmatrix} L_{Ohr} \\ R_{Ohr} \end{pmatrix} = \begin{pmatrix} F_1 & F_2 & F_3 & F_4 & F_5 & F_6 & F_7 \\ F_7 & F_6 & F_5 & F_4 & F_3 & F_2 & F_1 \end{pmatrix} \cdot \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_3 \end{bmatrix} = \mathbf{Filter} \cdot \mathbf{P} \quad (4.12)$$

Die Lautsprechersignale \mathbf{P} ergeben sich durch die Dekodierung der Ambisonic Signale.

$$\mathbf{P} = \mathbf{Decoder} \cdot \begin{bmatrix} W \\ X1 \\ Y1 \\ X2 \\ Y2 \\ X3 \\ Y3 \end{bmatrix} \quad (4.13)$$

Die zwei Komponenten der höchsten Ambisonic-Ordnung ($X3$ und $Y3$) werden geringer gewichtet als die übrigen Komponenten (räumliche Filterung). Dadurch wird die räumliche Welligkeit, die durch den Abbruch der unendlichen Reihe auftritt, minimiert (vgl. Filterdesign Frequenzabtastmethode mit Übergangssamples [Oppenheim, Schaffer, 1975]).

Bei einer direkten Berechnung von Gl. 4.12 werden 14 Filteroperationen benötigt, die jedoch durch geeignete Kombination und Berücksichtigung der Decodereigenschaften auf die Hälfte reduziert werden können.

Für ein spiegelsymmetrisches Lautsprecher-Layout gilt:

$$f_i = -f_{N-i} \quad \text{für gerades } N \text{ (} i=1,2,\dots,N/2 \text{)} \quad (4.14)$$

$$f_i = -f_{N+1-i} \quad \text{für ungerades } N \text{ (} i=1,2,\dots,(N-1)/2 \text{)} \quad (4.15)$$

wobei $f_{(N+1)/2} = \mathbf{p}$

Bei der Wahl von $N=7$ Lautsprecher ergibt das :

$$\begin{aligned} f_1 &= -f_7 \\ f_2 &= -f_6 \\ f_3 &= -f_5 \\ f_4 &= \mathbf{p} \end{aligned}$$

Im Folgenden wird gezeigt, dass die Eigenschaft des Decoders, durch die spezielle Anordnung der virtuellen Lautsprecher, zur Minimierung der Filteroperationen verwendet werden kann.

$$\mathbf{Filter} \cdot \mathbf{Decoder} = \left(\mathbf{Decoder}^T \cdot \mathbf{Filter}^T \right)^T \quad (4.16)$$

$$\mathbf{Decoder}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \cos(f_1) & \cos(f_2) & \cos(f_3) & -1 & \cos(f_3) & \cos(f_2) & \cos(f_1) \\ \sin(f_1) & \sin(f_2) & \sin(f_3) & 0 & -\sin(f_3) & -\sin(f_2) & -\sin(f_1) \\ \cos(2f_1) & \cos(2f_2) & \vdots & 1 & \vdots & \vdots & \cos(2f_1) \\ \sin(2f_1) & \sin(2f_2) & & 0 & & & \sin(2f_1) \\ \vdots & \vdots & & -1 & & & \vdots \\ \sin(3f_1) & & & 0 & & & -\sin(3f_1) \end{bmatrix} \quad (4.17)$$

und symbolisch

$$\mathbf{Decoder}_{\text{symp}}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ a & b & c & -1 & c & b & a \\ d & e & f & 0 & -d & -e & -f \\ g & h & i & 1 & i & h & g \\ j & k & l & 0 & -l & -k & -j \\ m & n & o & -1 & o & n & m \\ p & q & r & 0 & -r & -q & -p \end{bmatrix} \quad (4.18)$$

$$\mathbf{Filter}_{\text{symp, reduziert}} = \left(\mathbf{Decoder}_{\text{symp}}^T \cdot \mathbf{Filter}^T \right)^T \quad (4.19)$$

$$\mathbf{Filter}_{\text{symp, reduziert}} = \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ a & b & c & -1 & c & b & a \\ d & e & f & 0 & -d & -e & -f \\ g & h & i & 1 & i & h & g \\ j & k & l & 0 & -l & -k & -j \\ m & n & o & -1 & o & n & m \\ p & q & r & 0 & -r & -q & -p \end{bmatrix} \cdot \begin{bmatrix} F_1 & F_7 \\ F_2 & F_6 \\ F_3 & \vdots \\ \vdots & \\ F_3 & \\ F_6 & F_2 \\ F_7 & F_1 \end{bmatrix} \right)^T \quad (4.20)$$

Werden die Elemente der Matrix $\mathbf{Filter}_{\text{symb, reduziert}}$ algebraisch gelöst, so ergibt sich ein reduziertes Filterset wie folgt:

$$\mathbf{Filter}_{\text{symb, reduziert}}^T(1,1) = F_1 + F_2 + \dots + F_7 = \sum F = F_{1,1}$$

$$\mathbf{Filter}_{\text{symb, reduziert}}^T(2,1) = F_7 + F_6 + \dots + F_1 = \sum F = F_{1,1}$$

$$\mathbf{Filter}_{\text{symb, reduziert}}^T(1,2) = aF_1 + bF_2 + \dots + bF_6 + aF_7 = F_{1,2}$$

$$\mathbf{Filter}_{\text{symb, reduziert}}^T(2,2) = aF_7 + bF_6 + \dots + bF_2 + aF_1 = F_{1,2}$$

$$\mathbf{Filter}_{\text{symb, reduziert}}^T(1,3) = dF_1 + eF_2 + \dots - eF_6 - fF_7 = F_{1,3}$$

$$\mathbf{Filter}_{\text{symb, reduziert}}^T(2,3) = dF_7 + eF_6 + \dots - eF_2 - fF_1 = -F_{1,3}$$

usw.

Durch die Struktur der Decodermatrix reduziert sich die Anzahl der Filterungen um die Hälfte. Die Ambisonic-Signale können direkt mit dem reduzierten Filtersatz $\mathbf{Filter}_{\text{reduziert}}$ gefiltert werden. Die Ergebnisse diese Filteroperationen müssen für das linke und rechte Ohr entsprechend aufsummiert werden (siehe Abb. 4.4).

$$\mathbf{Filter}_{\text{reduziert}} = \begin{bmatrix} F_{1,1} & F_{1,2} & F_{1,3} & F_{1,4} & F_{1,5} & F_{1,6} & F_{1,7} \\ F_{1,1} & F_{1,2} & -F_{1,3} & F_{1,4} & -F_{1,5} & F_{1,6} & -F_{1,7} \end{bmatrix} \quad (4.21)$$

$$\begin{pmatrix} L_{\text{Ohr}} \\ R_{\text{Ohr}} \end{pmatrix} = \mathbf{Filter}_{\text{reduziert}} \cdot \mathbf{Amb} \quad (4.22)$$

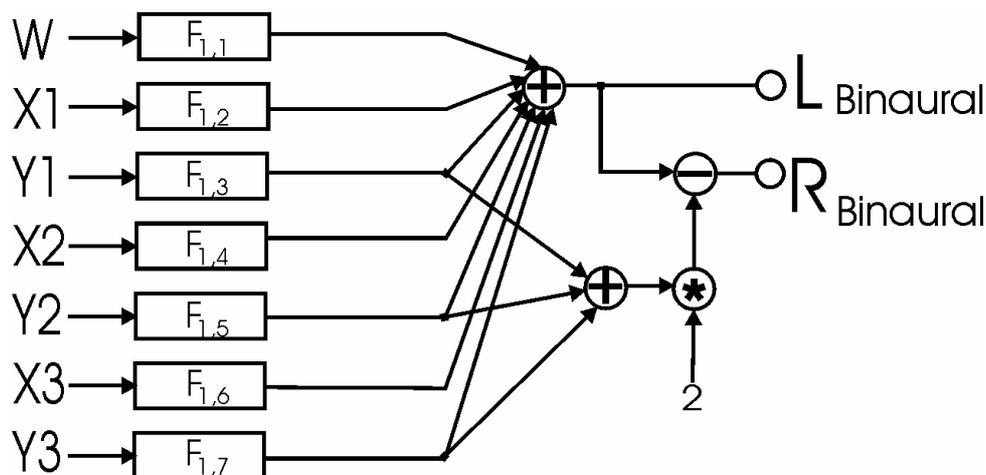


Abb. 4.4: Filterung der Ambisonic-Signale mit dem reduzierten Filterset.

4.2.2 Abbildungsfehler und Optimierung des Dekoders

Durch die Kodierung des Schallfeldes in Ambisonic und die damit verbundene endliche Anzahl von Übertragungskanälen und virtuellen Lautsprechern entsteht eine Verfälschung der ursprünglichen Außenohrübertragungsfunktionen. Abbildung 4.5 zeigt die Verteilung des relativen Fehlers über Frequenz und Winkel in Abhängigkeit von Ambisonic-Ordnung und Lautsprecheranzahl. Eine Untersuchung der perceptiven Unterschiede zwischen unterschiedlichen Parametersettings erfolgt in Kapitel 5.

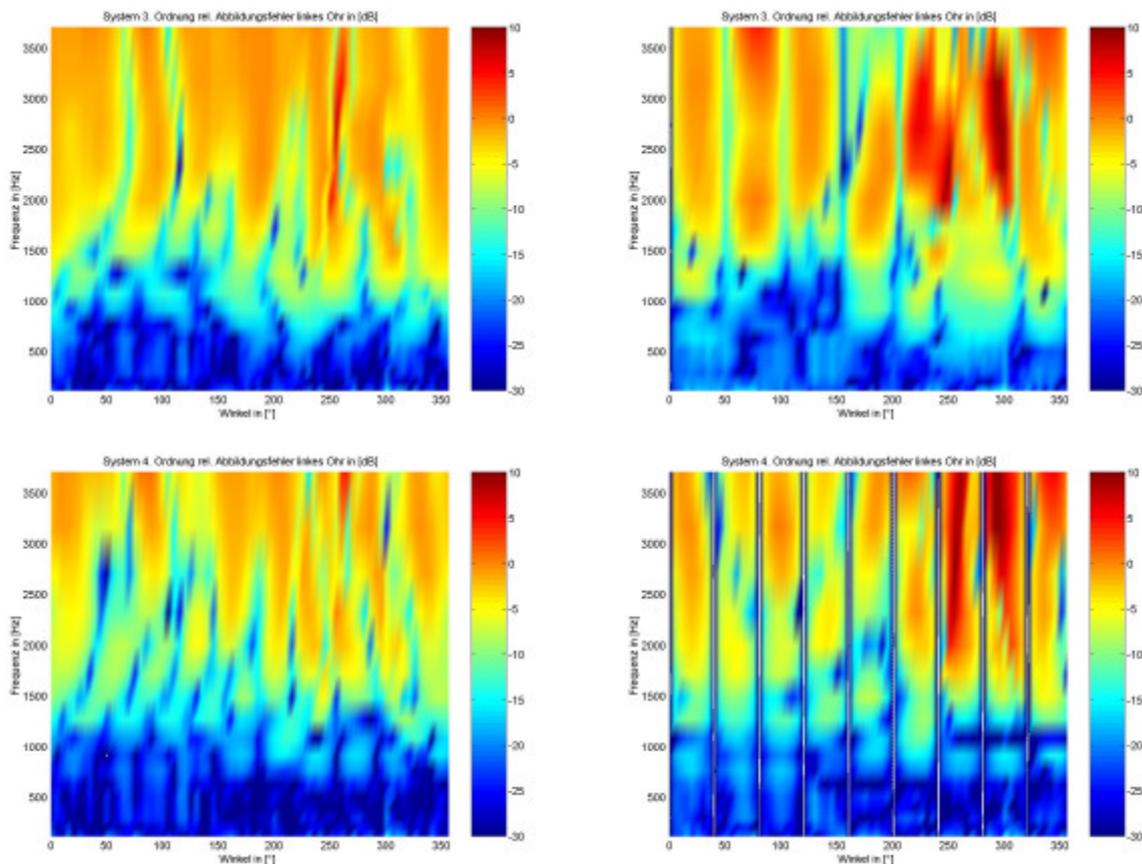


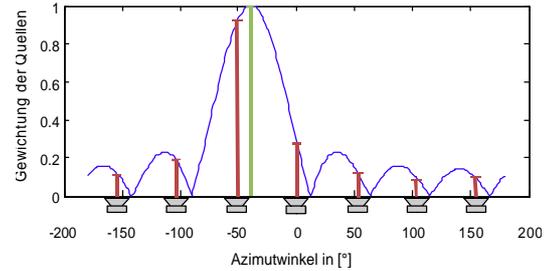
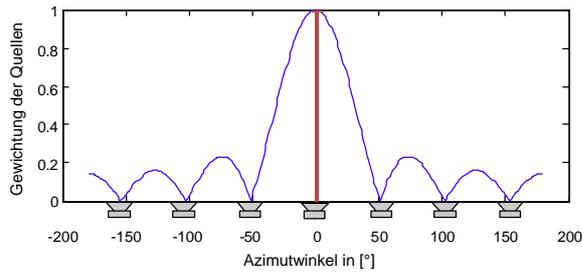
Abb. 4.5a: Relativer Fehler zwischen den Referenz-HRTFs (Kemar) und über Ambisonic synthetisierten HRTFs. Dekodiert auf jeweils 72 virtuellen Lautsprecher.

Oben: 3. Ordnung. Unten: 4. Ordnung .

Abb. 4.5b: Relativer Fehler zwischen den Referenz-HRTFs (Kemar) und über Ambisonic synthetisierten HRTFs. Dekodiert auf jeweils $N=2M+1$ virtuelle Lautsprecher.

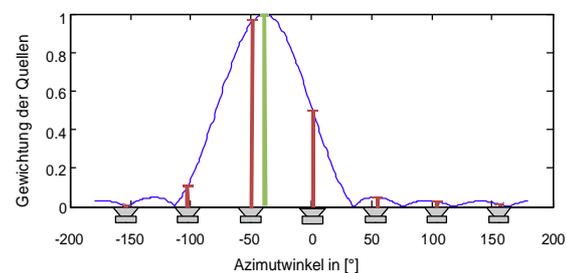
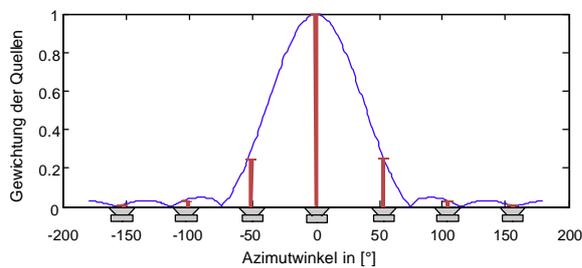
Oben: 3. Ordnung. Unten: 4. Ordnung .

Die Ambisonic-Kodierung erfolgt gemäß der Reihenentwicklung in die sphärischen Harmonischen bzw. im 2D-Fall in eine Fourierreihe des über den Azimut in $2p$ periodischen Schallfeldes. Der Abbruch der Reihenentwicklung verursacht - ähnlich wie bei der Verwendung eines Rechteckfenster im FIR-Filterentwurf - entsprechende Nebenkeulen abseits der Quellrichtung d.h. es kommt zum Übersprechen auf andere Richtungen (siehe Abb. 4.6, 1. Zeile).



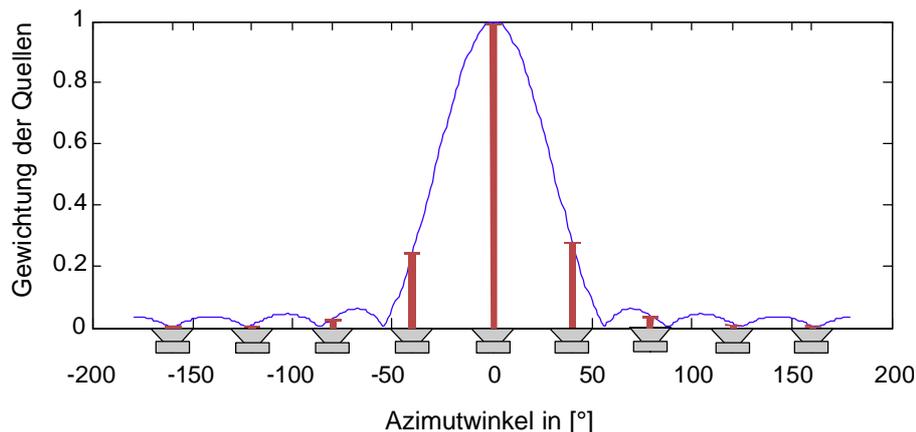
Ambisonic 3.Ordnung (2D) mit $N=7$ Lautsprecher an den Positionen $j_i=2p \cdot i/N$ $i=0, \dots, N-1$, ohne Gewichtung der Ordnungen (Rechteckfensterung). Quelle aus 0° abgebildet. Es erfolgt kein Übersprechen, da sich die Lautsprecher jeweils in den Nulldurchgängen der ASINC Funktionen befinden.

Ambisonic 3.Ordnung (2D) mit $N=7$ Lautsprecher an den Positionen $j_i=2p \cdot i/N$ $i=0, \dots, N-1$, ohne Gewichtung der Ordnungen (Rechteckfensterung). Quelle aus -40° abgebildet. Es erfolgt Übersprechen, da sich die Lautsprecher nicht in den Nulldurchgängen der ASINC Funktionen befinden. Die Vorzeichen der Gewichte der Nebenkeulen alternieren.



Ambisonic 3.Ordnung (2D) mit $N=7$ Lautsprecher an den Positionen $j_i=2p \cdot i/N$ $i=0, \dots, N-1$, Gewichtung der Ambisonic-Ordnungen mit Kaiser-Fenster. Quelle aus 0° abgebildet. Geringes Übersprechen und Verbreiterung der Hauptkeule.

Ambisonic 3.Ordnung (2D) mit $N=7$ Lautsprecher an den Positionen $j_i=2p \cdot i/N$ $i=0, \dots, N-1$, Gewichtung der Ambisonic-Ordnungen mit Kaiser-Fenster. Quelle aus -40° abgebildet. Kaum Übersprechen, aber Verbreiterung der Hauptkeule.



Ambisonic 4.Ordnung (2D) mit $N=9$ Lautsprecher an den Positionen $j_i=2p \cdot i/N$ $i=0, \dots, N-1$. Quelle aus 0° abgebildet. Gewichtung der Ordnungen mit Kaiserfenster. Die Winkelauflösung der verbreiterten Hauptkeule ist vergleichbar mit 3.Ordnung ohne Gewichtung.

Abb. 4.6: Auswirkung der begrenzten Ambisonic-Ordnung mit und ohne Fenstertechnik. (Vgl. Window Applied Decoder, Kap.2.2).

4.3 Erweiterungen des zeitvarianten Binauralsystems

Das im vorigen Abschnitt vorgestellte Basissystem ermöglicht die Quellspatialisierung für Kopfhörerwiedergabe. Die Binauralsignale werden unter Verwendung standardisierter HRTFs berechnet (hier wird vorerst nur ein 2D-System¹ betrachtet). Durch die Anbindung eines *Head Trackers* können die im Folgenden beschriebenen Funktionen erweitert werden, wobei durch entsprechendes Systemdesign Modularität und Skalierbarkeit gewährleistet werden kann.

4.3.1 Die „Rotation des Schallfeldes“

Um eine Vorne-Hinten Lokalisation zu ermöglichen, werden beim natürlichen Hören auch die Kopfbewegungen (vor allem die Rotation nach links und rechts, Peilbewegungen) und die damit verbundenen Ohrsignalveränderungen ausgewertet. Es gilt nun die Kopfhörersignale L und R entsprechend der Kopfposition so zu verändern, dass diese Signalveränderungen bei der Lokalisation zu Verfügung stehen. Die folgende Abbildung zeigt den Sachverhalt.

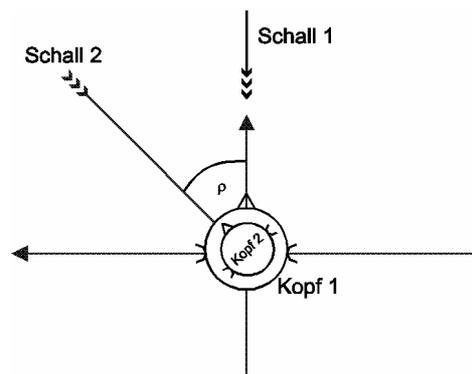


Abb. 4.7: Auswirkungen einer Kopfdrehung auf die Filtercharakteristik des Ohres.

Betrachtet werden die HRTF Filter für Schallereignisse aus zwei Richtungen ($0^\circ, 45^\circ$) und für zwei verschiedene Kopfpositionen (Kopf 1 und Kopf 2). In der Stellung Kopf 1 (nach vorne blickend) muss Schall1 mit der HRTF für 0° und Schall2 mit der HRTF für 45° gefiltert werden. Bei einer Kopfdrehung um ρ nach links (Kopf 2 $\rho=45^\circ$) muss Schall2 für 0° und Schall1 für -45° aufbereitet werden.

Aus Gleichung (4.22) folgt, dass die Kopfhörersignale direkt aus den gefilterten Ambisonic-Signalen berechnet werden können. Die Berechnung binauraler Kopfhörersignale unter Berücksichtigung von Kopfbewegungen erfolgt daher durch Rotation der Ambisonic-Signale. Der Rotationswinkel wird mittels *Head Tracker*, detektiert und führt zu einer zeitvarianten Rotationsmatrix, durch die die Ambisonic-Signale gedreht werden.

¹ Bei einer Berücksichtigung der Elevation ergeben sich aufgrund der standardisierten Übertragungsfunktionen Unsicherheiten bei der Lokalisation. Eine Verbesserung kann durch individualisierte Übertragungsfunktionen erreicht werden.

4.3.2 Die „akustische Linse“

Mit unseren Augen sehen wir - bedingt durch deren Anordnung am Kopf immer nur einen begrenzten Bereich, das sogenannte „Blickfeld“. Diese Eigenschaft kann für ein bewusstes „sich Hinwenden“ eingesetzt werden und unterstützt gezielte Aufmerksamkeit in eine bestimmte Richtung. Im Gegensatz dazu können wir mit unseren beiden Ohren nur bedingt einen bestimmten Ausschnitt selektiv auswählen, da wir Schall aus allen Richtungen gleichzeitig wahrnehmen¹. Die Fähigkeit sich in realen Umgebungen auf eine einzelne Schallquelle zu konzentrieren wird auch als „Cocktail Party“ - Effekt² bezeichnet. Dabei kann durch die Auswertung der beiden Ohrsignale im Gehirn diese spezielle Richtung selektiert werden.

Als akustische Linse wird im Folgenden ein Verfahren bezeichnet, das - ähnlich wie in der Optik - die gezielte Betrachtung eines Ausschnitts einer Schallsituation erlaubt. Sie ermöglicht im Hinblick auf Applikationen wie z.B. „Akustische Museumsführer“ oder Surround-Abmischungen mittels Kopfhörer die Hervorhebung der Schallquellen in Blickrichtung des Hörers. Dabei werden die übrigen Quellen unterdrückt, wobei deren Position unverändert bleibt. Die Breite des Ausschnittes kann voreingestellt werden, die Stärke der Unterdrückung der umgebenden Richtungen kann während des Betriebs verändert werden. Durch die geeignete Stellung in der Signalverarbeitungskette und kleine algorithmische Veränderungen bleibt die Anzahl der Übertragungskanäle unverändert und der zusätzliche Rechenaufwand beschränkt.

Die virtuellen Quellsignale werden - wie in 4.2 - zuerst in den erweiterten Ambisonic-Bereich transformiert (**Amb**). Nach Berücksichtigung der Kopfbewegung (Schallfeldrotation, **Rot**) werden durch Dekodierung (**Decoder**) die virtuellen Lautsprechersignale (**P**) gebildet mit:

$$\mathbf{P} = \mathbf{Decoder} \cdot \mathbf{Rot} \cdot \mathbf{Amb}$$

Die Ohrsignale **HP** werden durch HRTF-Filterung (**Filter**) der Lautsprechersignale erzeugt mit:

$$\mathbf{HP} = \mathbf{Filter} \cdot \mathbf{P}$$

Wobei Filter und Decoder als eine Gesamtheit betrachtet werden, um die symmetrische Struktur des Decoders für einen Effizienzgewinn auszunützen.

¹ Aus evolutionärer Sicht besitzt das Ohr im Gegensatz zum Auge eine umfassende Wächterfunktion. Es lässt sich anders als dieses auch nicht verschließen.

² Das menschliche Gehör hat die Fähigkeit sich in einem Stimmen- und Geräuschgewirr auf eine einzelne Stimme zu konzentrieren und dabei das akustische Rundherum weitgehend auszublenden. Eine Voraussetzung dafür ist das Hören mit zwei Ohren. Die unterschiedlichen Schallsignale der beiden Ohren werden im Gehirn ausgewertet. Dadurch entsteht eine akustische Abbildung der Umgebung, das auch die Ortung von Schallquellen an verschiedenen Punkten im Raum einschließt. So ist es möglich sich auf Schall aus einer bestimmten Richtung zu konzentrieren.

Vor der HRTF-Filterung wird nun die akustische Linse eingeführt, die durch eine entsprechende Gewichtung der virtuellen Lautsprecher signale realisiert wird mit:

$$\tilde{\mathbf{P}} = \mathbf{W}(\mathbf{a}) \cdot \mathbf{P} \quad (4.23)$$

mit $\tilde{\mathbf{P}}$, den modifizierten Lautsprecher signalen. Die Gewichtungsmatrix $\mathbf{W}(\mathbf{a})$ ist abhängig vom Steuerparameter \mathbf{a} , der die Linsenwirkung vorgibt, und kann wie folgt beschrieben werden.

$$\mathbf{W}(\mathbf{a}) = (1 - \mathbf{a}) \cdot \mathbf{V} + \mathbf{a} \cdot \mathbf{I} \quad (4.24)$$

für $\mathbf{a} = 0$ ist $\mathbf{W}(\mathbf{a}) = \mathbf{V}$

d.h. die akustische Linse wird vollständig wirksam.

für $\mathbf{a} = 1$ wird $\mathbf{W}(\mathbf{a}) = \mathbf{I}$

d.h. die akustischen Linse hat keinen Einfluss mehr.

Die akustische Linse wird in das zeitvariante Binauralsystem integriert durch:

$$\mathbf{HP} = \mathbf{Filter} \cdot \tilde{\mathbf{P}} = \mathbf{Filter} \cdot \mathbf{W}(\mathbf{a}) \cdot \mathbf{P} \quad (4.25)$$

$$\mathbf{HP} = \mathbf{Filter} \cdot [(1 - \mathbf{a}) \cdot \mathbf{V} + \mathbf{a} \cdot \mathbf{I}] \cdot \mathbf{Decoder} \cdot \mathbf{Rot} \cdot \mathbf{Amb} \quad (4.26)$$

Die Gewichtungsmatrix \mathbf{V} entspricht einer räumlichen Filterung. Die einfachste Lösung für die Filtermatrix \mathbf{V} wäre durch die Wahl nur eines einzigen virtuellen Lautsprechers (zB für Lautsprecher Nr. 3 aus $154,3^\circ$ in einem System 2. Ordnung mit 5 LS) gegeben.

$$\mathbf{V} = \begin{bmatrix} 0 & 0 & \dots & & \\ 0 & 0 & \ddots & & \\ \vdots & \ddots & 1 & & \vdots \\ & & & 0 & \\ & & & & \ddots & 0 \\ & & \dots & 0 & 0 \end{bmatrix} \quad (4.27)$$

Dadurch würden aber nicht nur Quellen direkt um die Lautsprecherposition aus $154,3^\circ$ hörbar werden, sondern durch die aus der Ambisonic-Dekodierung erstehenden Nebenkeulen auch andere Quellrichtungen. Abbildung 4.8 zeigt die Stärke des aus der Lautsprecherrichtung wiedergegebenen Anteils in Abhängigkeit vom Quellwinkel. Bei der Anwendung in einem Binauralsystem würden „fremde“ Quellrichtungen zwar abgesenkt, aber fälschlicherweise aus der Zielrichtung der akustischen Linse wiedergegeben, was eine weitere perzeptive Unterdrückung durch den Cocktailparty-Effekt verhindert.

Die minimale Breite der daraus resultierenden akustischen Linse ist durch das doppelte Winkelinkrement¹ Δf zwischen den virtuellen Lautsprecher gegeben, wenn keine Übergangssamples verwendet werden. Durch die Verwendung von Übergangssamples wird dieser Bereich ein wenig verbreitert. Ist dieser Bereich für eine Anwendung zu groß, so muss die Anzahl N der virtuellen Lautsprecher vergrößert werden, um Δf zu verringern. Diese Vorgangsweise hat zusätzliche Filteroperationen der neuen virtuellen Lautsprechersignale zur Folge.

Da die HRTFs **Filter** und die Decodermatrix **Decoder** aufgrund der Symmetrieeigenschaften nicht getrennt werden sollten, muss $\mathbf{W}(\mathbf{a})$ entsprechend umgeformt werden.

$$\mathbf{W}(\mathbf{a}) \cdot \mathbf{Decoder} \equiv \mathbf{Decoder} \cdot \tilde{\mathbf{W}}(\mathbf{a}) \quad (4.29)$$

$$\tilde{\mathbf{W}}(\mathbf{a}) = \mathbf{Decoder}^{-1} \cdot \mathbf{W}(\mathbf{a}) \cdot \mathbf{Decoder} \quad (4.30)$$

$$\tilde{\mathbf{W}}(\mathbf{a}) = (1 - \mathbf{a}) \cdot \mathbf{Decoder}^{-1} \cdot \mathbf{V} \cdot \mathbf{Decoder} + \mathbf{a} \cdot \mathbf{I} \quad (4.31)$$

Somit ergibt sich für die akustische Linse (Matrix **AL**) folgende Beziehung, die während der Laufzeit zeitlich unveränderlich bleibt.

$$\mathbf{AL} = \mathbf{Decoder}^{-1} \cdot \mathbf{V} \cdot \mathbf{Decoder} \quad (4.32)$$

$$\tilde{\mathbf{W}}(\mathbf{a}) = (1 - \mathbf{a}) \cdot \mathbf{AL} + \mathbf{a} \cdot \mathbf{I} \quad (4.33)$$

Es ergibt sich somit die neue Lösung:

$$\mathbf{HP} = \mathbf{Filter} \cdot \mathbf{Decoder} \cdot \tilde{\mathbf{W}}(\mathbf{a}) \cdot \mathbf{Rot} \cdot \mathbf{Amb} \quad (4.34)$$

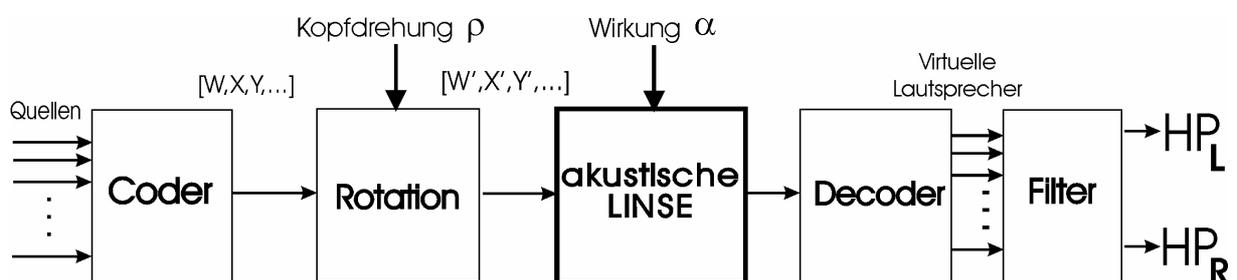


Abb. 4.10: Berücksichtigung der akustischen Linse im Strukturschaltbild.

¹ Das Winkelinkrement berechnet sich für eine symmetrische Anordnung wie folgt: $\Delta f = 2p/N$

4.3.3 Translatorische Bewegung

In diesem Abschnitt wird eine Kodierungsvorschrift zur Berücksichtigung der relativen Entfernung zwischen Quelle und Hörer bei Kopfhörerwiedergabe entwickelt. Ausgangspunkt ist eine Verteilung der virtuellen Lautsprecher entlang eines Kreises mit Radius R um die Abhörposition $(0,0)$. Bei einer translatorischen Bewegung des Hörers aus dem Ursprung zum Punkt (r, \mathbf{j}) bzw. (x,y) treten folgende zwei Phänomene auf:

Nähert man sich einer Schallquelle, so erhöht sich der Schalldruck. Da der Ambisonic-Ansatz von Freifeldbedingungen ausgeht, kann diese Abhängigkeit als Grundlage der Berechnung verwendet werden, wobei $R \gg r$ gewählt wird, d.h. die Bewegungen aus dem Ursprung sind klein im Vergleich zum Lautsprecherabstand R . Es ergibt sich somit ein die translatorische Bewegung begrenzender Bereich, der in Abbildung 4.11 für $r \leq 0.5 \cdot R$ strichliert dargestellt ist.

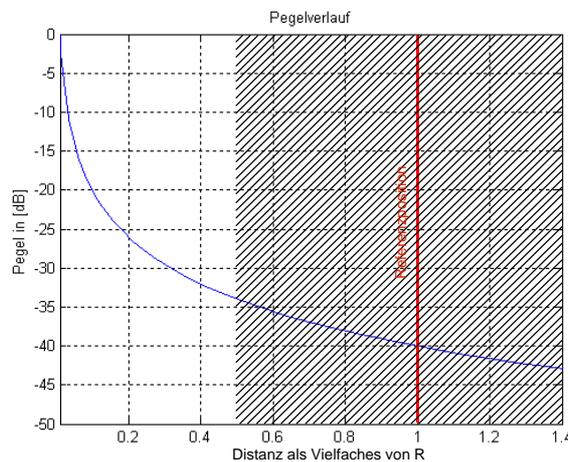


Abb. 4.11: Pegel für unterschiedliche Distanzen bei Freifeldbedingung (Referenzdruck bei 1m).

Darüber hinaus erfolgt infolge der neuen Abhörposition eine Änderung der wahrgenommenen Richtungen der räumlich verteilten Quellen. Durch die Annäherung an eine Quellenverteilung wird die wahrgenommene räumliche Ausdehnung bzw. der durch sie besetzte Winkelausschnitt vergrößert, wohingegen eine Entfernungszunahme eine Verkleinerung der räumlichen Ausdehnung zur Folge hat.

Die Quellen (virtuelle Lautsprecher) befinden sich auf einem Kreis mit dem Radius R um den Bezugspunkt. Wird nicht nur die Kopfdrehung \mathbf{r} bezüglich einer Referenzrichtung erfasst, sondern auch die relative Position (x,y) zum Referenzpunkt $(0,0)$, so kann die Änderung der Pegel und Winkel der einzelnen Quellen berücksichtigt werden.

$$r = \sqrt{x^2 + y^2} \quad (4.35)$$

$$f = \operatorname{atan} \frac{y}{x} \quad (4.36)$$

Nach einigen Umformungen¹ erhält man:

$$\sin \vartheta' \cong \left(1 + \frac{a^2}{8}\right) \sin \vartheta + \left(\frac{a}{2} - \frac{a^3}{8}\right) \sin 2\vartheta + \frac{3a^3}{8} \sin 3\vartheta \quad (4.44)$$

Die übrigen Komponenten werden mit Hilfe der folgenden Ansätze berechnet.

$$\cos \vartheta' = \frac{R \cdot \cos \vartheta - r}{R'} \quad (4.45)$$

$$\cos 2\vartheta' = 1 - 2 \cdot \sin \vartheta' \cdot \sin \vartheta' \quad (4.46)$$

$$\cos 3\vartheta' = \left(2 \cdot \cos 2\vartheta' - 1\right) \cdot \cos \vartheta' \quad (4.47)$$

$$\sin 2\vartheta' = 2 \cdot \sin \vartheta' \cdot \cos \vartheta' \quad (4.48)$$

$$\sin 3\vartheta' = \sin \vartheta' + 2 \cdot \cos 2\vartheta' \cdot \sin \vartheta' \quad (4.49)$$

Zusammenfassend ergibt sich für die Matrix **WV**:

$$\mathbf{WV} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\sqrt{2}\left(\frac{a}{2} + \frac{a^3}{16}\right) & 1 - \frac{3a^2}{8} & 0 & \frac{a}{2} - \frac{a^3}{8} & 0 & \frac{3a^2}{8} & 0 \\ 0 & 0 & 1 - \frac{a^2}{8} & 0 & \frac{a}{2} - \frac{a^3}{8} & 0 & \frac{3a^2}{8} \\ -\sqrt{2}\left(\frac{a^2}{8}\right) & -a & 0 & 1 - \frac{5a^2}{8} & 0 & a - \frac{3a^3}{8} & 0 \\ 0 & 0 & -a + \frac{3a^3}{16} & 0 & 1 - \frac{11a^2}{8} & 0 & a - a^3 \\ -\sqrt{2}\left(\frac{a}{2} - \frac{9a^3}{16}\right) & \frac{a^2}{2} & 0 & -\frac{3a}{2} - \frac{5a^3}{16} & 0 & 1 - \frac{23a^2}{8} & 0 \\ 0 & 0 & -\frac{a^2}{4} & 0 & -\frac{3a}{2} & 0 & 1 - \frac{5a^2}{8} \end{bmatrix} \quad (4.50)$$

Die Berücksichtigung der Pegelkorrektur und der Winkelverzerrung wird in der Matrixschreibweise in Gl. 4.51 beschrieben und ist in Abb. 4.14 dargestellt.

$$\mathbf{Amb}' = \mathbf{Rot}(f) \cdot \mathbf{WV}(a) \cdot \mathbf{PA}(a) \cdot \mathbf{Rot}(-f) \cdot \mathbf{Amb} \quad (4.51)$$

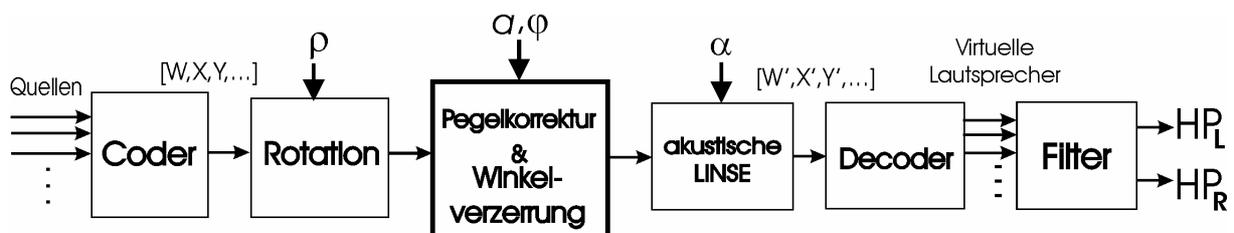


Abb. 4.14: Berücksichtigung der Pegelkorrektur und Winkelverzerrung im Strukturschaltbild.

¹ Der Term mit $\sin(4q)$, der mit einem a^3 -Glied auftritt, wird hier ebenfalls vernachlässigt.

4.3.4 Verbesserte Lokalisation in der Blickrichtung

Im Folgenden wird ein Ansatz für eine richtungsabhängig einstellbare räumliche Auflösung vorgestellt, der eine Winkelverzerrung der Ambisonic-Signale durchführt. Motiviert ist diese Vorgangsweise durch die Beschaffenheit des auditiven Lokalisationsvermögens. Bei Messungen der Lokalisationsschärfe in der Horizontalebene konnte gezeigt werden, dass diese in der Blickrichtung maximal ist und mit zunehmender Lateralisation abnimmt (vgl. [Blauert, 1997]). Bei der herkömmlichen Ambisonic-Kodierung ist die erzielbare Winkelauflösung zwar abhängig von der Ordnung, jedoch richtungsunabhängig, man kann von einer isotropen Kodierung sprechen.

Um der winkelabhängigen Lokalisationsschärfe Rechnung zu tragen, wäre es wünschenswert in Blickrichtung eine höhere wirksame Ambisonic-Ordnung zu verwenden als in anderen Richtungen. Die Breite der räumlichen Hauptkeule, die ein Maß für die Winkelauflösung darstellt, beträgt für ein 2D System m-ter Ordnung bei symmetrischer Lautsprecheraufstellung:

$$\Delta j_{2DAmb} = 2 \cdot \frac{2p}{(2m+1)} \quad (4.58)$$

Diese ergibt sich aus den Nulldurchgängen der angular sinc functions (reguläre Anordnung mit $2m+1$ Lautsprechern, vgl. Kapitel 2.2). Eine Reduktion der Hauptkeulenbreite in einer Vorzugsrichtung entspricht daher einer verbesserten räumlichen Auflösung und damit einer höheren effektiven Ambisonic Ordnung. Die einfachste Lösung dafür ergibt sich dadurch, dass die Wiedergabelautsprecher - beim Binauralsystem handelt es sich um virtuelle Lautsprecher – abweichend von regelmäßigen Aufteilung, für die der Decoder berechnet wurde, positioniert werden. Lautsprecher in Blickrichtung würden enger zueinander gestellt, Lautsprecher hinter dem Kopf weiter auseinander.

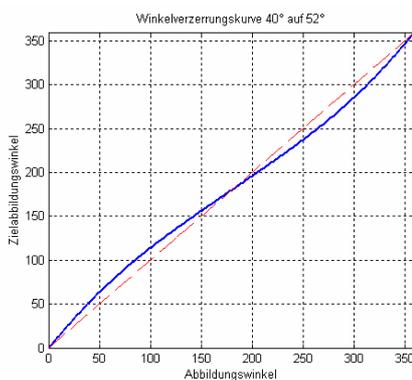
Um die daraus resultierende Winkelverzerrung des gesamten Schallfeldes auszugleichen, könnte das Schallfeld bereits bei der Kodierung entsprechend vorverzerrt werden, wodurch allerdings einer der größten Vorteile des Ambisonic-Ansatzes, nämlich die Unabhängigkeit von Kodierung und Dekodierung, aufgegeben würde. Geeigneter ist der Versuch, unter Vorgabe von räumlichen Zielfunktionen \tilde{Amb} die Winkelvorverzerrung auf die Ambisonic-Signale selbst anzuwenden. Für die Winkelvorverzerrung wird ein Winkeltransformationspaar vorgegeben, aus dem sich der Verzerrungskoeffizient a wie folgt ableitet (vgl. dazu die Translationslänge r im vorigen Abschnitt):

$$a = \cos j_{orig} - \frac{\sin j_{orig}}{\tan j_{ziel}} \quad (4.56)$$

wobei \tilde{j}_{ziel} jenen Quellwinkel in Vorzugsrichtung darstellt, auf den eine Quelle aus \tilde{j}_{orig} abgebildet werden soll. Die Vorverzerrungsfunktion für einen beliebigen Quellwinkel ergibt sich somit zu:

$$\tilde{j} = \arctan\left(\frac{\sin j}{\cos j - a}\right) \quad (4.57)$$

Die geeignete Wahl von \tilde{j}_{orig} und \tilde{j}_{ziel} hängt vom Lautsprecherwinkel der verwendeten Ambisonic Ordnung und der gewünschten verbesserten räumlichen Auflösung in der Vorzugsrichtung ab. In Abbildung 4.14 ist die Winkelvorverzerrung eines mit Ambisonic 3.Ordnung kodierten Schallfeldes dargestellt, das bei geeigneter Lautsprecheraufstellung in Vorzugsrichtung die Auflösung eines Systems 4. Ordnung erreicht. Zum Ausgleich der Winkelvorverzerrung werden die virtuellen Lautsprecher nun neu arrangiert siehe Abb. 4.14 (Tabelle rechts). Die Lautsprecher werden dabei in der Vorzugsrichtung im Abstand von 40° (typischer Lautsprecherwinkel von regelmäßiges Layout bei 4.Ordnung) anstatt von 52° (typisch für 3.Ordnung) aufgestellt.



LS#	Pos. [°]	HRTF Ohne Vorverz.	(Gl. 4.57)		HRTF Mit Vorverz.
			\tilde{j}_{orig}	\tilde{j}_{ziel}	
1	0	0	0	0	0
2	51,4	50	40	51,4	40
3	102,9	105	89	103,3	90
4	154,3	155	148	154,3	150
5	205,7	-155	212	205,7	-150
6	257,1	-105	271	256,7	-90
7	308,6	-50	320	308,6	-40

Abb. 4.14: (Links) Die Winkelvorverzerrung (dicke Linie) stellt den Zusammenhang zwischen dem Abbildungswinkel im ursprünglichen Ambisonic-Signal (Abszisse) und dem resultierenden Quellwinkel in Abbildungswinkel in \tilde{Amb} (Ordinate) für eine System 3.Ordnung und eine Erhöhung der Auflösung in 0° auf 4.Ordnung dar. (Rechts) Aufhebung der Winkelvorverzerrung durch neue virtuelle Lautsprecherpositionen.

Unter Verwendung einer Vorverzerrungsmatrix \mathbf{WV} erfolgt durch Berechnung von \tilde{Amb} mit:

$$\tilde{Amb} = \mathbf{WV} \cdot Amb \quad (4.59)$$

Diese wird durch Lösung (LMS Approximation) des überbestimmten Gleichungssystem (Gl. 4.59) bestimmt:

$$\mathbf{WV} = \tilde{Amb} \cdot \text{pinv}(Amb) \quad (4.60)$$

Ergebnisse zeigt Abbildung 4.15. Die Erhöhung der Auflösung im verzerrten Ambisonic Bereich bewirkt für den Bereich abseits der Vorzugsrichtung eine geringere Auflösung als im unverzerrten Signal.

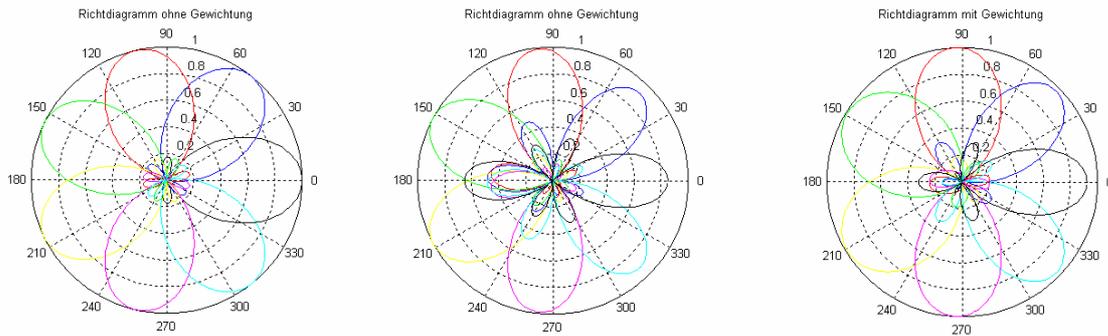


Abb. 4.15: Abbildungseigenschaften unter Verwendung der Winkelverzerrung.

Links: ohne Verzerrung (gleichverteilte Auflösung)

Mitte: mit Verzerrung (40° auf 52° Grad), ohne Gewichtung der Ambisonic Ordnungen.

Rechts: mit Verzerrung (40° auf 52° Grad) und Gewichtung der höchsten Ambisonic Ordnungen.

4.3.5 „Verräumlichung“

In [Begault et al., 2001] wurde anhand von Hörversuchen gezeigt, dass die Implementierung einer Raumsimulation (erste Reflexionen und Diffusanteil) in binauralen Wiedergabesystemen zu einer Verringerung des Azimuth-Fehlers bei der Lokalisation sowie zu einer verbesserten Außerkopf-Lokalisation bzw. Externität führt.

Die Einbindung von Raumsimulation kann durch Verwendung längerer Außenohrimpulsantworten, die zusätzlich zu den Übertragungseigenschaften von Ohren, Kopf und Torso auch Rauminformationen beinhalten, erreicht werden. Neben dem deutlich höheren Rechenaufwand stellt sich allerdings als nachteilig heraus, dass bei der Schallfeldrotation die in den HRTFs enthaltenen Raumreflexionen nicht mit dem Ambisonic-Schallfeld mitgedreht werden und so zwei virtuelle Räume inkonsistent ineinander greifen. Abhilfe kann durch die Modellierung des Raumes mit Spiegelquellen (z.B. durch ein Außenmauermodell, Spiegelquellen 1.Ordnung) geschaffen werden. Dabei wird jede Quelle durch entsprechenden Reflexionen, die zusätzlich zu kodierende Schallquellen darstellen, ergänzt [Sontacchi et al., 2002b].

Eine weitere Möglichkeit besteht in der Verwendung von Raumimpulsantworten in einem Ambisonic-Format (vgl. Kapitel 3.4). Die Ambisonic Signale müssen dabei mit den entsprechenden Impulsantworten gefiltert werden. Durch die Platzierung vor der Rotations (bzw. Translations-)matrix können auch Positionsänderungen im Raum berücksichtigt werden.

Selbst die Verhallung des fertigen Binauralsignals liefert als einfachste Lösung bereits einen wesentlich verbesserten realen Höreindruck.

Zusammenfassung

Gegenüber der direkten binauralen Implementierungstechniken ergibt sich bei der Verwendung des virtuellen Ambisonic-Ansatzes ein Vorteil, wenn mehrere Schallquellen mit zeitvarianten Positionen in einem dynamischen System berücksichtigt werden müssen. Die benötigte Rechenleistung ist im Allgemeinen durch die Wahl der Ambisonic-Ordnung und der verwendete Filterlänge der Außenohrübertragungsfunktionen festgelegt (variable Qualität der Lokalisation und Lokalisationsschärfe vgl. Kapitel 5). Durch Hinzufügen von weiteren Quellen wird der Berechnungsaufwand nicht sonderlich vergrößert. Es steigt nur der Kodierungsaufwand, der abhängig von der Ordnung M (im 3D-Fall) für jede weitere Quelle je Sample mit $(M+1)^2$ skalaren Multiplikationen und Additionen beziffert werden kann. Auf diesem Prinzip beruhend können weiters für alle Quellen diskrete Reflexionen (basierend auf dem Spiegelquellenmodell) berücksichtigt werden. Betrachtet man ein 3 dimensionales Raummodell mit Reflexionen 1.Ordnung, so ergeben sich pro Quelle zusätzlich 8 Spiegelquellen; bei Verwendung Reflexionen 2.Ordnung ergeben sich weitere 17 Spiegelquellen. Bei einer direkten Implementierung müsste jede Spiegelquelle eigens HRTF-gefiltert werden.

Eine vergleichbare Alternative stellt binaurale B-Format von Jot dar (siehe [Jot et al., 1998]). Die Außenohrübertragungsfunktionen werden zuerst in ein richtungsabhängiges Delay (excess phase) und ein minimalphasiges Filter zerlegt. Bei der Interpolation für beliebige Richtungen können Kammfiltereffekte, die im allgemeinen durch Überlagerung verschiedener Richtungen entstehen, vermieden werden. Bei den meisten binauralen Verfahren erfolgt eine Änderung der Quellrichtung durch Interpolation zwischen verschiedenen HRTFs. Beim binauralen B-Format wird eine gesuchte Übertragungsfunktion jedoch aus einer Linearkomposition eines orthogonalen Sets von raum- und frequenzabhängigen Basisfunktionen gebildet. Dabei wird jeweils für das linke und das rechte Ohr ein Set bestehend aus vier Basisfunktionen bereitgestellt. Trotz Ambisonic 1.Ordnung werden verglichen mit diskreten Interpolationstechniken wie VBAP, VBIP (vgl. Kapitel 1.2) wesentlich bessere Ergebnisse erzielt (vgl. [Jot et al., 1999]). Bei genauerer Betrachtung muss dieses System hinsichtlich des Berechnungsaufwands jedoch mindestens mit einem virtuellen Ambisonic System 2. bzw. 3.Ordnung (5 bzw. 7 Filter, vgl. Kapitel 4) verglichen werden, da beim binauralen B-Format auch im 2D Fall bereits 6 Filter benötigt werden. Ein wesentlicher Nachteil des binauralen B-Formats stellt die Trennung zwischen linkem und rechtem Ohr und somit die einheitliche von den HRTFs unabhängige Schallfeldrepräsentation dar. Dadurch können etwa Schallfeldrotation, Translation oder akustische Linse nur durch erheblich größeren Aufwand umgesetzt werden.

Kapitel 5

Hörversuche

Objektives Fehlermaß

Objektive Analyse der Systemparameter

Hörversuche zu den ausgewählten Systemen

Hörversuche sind ein elementarer Bestandteil zur Evaluierung von Wiedergabesystemen. Im folgenden werden die Ergebnisse umfangreicher Hörversuche am in Kapitel 4 vorgestellten zeitvarianten Binauralsystem vorgestellt, die im Rahmen dieser Arbeit durchgeführt wurden. Die Implementierung des notwendigen Echtzeitsystems erfolgte auf einem Texas Instruments TMS320C6711 (vgl. [Majdak, Noisternig, 2001]) sowie auf einem Motorola DSP 56367EVM (vgl. [Majdak, Noisternig, 2002]). Ausgangspunkt für den Entwurf der Versuchsreihe war die Entwicklung eines Fehlermaßes, welches die Wiedergabeeigenschaften bezüglich Lokalisation und Lokalisationsunschärfe objektiv beschreibt sollte. Ein Vergleich zwischen den Prognosen des Fehlermaßes und den subjektiven Ergebnissen findet sich ebenfalls in diesem Kapitel.

5.1 Objektives Fehlermaß

Im folgenden wird ein Fehlermaß zur objektiven Beurteilung binauraler Wiedergabesysteme vorgestellt (vgl. [Sontacchi et al., 2002c]). Die Definition dieses Fehlermaßes basiert auf der psychoakustisch bewerteten Abweichung der durch das Wiedergabesystem verfälschten HRIRs zu den direkt gemessenen HRIRs (Referenzkurven). Somit wird bei der Betrachtung des Fehlermaßes angenommen, dass die Referenz-HRIRs optimale Lokalisation bieten.

Für das in Kap. 4 vorgestellte binaurale Ambisonic-System wird die Auswirkung verschiedener Optimierungsverfahren auf die Abbildungsgenauigkeit virtueller Schallquellen untersucht. In Kap. 5.3 wird die Aussagekraft des Fehlermaßes in der Gegenüberstellung zu den Ergebnissen von Hörversuchen untersucht.

5.1.1 Berechnung von Lokalisationsfehlern

Die Optimierung eines Wiedergabesystems bedarf eines Maßes, mit dem sich das Qualität des Systems beschreiben lässt. Idealerweise handelt es sich um eine skalare Größe, deren Wert Rückschlüsse hinsichtlich der untersuchten Parameter zulässt. Die Untersuchung der Lokalisation lässt sich mittels der Lokalisationsfunktion und Lokalisationsunschärfe beschreiben. Beide Funktionen sind abhängig vom Azimutwinkel. Bei der Fehlerberechnung hängt die Auflösung der betrachteten Azimutwinkel von der Anzahl der Referenz-HRIRs ab und stellt somit eine diskrete Größe dar.

DEFINITIONEN

Lokalisationsfunktion $L(\mathbf{q})$, gibt den Azimutwinkel an, aus dem die virtuelle Schallquelle wahrgenommen wird.

Lokalisationsunschärfe¹ (*localisation blur*) $Bl(\mathbf{q})$, gibt die wahrgenommene Breite einer Punktquelle in Grad an.

Um im Sinne der Optimierung die winkelabhängigen Größen auf ein Skalar reduzieren zu können, erfolgt eine Mittelung über den Azimutwinkel \mathbf{q} . Dies führt zur Definition folgender Größen:

$$\text{Mittlerer Lokalisationsfehler } \bar{L} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N [L(\mathbf{q}_i) - \mathbf{q}_i]^2}$$

Stellt den Effektivwert der Abweichung der Lokalisationsfunktion von der Soll-Lokalisation dar.

Mittlere Lokalisationsunschärfe \overline{Bl}

Stellt den Mittelwert der Lokalisationsunschärfe dar und wird als Vielfaches des MAA^2 (*Minimal Audible Angle*) angegeben.

5.1.2 Berechnung der interauralen Laufzeitdifferenz

Aus psychoakustischer Sicht ist die Berechnung der mittleren Gruppenlaufzeit innerhalb der gehörspezifischen Frequenzbereiche (Bark-Bereiche) ausreichend. Die Impulsantworten werden mittels frequenzgruppenbreiter, nullphasiger Bandpaßfilter gefiltert. Nullphasigkeit wird hier durch die Filterung der HRIR mit der kausalen und anschließend antikausalen Impulsantwort eines Filters erreicht. Das Filterdesign erfolgt mittels 'Least-Square-Error-Minimalization'. Somit wird eine Anpassung der Bandbreite der jeweiligen Filter an die Frequenzgruppen der menschlichen Hörwahrnehmung gewährleistet. Die Bestimmung der Gruppenlaufzeit innerhalb der jeweiligen Frequenzgruppe erfolgt durch die Berechnung des Energieschwerpunktes für die frequenzgruppenbreiten gefilterten Impulsantworten des Systems.

$$h_F(n, z, \mathbf{q}) = h(n, \mathbf{q}) * f(n, z) * f(-n, z) \quad (5.1)$$

¹ Beschreibt die Konzentration der Lokalisation *einer* Schallquelle (nicht auf eine Referenzquelle bezogen). Die Lokalisationsunschärfe hängt sehr stark vom Quellspektrum ab.

² Der minimal hörbare Winkel beschreibt den kleinsten Winkelunterschied, für den *zwei* Schallquellen als aus unterschiedlichen Richtungen kommend wahrgenommen werden (vgl. [Blauert 1997], [Gilkey et al. 1997]). Der MAA hängt sowohl von der Richtung als auch vom Quellspektrum ab.

Wobei $h_F(n, z, \mathbf{q})$ die gefilterten Impulsantworten, $h(n, \mathbf{q})$ die HRIR und $f(n, z)$ die frequenzgruppenbreiten Filter darstellt. Die Berechnung des Energieschwerpunktes und somit der Gruppenlaufzeit innerhalb der jeweiligen Frequenzgruppe, kann wie folgt durchgeführt werden:

$$T_g(z, \mathbf{q}) = \frac{\sum_n n \cdot h_F^2(n, z, \mathbf{q})}{\sum_n h_F^2(n, z, \mathbf{q})} \quad (5.2)$$

Mit diesem Verfahren werden nun folgende Gruppenlaufzeiten berechnet:

$$\begin{aligned} T_{g,REF,links}(z, \mathbf{q}) & \dots \text{Referenz-HRIRs, für das linke Ohr} \\ T_{g,REF,rechts}(z, \mathbf{q}) & \dots \text{Referenz-HRIRs, für das rechte Ohr} \\ T_{g,AMB,links}(z, \mathbf{q}) & \dots \text{Ambisonic-System, für das linke Ohr} \\ T_{g,AMB,rechts}(z, \mathbf{q}) & \dots \text{Ambisonic-System, für das rechte Ohr} \end{aligned}$$

Die Interaurale Laufzeitdifferenz (*ITD*, *Interaural Time Difference*) wird durch die Berechnung der Laufzeitdifferenz zwischen linker und rechter HRIR in jeder Frequenzgruppe ermittelt. Dies erfolgt sowohl für die Referenz-HRIRs,

$$IDT_{Ref}(z, \mathbf{q}) = T_{g,REF,links}(z, \mathbf{q}) - T_{g,REF,rechts}(z, \mathbf{q}) \quad (5.3)$$

als auch für die durch das Ambisonicsystem verzerrten HRIRs:

$$IDT_{AMB}(z, \mathbf{q}) = T_{g,AMB,links}(z, \mathbf{q}) - T_{g,AMB,rechts}(z, \mathbf{q}) \quad (5.4)$$

5.1.3 Lokalisation aufgrund interauraler Laufzeitdifferenzen

Die Laufzeitdifferenzen (ITDs) sind vorwiegend für die Lokalisation im tieffrequenten Bereich verantwortlich. Bei gegenüber den Referenz-HRIRs größerer ITD, wird die Schallquelle bis zu einer bestimmten Grenze weiter auslateralisiert empfunden. Um die Verfälschung der Lokalisation durch das System ermitteln zu können, wird die Änderung der Laufzeitdifferenz betrachtet. Das Ambisonic-System erzeugt für jeden Azimutwinkel und jede Frequenzgruppe eine bestimmte ITD, welche durch die Größe $ITD_{AMB}(z, \mathbf{q})$ dargestellt wird.

$IDT_{REF}(z, \mathbf{q})$ folgt aus den Referenz-HRIRs. Dieser allgemeine Ansatz geht davon aus, dass die Lokalisation der Schallquellen sich aus den Referenz-HRIRs ableiten lässt und Verzerrungen der ITD eine Veränderung der Lokalisation hervorrufen. Um die Systemlokalisierung \mathbf{q}_{ITD} berechnen zu können, wird jener Azimutwinkel \mathbf{q} ermittelt, bei dem die Referenz-HRIRs und die HRIRs am Ausgang des Systems denselben Wert für die ITD aufweisen.

$$\mathbf{q}_{ITD}(z, \mathbf{q}_0) = \arg \min_{\mathbf{q}} |IDT_{REF}(z, \mathbf{q}) - IDT_{AMB}(z, \mathbf{q}_0)| \quad (5.5)$$

Die Suche nach der passenden ITD erfolgt allerdings nur in der Umgebung des gegebenen Azimutwinkels \mathbf{q}_0 :

$$\forall \mathbf{q} \in \left[\mathbf{q}_0 - \frac{\mathbf{q}_{WND}}{2}; \mathbf{q}_0 + \frac{\mathbf{q}_{WND}}{2} \right] \quad (5.6)$$

wobei \mathbf{q}_{WND} die Breite des Suchfensters (hier meist 180°) darstellt.

Ist der gesuchte Wert für die ITD jedoch größer als die maximale ITD der Referenz-HRIRs, terminiert die Suche nicht. Ein $ITD > 1\text{ms}$ zerstört die Konsistenz des Binauralsignals. In solchen Fällen wird die Systemlokalisierung dennoch auf $\pm 90^\circ$ gesetzt:

$$\max_{\mathbf{q}} (ITD_{REF}(z, \mathbf{q})) < ITD_{AMB}(z, \mathbf{q}_0) \Rightarrow \mathbf{q}_{ITD}(z, \mathbf{q}_0) = \begin{cases} 90^\circ \text{ für } \mathbf{q}_0 \in [0^\circ; 180^\circ] \\ -90^\circ \text{ für } \mathbf{q}_0 \in (0^\circ; 180^\circ) \end{cases} \quad (5.7)$$

Die Berechnung erfolgt für: $\mathbf{q}_0 \in [0^\circ; 360^\circ]$ und $z \in [1; 24]$

Dies ergibt den auf der ITD basierenden Datensatz an Lokalisationspunkten, welcher auch als ITDA (*Interaural Time Difference Angle*) bezeichnet wird (vgl. [Pulkki, Karjalainen, 2001a], [Pulkki, Karjalainen, 2001b]).

5.1.4 Berechnung der interauralen Pegeldifferenz

Die Differenz der Pegel am linken und rechten Ohr wird als ILD (*Interaural Level Difference*) bezeichnet und in dB dargestellt. Für die Berechnung der ILD werden die Impulsantworten mittels Fouriertransformation in den Frequenzbereich transformiert und die Beträge der Übertragungsfunktionen logarithmisch über kritische Bandbreiten gemittelt dargestellt (vgl. [Blauert, 1997]). Somit erhält man die Pegeldifferenzen der Referenz-HRIRs

$$ILD_{REF}(z, \mathbf{q}) = 10 \log \sum_{f(z-1)}^{f(z)-1} [FFT \{h_{REF,L}(n, \mathbf{q})\}]^2 - 10 \log \sum_{f(z-1)}^{f(z)-1} [FFT \{h_{REF,R}(n, \mathbf{q})\}]^2 \quad (5.8)$$

sowie die des Ambisonic-Systems:

$$ILD_{AMB}(z, \mathbf{q}) = 10 \log \sum_{f(z-1)}^{f(z)-1} [FFT \{h_{AMB,L}(n, \mathbf{q})\}]^2 - 10 \log \sum_{f(z-1)}^{f(z)-1} [FFT \{h_{AMB,R}(n, \mathbf{q})\}]^2 \quad (5.9)$$

5.1.5 Lokalisation aufgrund interauraler Pegeldifferenzen

Die Berechnung der Lokalisation aufgrund der Pegeldifferenzen erfolgt ebenfalls nach dem genannten Suchalgorithmus. Dies führt auf die Lokalisationsfunktion $\mathbf{q}_{ILD}(z, \mathbf{q}_0)$:

$$\mathbf{q}_{ILD}(z, \mathbf{q}_0) = \arg \min_{\mathbf{q}} |ILD_{REF}(z, \mathbf{q}) - ILD_{AMB}(z, \mathbf{q}_0)| \quad (5.10)$$

Auch hier wird die Suche auf den Bereich $\mathbf{q}_0 \pm \frac{\mathbf{q}_{WDN}}{2}$ eingeschränkt. Bei Pegeldifferenzen größer als die maximale ILD der Referenz-HRIRs kann kein passender Azimutwinkel gefunden werden. Für die Lokalisation bedeuten größere Pegeldifferenzen stärkere Auslateralisierung der wahrgenommenen Schallquelle. Die Systemlokalisierung wird in solchen Fällen auf $\pm 90^\circ$ gesetzt, was einer dichotischen¹ Wahrnehmung entspricht.

$$\max_{\mathbf{q}} (ILD_{REF}(z, \mathbf{q})) < ILD_{AMB}(z, \mathbf{q}_0) \Rightarrow \mathbf{q}_{ILD}(z, \mathbf{q}_0) = \begin{cases} 90^\circ \text{ für } \mathbf{q}_0 \in [0^\circ; 180^\circ] \\ -90^\circ \text{ für } \mathbf{q}_0 \in (0^\circ; 180^\circ) \end{cases} \quad (5.11)$$

Dies ergibt den auf der ILD basierenden Datensatz an Lokalisationspunkten $\mathbf{q}_{ILD}(z, \mathbf{q}_0)$, welcher als ILDA (*Interaural Level Difference Angle*) bezeichnet wird (vgl. [Pulkki, Karjalainen, 2001a], [Pulkki, Karjalainen, 2001b]).

5.1.6 Zusammenführung der Lokalisationsparameter

Sowohl die Lokalisation aufgrund der Pegeldifferenzen als auch der Laufzeitdifferenzen sind für die Richtungswahrnehmung der virtuellen Schallquelle verantwortlich. Bei der Zusammenführung der beiden Größe müssen diese je nach Frequenzbereich unterschiedlich gewichtet werden. Die Lokalisation aufgrund der Laufzeitdifferenz $\mathbf{q}_{ITD}(z, \mathbf{q})$ wirkt vorwiegend im tieffrequenten Bereich (vgl. [Blauert, 1997]).

Nach tiefen Frequenzen hin ist jedoch die Übertragung des Gesamtsystem (gemessene HRIR, Kopfhörer, etc.) begrenzt, so dass das erste Barkband ($f < 100\text{Hz}$) für die weiteren Berechnungen nicht berücksichtigt wird. Der Einfluss der ITD auf die Lokalisation wird für Frequenzen oberhalb von 800Hz geringer², dies wird durch die Gewichtung $w_{ITD}(z)$ modelliert (siehe Abb.5.1).

¹ Binaurale Verschmelzung der beidseitigen frequenz- u. intensitätsverschiedenen Hörempfindung.

² Bei dem vorgestellten Fehlermaß werden nur die Laufzeit bzw. Phasenunterschiede der Frequenzkomponenten (Trägerfrequenzen) nicht jedoch der Einhüllenden betrachtet.

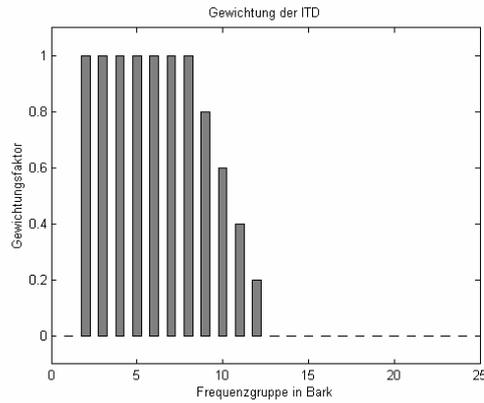


Abb. 5.1: Gewichtung der ITD über die Bark-Bänder

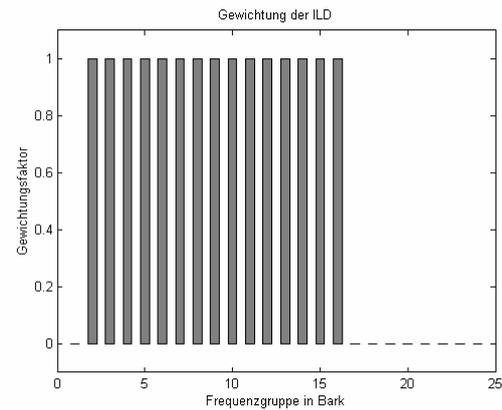


Abb. 5.2: Gewichtung der ILD über die Bark-Bänder

Die Auswirkung der Pegeldifferenzen auf die Lokalisation von Schallquellen ist im gesamten Frequenzbereich vorhanden. Jedoch wird auch hier die erste Frequenzgruppe aus den oben genannten Gründen ausgeblendet (siehe Abb 5.2).

Das in Kap. 4.2 beschriebene zweidimensionale binaurale Wiedergabesystem liefert aufgrund der Ambisonic-Ordnung nur bis annähernd 3 kHz eine ausreichende Abbildungsgenauigkeit. Aus diesem Grund werden alle Frequenzgruppen oberhalb von 16 Bark ($f > 3100$ Hz) ausgeblendet. Die Grenzfrequenz kann durch Erhöhung der Ambisonic-Ordnung erhöht werden (vgl. Kap. 4.2, Abb. 4.2).

Die Lokalisationsfunktion ergibt sich durch Verknüpfung der frequenzgewichtet gemittelten ITD- und ILD-Lokalisation. Dabei werden beide Mechanismen gleichstark gewichtet.

$$L(\mathbf{q}) = \frac{1}{2} \cdot \left(\frac{1}{\sum w_{ITD}(z)} \cdot \sum_{z=1}^{24} w_{ITD}(z) \cdot \mathbf{q}_{ITD}(z, \mathbf{q}) + \frac{1}{\sum w_{ILD}(z)} \cdot \sum_{z=1}^{24} w_{ILD}(z) \cdot \mathbf{q}_{ILD}(z, \mathbf{q}) \right) \quad (5.12)$$

5.1.7 Lokalisationsunschärfe

Die Lokalisationsunschärfe (*localisation blur*) gibt die Breite des Hörereignisses an und wird als Standardabweichung der Lokalisationsfunktion über die Frequenzgruppen bei gegebenem Azimutwinkel definiert. Unter Verwendung der oben eingeführten Frequenzgewichte wird ein Einzahlwert für die Lokalisationsunschärfe als Funktion des Azimutwinkels berechnet mit:

$$Bl(\mathbf{q}) = \sqrt{\frac{1}{2} \cdot \frac{\sum_{i=ILD}^{ITD} \frac{1}{\sum w_i(z)} \cdot \sum_{z=1}^{24} w_i(z) \cdot [\mathbf{q}_i(z, \mathbf{q}) - L(\mathbf{q})]^2}{\sum_{i=ILD}^{ITD} \frac{1}{\sum w_i(z)} \cdot \sum_{z=1}^{24} w_i(z)}} \quad (5.13)$$

wobei für $i=ILD$ gilt: $\mathbf{q}_i \equiv \mathbf{q}_{ILD}$ $w_i \equiv w_{ILD}$ und für $i=ITD$ gilt: $\mathbf{q}_i \equiv \mathbf{q}_{ITD}$ $w_i \equiv w_{ITD}$.

5.1.8 Mittlere Fehler

Mit der berechneten Lokalisationsfunktion und der Lokalisationsunschärfe kann nun eine Aussage über die Lokalisation von Ambisonic-kodierten Quellen gegenüber den Referenz-HRIRs getroffen werden. Um das Ambisonic-System einfacher beurteilen zu können, werden mittlere Fehler gemäß Kap. 5.1.1 berechnet. Aufgrund der symmetrischen Eigenschaften der ITD und ILD stellt die Lokalisationsfunktion eine schiefsymmetrische, die Lokalisationsunschärfe eine symmetrische Funktion bezüglich der \mathbf{q} -Achse dar. Die Berechnung entlang eines Halbkreises (in den ersten beiden Quadranten) beinhaltet somit alle relevanten Daten.

Der Lokalisationsfehler \bar{L} berechnet sich somit folgendermaßen:

$$\bar{L} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N [L(\mathbf{q}_i) - \mathbf{q}_i]^2} \quad (5.14)$$

wobei N der Anzahl der Referenz-Aufnahmen zwischen 0° und 180° entspricht.

Um die mittlere Lokalisationsunschärfe zu berechnen, wird der jeweilige Wert der Lokalisationsunschärfe auf den $MAA(\mathbf{q})$ bezogen. Zwischen den einzelnen der Literatur entnommenen Werten wird linear interpoliert.

Die mittlere Lokalisationsunschärfe berechnet sich folgendermaßen:

$$\overline{Bl} = \frac{1}{N} \cdot \sum_{q=0^\circ}^{180^\circ} \frac{Bl(\mathbf{q})}{MAA(\mathbf{q})} \quad (5.15)$$

- wobei N wiederum der Anzahl der Referenz-Aufnahmen zwischen 0° und 180° entspricht.

Im Idealfall ergeben sich Lokalisationsfehler und Lokalisationsunschärfe zu Null. Somit besteht die Möglichkeit Optimierungen der Systemparameter (Ambisonic Ordnung, Dekoder Gewichte, Länge der verwendeten HRIRs, etc.) hinsichtlich eines vorgegebenen Ziels durchzuführen – z.B. Minimierung der Rechenleistung bei konstantem Lokalisationsfehler.

5.2 Objektive Analyse der Systemparameter

Im folgenden Abschnitt erfolgt die Analyse und Beurteilung wesentlicher Systemparameter des binauralen Wiedergabesystems auf der Basis der Fehlermaes, wobei in einer Optimierung eher die Lokalisationsunschrfe reduziert werden sollte, da der Lokalisationsfehler durch eine Vorverzerrung in der Ambisonic-Kodierung kompensiert werden kann. Bei den Abbildungen wird folgende Kurzbezeichnung verwendet: #xxx, wobei xxx fr die Nummer des jeweiligen Settings steht. Die genaue Beschreibung der jeweiligen Settings kann dem Anhang entnommen werden.

5.2.1 Ordnung des Ambisonic-Systems

Wie bereits in Kap. 4.3.1 beschrieben erfolgt eine Dekodierung eines 2D Ambisonic-Systems bei $N = 2 \cdot M + 1$ Lautsprechern den geringsten mittleren Abbildungsfehler (vgl. [Poletti, 1996a]). Bei Systemen 3. Ordnungen tritt eine wesentliche Verbesserung der Lokalisationseigenschaften gegenber der Ordnung 2 auf. Durch Erhhung der Ordnung auf 4 kann jedoch keine wesentliche Verbesserung erzielt werden. Vor allem die Reduktion der Lokalisationsunschrfe im fr Binauralwiedergabe primr wichtigen 1. Quadranten rechtfertigt den erhhten Rechenaufwand von Systemen 4. Ordnung (vgl. Abb. 5.3 und Abb. 5.4).

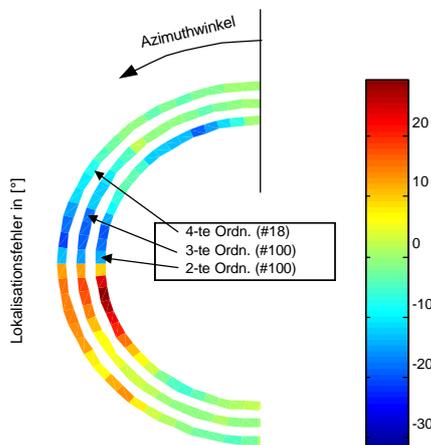


Abb. 5.3: Lokalisationsfehler verschiedener Ordnungen bei Filterlngen von 128 Samples und optimaler Dekodierung

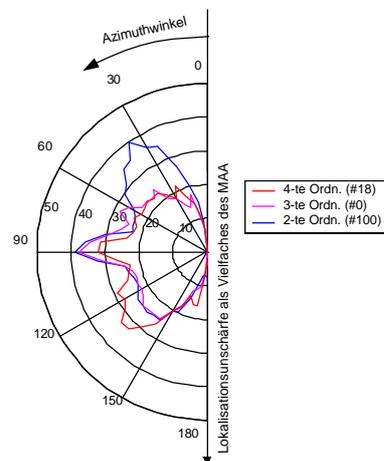


Abb. 5.4: Lokalisationsunschrfe verschiedener Ordnungen bei Filterlngen von 128 Samples und optimaler Dekodierung

5.2.2 Anzahl und Anordnung der Lautsprecher

Durch eine Erhhung der Lautsprecheranzahl $N > 2 \cdot M + 1$ erfolgt keine optimale Dekodierung. Wird eine virtuelle Schallquelle genau aus der Richtung eines Lautsprechers abgebildet, liegen die anderen Lautsprecher nicht mehr in den Nullpunkten der ASINC -

Funktionen (vgl. Kap. 1.2). Dadurch erfolgt ein Übersprechen auf nebenliegende Lautsprecher, die virtuelle Schallquelle wird aufgeweitet. Die Untersuchung unterschiedlicher Lautsprecheranordnung reduziert sich unter Ausnutzung der Symmetrieeigenschaften (Halbierung der benötigten HRIR-Filter Anzahl, vgl. Kapitel 4.2.1) auf 2 Fälle:

- 3 virtuelle Lautsprecherpositionen vorne
(Drehung des Systems äquidistanter Lautsprecher, so dass in 0° Richtung genau ein Center-Lautsprecher steht).
- 2 virtuelle Lautsprecherpositionen vorne
(Anordnung des Systems äquidistanter Lautsprecher, so dass links und rechts von der 0° Richtung genau ein Lautsprecher steht, kein Center-Lautsprecher).

Der konkrete Verhalten von Lokalisationsunschärfe und –fehler hängt – wie die Analyse in Abbildung 5.5-8 zeigt - nicht nur von der Lautsprecheranordnung, sondern auch von den verwendeten HRIRs bzw. deren Länge ab. Während für längere HRIRs die mittleren Fehler für die „3 Lautsprecher vorne“ - Konfiguration geringer ausfallen, zeigt eine richtungsabhängige Betrachtung durchaus Vorteile der „2 Lautsprecher vorne“ - Aufstellung im vorderen Quadranten. Im hinteren Bereich weist diese Anordnung jedoch größere Abbildungsfehler auf. In Abb. 5.5 und 5.6 sind für die beiden Lautsprecherordnungen und zwei Referenz-HRIRs (Kemar¹, AKG²) der Lokalisationsfehler und die Lokalisationsunschärfe abhängig von der Filterlängen dargestellt.

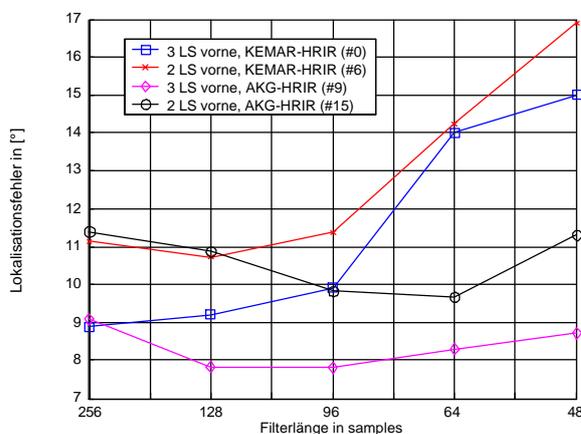


Abb. 5.5: Lokalisationsfehler verschiedener Lautsprecheranordnungen bei unterschiedlichen Referenz HRIRs.

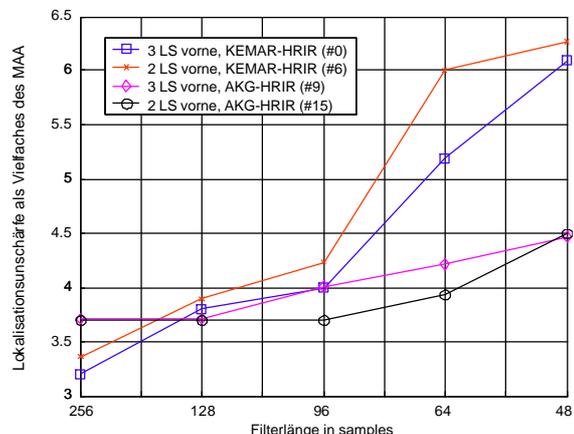


Abb. 5.6: Lokalisationsunschärfe verschiedener Lautsprecheranordnungen bei unterschiedlichen Referenz HRIRs.

¹ HRIR Kurven, gemessen am Kemar Kunstkopf von Gardner und Martin
[<http://sound.media.mit.edu/KEMAR.html>].

² HRIR Kurven, zur Verfügung gestellt von der Fa. AKG Acoustics, Wien (Kooperationspartner).

In Abb. 5.7 und 5.8 sind für die beiden Lautsprecherordnungen jeweils der Lokalisationsfehler und die Lokalisationsunschärfe abhängig vom Azimutwinkel dargestellt.

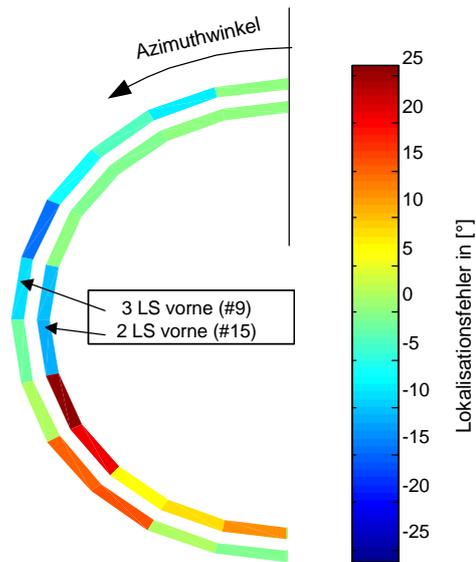


Abb. 5.7: Lokalisationsfehler bei verschiedenen Lautsprecheranordnungen für unterschiedliche Referenz HRIRs mit 128 Samples Länge.

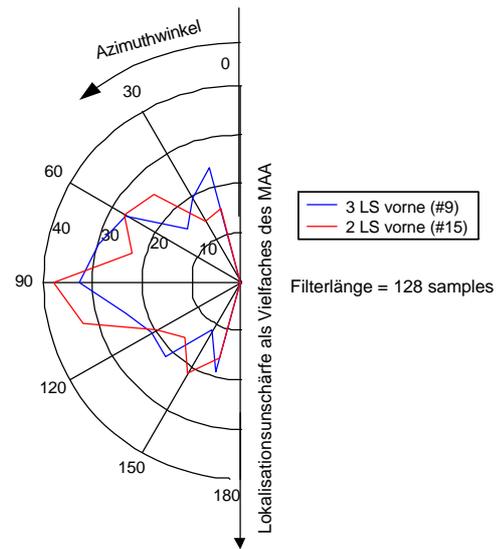


Abb. 5.8: Lokalisationsunschärfe bei verschiedenen Lautsprecheranordnungen für unterschiedliche Referenz HRIRs mit 128 Samples Länge.

5.2.3 Filterlänge

Eine Untersuchung der verwendeten Filterlänge zeigt, dass bei höheren Systemordnungen Lokalisationsfehler und –unschärfe mit der Filterlänge abnehmen, bei Systemen 2. Ordnung jedoch der Lokalisationfehler mit zunehmender Filterlängen wieder ansteigt (vgl. Abb. 5.9 und 5.10). Dieser Sachverhalt ist durch die Theorie des Ambisonic-Ansatzes nicht erklärbar. Eine mögliche Erklärung liegt im Ausschwingverhalten der HRIRs selbst.

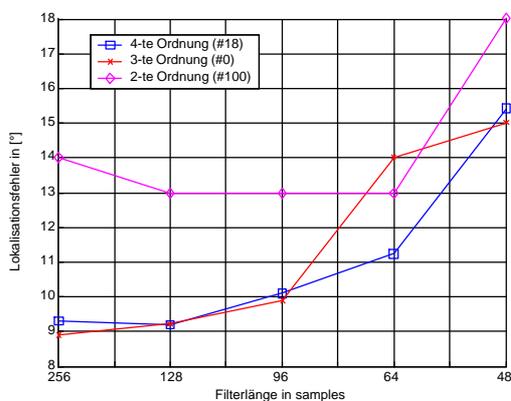


Abb. 5.9: Lokalisationsfehler verschiedener Filterlängen für unterschiedliche Ambisonic-Ordnungen bei optimaler Dekodierung.

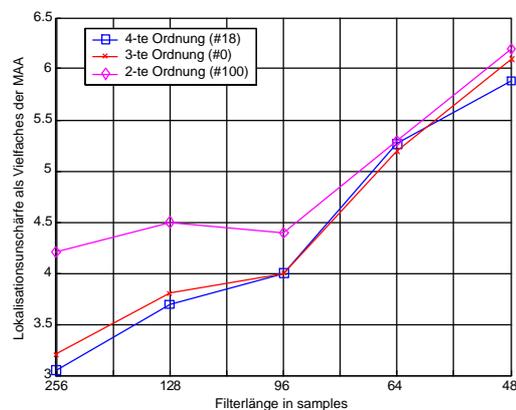


Abb. 5.10: Lokalisationsunschärfe verschiedener Filterlängen für unterschiedliche Ambisonic-Ordnungen bei optimaler Dekodierung.

5.2.4 Gewichtung der höchsten Ordnung

Bei Gewichtung der höchsten Ambisonic-Ordnung bleibt die grundlegende Tendenz erhalten, dass kürzere Filter größere Lokalisationsfehler produzieren (vgl. Abb. 5.11 und 5.12). Jedoch macht sich vor allem bei längeren Filtern der Effekt der Verbreiterung der Hauptkeule bemerkbar was zu einem Anstieg der Lokalisationsunschärfe führt. Durch die damit verbundene größere Dämpfung der Nebenkeulen verringert sich der Lokalisationsfehler.

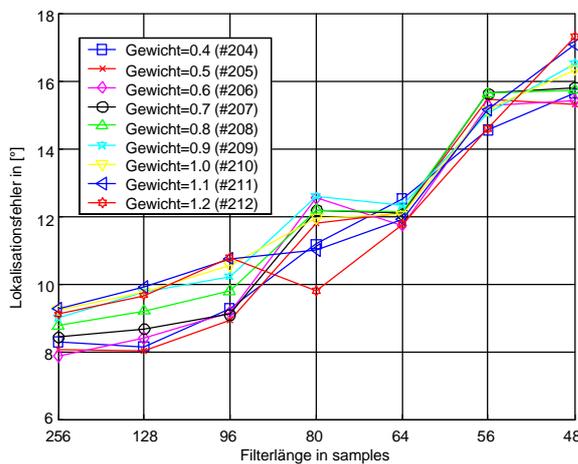


Abb. 5.11: Lokalisationsfehler bei unterschiedlicher Gewichtung der 3. Ordnung für verschiedene Filterlängen.

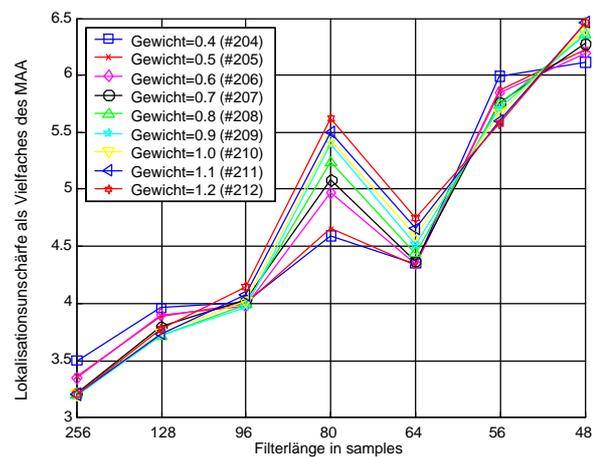


Abb. 5.12: Lokalisationsunschärfe bei unterschiedlicher Gewichtung der 3. Ordnung für verschiedene Filterlängen.

5.2.5 Systeme 3. und 4. Ordnung mit Kaiserfenster

Durch Verwendung verschiedener Fenstertechniken lässt sich eine weitere Verbesserung der Lokalisationseigenschaften erreichen. In Abb. 5.13 ist ersichtlich, dass die Gewichtung der 3. Ordnung mit 0.4 zu einer Verschlechterung des Lokalisationsfehlers führt. Eine geringe Gewichtung der 3. Ordnung entspricht annähernd einem voll gewichtetem System 2. Ordnung. Um dieses Problem zu umgehen, wird nicht nur die 3. Ordnung gewichtet, sondern eine Fensterfunktion über alle Ordnungen verwendet. Aus der signaltheoretischen Betrachtung liefert die Verwendung eines Kaiserfensters für ein System 3. Ordnung bessere Werte als ein voll gewichtetes System 4. Ordnung (siehe Abb. 5.13). Abbildung 5.15 zeigt den Lokalisationsfehler bei einer Filterlänge von 128 Samples. Es ist gut ersichtlich, dass der Lokalisationsfehler eines Systems 3. Ordnung im vorderen Bereich annähernd einem System 4. Ordnung gleichkommt. Bei 90° tritt eine Verkleinerung des schwer lokalisierbaren Bereiches ein. Die Gewichtung des Systems 4. Ordnung mit einem Kaiserfenster führt zu einem kleineren Lokalisationsfehler über den gesamten Bereich. Abbildung 5.16 zeigt für Systeme 4. Ordnung keine wesentliche Verschlechterung der Lokalisationsunschärfe durch die Fensterung.

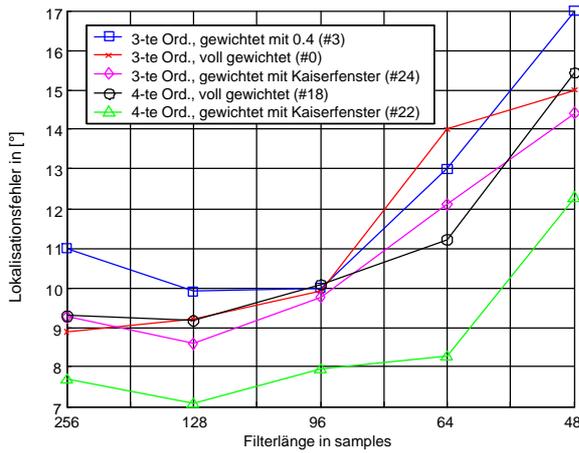


Abb. 5.13: Lokalisationsfehler bei Gewichtung der höchsten Ordnung und Kaiserfenster für verschiedene Filterlängen.

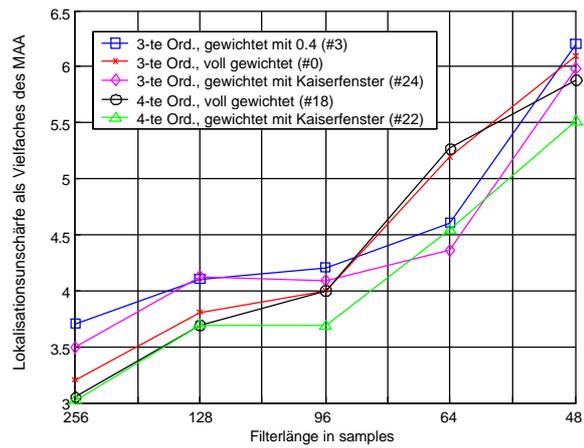


Abb. 5.14: Lokalisationsunschärfe bei Gewichtung der höchsten Ordnung und Kaiserfenster.

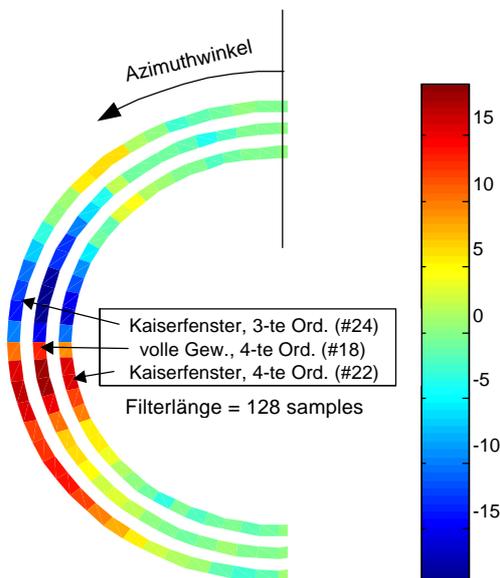


Abb. 5.15: Lokalisationsfehler bei Gewichtung der höchsten Ordnung und Kaiserfenster für eine Filterlänge von 128 Samples für verschiedene Filterlängen.

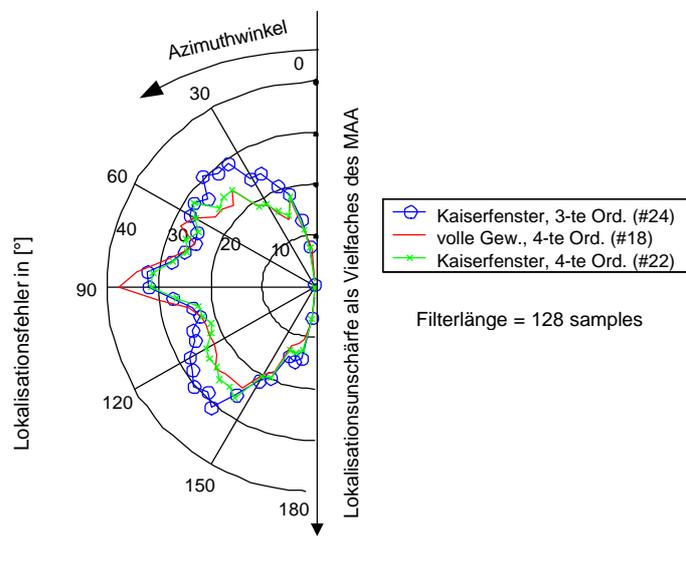


Abb. 5.16: Lokalisationsunschärfe bei Gewichtung der höchsten Ordnung und Kaiserfenster für eine Filterlänge von 128 Samples für verschiedene Filterlängen.

5.2.6 Erweiterungsmodul für das objektive Fehlermaß

Das mathematische Modell zur objektiven Evaluierung der Lokalisationsfunktion und der Lokalisationsunschärfe liefert Werte, die sich mit den Ergebnissen der subjektiven Untersuchung weitgehend decken. Aussagen über die Klangtreue bzw. die Wahrnehmbarkeit von richtungsabhängigen Klangfärbungen können aus dem Fehlermaß allerdings nicht gewonnen werden, sondern erst durch den Hörversuch evaluiert werden.

In [Pulkki, Karjalainen, 2001a, 2001b] wird eine Modell zur objektiven Beurteilung der Lokalisation und der Klangqualität von Phantomschallquellen beschrieben, dass auf der auditiven Verarbeitung von binauralen Schallsignalen basiert. Bei diesem *Auditory Model*

(vgl. auch [Pulkki et al., 1999], [Pulkki, 1999] werden die gehörrelevanten Verarbeitungsstufen - d.h. physiologische und neuronale Komponenten – weitgehend berücksichtigt. Die Frequenzselektivität des Innenohrs wird durch eine Gammaton-Filterbank modelliert, die die Simultanmaskierung nachbildet (vgl. [Zwicker, Fastl, 1990]). Für breitbandige Eingangssignale (z.B. rosa Rauschen) bildet das Modell die natürliche Hörwahrnehmung ausreichend nach. Da jedoch durch die Maskierung schmalbandige Einbrüche in der Außenohrübertragungsfunktion, die für die Beurteilung der Wiedergabenatürlichkeit besonders von elevierten Schallquellen von zentraler Bedeutung sind [Blauert, 1969], [Møller et al., 1995], geglättet werden, muss das zitierte Modell hier versagen.

Zur objektiven Beurteilung von Klangfärbungen kann das bestehende Modell aus 5.1 um die Berechnung der Composite Loudness Levels (CLL) erweitert werden. Diese erfolgt über eine Approximation der Lautheitspektren (vgl. [Zwicker, Fastl, 1990]), die für das linke und rechte Ohr addiert werden und so das CLL-Spektrum eines jeweils mit der Referenz- und der Systemübertragungsfunktion gefilterter Quellsignals ergeben. Als Grundlage für die Beurteilung von Klangfärbungen kann die Unterschiedsschwelle (JND, just noticeable difference) für die Lautheit (hier in CLL) verwendet werden. Unterscheiden sich die normierten Lautheitspektren zwischen realer und virtueller Schallquelle um mehr als 1 phon (dB), so ist dies ein Hinweis auf eine Klangfärbung (vgl. sinngemäß [Pulkki et al., 1999]).

5.3 Hörversuche zu den ausgewählten Systemen

Im folgenden wird versucht die auf dem Fehlermaß basierende Parameteroptimierung anhand von Hörversuchen auf deren Relevanz zu verifizieren. Um die Anzahl der zu untersuchenden Systeme zu reduzieren und den zeitlichen Rahmen der Hörversuche zu begrenzen, wurden die im Fehlermaß viel versprechenden Parametersettings getestet.

5.3.1 Versuchsdesign

Zur Durchführung der Hörversuche wurde das Gesamtsystem auf einem DSP implementiert, die Steuerung der Parameter erfolgte über ein Benutzerinterface durch Einbindung eines PCs (siehe [Majdak, Noisternig 2001]). Um die wahrgenommene Position der virtuellen Schallquelle (Kopfhörerdarbietung) bestimmen zu können, mussten die Versuchspersonen (Vpn¹) diese mit einer über Lautsprecher zugespielten „realen“ Schallquelle zur Deckung bringen, wobei auf die Bildung von Phantomschallquellen zwischen den Lautsprechern verzichtet wurde. Da die Vpn über diesen Sachverhalt instruiert wurden, konnte die Wahrnehmung der Position der realen Schallquelle auf einen Punkt konzentriert werden (vgl.

¹ Im folgenden wird Vpn als Abkürzung für den Begriff Versuchsperson(en) verwendet

„Räumliches Sampling“ in [Carlile et al., 1999]). Die virtuelle Schallquelle wurde mit einem Joystick bewegt. Wurde die Position der virtuellen Schallquelle als deckungsgleich mit der realen wahrgenommen, erfolgte die Bestätigung durch einen Joystick-Button. Die so ermittelte Position wurde automatisch protokolliert. Um der Vorne/Hinten Ambiguität entgegen wirken zu können, wurden Kopfdrehungen mit einem Headtracker erfasst und das Binauralsignal entsprechend gedreht. Die Vpn durften den Kopf nur in einem Bereich von $\pm 5^\circ$ bewegen, zu große Bewegungen wurde durch Anzeige auf einem Display signalisiert. Die Kopfbewegungen wurden protokolliert¹. Zur Vermeidung einer Vorne/Hinten Verwechslung bei den realen Schallquellen erfolgte die Zuspiegelung nur über Lautsprecher aus den beiden vorderen Quadranten ($0^\circ \pm 90^\circ$). Die Vpn wurden auf diesen Sachverhalt explizit hingewiesen. Damit konnte der Einfluss der Vorne/Hinten Verwechslung bei reale Quellen weitgehend ausgeschlossen werden.

Weiters wurden folgende Eigenschaften verbal bewertet:

- Klangfarbe der virtuellen Schallquelle
- Externität der Schallquelle (Außerkopflokalisation)
- Lokalisationsunschärfe
- Konsistenz der Bewegung der virtuellen Schallquellen

Der Vpn wurden die verwendeten Begriffe und deren Bedeutung erläutert. Für die Bewertung konnten die Vpn die virtuelle Schallquelle frei bewegen, wobei am Bildschirm die Quellposition angezeigt wurde. Die zu bewertenden Begriffe und eine dazugehörige Bewertungsskala wurden ebenfalls angezeigt. Während dieser Versuchsphase erfolgte keine Zuspiegelung über die Lautsprecher.

5.3.1.1 Versuchspersonen

Der Versuch wurde an 10 freiwilligen Versuchspersonen (4 weiblich, 6 männlich, Alter zwischen 20 und 40 Jahren) mit uneingeschränktem Hörvermögen durchgeführt. Von den 10 Personen hatten 6 bereits Erfahrung mit Hörversuchen. Vor der Durchführung der Versuchsreihe wurden die Vpn mit der Steuerung der virtuellen Schallquelle, sowie dem Softwareinterface (GUI, *Graphical User Interface*) vertraut gemacht. Darüber hinaus wurden die Vpn instruiert Kopfbewegungen nur in dem durch das GUI vorgegebenen Bereich durchzuführen. Dabei erfolgte kein Training, den Kopf in einer bestimmten Weise zu bewegen. Um Fehlern durch Ermüdung vorzubeugen, wurden die Versuchsreihen auf mehrere Tage verteilt durchgeführt und zwischen den unterschiedlichen Parametersettings Pausen eingehalten.

¹ Die aufgezeichneten Kopfdrehungen wurden im Rahmen dieser Untersuchungen nicht ausgewertet. Sie könnten jedoch zusätzliche Information über die Plausibilität und „Sicherheit“ bei der Auffindung einer Position geben.

5.3.1.2 Testsignale

Laut Blauert ([Blauert, 1997], S. 32 ff.) eignen sich bekannte Signale gut zur Untersuchung natürlicher Lokalisation, da monaurale Faktoren der Lokalisation nicht ausgeschlossen werden. Darüber hinaus sind typische Applikationen des Binauralsystems in der Spatialisation von Sprach- und Musiksignalen zu sehen. Ergebnisse von Hörversuchen, welche künstlich generierte Signale wie Rauschen¹ und Klicks verwenden, können laut [Begault et al., 2001] nicht mit Studien verglichen werden, die natürliche Signale verwenden. Aus diesen Gründen wurden für die Durchführung des Hörversuchs Sprachsignale gewählt. Dabei wurden jeweils eine männliche und eine weibliche Stimme als reales bzw. virtuelles Signal zugespielt, um eine gute Unterscheidbarkeit zu gewährleisten.

5.3.1.3 Versuchsaufbau und Hardware

Der grundlegende Versuchsaufbau kann der Abbildung 5.17 entnommen werden. Die zu untersuchenden Algorithmen zur Berechnung der Binauralsignale wurden auf einem DSP (Texas Instruments TMS320C6711) in Verbindung mit einem Audio-Codec (Texas Instruments TLV320AIC27) implementiert (Details finden sich bei [Majdak, Noisternig, 2001]). Den Versuchspersonen wurden die binauralen Signale über einen elektrostatischen Kopfhörer (Sennheiser HEV70) zugespielt, welcher fix mit einem Headtracker (Ascension Technology, Flock of Birds) verbunden war. Die gesamte Ablaufsteuerung sowie die Benutzeroberflächen (GUIs, *graphical user interfaces*) wurden in Visual Basic auf einem handelsüblichen PC realisiert. Bei der Aufstellung der Lautsprecher (Tannoy System 800, Verstärker: t-amp S150) wurde eine Winkelauflösung von 10° gewählt. Unter Ausnutzung der Symmetrieeigenschaften des zu untersuchenden Systems wurden die Lautsprecher wie folgt angeordnet (Abb. 5.18).

¹ Mit Rauschsignalen kann jedoch auch die Hörwahrnehmung im hohen Frequenzbereich erfasst werden. Zum Vergleich: Bei Sprachsignalen liegt die maximale Energie im Oktavband bei 500Hz, wobei zu höheren und tieferen Frequenzen ein Abfall von 6dB/Oktave stattfindet [Begault et al., 2001]

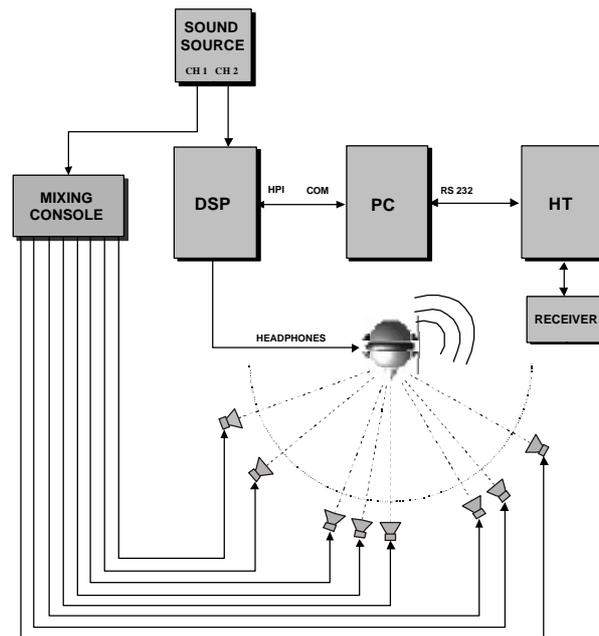


Abb. 5.17: Gesamtaufbau zur Durchführung der Hörversuche (vgl. [Majdak, Noisternig, 2002]).

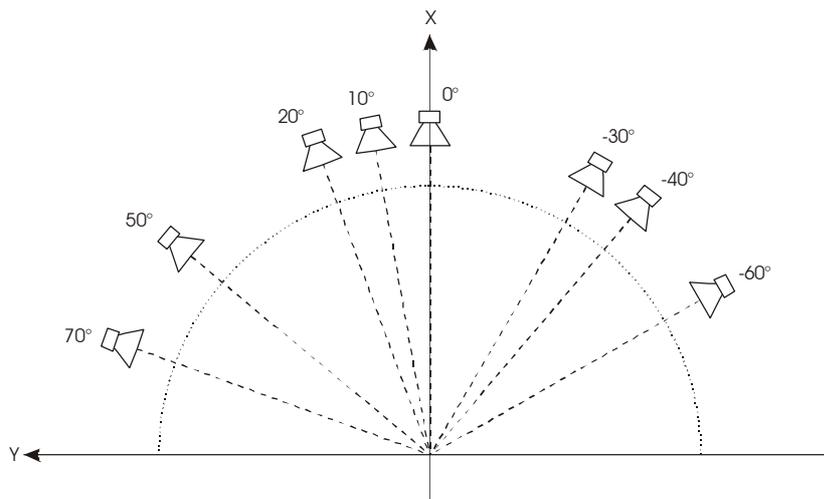


Abb. 5.18: Aufstellung der Lautsprecher mit einer Winkelauflösung von 10° unter Ausnutzung der Symmetrieeigenschaften des Systems (vgl. [Majdak, Noisternig, 2002]).

Die Durchführung erfolgte in einem Tonstudio, wobei durch geeignete zusätzliche Aufstellung variabler Schallabsorberwände (siehe Abb. 5.19) eine Nachhallzeit¹ von 0,27s erreicht wurde.

¹Die Messung erfolgte mittels einer Maximallängenfolge [Rife et al. 1989].

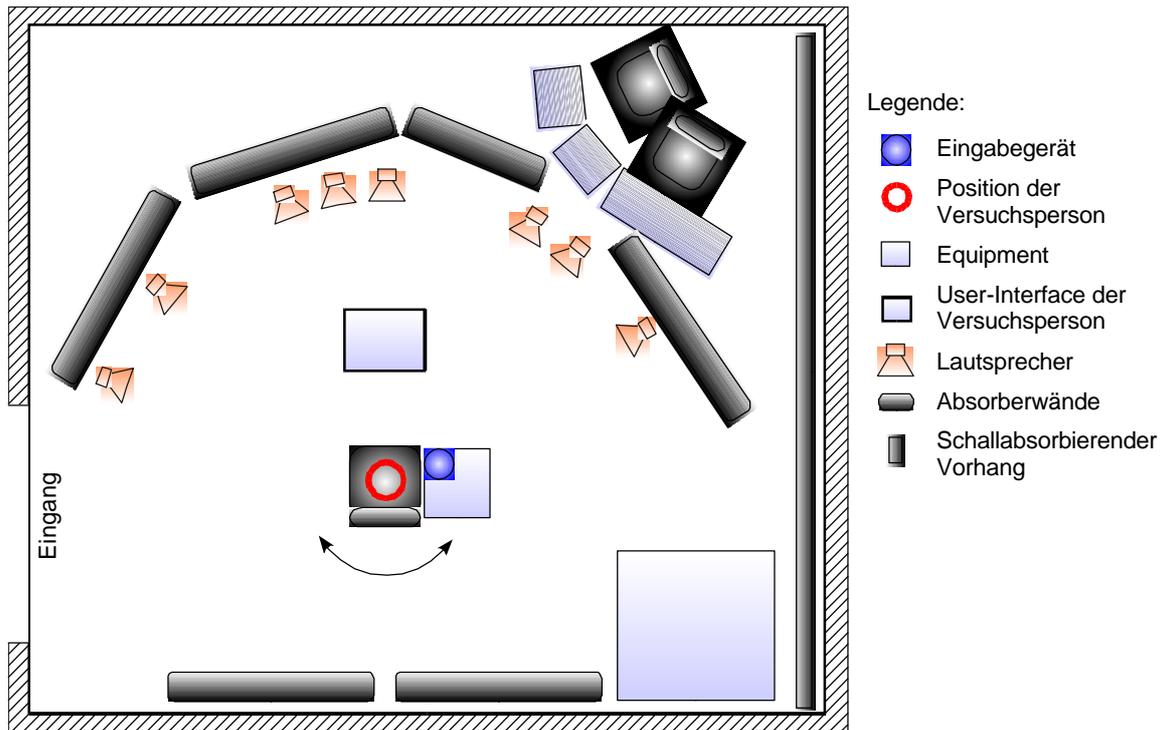


Abb. 5.19: Schematischer Aufbau und Raumadaptierung
(vgl. [Majdak, Noisternig, 2002]).

5.3.1.4 Versuchsablauf

Zu Beginn wurden die Versuchspersonen (V_{pn}) über den Gesamtprozess und die einzelnen Aufgabestellungen instruiert. Danach erfolgte eine kurze Trainings- und Eingewöhnungsphase (Interface-Handling) und eine Kalibrierung des Gesamtsystems auf die 0° Richtung. Zu Beginn der subjektiven Datenerhebung erfolgte die verbale Befragung zu den Eigenschaften: Klangfarbe, Lokalisationsunschärfe, Externität und Konsistenz der Bewegung der virtuellen Schallquellen. Für die Messung der Lokalisationsabweichung (*Offset*) zwischen virtueller und realer Quelle wurde pro Parametersetting jede Lautsprecherposition n Mal präsentiert, wobei die Abfolge Lautsprecherpositionen zufällig erfolgte und die Startposition der virtuellen Schallquelle mindestens 30° von der Position der realen Quelle entfernt war.

5.3.2 Auswertung der Hörversuche

Die Werte der Offsetmessung wurden nach folgenden Kriterien klassifiziert:

- Versuchsperson (V_{pn})
- Nummer des Settings ($SetNr$, ist dem Anhang zu entnehmen)
- Position der realen Quelle ($QPos$)
- Nummer der Messung ($VersNr$: (1 .. n), da pro Position n Messungen durchgeführt wurden)

und zu folgender vierdimensionalen Matrix zusammengefasst:

$$OS(V_{pn}, SetNr, VersNr, Qpos) \quad (5.16)$$

In der Matrix der Offsetmessung wird jene Datenpunkte markiert, die in den hinteren beiden Quadranten liegen. Da die realen Schallquellen nur aus den vorderen Quadranten zugespielt wurden, liegt bei diesen Werten eine Vorne/Hinten-Verwechslung vor. Dadurch entsteht eine vierdimensionale Ambiguitätsmatrix:

$$\mathbf{AG}(Vpn, SetNr, VersNr, QPos) \quad (5.17)$$

die Ausgangspunkt folgender Untersuchungen ist:

- Vorne/Hinten Ambiguität einer Versuchspersonen

Der Querstrich kennzeichnet jene Parameter über die gemittelt wird.

$$\mathbf{AG}(Vpn) = \mathbf{AG}(Vpn, \overline{SetNr}, \overline{VersNr}, \overline{Qpos}) \quad (5.18)$$

- Vorne/Hinten Ambiguität für die jeweiligen Quellenpositionen

$$\mathbf{AG}(Vpn, QPos) = \mathbf{AG}(Vpn, \overline{SetNr}, \overline{VersNr}, Qpos) \quad (5.19)$$

- Vorne/Hinten Ambiguität für ein Parametersetting

Es kann eine Aussage darüber getroffen werden, bei welchen Parametersetting vermehrt Vorne/Hinten Verwechslungen auftreten. Hinsichtlich der Optimierung des Systems stellt dies die wichtigste Untersuchung der Ambiguität dar.

$$\mathbf{AG}(SetNr) = \mathbf{AG}(\overline{Vpn}, \overline{SetNr}, \overline{VersNr}, \overline{Qpos}) \quad (5.20)$$

Im oben beschriebenen Vorgang zur Erzeugung der Ambiguitätsmatrix \mathbf{AG} wird gleichzeitig die ambiguitätsbefreite Offsetmatrix $\hat{\mathbf{OS}}(Vpn, SetNr, VersNr, Qpos)$ berechnet. Dabei werden alle Datenpunkte in die vorderen beiden Quadranten ($\mathbf{j} = \pm 90^\circ$) transformiert, da die Zuspiegelung der „realen“ Schallquelle über Lautsprecher nur von vorne durchgeführt wurde.

Bei der Durchführung des Hörversuches wurden für jede Quellposition n Messpunkte pro Versuchsperson und Setting aufgenommen. Wird nun für jede Position aus den n Messpunkten die Standardabweichung \mathbf{s} berechnet, entsteht aus der vierdimensionalen ambiguitätsbefreiten Matrix der Offsetmessung $\hat{\mathbf{OS}}$ die dreidimensionale Matrix der Standardabweichung \mathbf{STD} .

$$\mathbf{STD}(Vpn, SetNr, Qpos) = std \left[\hat{\mathbf{OS}}(Vpn, SteNr, VersNr, Qpos) \right] \quad (5.21)$$

Je genauer sich eine Vpn bei der Lokalisation einer virtueller Schallquelle reproduzieren konnte, desto kleiner ist die Standardabweichung. Da das Lokalisationsvermögen mit steigender Lateralisation sinkt [Blauert, 1997], stellt die Standardabweichung allein kein repräsentatives Maß dar, da sich deren Einzelwerte nicht direkt auf ein psychoakustisch plausibles Maß zusammenfassen lassen. Daher wird die Standardabweichung auf den MAA für die entsprechende Quellrichtung bezogen und ergibt die Streuparametermatrix $\mathbf{S}(Vpn, SetNr, QPos)$, deren Einzelwerte über die Quellpositionen sinnvoll gemittelt werden

können und die den Ausgangspunkt für die weiteren statistischen Untersuchungen darstellt. Ein wesentlicher Punkt für die Interpretation der Versuchsergebnisse ist die Frage nach der Signifikanz der Unterschiede zwischen den Datensätzen, die mittels der einfaktoriellen Varianzanalyse (ANOVA, *Uni-variate Analysis of Variance*) beantwortet wurde:

(1) Bewertung der Lokalisationsfähigkeit der jeweiligen Versuchspersonen

Die Berechnung der ANOVA über die Vpn ergibt, dass signifikante Lokalisationsunterschiede zwischen den einzelnen Vpn bestehen. Somit führt eine Mittelung der Streuparametermatrix über alle Quellpositionen und Settings zu einer Aussage über die durchschnittliche Lokalisationsfähigkeit einer Vpn (vgl. Abb. 5.20).

(2) Bewertung der Lokalisation der jeweiligen Parameter-Settings

Die ANOVA über die SetNr ergibt, dass kein signifikanter Unterschied zwischen allen Kombinationen von Parameter-Settings besteht. Um trotzdem Vergleiche der Settings zueinander durchführen zu können, wurde wie folgt vorgegangen:

Die Mittelung der Streuparametermatrix in (1) ergibt eine mittlere Lokalisationsunsicherheit für jede Vpn. Diese kann als Gleichanteil (*Bias*) der Streuparametermatrix aufgefasst werden. Durch Subtraktion des Gleichanteils von den zur jeweiligen Vpn gehörenden Streuparametern erhält man die vom Gleichanteil befreite Streuparametermatrix $\mathbf{S}_{unbias}(Vpn, SetNr, QPos)$. Eine Mittelung über die Vpn und QPos ergibt den Vektor

$$\bar{\mathbf{S}}_{unbias}(SetNr) = \mathbf{S}_{unbias}(\overline{Vpn}, SetNr, \overline{QPos}) \quad (5.22)$$

dessen Darstellung gemeinsam mit der Streuung über die Vpn einen Vergleich von jeweils zwei Settings erlaubt (vgl. Abb. 5.21).

5.3.3 Interpretation der Ergebnisse

Es erfolgt eine Interpretation der Hörversuchsergebnisse und ein Vergleich zwischen den theoretischen Prognosen des objektiven Fehlermaßes und den subjektiven Beurteilungen aus dem Hörversuch (vgl. [Sontacchi et al., 2002d]). Die statistische Auswertung des Hörversuches erfolgt mit Hilfe der einfaktoriellen Varianzanalyse (Analysis of Variance ANOVA). Die Hypothese der unterscheidbaren Mittelwerte in den einzelnen Gruppen (Settings bzw. Vpns) wurde auf einem Signifikanzniveau von 5% getestet. Bei den Untersuchungen sind jeweils die Ergebnisse des F-Tests F und der Irrtumswahrscheinlichkeiten p angegeben.

Beurteilung der Versuchspersonen

Der Mittelwert der Streuparameter über alle Settings und Positionen gibt Auskunft darüber, wie sicher eine Vpn lokalisiert. Aus der Abb. 5.20 ist zu entnehmen, dass Vpn 7 große Probleme bei der Lokalisierung hat. Vpn 2 weist hingegen eine sehr sichere Lokalisation auf, was durch die größere Erfahrung von Vpn 2 mit binauralen Wiedergabesystemen erklärt werden kann. Die konkrete Auswirkung der Hörerfahrung auf die Lokalisationszuverlässigkeit müsste getrennt untersucht werden. Um in weiterer Folge Vergleiche zwischen den Settings durchführen zu können, dürfen die Streuparameter nicht von der unterschiedlichen Lokalisationsfähigkeit der jeweiligen Vpn beeinflusst werden. Es erfolgt die oben beschriebene Mittelwertbefreiung.

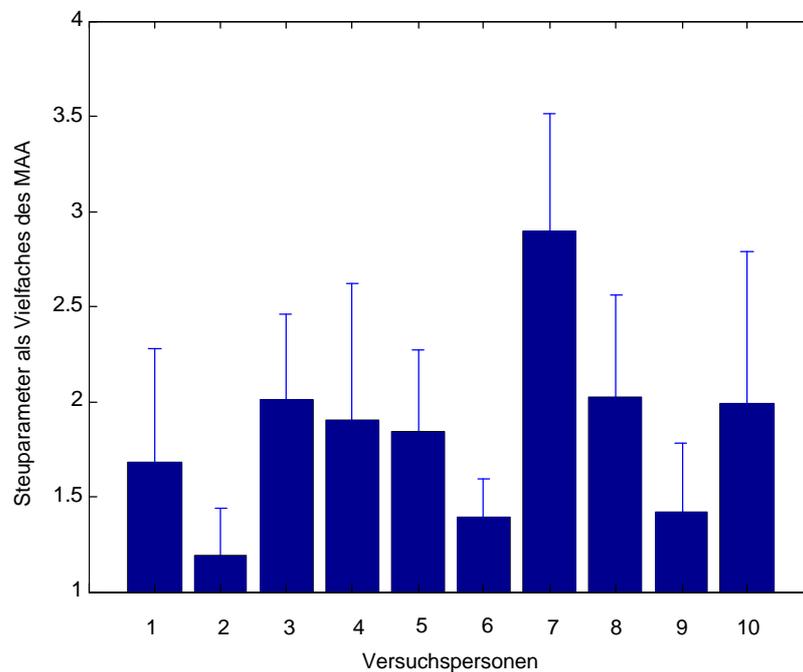


Abb. 5.20: Mittelwert des Streuparameters für die jeweiligen Versuchspersonen (Mittelwert und Standardabweichung über die Settings).

Unterschiede zwischen den Settings

Die über die Vpn gemittelten Streuparameter für jedes Setting zeigt die Abbildung 5.21, aus der keine globalen Aussagen über die unterschiedliche Qualität der Settings gewonnen werden können. Ein geringer Mittelwert (kleiner Streuparameter) bei einem Setting bedeutet zuverlässiges Lokalisieren durch die Vpn. Die Standardabweichung in der Darstellung ist ein Maß für die Personenabhängigkeit des jeweiligen Settings. Es ist somit deutlich zu erkennen, dass die beiden Referenz-HRIRs (Kemar, AKG) unterschiedliche Personenabhängigkeit aufweisen.

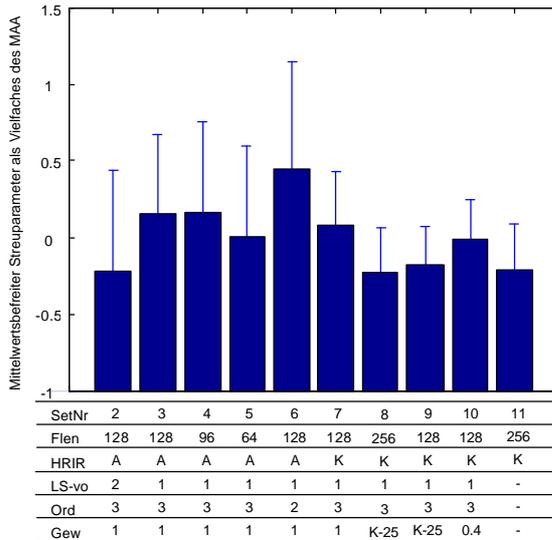


Abb. 5.21: Mittelwertsbefreiter Streuparameter für die jeweiligen Settings (Mittelwert und Standardabweichung)

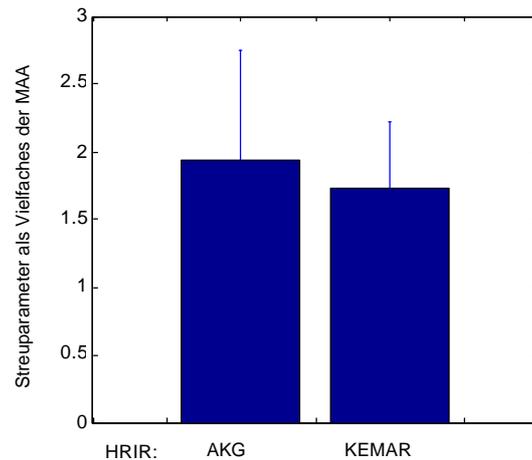


Abb. 5.22: Lokalisationsunsicherheit für verschiedene HRIR-Kurven als Mittelwert der Streuparameter (Mittelwert und Standardabweichung)

Bei der Untersuchung des Einflusses der jeweiligen HRIR-Kurven¹ auf die Lokalisationsunsicherheit zeigt sich ein signifikanter Unterschied ($F_{(1,98)} = 4.94$, $p=0.029$). Die Kemar-HRIRs weisen eine höhere Lokalisationssicherheit auf (Abb.5.22).

Weiters wurde der Einfluss der Ambisonic-Ordnung auf die Lokalisationssicherheit untersucht (Abb. 5.23). Die Unterschiede zwischen Systemen 2. und 3. Ordnung bestätigen zwar die theoretischen Betrachtungen, dass eine geringere Ordnung eine höhere Lokalisationsunsicherheit nach sich zieht, sie sind jedoch nicht ausreichend signifikant ($F_{(1,48)} = 3.72$, $p=0.06$).

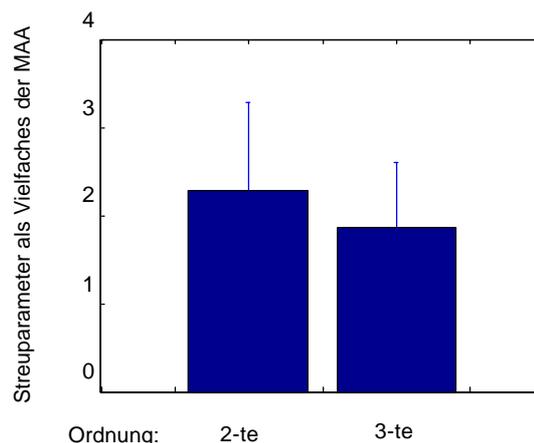


Abb. 5.23: Lokalisationsunsicherheit für verschiedene Ambisonic -Ordnungen als Mittelwert der Streuparameter (Mittelwert und Standardabweichung).

Der Einfluss der Gewichtung der Ambisonic-Ordnungen (Fensterung) auf die Lokalisationsunsicherheit weist hingegen signifikante Unterschiede auf ($F_{(1,38)} = 7.61$, $p=0.012$). Die Verwendung eines Kaiserfensters erbringt eine eindeutige Verbesserung.

¹Vergleich der Settings (7, 8, 9, 10, 11) mit (1, 2, 3, 4, 5, 6)

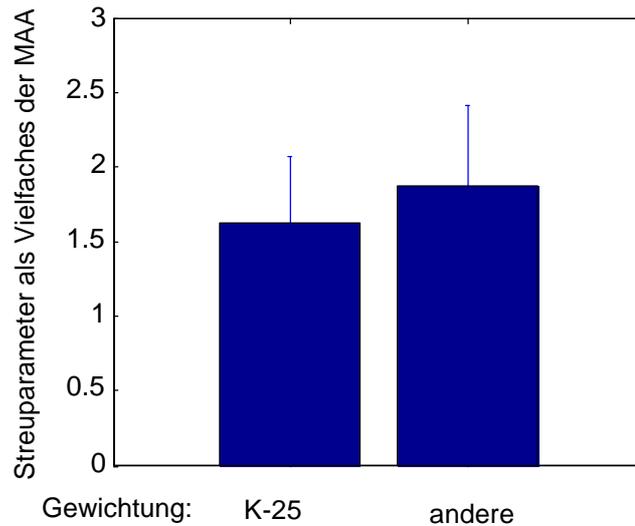


Abb. 5.24: Lokalisationsunsicherheit für verschiedene Gewichtungen der Ambisonic-Ordnungen (Fensterung) als Mittelwert der Streuparameter (Mittelwert und Standardabweichung).

Vorne/Hinten Ambiguität

Wie Abbildung 5.25 zu entnehmen ist, weist die Vorne/Hinten Ambiguität eine starke Personenabhängigkeit auf. Vpn 7 sowie Vpn 3 hatten große Probleme bei der Unterscheidung zwischen vorne und hinten. Die Vorne/Hinten Ambiguität für die verschiedenen Lautsprecherpositionen ist in Abb. 5.26 dargestellt. Je größer die Lateralisation der virtuellen Schallquelle ist, desto schwieriger wird eine zuverlässige Vorne/Hinten Unterscheidung. Dies lässt sich auf die zur Seite hin steigende Lokalisationsunschärfe zurückführen.

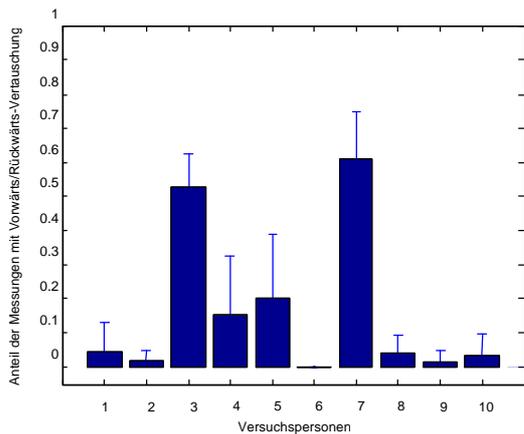


Abb. 5.25: Vorne/Hinten Ambiguität der Versuchspersonen.

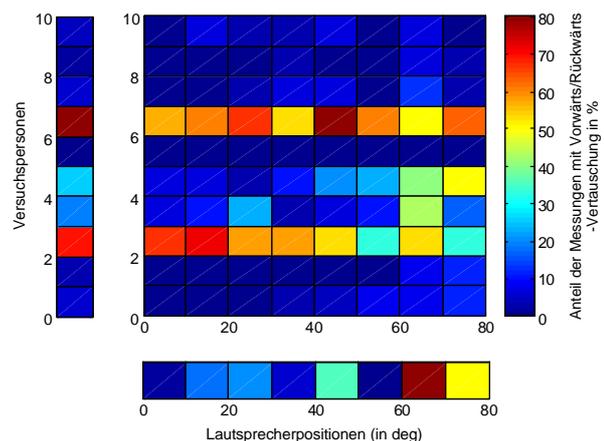


Abb. 5.26: Vorne/Hinten Ambiguität für die Lautsprecherposition und die Versuchspersonen.

Die Untersuchung der Vorne/Hinten Ambiguität bezüglich der Settings ergibt eine erhöhte Vertauschungswahrscheinlichkeit bei Systemen 2. Ordnung. Bei Gewichtung der 3. Ordnung mit 0.4 ergibt sich ebenfalls eine erhöhte Verwechslung, was sich darauf zurückführen lässt, dass diese geringe Gewichtung annähernd eine Reduzierung der Ordnung des Gesamtsystems auf 2 bedeutet.

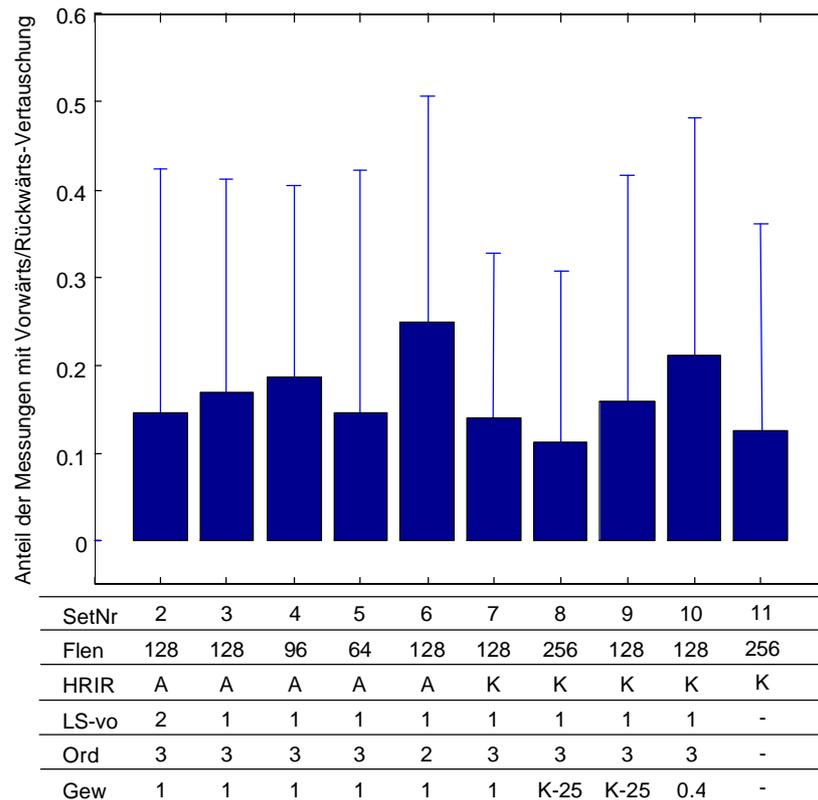


Abb. 5.27: Vorne/Hinten Ambiguität für verschiedene Settings (Mittelwerte und Standardabweichung).

Auswertung der verbalen Beschreibung

Um die Ergebnisse der verbalen Befragung statistisch verarbeiten zu können, mussten den Aussagen der Versuchspersonen Zahlenwerte zugewiesen werden (siehe Anhang).

- **Lokalisationsunschärfe**

Bei Verwendung eines Kaiserfensters ist eine Vergrößerung der Lokalisationsunschärfe erkennbar. Dies lässt sich durch die verbreiterte Hauptkeule der sinc-Funktion erklären. Die geringste Lokalisationsunschärfe entsteht bei direkter Implementierung der Referenz HRIRs, da keine Verwischung der virtuellen Schallquelle durch das Ambisonic-System entsteht. Erwartungsgemäß liegt die Lokalisationsunschärfe von Ambisonic-Systemen 3. Ordnung zwischen der direkten Implementation und Systemen 2. Ordnung.

- **Konsistenz der Bewegung**

Die größte Inkonsistenz weist die direkte Implementierung der AKG-HRIRs ohne Interpolation zwischen den diskreten Richtungen (15° Auflösung) auf. Dieses Setting wurde bewusst in die Versuchsreihe aufgenommen, um die Aufmerksamkeit der Vpn zu testen. Die Inkonsistenz der Bewegung wurde von allen Vpn richtig erkannt.

Die Verkürzung der Filterlänge führt zu einer signifikanten Verschlechterung der Konsistenz der Bewegung. Die direkte Implementierung der Kemar-HRIRs liefert zufriedenstellende Ergebnisse, d.h. die Winkelauflösung von 5° scheint weitgehend ausreichend. Weiters wurde festgestellt, dass durch die Fensterung der Ambisonic-Ordnungen die Konsistenz der Bewegung erhöht werden kann. Die breitere Hauptkeule führt zu einer Aufweitung der Quelle und somit zu einer als „weich“ wahrgenommenen Bewegung.

- **Externität**

Bei geringer Lateralisation wandert die virtuelle Schallquelle in den Kopf. Es entsteht eine elliptische Bewegungsbahn. Durch Fensterung wird der Externitätsunterschied zwischen vorne und seitlich geringer. Die bewegte Quelle beschreibt eine annähernd kreisförmige Bahn. Die größte Externität wurde bei der direkten Implementierung der Kemar-HRIRs erzielt. Das Ambisonic-System verschlechtert somit die Außerkopflokalisation der virtuellen Quelle.

- **Klangfarbe**

Die Bewertung der Klangfarbe erfolgte in der Gegenüberstellung und Ausgewogenheit von Bässen, Mitten und Höhen. Kurze Filterlängen ergaben eine Dämpfung des Bassbereiches. Es konnte nachgewiesen werden, dass starke Bässe grundsätzlich als wesentlich realistischer empfunden werden.

Zusammenfassung 6

In dieser Arbeit werden Lautsprecher- und Kopfhöreranwendungen für eine dreidimensionale Schallfeldreproduktion vorgestellt. Diese Anwendungen basieren auf der erweiterten Ambisonic Theorie und der Wellenfeldsynthese. Während Ambisonic als ein mathematisch motiviertes Verfahren bezeichnet werden kann, handelt es sich bei der Wellenfeldsynthese um eine physikalische Modellierung des Schallfeldes. Die beiden Beschreibungsformen sind äquivalent und können unter sehr allgemeinen Bedingungen ineinander übergeführt werden. Sie weisen abhängig vom Anwendungsbereich unterschiedliche Vorzüge und Mängel auf, die eingehend diskutiert werden.

Der Ambisonic Ansatz basiert auf einer richtungsabhängigen Kodierung und Dekodierung eines Schallfeldes auf ein N-dimensionales orthogonales Set von Basisvektoren. Die Basisvektoren werden im dreidimensionalen Fall durch die Kugelflächenfunktionen¹ beschrieben. Durch diese Kodierungsvorschrift kann eine Schallquelle bzw. ihr dazugehöriges Schallfeld wie ein grafisches Objekt behandelt werden. Rotationsmatrizen² und räumliche Filterfunktionen können dadurch gezielt eingesetzt werden, um das gewünschte Schallfeld zu synthetisieren. Der Wiedergabebereich bei der Synthese wird abhängig vom Wiedergabebereich (mit $|r| < x_0$) und der Signalbandbreite (mit $k < k_0$) durch die verwendete Systemordnung M bestimmt. Für die Approximation $M = \lceil k_0 x_0 \rceil$ ergibt sich im definierten Bereich ein relativer Amplitudenfehler von 4%. Durch geeignete Wahl der Lautsprecherpositionen und Gewichtung der Ambisonic-Signale kann der Wiedergabebereich bzgl. Raum und Frequenz optimiert verteilt werden. Eine Steigerung der Systemordnung impliziert eine Zunahme der richtungsabhängigen Komponenten, der Übertragungskanäle und somit auch der erforderlichen Anzahl von Lautsprechern.

Objektive Schallfeldindikatoren wie die mittlere Geschwindigkeit der gerichteten Wirkleistung (*Active Velocity*) können zur Beurteilung der Wiedergabeeigenschaften des Ambisonic Systems herangezogen werden. Die allgemeinen Eigenschaften unterschiedlicher Dekodierungsvorschriften (vor allem der *window applied decoder*) können dadurch in beliebigen Raumpunkten beschrieben und beurteilt werden.

Da es sich bei Ambisonic um ein koinzidentes Aufnahme- (Kodierungs-) verfahren handelt, komplexe Mikrofoncharakteristiken aber nur begrenzt existieren, ist dieses Verfahren für Aufnahmen von realen Schallfelder für höhere Ordnungen nur begrenzt einsetzbar³. Bei synthetisch erzeugten Schallfelder existiert theoretisch keine Grenze. Im Falle einer Kopfhörerwiedergabe ergeben sich Grenzen nur bedingt durch die Rechenleistung. Hingegen

¹ Eine Kombination von normalisierten Legendre Polynomen und der Fourier Reihe, die sich aus der Lösung der homogenen Wellengleichung für die Entwicklung im Koordinatenursprung ergibt (engl. spherical harmonics). Diese Lösung reduziert sich im zweidimensionalen auf die Fourier Reihe.

² Beliebige Schallfelder können gedreht, gekippt oder geneigt werden.

³ Auf die Realisierung von Mikrofonarrays zur Generierung von Charakteristika höherer Ordnung wird hier nicht näher eingegangen.

ergibt sich bei Lautsprecherwiedergabe eine zusätzlich Einschränkung durch die Anzahl der Übertragungskanäle und der physikalisch vorhandenen Lautsprecher.

Die Wellenfeldsynthese basiert auf dem Huygens'schen Prinzip, welches besagt, dass ein Punkt einer Wellenfront als Ausgangspunkt einer Elementarwelle betrachtet werden kann, die sich infolge der Ausbreitung mit benachbarten Elementarwellen zu einer resultierenden fortschreitenden Wellenfront überlagert. Mathematisch wird dies durch das Kirchhoff-Helmholtz Integral beschrieben. Es werden die theoretischen Grundlagen der Wellenfeldsynthese für beliebige Wiedergabeanordnungen vorgestellt. Mit einer geeigneten Wahl von verteilten sekundären Quellen, die durch einen Syntheseoperator angetrieben werden, kann das Schallfeld einer primären Quelle eindeutig innerhalb eines definierten Bereiches reproduziert werden. Der Syntheseoperator berechnet sich aus der Geometrie der Wiedergabeanordnung sowie aus der Position der primären Quelle und der Senke.

Eine Lautsprecherapplikation basierend auf dem Ambisonic Verfahren wird anhand der Implementierung des IEM CUBE (Computer Unterstützte BeschallungsEinheit) vorgestellt. Die vorgestellte Realisierung ermöglicht eine dreidimensionale Wiedergabe 5.Ordnung für eine beliebige Anzahl von virtuellen Quellen. Die gängigen standardisierten Wiedergabeformate können sowohl diskret wie auch virtuell realisiert werden. Neben der Live-Verstärkung von natürlichen Quellen und der Wiedergabe von gespeicherten Audiodaten können raumakustisch relevante Parameter gezielt verändert werden. Die Lautsprecher sind auf einer Halbkugel über dem Publikum angeordnet. Die Wiedergabe erfolgt somit in der oberen Hemisphäre, d.h. Schallquellen von unten können nicht realisiert werden. Anhand der eingeführten Schallfeldindikatoren wird der IEM CUBE stellvertretend für einen Parametersatz untersucht.

Zur Erzeugung einer virtuellen Akustik müssen die charakteristischen Eigenschaften des vorhandenen physikalischen Raumes ausgeblendet werden. Dazu wird ein neuer Ansatz für die Kompensation von Raumreflexionen basierend auf dem erweiterten Ambisonic Ansatz und den gemessenen direktionalen Raumimpulsantworten vorgestellt. Der Vorschlag beruht darauf, dass jede Reflexion als eine Schallquelle interpretiert werden kann. Da diese wie eine virtuelle Quelle synthetisiert werden können, ist es auch möglich, durch entsprechende Wahl der Ambisonic Signale Komponenten zu generieren, die diese Reflexionen in Summe eliminieren. Der Vorteil dieser Lösungsmethode beruht auf einer festen Filteranzahl, die unabhängig von der Anzahl der Lautsprecher ist. In diesem neuen Kontext werden zusätzlich bereits bekannte Lösungswege für zeitinvariante und zeitvariante Kompensation von frühen Reflexionen im Zeit- und Frequenzbereich vorgestellt. Für die Berechnung der Kompensationsfilter ist die direkte Lösung im Frequenzbereich der Lösung im Zeitbereich aufgrund der Recheneffizienz vorzuziehen, es sei denn die Filter werden adaptiv berechnet. Allgemein ist zu bemerken, dass alle vorgestellten Lösungsansätze nur für stark höhenbedämpfte Räume wirklich realistisch wirksame Ergebnisse erzielen lassen.

Als ein weiteres Novum gilt die Einbindung der Distanzkodierung in ein Ambisonic basiertes Gesamtsystem. Die Distanzkodierung wird durch die Verbindung von Ambisonic mit einem Derivat der Wellenfeldsynthese realisiert. Das vorgestellte Modell ist skalierbar, und die Vorteile der Ambisonic Kodierung bleiben weitgehend erhalten. Für zufriedenstellende Ergebnisse muss bei der praktischen Realisierung entsprechend großer Aufwand bzgl. der Anzahl der Wiedergabelautsprecher geleistet werden.

Des Weiteren wird die Implementierung und Erweiterung einer umfassenden binauralen Kopfhörerapplikation vorgestellt. Die Berechnung eines binauralen Kopfhörersignals basierend auf Ambisonic wird für den 2D-Fall vorgestellt. Im Gegensatz zu herkömmlichen Systemen wird hier zur Berücksichtigung der Positionsänderung der Quellen und der Orientierung des Hörers nicht die Außenohrübertragungsfunktionen interpoliert, sondern das Schallfeld verändert. Dadurch kann der üblicherweise hohe Berechnungsaufwand und Speicherbedarf ohne Einbußen der Wiedergabequalität stark reduziert werden. Abhängig von der Schallfeldkodierung und der Dekodierung entstehen Abbildungsfehler, die sowohl objektiv als auch subjektiv untersucht werden.

Ein Fehlermaß zur objektiven Beschreibung der Wiedergabeeigenschaften bezüglich der Lokalisation und der Lokalisationsunschärfe wird vorgestellt, das auf der psychoakustisch bewerteten Abweichung der durch das Wiedergabesystem verfälschten Außenohrimpulsantworten (Head Related Impuls Responses, HRIRs) zu den direkt gemessenen HRIRs (Referenzkurven) basiert. Untersuchungen der perceptiven Unterschiede zwischen unterschiedlichen Parametersätzen werden anhand von Hörversuchen durchgeführt. Mit dem objektiven Maß kann das subjektive Empfinden im Allgemeinen gut nachgebildet werden. Als Ergebnis der Studie kann berichtet werden, dass ein System 4. Ordnung mit Kaiserfenster Gewichtung (*window-applied decoder*) weitgehend gleichwertig mit der direkten Verwendung der Referenz HRIRs bewertet wird. Durch eine einfache Verhallung des fertigen Binauralsignals kann der Gesamteindruck noch zusätzlich verbessert werden. Die Vorgangsweise und die Ergebnisse dieser Studie werden hier zusammengefasst vorgestellt. Die Bewertung der Klangqualität konnte erst durch den Hörversuch erfolgen. Vorschläge zur objektiven Beurteilung der Wahrnehmbarkeit von Klangfärbungen werden anhand eines erweiterten Modells vorgestellt.

Literaturverzeichnis

- [Ahnert, 1975], Ahnert, W.: „Einsatz elektroakustischer Hilfsmittel zur Räumlichkeitssteigerung, Schallverstärkung und Vermeidung der akustischen Rückkopplung“, Dissertation, Technische Universität Dresden, 1975.
- [Ahnert, Steffen, 1993], Ahnert, W., Steffen, F.: „Beschallungstechnik – Grundlagen und Praxis“, S. Hirzel Verlag, Leipzig, 1993.
- [Ahnert, Reichardt, 1981], Ahnert, W.; Reichardt, W.: „Grundlagen der Beschallungstechnik“, S. Hirzel Verlag, Stuttgart 1981. ISBN 3-7776-0373-2.
- [Allen, Berkley, 1979], Allen, J.B.; Berkley, D.A.: „Image Method for Efficiently Simulating Small-Room Acoustics“, J. Acoust. Soc. Am., Vol. 65, pp. 943-950, 1979.
- [Atal, Schroeder, 1966], Atal, B. S.; Schroeder M.: „Apparent sound source translator“, United States Patent No. 3,236,949 (22 Februar 1966).
- [Bamford, 1995], Bamford, J.S.: „An Analysis of Ambisonic Sound Systems of First and Second Order“, Thesis presented to the University of Waterloo, Waterloo, Ontario, Canada; 1995.
- [Bronstein et al., 1999], Bronstein, I.N., Semendjajew, K.A.; Musiol G., Mühlig H.: „Taschenbuch der Mathematik“, 4. überarb. U. erw. Auflage d. Neubearbeitung, Verlag Hari Deutsch, 1999.
- [Batteau, 1967], Batteau, D.W.: „The rule of the pinna in human localization“. Proc. R. Soc. Lond. B. Vol.168, pp.158-180, 1967.
- [Begault, 1994], Begault, D.R.: „3-D Sound for Virtual Reality and Multimedia“, Academic Pree Professional, ISBN 0-12-084735-3, 1994.
- [Begault et al., 2001], Begault, D.R.; Wenzel, E.M.; Anderson, M.R.: „Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions of Spatial Perception of a Virtual Speech Source.“, J. Audio Eng. Soc. Vol. 49, No. 10, October, 2001.
- [Beranek, 1996], Beranek: „Concert and Opera Halls: how they sound“, Published by the American Institute of Physics, 1996.
- [Berkhout, 1982], Berkhout A.J.: „Seismic Migration: imaging of acoustics energy by wave field extrapolation. A: Theoretical Aspects“. Elsevier, Amsterdam, Niederlande, 1982.
- [Berkhout, 1988], Berkhout, A.J.: „Holographic approach to acoustic control“, J. Audio Eng. Soc., Vol. 36, pp.977-995, 1988.
- [Berkhout et al., 1993], Berkhout, A.J.; de Vries, D.; Vogel, P.: „Acoustic control by wave field synthesis“, J. Acoust. Soc. Am., Vol. 93, No. 5, p.2764, 1993.
- [Blauert, 1969], Blauert, J.: „Sound localization in the median plane“, Acoustica 22, S. 205-213, 1969.
- [Blauert, 1974], Blauert, J.: „Räumliches Hören“, S. Hirzel Verlag, 1974.
- [Blauert, 1997], Blauert, J.: „Spatial Hearing“, 2nd ed., MIT Press, Cambridge, MA, 1997.
- [Bleistein, 1984], Bleistein N.: „Mathematical Methods for Wave Phenomena“. Acad. Press, New York, USA, 1984.
- [Blumlein, 1931], Blumlein, A.D.: British patent 394,325 (field 1931 Dec.,14), reprinted in J. Audio. Eng. Soc. Vol.6 No. 2, April, 1958.
- [Boone, 2001], Boone, M.: „Acoustic Rendering with Wave Field Synthesis“, ACM Siggraph and Eurographics Campfire: Acoustic Rendering for Virtual Environments, Snowbird, Utah, May 2001.
- [Borish, 1984], Borish, J.: „Extension of the Image Model to Arbitrary Polyhedra“, J. Acoust. Soc. Am., Vol. 75, pp. 1827-1836, 1984.
- [Bortz, 1993], Bortz, J.: „Statistik“, Springer Verlag, 1993.
- [Bosun, 2001], Bosun, X.: „Signal Mixing for a 5.1-Channel Surround Sound System – Analysis and Experiment“, J. Audio Eng. Soc., Vol. 49, No. 4, April, 2001.
- [Bouchard, Quednau, 2000], Bouchard, M.; Quednau, S.: „Multichannel Recursive Least-Squares Algorithms and Fast Transversal Filter Algorithms for Active Noise Control and Sound Reproduction Systems“, IEEE Transactions on Speech and Audio Processing, Vol.8, No.5, pp.606-618, September, 2000.
- [Botteldooren, 1995], Botteldooren, D.: „Finite-Difference Time-Domain Simulation of Room Acoustic Problems“, J. Acoust. Soc. Am., Vol. 98, pp. 3302-3308, 1995.
- [Brandenstein, Unbehauen, 1992], Brandenstein, H.; Unbehauen, R.: „Least-Squares Approximation of FIR by IIR Digital Filters“, IEEE Transaction on Signal Processing, Vol. 46, No. 1, pp. 21-30, 1992.
- [Braun, 1996], Braun, D.: „Ein neues Hörversuchskonzept“, DAGA'96, 1996.
- [Brown, Duda, 1998], Brown, C.P.; Duda, R. O.: „A Structural Model for Binaural Sound Synthesis“, IEEE Transactions on Speech and Audio Processing, Vol. 6 No. 5, Sept., 1998.

- [Burkhard, 1963], Burkhard, M.D.: „A simplified frequency shifter for improving acoustic feedback stability.“, J. Audio Eng. Soc., Vol. 11, pp.234-237, 1963.
- [Carlile et al., 1999], Carlile, S.; Delaney, S.; Corderoy, A.: „The localisation of spectrally restricted sounds by human listeners“, Hear. Res., Vol. 128, pp 175-189, 1999.
- [Chen et al., 1995], Chen, J. B. ; Van Veen, B. D. ; and Hecox, K.E.: „A spatial feature extraction and regularization model for the head-related transfer function“, J. Acoust. Soc. Am., Vol. 97, pp. 439-452, January, 1995.
- [Cheng, Wakefield, 2001], Cheng, C.I.; Wakefield, G.H.: „Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space.“, J. Audio Eng. Soc., Vol. 49, No. 4, April, 2001.
- [Cremer, Müller, 1978], Cremer, L.; Müller, H.A.: „Die wissenschaftlichen Grundlagen der Raumakustik, Band 1“, Hirzel, Stuttgart, 1978.
- [Curtis, 1969], Curtis, G.J.: „Acoustic feedback as a modal investigation tool.“, J. Sound Vibr., Vol. 9, No.1, pp.101-120, 1969.
- [Damaske, 1967], Damaske, P.: „Subjective Investigation of Sound Fields“, Acustica, Vol.19, pp.199-213, 1967/68.
- [Daniel, et al., 1998], Daniel, J.; Rault, J.B.; and Polack, J.D.: „Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions“, AES ???, 1998.
- [Daniel et a., 1999], Daniel, J. ; Rault, J.B. ; Polack, J.D.: „Acoustic Properties and Perceptive Implications of Stereophonic Phenomena“, in Proc. AES 16th Int. Conf. On Spatial Sound Reproduction, pp. 91-102, 1999.
- [Daniel, 2000], Daniel, J.: „Acoustic field representation, application to the transmission and the reproduction of complex sound environments in a multimedia context“ (engl. translation), Thesis, 2000.
- [Dickins et al.,1999], Dickins, G. ; Flax, M. ; McKeag, A., and McGrath, D.: „Optimal 3D Speaker Panning“, 16th AES International Conference on Spatial Sound Reproduction, Finland, 1999.
- [Duda, 1996], Duda, R.O.: „A general sound localisation model“, Sound Localisation Research, <http://www.engr.sjsu.edu/~duda/Duda.R.GSLM.html>.
- [Douglas, 1997], Douglas, S.C.: „An Efficient Implementation of the Modified Filtered-X LMS Algorithm“, IEEE Signal Processing Letters, Vol. 4, No.10, pp.286-288, October, 1997.
- [Dolby-Link], <http://www.dolby.com/>
- [Farina, 1998], Farina, A.: „Software Implementation of B-Format Encoding and Decoding“,104th AES Convention, Amsterdam, Preprint Nr. 4691, 1998.
- [Farina, Ugolotti, 1998], Farina, A.; Ugonotti, E.: „Subjective Comparison between Stereo Dipole and 3d Ambisonic Surround Systems for Automotive Applications“, in Proc. AES Int. Conf. On Spatial Sound Reproduction, pp. 532-543, 1998.
- [Farina, Tronchin, 1998], Farina, A.; Tronchin, L.:“3D Impulse Response measurements on S. Maria dl Fiore Church, Florence, Italy“, Proceedings of ICA98, International Conference on Acoustics, Seattle, 1998.
- [Fastl, 1982], Fastl, H.: „Beschreibung dynamischer Hörempfindungen anhand von Mithörschwellen-Mustern“, Hochschulverlag, München , ISBN 3-8107-2171-9, 1982.
- [Felderhoff, Reichenauer und Theile, 1998], Felderhoff, U.; Reichenauer, K. und Theile, G.: „Stabilität der Lokalisation bei verfälschter Reproduktion verschiedener Merkmale der binauralen Signale“,20. Tonmeistertagung, VDT, TMT-Tagungsband, 1998.
- [Feldbauer, 1998], Feldbauer, C.: „Echtzeit-Spatialisation für beliebige Lautsprecheranordnungen“, Projektbeschreibung: http://www.kug.ac.at/iem/f_e/projektbeschreibung/pb_echtzeitspat.htm, 1998.
- [Fellner, 1998], Fellner, M.: „Recheneffiziente Modellierung der Außenohrübertragungsfunktionen“, Diplomarbeit durchgeführt am IEM Graz, eingereicht TU Graz, 1998.
- [Furse, 1999], Furse, R.: „3D Audio links and information“, <http://www.muse.demon.co.uk/3daudio.html>.
- [Gardner, Martin, 1994], Gardner, W. G.; Martin, K.D.: „HRTF measurements of a KEMAR Dummy-Head Microphone“, M.I.T. Media Lab, Perceptual Computing Section, Technical Report No. 280, 1994. <http://sound.media.mit.edu/KEMAR.html>
- [Gerzon, 1992a], Gerzon, M.A.: „General Metatheory of Auditory Localisation“,92nd AES Convention, Preprint No. 3306, March, Vienna, 1992.
- [Gerzon, 1992b], Gerzon, M.A.: „, Panpot Laws for Multispeaker Stereo“, 92nd AES Convention, Preprint No. 3309, March, Vienna, 1992.
- [Gerzon, 1992c], Gerzon, M.A.: „Problems of Upward and Downward Compatibility in Multichannel Stereo Systems“, 93rd AES Convention, Preprint No. 3404 (J-6), October, San Francisco, 1992.

- [Gerzon, 1992d], Gerzon, M.A.: „Optimum Reproduction Matrices for Multispeaker Stereo“, J. Audio Eng. Soc., Vol. 40, No.7/8, pp.571-589, July/August, 1992.
- [Gerzon, 1992f], Gerzon, M.A.: „Ambisonic Decoders for HDTV“, 92nd AES Convention, Preprint No. 3345, March, Vienna, 1992.
- [Gerzon, 1992g], Gerzon, M.A.: „Hierarchical System of Surround Sound Transmission for HDTV“, 92nd AES Convention, Preprint No. 3339, March, Vienna, 1992.
- [Gerzon, 1992h], Gerzon, M.A.: „Psychoacoustic Decoders for Multispeaker Stereo and Surround Sound“, 93rd AES Convention, Preprint No. 3406 (J-8), October, San Francisco, 1992.
- [Gerzon, 1992i], Gerzon, M.A.: „Signal Processing for Simulating Realistic Stereo Images“, 93rd AES Convention, Preprint No. 3406 (J-8), October, San Francisco, 1992.
- [Gilkey et al., 1997], Gilkey, R.H.; Anderson, T.R.: „Binaural and spatial hearing in real and virtual environments“, Lawrence Erlbaum Association, New Jersey, 1997.
- [González, López, 2001], González, A.; López, J.J.: „Fast Transversal Filters for Deconvolution in Multichannel Sound Reproduction“, IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 4, May 2001.
- [Graber, 1997], Graber, G.: „Tontechnik und interdisziplinäres Sinnen. Eine grundlegende Fragestellung“, Habilitationsschrift, TU-Graz, 1997.
- [Griesinger, 1991], Griesinger, D.: „Improving room acoustics through time variant synthetic reverberation“, 90th AES Convention, preprint 3014, Paris, 1991.
- [Griesinger, 1993], Griesinger, D.: „On controlling the apparent absorption and volume in assisted reverberation systems“, Acta acoustica, Vol.78, p. 61-73, 1993.
- [Griesinger, 1995], Griesinger, D.: „Design and performance of multichannel time variant reverberation systems“, Proceedings of the Active 95 conference, p1203-1212, Newport Beach CA, 1995.
- [Griesinger, 1997], Griesinger, D.: „The psychoacoustics of apparent source width, spaciousness and envelopment in performing spaces“, Acta Acoustica, Vol.83, pp. 721-31, 1997.
- [Hebrank, Wright, 1974], Hebrank, J., Wright, D.: „Spectral Cues Used in the Localization of Sound Source on the Median Plane“, J. Acoust. Soc. Am., Vol. 56, pp. 1829-1834, 1974.
- [Heil et al., 1992], Heil, C.; Urban, M.: „Sound Fields Radiated by Multiple Source Arrays“. Preprint Nr. 3269, presented at the 92nd AES Convention, Vienna, Austria, 1992.
- [Heyser, 1986], Heyser, R.C.: „Instantaneous Intensity“, 81st AES Convention, Preprint 2399, J. Audio Eng. Soc., Vol. 34, p.1038, December, 1986.
- [Hull, 1992], Hull, J.: „Surround Sound Past, Present and Future“, Dolby Laboratories Inc., 100 Potrero Ave., San Francisco, CA 94103, Pub. S92/9295/9514, 1992.
- [Huygens, 1690], Wolf, E.: „The life and work of Christiaan Huygens, in Huygens' principle 1690-1990: theory and applications“, Amsterdam, 1992.
- [Ise, 1999], Ise, S.: „A Principle of Sound Field Control Based on the Kirchhoff-Helmholtz Integral Equation and the Theory of Inverse Systems“, ACUSTICA - acta acoustica, Vol. 85, pp.78-87, 1999.
- [Jacobsen, 1989], Jacobsen, F.: „Active and Reactive, Coherent and Incoherent Sound Fields“, J. Sound Vib., Vol. 130, pp. 493-507, 1989.
- [Jecklin, 1980], Jecklin, J.: „Musikaufnahmen : Grundlagen ; Technik ; Praxis“, Franzis-Verlag, München, 1980.
- [Jot et al., 1995], Jot, J. M.; Larcher, V.; and Warusfel, O.: „Digital Signal Processing Issues in the Context of Minaural and Transaural“, Audio Eng. Soc. 98th Conv. Paris, Preprint Nr. 3980, 1995.
- [Jot et al., 1998], Jot, J. M.; Wardle, S.; and Larcher, V.: „Approaches to Binaural Synthesis.“, 105th AES Convention, Preprint 4861, San Francisco, California, September, 1998.
- [Jot et al., 1999], Jot, J. M.; Larcher, V. and Pernaux J. M.: „A Comparative Study of 3-D Audio Encoding and Rendering Techniques“, 16th AES International Conference on Spatial Sound Reproduction, Finland, 1999.
- [Kirkeby, Nelson 1993], Kirkeby, O.; Nelson, P.A.: „Reproduction of plane wave sound fields.“, J. Acoust. Soc. Am., Vol. 94, No. 5, pp.2992-3000, November, 1993.
- [Kirkeby et al. 1996a], Kirkeby, O.; Nelson, P.A.; Hamada, H.: „Stereo Dipole“, British Patent Application, No. 9603236.2.
- [Kirkeby et al., 1996b], Kirkeby, O.; Nelson, P.A.; Orduna-Bustamante, F.; Hamada, H.: „Local sound field reproduction using digital signal processing“, J. Acoust. Soc. Am., Vol 100, No. 3, pp.1584-1593, September, 1996.
- [Kirkeby et al. 1997], Kirkeby, O.; Nelson, P.A.; Hamada, H.: „The ‘Stereo Dipole’ – Binaural Sound Reproduction using two closely spaced Loudspeakers.“, 102nd AES Convention, Preprint 4463 (I6), 1997.

- [Kistler, Wightman, 1992], Kistler, D. J.; Wightman, F. L.: „A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction“, J. Acoust. Soc. Am., Vol. 91, pp.1637-1647, March, 1992.
- [Krejci, 2001], Krejci, P.M.: „Entwicklung eines Mehrkanaligen Systems für eine virtuelle Raumakustik“, Diplomarbeit, Technische Universität Graz, 2001.
- [Krokstad et al., 1968], Krokstad, A.; Strøm, S.; and Sørsdal, S.: „Calculating the Acoustical Room Response by the Use of a Ray Tracing Technique“, J. Sound Vib., Vol. 8, pp.118-125, 1968.
- [Kulowski, 1985], Kulowski, A.: „Algorithmic Representation of the Ray Tracing Technique“, Appl. Acoust., Vol. 18, pp. 449-469, 1985.
- [Kuo, Morgan, 1999], Kuo, S. M.; Morgan, D. R.: „Active Noise Control“, Proceedings of the IEEE, Vol. 87, No. 6, June, 1999.
- [Kuttruff, 1995], Kuttruff, H.: „Sound Field Prediction in Rooms“, in Proc. 15th Int. Congr. On Acoustics (ICA'95), Vol. 2, pp. 545-552, Trondheim, Norway, June, 1995.
- [Kyriakakis, 1998], Kyriakakis, C.: „Fundamental and Technological Limitations of Immersive Audio Systems“, Proceedings of the IEEE, Vol.86 ,No. 5, 1998.
- [Lokki, Hiipakka, 2001], Lokki, T.; Hiipakka, J.: „A Time-Variant Reverberation Algorithm for Reverberation Enhancement Systems“, DAFX Proc. of the COST G-6 Conference DAFX-01, Limerick; 2001.
- [Link, Mandell, 1993], Link, B.; Mandell D.: „A DSP Implementation of a Pro Logic Surround Decoder“, 95th AES Convention, Preprint Nr. 3758 (B3-PM-6), New York, October 7-10, 1993.
- [Mackensen, Reichenauer und Theile, 1998], Mackensen, P.; Reichenauer, K. und Theile, G.: „Einfluß der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören“, 20. Tonmeistertagung, VDT, TMT-Tagungsband, 1998.
- [Mackenzie et al., 1997], Mackenzie, J.; Huopaniemi J.; Välimäki, V.; and Kale, I.: „Low-Order Modeling of Head-Related Transfer Functions Using Balanced Model Truncation“, IEEE Signal Processing Letters, Vol. 4, No. 2, pp. 39-41, 1997.
- [Makita, 1962], Makita, Y.: „On the Directional Localization of Sound in the Stereophonic Sound Field“. E.B.U. Review, Part A, (73):102-108, June 1962.
- [Majdak, Noisternig, 2001], Majdak, P.; Noisternig, M.: „A head position related binaural sound reproduction system“, Toningenieur - Projekt, Institut für Elektronische Musik und Akustik, Kunstuniversität Graz, 2001.
- [Majdak, Noisternig, 2002], Majdak, P.; Noisternig, M.: „Implementation Kopfpositionsbezogener Binauraltechnik“, Diplomarbeit, Technische Universität Graz, 2002.
- [Malham, Myatt, 1995], Malham, D. G.; Myatt, A.: „3-D Sound Spatialization using Ambisonic Techniques“, Computer Music Journal, Vol. 19, No. 4, pp.58-70, 1995.
- [Mann et al., 1987], Mann, J.A.; Tichy, J.; Romano, J.: „Instantaneous and Time Averaged Energy Transfer in Acoustic Fields“, J. Acoust. Soc. Am., Vol. 82, pp. 17-30, 1987.
- [Martens, 1987], Martens, W. L.: „Principal components analysis and resynthesis of spectral cues to perceived direction“, Proc. Int. Computer Music Conference, pp. 274-281, 1987.
- [Miles, 1996], Miles, T. M.: „An Optimum Linear-Matrix Stereo Imaging System“, 101st AES Convention, Preprint No. 4364 (J-4), November, Los Angeles, 1996.
- [Mishin, 1958], Mishin, L.N.: „O metode uvelicenija ustojcivosti sistem zvukousilenija“(Übersetzt: Über eine Methode der Erhöhung der Stabilität von Schallverstärkersystemen), Akust. Žurnal, Vol. 4, No.1, S.64-68, 1958.
- [Møller et al., 1995], Møller, H.; Sørensen, M.F.; Hammershøi, D.; und Jensen, C.B.: „Head-Related Transfer Functions of Human Subjects“, J. Audio Eng. Soc., Vol. 43, Nr. 5, pp.300-321, May, 1995.
- [Moore, 1982], Moore, F.R.: „A general model for spatial processing of sounds“, Computer Music Journal, Vol. 7, No. 3, p. 6-15, 1982.
- [Mc Namara et al., 1990], Mc Namara, D.A.; Pistorius, C.W.I.; Malherbe, J.A.G.: „Introduction to the Uniform Geometrical Theory of diffraction“, Artech House, 1990.
- [Nelson, Elliott, 1992], Nelson, P.A.; Elliott, S.J.: „Active Control of Sound“, Academic, London, 1992.
- [Nicol, Emerit, 1999], Nicol, R.; Emeritt, M.: „3D-sound reproduction over an extensive listening area: A hybrid method derived from holophony and ambisonic“, in Proc. AES 16th Conf. Spatial Sound Reproduction, Rovaniemi, Finland, April, 1999.
- [Oppenheim, Schafer, 1975], Oppenheim, A.V.; Schafer, R.W.: „Digital signal processing“, Englewood Cliffs, New Jersey, 1975.
- [Poletti, 1993], Poletti: „On controlling the apparent absorption and volume in assisted reverberation systems“, *Acustica*, vol. 78, pp 61-73, 1993.

- [Poletti, 1996a], Poletti, M.A.: „The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems”, J. Audio Eng. Soc., Vol. 44, No. 11, pp. 948-963, 1996.
- [Poletti, 1996b], Poletti: „An assisted reverberation system for controlling apparent room absorption and volume,“ *Audio Eng. Soc. 101st Convention*, Preprint 4365, 1996.
- [Poletti, 2000], Poletti, M.A.: „A unified theory of Horizontal Holographic Sound Systems“, J. Audio Eng. Soc., Vol. 48, No. 12, December, 2000.
- [Pralong, Carlile, 1996], Pralong, D.; Carlile, S.: „The Role of Individualized Headphone Calibration for the Generation of High Fidelity Virtual Auditory Space”, J. Acoust. Soc. Am., Vol. 100, December, 1996.
- [Pulkki, 1997], Pulkki, V.: „Virtual Sound Source Positioning Using Vector Base Amplitude Panning”, J. Audio Eng. Soc., Vol. 45, No. 6, pp. 944-951, 1997.
- [Pulkki, 1999], Pulkki V.: „MATLAB Source Codes for a simple Binaural Auditory Model”, URL: www.acoustics.hut.fi/~ville/software/auditorymodel/, 1999.
- [Pulkki et al. 1999], Pulkki V., Karjalainen M., Huopaniemi J.; „Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model.”, J. Audio Eng. Soc., Vol.47, No. 4, pp. 203-217, April, 1999.
- [Pulkki, 2001], V. Pulkki. *Evaluating spatial sound with binaural auditory model.* Proc. International Computer Music Conference ICMC 2001 Havana, Cuba September 2001.
- [Pulkki, Karjalainen, 2001a], Pulkki V., Karjalainen M.; Localization of Amplitude-Panned Virtual Sources I: Stereophonic Panning; Audio Eng. Soc. Vol. 49, No. 9; 2001.
- [Pulkki, Karjalainen, 2001b], Pulkki V.; Localization of Amplitude-Panned Virtual Sources II: 2D-3D Panning; J. Audio Eng. Soc. Vol. 49, No. 9; 2001.
- [Rife, Vanderkooy, 1989], Rife, D.D.; Vanderkooy, J.: „Transfer-Function Measurement with Maximum-Length Sequences”, J. Audio Eng. Soc., Vol. 37, pp.419-443, June, 1989.
- [Sarti, Tubaro, 2001], Sarti A., Tubaro S.; Low-cost geometry-based acoustic rendering; DAFX Proc. of the COST G-6 Conference DAFX-01, Limerick; 2001
- [Savioja et al., 1995], Savioja, L.; Backman, J.; Järvinen, A.; and Takala, T.: „Waveguide Mesh Method for Low-Frequency Simulation of Room Acoustics“, in Proc. 15th Int. Congr. On Acoustics (ICA'95), Vol. 2, pp. 637-640, Trondheim, Norway, June, 1995.
- [Savioja et al., 1999], Savioja, L.; Huopaniemi, J.; Lokki, T.; and Väänänen, R.: „Creating Interactive Virtual Acoustic Environments“, J. Audio Eng. Soc., Vol. 47, No. 9, pp. 675-705, September, 1999.
- [Schiffre, Stanzial, 1994], Schiffre, G.; Stanzial, D.: „Energetic Properties of Acoustic Fields”, J. Acoust. Soc. Am., Vol. 96, pp. 3645-3653, 1994.
- [Schröder, 1959], Schröder, M.R.: „Improvement of acoustic feedback stability in P.A. systems“, Proc. 3rd International Congress on Acoustics, Stuttgart, 1959. Amsterdam: Elsevier Publ. Comp. 1961, S. 771 ff. (P.II).
- [Schröder, 1962], Schröder, M.R.; Kuttruff, H.: „On frequency response curves in rooms.” J. Acoust. Soc. Am. Vol. 34, S.76 ff., 1962.
- [Sonke et al., 1998], Sonke, J.J.; de Vries, D.; Labeeuw, J.: „Variable acoustics by wave field synthesis: a closer look at amplitude effects”, 104th AES Convention, Preprint No. 4712, Amsterdam, 1998.
- [Sontacchi, Höldrich, 2002], Sontacchi, A.; Höldrich, R.: „Distance Coding in 3D Sound Fields”, AES 21st Int. Conference, St. Petersburg, Russia, June, 2002.
- [Sontacchi et al., 2002a], Sontacchi, A.; Strauß, M.; Höldrich, R.: „3D Audio Interface für Bildschirmbenutzer in Wellenfeldsynthese“, 22.TMT, Postersession, Jahrestagung des Vereins Deutscher Tonmeister, Hannover, November, 2002.
- [Sontacchi et al., 2002b], Sontacchi, A.; Noisternig, M.; Majdak, P.; Irnberger, W.; Hölderich, R.: „DSP-Implementation eines Kopfpositionsbezogenen binauralen Wiedergabesystems“, 22.TMT, Produktforum, Jahrestagung des Vereins Deutscher Tonmeister, Hannover, November, 2002.
- [Sontacchi et al., 2002c], Sontacchi, A.; Noisternig, M.; Majdak, P.; Höldrich, R.: „An Objective Model of Localisation in Binaural Sound Reproduction Systems“, AES 21st Int. Conference, St. Petersburg, Russia, June, 2002.
- [Sontacchi et al., 2002d], Sontacchi, A.; Majdak, P.; Noisternig, M.; Höldrich, R.: „Subjective Validation of Perception Properties in Binaural Sound Reproduction Systems“, AES 21st Int. Conference, St. Petersburg, Russia, June, 2002.
- [Stanzial et al., 1996], Stanzial, D.; Prodi, N.; and Schiffre, G.: „Reactive Acoustic Intensity for General Fields and Energy Polarization”, J. Acoust. Soc. Am., Vol. 99, pp. 1868-1876, 1996.
- [Start et al., 1995], Start, E. W.; Valstar, V. G.; de Vries, D.: „Application of Spatial Bandwidth Reduction in Wave Field Synthesis”, Preprint Nr. 3972, presented at the 98th AES Convention, Paris, France, 1995.
- [Start, 1996], Start, E. W.: „Application of Curved Arrays in Wave Field Synthesis”, Preprint Nr. 4143, presented at the 100th AES Convention, Copenhagen, Denmark, 1996.

- [Start, 1997], Start, E. W.: „Direct sound enhancement by wave field synthesis“, Thesis, TU-Delft, Niederlande, 1997.
- [Start et al., 1999], Start, E.W.; de Vries, D.; Berkhout, A. J.: „Wave Field Synthesis Operators for Bent Line Arrays in a 3D Space“, ACUSTICA – acta acoustica, Vol. 85, 1999.
- [Stearns, 1988], Stearns, S. D.: „Fundamentals of Adaptive Signal Processing“, *Advanced Topics in Signal Processing*, chapter 5, p.246-288, edited by J.S: Lim, A. V. Oppenheim, ISBN 0-13-013129-6 025, 1988.
- [Strauß, 2002], Strauß, M.: „Simulation und Implementierung eines Audio Interface in Wellenfeldsynthese“, Diplomarbeit, Technische Universität Graz, 2002.
- [Strutt, 1907], Strutt, J.W.: „On our perception of sound direction“, *Philos. Mag.* 13, pp.214-232, 1907, (Lord Rayleigh, 3rd Baron of Rayleigh).
- [Takane et al., 1999], Takane, S.; Suzuki, Y.; and Sone, T.: „A New Method for Global Sound Field Reproduction Based on Kirchhoff's Integral Equation“, ACUSTICA – acta acoustica, Vol. 85, pp.250-257, 1999.
- [Takeuchi, Nelson, 2000a], Takeuchi, T.; Nelson, P.A.: „Optimal Source Distribution for Virtual Acoustic Imaging“, British Patent application, No.0015419.5.
- [Takeuchi, Nelson, 2000b], Takeuchi, T.; Nelson, P.A.: „Optimal Source Distribution for Virtual Acoustic Imaging“, ISVR Technical Report, No. 288, Institute of Sound and Vibration Research, University of Southampton, 2000.
- [Täschl, 2000], Täschl, M.: „Binaural Sound Reproduction via Distributed Loudspeaker Systems“, Diplomarbeit, eingereicht am Institut für Elektronische Musik und Akustik, Kunstuniversität Graz, durchgeführt am Institute of Sound and Vibration Research, University of Southampton, 2000.
- [Theile, 1980], Theile, G.: „Über die Lokalisation im Überlagerten Schallfeld“, Dissertation, Technische Universität Berlin, 1980.
- [Theile, 1981], Theile, G.: „Zur Theorie der optimalen Wiedergabe von stereofonen Signalen über Lautsprecher und Kopfhörer“, *Rundfunktechnik, Mitteilungen* 4, S. 155-170, 1981.
- [Theile, 1996], Theile, G.: „Capabilities and limits of 3/2-stereo recording.“, 19. Tonmeistertagung, VDT, Proceedings, pp. 139-158, ISBN 3-598-20360-8, 1996.
- [Theile, 2000], Theile, G.: „Multichannel Natural Music Recording Based On Psychoacoustic Principles“, Handout, 108th AES Convention, Paris, 2000.
- [Thurlow und Runge, 1967], Thurlow, W.; Runge, P.: „Effect of induced head movements on localization of direction of sounds“, *J. Audio Soc. Am.*, Vol. 42, No. 2, p. 480-488, 1967.
- [Thurlow, Mangels und Runge, 1967], Thurlow, W.; Mangels, J. und Runge, P.: „Head movements during sound localization“, *J. Audio Soc. Am.*, Vol. 42, No. 2, p. 489-493, 1967.
- [Trebuchet, 1997], Trebuchet, L.C.: „Etude et mise en oeuvre des techniques ambisoniques pour la spatialisation du son“. Technical report, ENST Paris, CCETT, 1997.
- [Tsingos et al., 2001], Tsingos N., Funkhouser T., Ngan A., Carlbom I.: „Modeling Acoustics in Virtual Environments Using the Uniform Theory of Diffraction“, *Computer Graphics (SIGGRAPH 2001)*, August, 2001.
- [Verheijen, 1997], Verheijen, E.: „Sound Reproduction by Wave Field Synthesis“, Dissertation, Delft University of Technology, The Netherlands, 1997.
- [Vogel, 1993], Vogel, P.: „Application of wave field synthesis in room acoustics“. Thesis, TU-Delft, Netherlands, 1993.
- [de Vries et al., 1994], de Vries, D.; Start, E.W.; Valstar, V.G.: „The Wave Field Synthesis Concept Applied to Sound Reinforcement: Restrictions and Solutions“, 96th AES Convention, Preprint Nr. 3812, Amsterdam, Netherlands, 1994.
- [de Vries, 1996], de Vries, D.: „Sound Reinforcement by Wave Field Synthesis: Adaption of the Synthesis Operator to the Loudspeaker Directivity Characteristics“, *J. Audio Eng. Soc.*, Vol. 44, No. 12, p. 1120, 1996.
- [Ward, Abhayapala, 2001], Ward, B. und Abhayapala, T.D.: „Reproduction of a plane-wave sound field using an array of loudspeakers“ *IEEE Transactions on speech and audio processing*, Vol. 9, No. 6, Sept., 2001.
- [Wenzel et al., 1993], Wenzel, E.M.; Arruda M.; Kistler, D. und Wightman, F.L.: „Localisation Using Nonindividualised Head-Related Transfer Functions.“, Prentice-Hall Inc., *J. Acoust. Soc. of Am.*, Vol. 94, No.1, pp. 111-123, July, 1993.
- [Wenzel, 1997], Wenzel, E.M.: „Analysis of the role of update rate and system latency in interactive virtual acoustic environments.“, 103rd AES Convention, Preprint 4633, New York, 1997.
- [Widrow, Stearns, 1985], Widrow, B.; Stearns, S.D.: „Adaptive Signal Processing“, Prentice-Hall, Englewood Cliffs, NJ 1985.
- [Wightman, Kistler 1989a], Wightman, F.L.; Kistler, D.J.: „Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis“, *J. Acoust. Soc. Am.*, Vol. 85, pp.271-282, Februar, 1989.

[Wightman, Kistler, 1989b], Wightman, F.L.; Kistler, D.J.: „Headphone Simulation of Free-Field Listening. II: Psychophysical Validation“, J. Acoust. Soc. Am., Vol. 85, Februar, 1989.

[Wright et al., 1974], Wright, D.; Hebrank, J.H.; und Wilson, B.: „Pinna reflections as cues for localization“, J. Acoust. Soc. Am., Vol. 56, pp. 957-962, 1974.

[Ziomek, 1995], Ziomek, L. J.: „Fundamentals of Acoustic Field Theory and Space-Time Signal Processing“, CRC Press Inc., Monterey, California, USA, 1995.

[Zmölnig, 2002], Zmölnig, J.M.: „Entwurf und Implementation einer Mehrkanal- Beschallungsanlage“, Diplomarbeit, Technische Universität Graz, 2002.

[Zwicker, 1982], Zwicker, E.: „Psychoakustik“, Springer Verlag, Berlin (u.a.), ISBN 3-540-11301-7, 1982.

[Zwicker, Fastl, 1990], Zwicker E. ; Fastl, H. : „Psychoacoustics - Facts and Models“, Springer-Verlag, Berlin (u.a.), ISBN 3-540-52600-5, 1990.

[Zwicker], Zwicker, C.: Französisches Patent Nr. 712.588.

Normen

[EBU-Tech. 3276] EBU document Tech 3276 Supplement 1 (1999): „Listening Conditions for the assessment of sound programme material: multichannel sound“.

[ITU-R BS. 775-1] ITU-Empfehlung: Recommendation ITU-R BS. 775-1: Multichannel Stereophonic Sound System with and without accompanying Picture (Genf, 1992-1994).

[SSF– 01-SV-E1] Surround-Sound-Forum – 01-SV-E1, Recommended Practice, Listening Conditions and Reproduction Arrangements for Multichannel Stereophony (01/00), bzw. deutsch: Empfehlungen für die Praxis SSF – 01-10/98

Anhang

Verteilung und Positionierung der Lautsprecherringe im IEM CUBE (vgl. dazu 3. Kapitel).

Für das Beschallungssystem wurden drei Lautsprecherringe fixiert. Die Anzahl der Lautsprecher je Ring wurden wie folgt festgelegt:

- 1.Ring:** 12 Lautsprecher (0° Elevation, in einer Höhe von 1,4m)
- 2.Ring:** 8 Lautsprecher (Höhe H_1 wird berechnet)
- 3.Ring:** 4 Lautsprecher (Höhe H_2 wird berechnet)

Die Aufteilung der Teilflächen, die sich durch die Lautsprecherringe ergeben, erfolgt in jeweils zwei gleich große Flächen, so dass die Lautsprecher auf deren Schwerpunktlinie liegen.

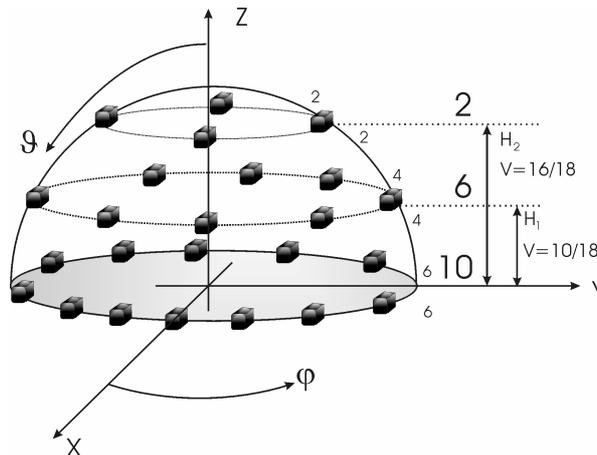


Abb. A.1: Halbkugel mit Lautsprecherringe und Teilflächen.

Der erste Ring liegt auf der gedachten Äquatorlinie. Von den 12 Lautsprecher sind je sechs für die obere und untere Teilfläche - die ja der unteren Hemisphäre zuzuordnen ist - zuständig. Für die Teilflächen ergibt sich durch die zusätzliche Aufteilung und Zuordnung der Lautsprecher folgende Flächengewichte, die für die Bestimmung der Höhen H_1 und H_2 benötigt werden:

$$A_1 : A_2 : A_3 = 2 : 6 : 10$$

Die Lautsprecher liegen auf einen gedachten Rotationsellipsoiden.

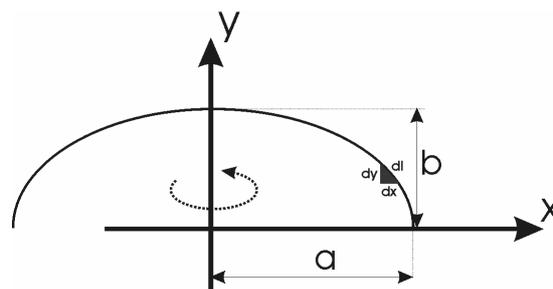


Abb. A.2: Skizze des Rotationsellipsoids mit Bezeichnung.

Die Oberfläche A eines Rotationskörpers berechnet sich wie folgt:

$$A = 2\pi \int_{y_1}^{y_2} x(y) \cdot dl \quad (\text{A.1})$$

wobei eine Ellipse um die y -Achse rotiert. Durch Umschreiben der Ellipsengleichung ergibt sich die Funktionsdarstellung für $x=x(y)$ wie folgt:

$$x^2 = a^2 \left(1 - \left(\frac{y}{b} \right)^2 \right) \quad (\text{A.2})$$

Das Kurvenelement dl für die Integration in Gl. A.1 ergibt sich aus der Bogenlänge einer Kurve zwischen zweier Punkte mit:

$$dl = \sqrt{1 + \left(\frac{dx}{dy} \right)^2} dy \quad (\text{A.3})$$

Setzt man Gl. A.3 in A.1 unter Verwendung von A.2 ein, so erhält man nach Umformung folgenden Ausdruck:

$$A = 2\pi \int_{y_1}^{y_2} \sqrt{a^2 - \left[\frac{a^2(b^2 - a^2)}{b^2} \right] y^2} \cdot dy \quad (\text{A.4})$$

Setzt man den unabhängigen Term in den eckigen Klammer in der Gl. A.4 gleich Δ und zieht diese Konstante vor das Integral, so ergibt sich ein bereits bekannter unbestimmter Integrationstypus der Form $\int \sqrt{X} dx$ mit $X = \text{const}^2 - x^2$. Die Lösung erfolgt durch Substitution von $\sqrt{\Delta} y = z$ und durch die Wahl $z = a \cdot \sin(t)$ (das Ergebnis findet sich auch bei ([Bronstein, 1993], S.1029, Typ 157). Die Oberflächenlösung mit den variablen Integrationsgrenzen y_1 und y_2 (die den entsprechenden Ringhöhen entsprechen) ergibt sich somit zu:

$$A = 2\pi a^2 \left[\frac{\arcsin\left(\frac{\sqrt{\Delta}}{a} \cdot y\right)}{2\sqrt{\Delta}} + \frac{y}{2a} \sqrt{1 - \frac{\Delta y^2}{a^2}} \right]_{y_1}^{y_2} \quad (\text{A.5})$$

Die beiden Parameter a und b für die Ellipse sind durch die Raumform des IEM-CUBE vorgegeben. Die gesuchten Höhen für die vorgegebenen Flächenteilungsverhältnisse V sind $H_2=16/18$ für den obersten Ring und $H_1=10/18$ für die Höhe des zweiten Ring. Sie können durch iterative Näherung (siehe Pseudocode) gefunden werden.

Pseudocode zur Bestimmung der Höhe $H_i=y$ für das Teilungsverhältnis V_i

```

y = b ;           # für die gesuchten Oberflächenverhältnisse gilt  $y \in [0, b]$ 
step = 0.1 · y ;
e = V · S - S(y,0); # Fehlerfunktion, S ist die Gesamtfläche mit  $S=S(b,0)$ 

while abs(e) > d   # d ist die tolerierte Fehlerschranke.
  if e < 0
    y = y - step;   # dekrementiere y, da Fläche zu groß ist.
    e = V · S - S(y,0);
  elseif e > 0
    y = y + step/2; # inkrementier y, da Fläche zu klein ist.
    step = step/2;  # iterative Schrittgröße wird halbiert.
    e = V · S - S(y,0);
  else
    break # Abbruch der while-Schleife, da der Fehler gleich Null ist.
  end
end # end if

end # end while

```

Beschreibung der verwendeten Systemparameter in Kapitel 5.3.

SetNr	Set#	Flen	HRIR	LS-vorne	Ord	Gew
1	#301	256	AKG	-	-	-
2	#15	128	AKG	2	3	1
3	#9	128	AKG	1	3	1
4	#9	96	AKG	1	3	1
5	#9	64	AKG	1	3	1
6	#10	128	AKG	1	2	1
7	#0	128	KEMAR	1	3	1
8	#24	256	KEMAR	1	3	K-25
9	#24	128	KEMAR	1	3	K-25
10	#204	128	KEMAR	1	3	0.4
11	#300	256	KEMAR	-	-	-

Tabelle A.1: verwendete Settings, Legende dazu siehe Tabelle A.2

<i>SetNr</i>	Nummer der verwendeten Settings, wird bei sowohl bei den Abbildungen als auch in den mathematischen Formulierung benutzt	
<i>Set#</i>	Nummer der entsprechenden Settings bei der Fehlermaßberechnung. Die Settings mit der direkten Implementation von HRIRs (Set#: #300 und #301) wurden nicht simuliert und sind deshalb bei der Fehlermaßberechnung nicht enthalten	
<i>Flen</i>	Verwendete Länge der HRIRs in Samples	
<i>LS-vorne</i>	Anzahl der virtuellen Lautsprecher vorne, entspricht der jeweiligen Lautsprecheranordnung von Fehlermaß, siehe Tabelle B. 3	
<i>Ord</i>	Ordnung des Ambisonic-Systems	
<i>Gew</i>	Gewichtung der jeweiligen Ordnungen	
	1	Alle Ordnungen voll gewichtet
	0.4	Die 3-te Ordnung mit dem Faktor 0.4 gewichtet
	K-25	Gewichtung aller Ordnungen mit einem Kaiserfenster mit 25dB-Dämpfung bei der ersten Nebenkeulen

Tabelle A.2: Legende zu Tabelle A.1

Schlüssel für die verbalen Beschreibungen:

Aussage	Bewertung
keine Aussage	-
sehr wenig	-2
zu wenig	-1
<u>ausgewogen</u>	<u>0</u>
zu viel	+1
sehr viel	+2

Tabelle A.3: Klangfarbe

Aussage	Bewertung
keine Aussage	-
Im Kopf	0
Am Kopf	1
leicht Außerkopf, Radius von ca. 10cm	2
Außerkopf, Radius von ca. 30cm	3
<i>weit bzw. sehr <u>weit Außerkopf</u>, Radius >50cm</i>	<u>4</u>

Tabelle A.4: Externität

Aussage	Bewertung
keine Aussage	-
Stark verschwommen	0
Verschwommen, Personenbreit	1
Leicht verschwommen	2
<i><u>Punktuell</u></i>	<u>3</u>

Tabelle A.5: Lokalisationsunschärfe

Aussage	Bewertung
keine Aussage	-
Sprunghaft	0
<i><u>Stetig</u></i>	<u>1</u>

Tabelle A.6: Konsistenz