

# LEIBNIZ UNIVERSITÄT HANNOVER

FAKULTÄT FÜR ELEKTROTECHNIK UND INFORMATIK  
INSTITUT FÜR KOMMUNIKATIONSTECHNIK

## Wahrnehmungsbasierte Analyse von Higher-Order-Ambisonics-Signalen bei binauraler Synthese

### Diplomarbeit

eingereicht von

STEFAN BALKE

am 04. November 2013

Erstprüfer : Prof. Dr. Jürgen Peissig  
Zweitprüfer : Prof. Dr.-Ing. Markus Fidler  
Betreuer : Dipl.-Ing. Stefan Abeling  
Dr.-Ing. Johann-Markus Batke



## EHRENWÖRTLICHE ERKLÄRUNG

---

Hiermit versichere ich, die vorliegende Diplomarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die wörtlich oder inhaltlich aus den Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

*Hannover, 04. November 2013*

---

Stefan Balke



## ABSTRACT

---

This diploma-thesis deals with the binaural synthesis of soundfields which are described with the use of Higher-Order Ambisonics. After introducing the needed background information, three listening-tests with a focus on different problems of binaural synthesis are presented. The first two tests examine the localization of sources in the horizontal plane using different synthesis approaches and different Head-Related Impulse responses. The second test limits the spatial resolution and evaluates the emerging effects. The third test deals with an approach for enhancing the externalization by adding different rooms to the ear signals. In this test, the signal-processing is done using spherical harmonics exclusively.

## ZUSAMMENFASSUNG

---

Diese Diplomarbeit beschäftigt sich mit der binauralen Synthese von Schallfeldern, die mithilfe von Higher-Order Ambisonics beschrieben werden. Nach der Beschreibung der benötigten Grundlagen werden drei Hörtests vorgestellt, die verschiedene Probleme der binauralen Synthese untersuchen. Die ersten beiden Tests legen den Fokus auf die Lokalisierungsfähigkeit in der Ebene unter Verwendung verschiedener Syntheseverfahren und unterschiedlicher kopfbezogener Impulsantworten. Im zweiten Test wird zusätzlich die räumliche Auflösung reduziert. Der dritte Test beschäftigt sich mit der Verbesserung der Externalisierung durch Hinzufügen verschiedener Räume. Die Signalverarbeitung findet bei diesem Test ausschließlich unter Verwendung sphärischer Harmonischer statt.



# INHALTSVERZEICHNIS

---

1	EINLEITUNG	1
2	GRUNDLAGEN	3
2.1	Technologieüberblick	3
2.1.1	Kanalbasierte Verfahren	3
2.1.2	Szenenbasierte Verfahren	4
2.1.3	Objektbasierte Verfahren	4
2.1.4	Mischverfahren	5
2.2	Binaurales Hören	6
2.2.1	Quellenlokalisation	7
2.3	Binaurale Synthese	8
2.3.1	HRTF Messung	11
2.4	Higher-Order-Ambisonics	13
2.4.1	Transformationsvorschriften	15
2.4.2	Weitere Eigenschaften	16
2.4.3	Darstellung als reelle Koeffizienten	16
2.4.4	Ebene-Wellen-Zerlegung	17
2.4.5	Schallfeldsynthese	18
2.5	Teststatistik	18
2.5.1	Terminologie	19
2.5.2	Schätzfunktionen	20
2.5.3	Regressionsanalyse	23
2.5.4	Varianzanalyse	26
3	TESTKONZEPTE	29
3.1	Lokalisierungstest	29
3.1.1	Testablauf	29
3.2	Modifizierter Lokalisierungstest	32
3.2.1	Ordnungsvariation	32
3.2.2	Stimuluswahl	33
3.2.3	Stimuluswiederholung	33
3.2.4	Zeigerbehandlung	34
3.2.5	Ergonomie	34
3.2.6	Testablauf	34
3.2.7	Statistisches Modell	34
3.3	Externalisierungstest	35
3.3.1	Messung der Raumimpulsantwort	35
3.3.2	Synthetische Raumimpulsantwort	37
3.3.3	HOA-Transformation der Quelldaten	38
3.3.4	HOA-Transformation der HRTF	38
3.3.5	Synthese der Audiodaten	39
3.3.6	Testablauf	40
3.3.7	Implementierung	41
4	ERGEBNISANALYSE	45

4.1	Modifizierter Lokalisierungstest	45	
4.1.1	Erster Testdurchlauf mit trainierten Hörern	45	45
4.1.2	Ergänzender Testdurchlauf	49	
4.2	Externalisierungstest	53	
5	ZUSAMMENFASSUNG	57	
6	AUSBLICK	59	
A	MATHEMATISCHER ANHANG	61	
A.1	Lösung der sphärischen Wellen-Differentialgleichung	61	61
B	IMPLEMENTIERUNGSDetails	65	
B.1	SPAT-Parameter	65	
C	HÖRTEST-HANDOUTS	91	
C.1	Lokalisierungstest	91	
C.1.1	Testablauf	91	
C.2	Externalisierungstest	92	
C.2.1	Testablauf	92	
C.2.2	Bewertung	92	
C.2.3	Optionale Anmerkungen	93	
C.3	Kommentare Externalisierungstest	94	
	LITERATUR	105	



## ABBILDUNGSVERZEICHNIS

---

Abbildung 1	Menschliches Hörorgan	6
Abbildung 2	Festlegung des Koordinatensystems	7
Abbildung 3	Binaurales Hören einer Schallquelle	9
Abbildung 4	Head-Related Transfer Function (HRTF) für das linke und rechte Ohr für $\vartheta = 0^\circ$	9
Abbildung 5	Wiedergabe über Lautsprecher	10
Abbildung 6	Wiedergabe über Kopfhörer	11
Abbildung 7	Sphärische Harmonische $\mathfrak{R}\{Y_n^m(\vartheta, \varphi)\}$	14
Abbildung 8	Wahrscheinlichkeitsdichtefunktion der F-Verteilung für verschiedene Freiheitsgrade	28
Abbildung 9	System 1: HRIR	30
Abbildung 10	System 2: HOA	31
Abbildung 11	System 3: HOA-Rotation	31
Abbildung 12	Betrag des Winkelspektrums für das linke Ohr	33
Abbildung 13	Residuenuntersuchung	45
Abbildung 14	Interaktions-Plot der Haupteffekte	46
Abbildung 15	Boxplots erster Testdurchgang	48
Abbildung 16	Residuenuntersuchung	50
Abbildung 17	Interaktions-Plot der Haupteffekte mit neun Hörern	51
Abbildung 18	Boxplots ergänzender Testdurchlauf	52
Abbildung 19	Überblick der Testergebnisse	53
Abbildung 20	Bewertungen aufgeteilt nach Stimulus	55

## TABELLENVERZEICHNIS

---

Tabelle 1	Mittlerer Winkelfehler und Standardabweichung	46
Tabelle 2	ANOVA Analyse I(a): Alle Interaktionen	49
Tabelle 3	ANOVA Analyse I(b): Nur signifikante Interaktionen	49
Tabelle 4	ANOVA Analyse I(c): Ohne Unterscheidung der Hörer	50
Tabelle 5	Mittlerer Winkelfehler und Standardabweichung	50
Tabelle 6	ANOVA Analyse II: Ohne Unterscheidung der Hörer	51
Tabelle 7	ANOVA: Ohne Unterscheidung der Hörer	54

Tabelle 8	Paarweiser T-Test	54
-----------	-------------------	----

## ALGORITHMENVERZEICHNIS

---

Quelltext 1	Kommandozeilenparameter main.py	43
-------------	---------------------------------	----

## ABKÜRZUNGSVERZEICHNIS

---

ANOVA	Analysis of variance
ASA	Auditory Scene Analysis
BRIR	Binaural Room Impulse Response
CASA	Computational Auditory Scene Analysis
DAW	Digital Audio Workstation
FFT	Fast-Fourier Transform
HOA	Higher-Order-Ambisonics
HRIR	Head-Related Impulse Response
HRTF	Head-Related Transfer Function
ILD	Interaural level difference
ITD	Interaural time difference
IRCAM	Institut de Recherche et Coordination Acoustique/Musique
JSON	JavaScript Object Notation
MPEG	Moving Picture Experts Group
MUSHRA	Multi-Stimulus Test with Hidden Reference and Anchor
NLMS	Normalize Least Mean Squares
VBAP	Vector Base Amplitude Panning
WFS	Wellenfeldsynthese
XML	Extensible Markup Language



## NOMENKLATUR

---

$(\cdot)^*$	komplex konjugiert
$(\cdot)^H$	Hermitesche Matrix (transponierte, komplex konjugierte Matrix)
$\langle \cdot \rangle$	inneres Produkt
$\mathring{A}_n^m$	Higher-Order-Ambisonics Koeffizienten
$\nabla$	Nabla-Operator, enthält partielle Ableitungen aller Komponenten: $\nabla = \left( \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right)$
$\omega$	Kreisfrequenz
$\vec{Q}$	Raumpunkt mit $(r, \vartheta, \varphi)^T$
$c$	Schallgeschwindigkeit, $\sim 343 \text{ m s}^{-1}$ in Luft bei $20^\circ\text{C}$
$i$	imaginäre Einheit
$j_n$	sphärische Besselfunktion
$k$	Wellenzahl, Wellenperioden pro Meter
$N$	Schallfeldordnung
$O$	Anzahl der Schallfeld-Komponenten
$P_n^m$	assoziierte Legendre Funktionen erster Art
$y_n$	sphärische Hankelfunktion
$df$	Degrees of freedom, Freiheitsgrade
$F$	F-Statistik
$MS$	Mittlere Quadratsumme
$QS$	Quadratsumme



## Diplomarbeit

für

Herrn cand. el. Stefan Balke  
Matrikel-Nr. 2677830

### **Wahrnehmungsbasierte Analyse von Higher-Order-Ambisonics-Signalen bei binauraler Synthese**

Um einen räumlichen Eindruck bei der Wiedergabe von Tonsignalen zu erzeugen, werden Formate wie Stereo (2.0) oder Surround-Sound (5.1) eingesetzt. Technicolor untersucht die Anwendung von Tonwiedergabeverfahren, die mit deutlich mehr als zwei oder fünf Kanälen betrieben werden und im Gegensatz zu Stereo oder Surround auch Höheninformation wiedergeben können. Als Format wird eine Schallfeldbeschreibung auf Basis von Higher-Order-Ambisonics verwendet. Zur Wiedergabe werden räumlich verteilte Lautsprecheranordnungen benötigt, die auch erhöht platzierte Lautsprecherpositionen enthalten. Um den technischen Aufwand zu begrenzen, kann die binaurale Synthese zur Erzeugung virtueller Lautsprecher signale für die Kopfhörerbildung herangezogen werden.

Ziel der binauralen Synthese ist es, den Schalldruck an den Ohrtrommelfeldern, der durch reale Schallquellen erzeugt wird, nur durch die Verwendung von Kopfhörern exakt zu reproduzieren. Dabei wird der Übertragungsweg von der realen Schallquelle zu den beiden Ohren durch „Kopfbezogene Übertragungsfunktionen“ (head related transfer functions, HRTF), nachgebildet. Die Verwendung von HRTFs bei Kopfhörerbildung lässt ein Hörereignis als virtuelle Quelle außerhalb des Kopfes des Hörers erscheinen. Diese virtuellen Quellen werden für die Wiedergabe eines 3D-Schallfelds verwendet. Der hervorgerufene Höreindruck hängt sowohl von den verwendeten HRTFs wie auch den Eigenschaften des wiedergegebenen Schallfelds ab.

Zur Darstellung des Schallfelds werden Koeffizienten der sphärischen Fouriertransformation berechnet; diese Signaldarstellung ist auch als Higher-Order-Ambisonics-Format (HOA-Format) bekannt. Wesentliches Merkmal dieser Darstellung ist die räumliche Bandbegrenzung, die aus der endlichen Ordnung der HOA- Koeffizienten entsteht.

Herr Balke bekommt die Aufgabe, die Qualität der räumlichen Wiedergabe zu untersuchen und zu bewerten. Ein Kriterium ist die Lokalisierbarkeit virtueller Quellen, die sich abhängig von der gewählten HOA-Ordnung, den HRTFs und dem Wiedergabesystem ändern kann. Neben objektiven Kriterien sollen insbesondere subjektive Hörtests durchgeführt werden.

Weitere Angaben erfolgen mündlich in Absprache mit dem Betreuer.





## EINLEITUNG

---

Raumklang wird insbesondere von Filmproduzenten und Musikern benutzt, um ihre künstlerische Intention nicht nur über visuellem, sondern auch über auditivem Wege dem Publikum zu präsentieren. So spielen beispielsweise in einer Filmszene Geräusche und Hintergrundmusik bei der Erzeugung einer gewünschten Stimmung eine Hauptrolle. Musiker verwenden die Möglichkeiten der Aufnahme von Raumklang, um die Stimmung ihrer Konzerte einzufangen.

Längst sind Lautsprecheraufstellungen mit fünf oder mehr Lautsprechern zur Wiedergabe von Raumklang etabliert und beim Endkunden angekommen. Mit der steigenden Verwendung von Smartphones erschließt sich ein neues mögliches Zielmedium für die Wiedergabe der oben genannten Inhalte. Aufgrund ihrer Leistungsfähigkeit könnte das Erlebnis des Raumklangs unterwegs verfügbar gemacht werden. Im Oktober 2012 besitzt fast jeder zweite Deutsche ein Smartphone [19]. Allein in Deutschland wurden 2013 rund 26 Millionen Smartphones verkauft, knapp 23 % mehr, als im Vorjahr [16]. 47,1 % der Smartphone-Besitzer nutzen ihr Smartphone zum regelmäßigen Konsum von Musik [49]. Die Musikwiedergabe erfolgt hierbei größtenteils über Kopfhörer. Bei den wiedergegebenen Inhalten handelt es sich meistens um Aufnahmen, die für die Lautsprecherwiedergabe erstellt worden sind. Die Wiedergabe dieser Signale über Kopfhörer erzeugt den Eindruck, dass sich das dargebotene Signal im Kopf befindet und nicht externalisiert wahrgenommen wird.

Diese Arbeit untersucht die Möglichkeit, mittels sogenannter binauraler Synthese, ein neues Format des Raumklangs, genannt Higher-Order-Ambisonics (HOA), auf Kopfhörern wiederzugeben. Es handelt sich bei HOA um ein dreidimensionales Verfahren, das das Schallfeld auf einer Kugeloberfläche beschreibt. Diese allgemeine Form der Schallfeldbeschreibung ist sehr flexibel, da sie auf unterschiedlichen Zielmedien und Lautsprecheranordnungen ausgegeben werden kann.

Zunächst werden die Grundlagen zu den Kernthemen binaurales Hören, binaurale Synthese und der verwendeten HOA-Technologie erläutert. Es folgen statistische Methoden zur Auswertung der später durchgeführten Hörtests. Die Hörtests untersuchen die Quellenlokalisierung und Quellenexternalisierung bei der Verwendung von binauraler Synthese. Die Quellenexternalisierung wird mithilfe eines Syntheseverfahrens untersucht, welches ausschließlich im Bereich der sphärischen Harmonischen arbeitet. Nach der Analyse der Hörtestergebnisse schließt diese Arbeit mit einer Zusammenfassung und einem Ausblick ab.



Zur binauralen Synthese werden verschiedene Aspekte der digitalen Signalverarbeitung benötigt. Dieses Kapitel beginnt mit einer Überblicksdarstellung von Verfahren zur Darstellung von Raumklang auf Lautsprechern. Anschließend folgt eine Einführung in das binaurale Hören des Menschen. Die binaurale Synthese greift dies auf, um eine räumliche Kopfhörerwiedergabe zu ermöglichen. Es folgt die Beschreibung des hier verwendeten Verfahrens zur Messung der kopfbezogenen Übertragungsfunktionen.

Anschließend wird die Theorie der Higher-Order-Ambisonics (HOA) vorgestellt. Das Kapitel schließt mit einem Überblick über statistische Analyseverfahren zur Bewertung von Hörtests.

## 2.1 TECHNOLOGIEÜBERBLICK

Zur lautsprecherbasierten Wiedergabe von Schallfeldern existieren mehrere technische Ansätze. Die Umstellung von zweidimensionaler auf eine dreidimensionale Wiedergabe erfordert die Erweiterung bereits benutzter Techniken oder die Erschließung neuer. Zweidimensional bedeutet, dass das Schallfeld nur in einer Ebene dargeboten werden kann. Dreidimensionale Verfahren ergänzen die Wiedergabe um eine Höhenkomponente und können ein Schallfeld in allen drei Raumrichtungen wiedergeben.

Neben der traditionellen kanalbasierten Wiedergabe, werden hier szenenbasierte, objektbasierte und Mischformen der Verfahren vorgestellt.

### 2.1.1 Kanalbasierte Verfahren

Der kanalbasierte Ansatz ist das wohl am weitesten verbreitetste Verfahren zur Generierung von Raumklang. Dabei sind die Informationen des Schallfelds als PCM-Signale in Kanälen gespeichert. Jeder Lautsprecher repräsentiert einen Kanal. Die Bandbreite der angebotenen Aufstellungen ist groß: Angefangen mit einer reinen Mono-Wiedergabe (1.0, ein Lautsprecher), über Stereo-Wiedergabe (2.0, zwei Lautsprecher) bis hin zu Mehrkanalsystemen wie 5.1, 7.1 oder 22.2.

Im Falle von Stereo oder 5.1 existieren beispielsweise spezielle Aufnahmetechniken, die auf die in der Wiedergabe verwendete Lautsprecheranordnung abgestimmt sind, z.B. für Stereo AB, XY und ORTF oder für 5.1 Decca-Tree oder Hamasaki-Square. Die Auswahl der Mi-

krofonaufstellung obliegt dem künstlerischem Empfinden des Aufnahmeleiters und trägt maßgeblich zum späteren Ergebnis bei.

Im Kinobereich ist kanalbasierter Raumklang in der Produktion und in den Kinosälen Standard. Etablierte kommerzielle Vertreter sind Dolby<sup>®</sup> Laboratories, Inc. und DTS<sup>®</sup>, Inc. Auro Technologies stellt mit AURO-3D<sup>®</sup> einen neuen Konkurrenten dar [3].

### 2.1.2 Szenenbasierte Verfahren

Mit szenenbasierten Verfahren ist es möglich, Schallfelder physikalisch zu beschreiben und zu rekonstruieren. Zwei wichtige Vertreter sind die Wellenfeldsynthese (WFS) und Higher-Order-Ambisonics (HOA).

WFS reproduziert das Feld in einem Raum als Lösung einer Randwertaufgabe. Das Kirchhoff-Helmholtz Integral als Lösung der Wellen-Differentialgleichung wird zugrunde gelegt, um das Feld zu beschreiben [45]. Lautsprecher am Rand des Wiedergabebereiches ermöglichen die flexible Steuerung der Randwerte und somit das gesamte Feld innerhalb der Lautsprecheranordnung. Dieses Verfahren erfordert je nach Größe des zu beschallenden Raumes eine große Anzahl von Lautsprechern.

Ein kommerzieller Anbieter von WFS-Systemen ist die IOSONO GmbH. Sie rüstet Räume, insbesondere Kinos, mit WFS-Systemen aus und bietet passende Produktionswerkzeuge an. Die meisten Installationen haben eine Lautsprecherzeile, die sich um den gesamten Raum spannt. Damit ist die Feldreproduktion in einer Ebene möglich. Es handelt sich um ein zweidimensionales Verfahren. Eine weitere Ebene würde die Installation einer zusätzlichen Lautsprecherzeile bedeuten. Die Beschreibung der wiederzugebenden Szenen erfolgt bei WFS über Objekte. Ein Objekt wird hier definiert als monoaurales Signal, zu dem weitere Metadaten vorliegen, z.B. Richtungsinformationen.

HOA benutzt die Winkelkomponenten der Lösung der sphärischen Wellendifferentialgleichung zur Darstellung von Schallfeldern auf einer Kugeloberfläche (s. Anhang A.1). Es ist im Gegensatz zur WFS ein dreidimensionales Verfahren und unabhängig von dem Lautsprecheraufbau auf der Ausgabeseite. Weitere Details zu HOA folgen in Abschnitt 2.4.

### 2.1.3 Objektbasierte Verfahren

In diesem Verfahren werden zum Aufbau einer Szene Objekte übertragen. Auf der Ausgabeseite wird diese Information verwendet um die mit Objekten beschriebene Szene zu synthetisieren.

Ein mögliches Verfahren zur Wiedergabe von Objekten ist Vector Base Amplitude Panning (VBAP). Dieses Verfahren ermöglicht es Quellen unabhängig vom Lautsprecheraufbau im Raum zu positionieren.

Im dreidimensionalen Fall wird dazu ein Vektorraum mithilfe der drei am nächsten zur Wunschposition liegenden Lautsprecher aufgespannt. Das Verfahren passt die Lautstärken für jeden der drei beteiligten Lautsprecher an, damit das Objekt an der Wunschposition wahrgenommen wird [40]. VBAP wird nicht nur für die Synthese von Objekten verwendet, sondern allgemein als Verfahren, um Quellen im Raum abzubilden.

Objekte sind Gegenstand aktueller und vergangener Diskussionen in den Standardisierungsgremien. Die Moving Picture Experts Group (MPEG) beschäftigt sich mit einem Objektansatz seit dem MPEG-4 Standard [33].

#### 2.1.4 Mischverfahren

Die oben beschriebenen Verfahren lassen sich im Mischbetrieb verwenden, um ein Maximum an kreativer Flexibilität zu erhalten.

Prominentes Beispiel ist Dolby® Atmos™. Neben der Einführung von Deckenlautsprechern erlaubt das Format 128 Spuren, die jeweils entweder einen Kanal repräsentieren oder ein Objekt enthalten. Zur Wiedergabe werden diese dann zu maximal 64 Ausgangskanälen zusammgeführt und an die Lautsprecher ausgegeben [20].

Bei den Verfahren WFS und HOA ist es ebenfalls möglich Objekte einzufügen und anschließend zu einem Gesamtschallfeld zu superpositionieren.

Je weiter die Funktionalitäten der kanalbasierten Verfahren erhöht werden, desto deutlicher stellt sich die flexible Reproduktion eines Schallfelds als gemeinsames Ziel aller Verfahren dar.

Hier spielt sich die Stärke der szenenbasierten Verfahren aus: Sie besitzen einen maximalen Abstraktionsgrad vom Zielmedium und sind damit sehr flexibel verwendbar.

HOA hat den besonderen Vorteil, dass es im gesamten Prozess einer Produktion verwendet werden kann: Angefangen mit der Aufnahme des Schallfelds mit sphärischen Mikrofon-Arrays, in der Produktion durch Platzierung virtueller Objekte im Raum mittels sphärischer Harmonischer, über die Übertragung des Schallfelds und schlussendlich bei der Wiedergabe auf beliebigen Lautsprecheranordnungen.

Ein mögliches Zielsystem von HOA sind Kopfhörer. Vorteil bei der Wiedergabe über Kopfhörer ist, dass alle Randbedingungen der Wiedergabe vorgegeben werden können oder a priori bekannt sind.

Im Folgenden sollen die prinzipielle Mechanismen des menschlichen Hörens betrachtet werden. Im Anschluss wird die binaurale Synthese verwendet, um Raumklang auf Kopfhörern abzubilden.

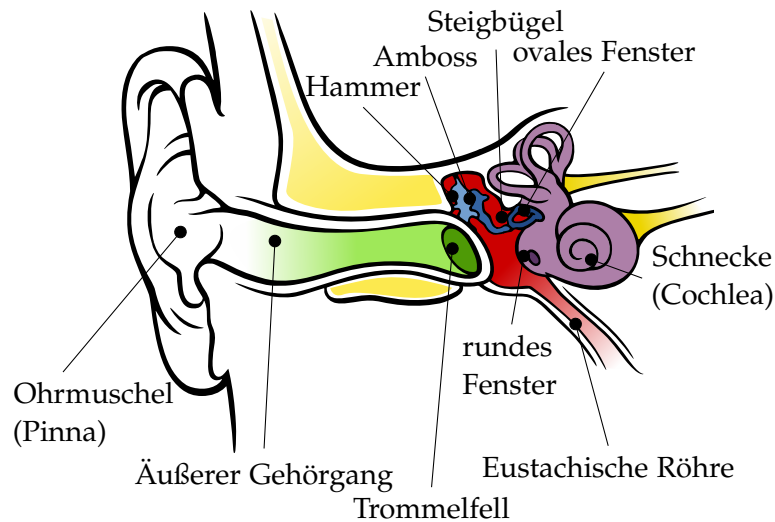


Abbildung 1: Menschliches Hörorgan (Quelle: Basiert auf Illustration vom Wikipedia-Autor Sgbeer)

## 2.2 BINAURALES HÖREN

Das menschliche Hörorgan kann als komplexer elektroakustischer Wandler angesehen werden. Schallwellen treffen über das Außenohr, bestehend aus Ohrmuschel und dem äußeren Gehörgang, aufs Trommelfell. Das Trommelfell koppelt das Außenohr mit dem Mittelohr. Drei kleine, mechanisch verbundene Knochen – Hammer, Amboss und Steigbügel – leiten die Auslenkungen des Trommelfells an den Schalleingang des Innenohrs – dem ovalen Fenster – weiter und verstärken sie [30, S. 213]. Das Innenohr ist mit einer Flüssigkeit ausgefüllt, die die Schwingungen des ovalen Fensters an die Schnecke, auch bekannt als Cochlea, weiterleitet.

Die Cochlea besteht aus einer aufgewickelten Röhre, die in ihrer Mitte durch die häutige Schnecke (Ductus cochlearis) getrennt wird. Der obere Teil, begrenzt durch die Reissner-Membran, wird als Vorhoftrappe (Scala vestibuli), der untere Teil, begrenzt durch die Basilarmembran, als Paukentreppe (Scala tympani), bezeichnet. Beide sind an der Schneckenspitze (Helicotrema) miteinander verbunden. Eine einfallende Schallwelle wandert vom ovalen Fenster durch die Vorhoftrappe und die Paukentreppe. Sie endet am runden Fenster, das die Welle wieder ins Mittelohr auskoppelt.

Der elektrische Schallwandler, um das eingangs erwähnte Prinzip wieder aufzunehmen, wird durch die Basilarmembran repräsentiert. Die Basilarmembran entspricht einer frequenz-kontinuierlichen Filterbank.

Durch eine ortsabhängige Kombination von Masse und Steifigkeit, resoniert an jeder Stelle der Membran eine andere Frequenz. Tiefe Frequenzen resonieren in der Nähe der Schneckenspitze, hohe am Anfang der Membran. Die Wandlung zu elektrischen Signalen, findet

durch feine Härchen statt, die auf der Basilarmembran sitzen. Durch die Schwingung der Membran ändern diese ihre Länge und leiten eine Potenzialänderung an die inneren Härchen weiter, die ihrerseits die Aktionspotenziale für das zentrale Nervensystem generieren.

Es findet somit eine kontinuierliche Abtastung im Frequenzbereich statt. Das Gehirn verarbeitet diese elektrischen Signale weiter. Diese folgende Verarbeitung ist noch nicht komplett verstanden. Die Auditory Scene Analysis (ASA) [14] beschäftigt sich ausgiebig mit dem Verständnis dieser Vorgänge. Computational Auditory Scene Analysis (CASA) [52] fasst diese in Modellen zusammen, die zur digitalen Signalverarbeitung eingesetzt werden können. Gerade im Bereich der verlustbehafteten Signalkodierung (z.B. MP3 oder AAC) ist dieses Wissen enorm hilfreich. Leider gibt es noch kein allgemein gültiges Modell des Hörens, weshalb die einzige Möglichkeit der Evaluation nach wie vor die Durchführung von Hörtests ist.

### 2.2.1 Quellenlokalisierung

Zur weiteren Beschreibung der Effekte wird das Koordinatensystem aus Abbildung 2 verwendet.  $\vartheta$  beschreibt die Elevation beginnend von der Horizontalebene (Elevation  $\vartheta = 0$ ) und  $\varphi$  die azimutale Komponente mit Nullstellung in Blickrichtung nach vorn.

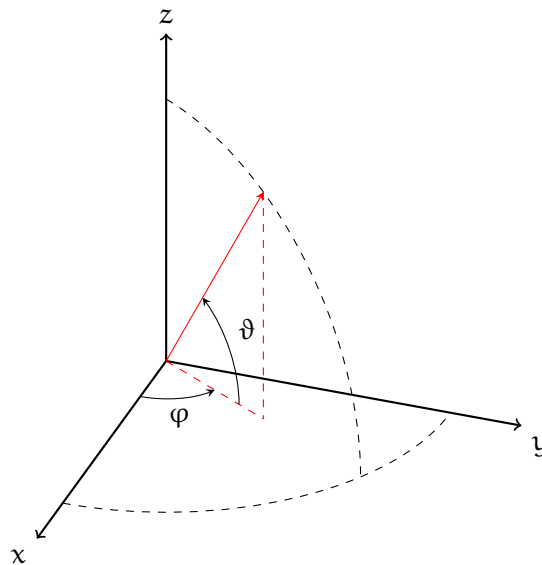


Abbildung 2: Festlegung des Koordinatensystems

Lokalisierung von Quellen in der Horizontalebene (Elevation  $\vartheta = 0^\circ$ ) lässt sich mit der von Lord Rayleigh benannten Duplex-Theorie erklären [44]. Diese besteht im Wesentlichen aus den beiden Komponenten Interaural level difference (ILD) und Interaural time difference (ITD). Zur Lokalisation tiefer Frequenzen wird primär die ITD benutzt. Ab einer Frequenz von 1500 Hz wird die Abschattung des

Kopfes immer dominanter und die ILD wird primär ausgewertet. Die Ohren können als Abtastpunkte im Raum angesehen werden. Bei gegebenem Abstand  $d$  ist die Abtastung für einfallende Wellen mit Wellenlänge  $\lambda$  für

$$d < \frac{\lambda}{2} < \frac{c}{2f} \quad (1)$$

eindeutig. Bei höheren Frequenzen entsteht analog zur Abtastung im Zeitbereich Aliasing. Für einen angenommenen Kopfdurchmesser  $d = 11$  cm ergibt sich für die kritische Abtastung eine maximale Frequenz von 1559 Hz, welche der Duplex-Theorie annähernd entspricht.

Für die Beschreibung von Quellpunkten außerhalb der Horizontalebene ist die Duplex-Theorie nicht mehr ausreichend. Es existieren Punkte gleicher ITD und ILD, die auf Kreisbahnen liegen. Der Radius dieser Kreisbahnen nimmt proportional zum Ohrabstand zu. Da auf diesen entstehenden Zylindern keine eindeutige Ortung nach der Duplex-Theorie mehr möglich ist, werden Sie „Cones of Confusion“ genannt.

Um trotzdem eine Lokalisation zu erzeugen, benutzt das Gehirn neben den Zeit- und Amplitudenunterschieden an den Ohren noch weitere spektrale Eigenschaften. Diese werden maßgeblich durch Reflexionen am Körper, dem Außenohr und den Übertragungsfunktionen der weiteren Elemente im Ohr definiert. Diese Charakteristiken werden in den Head-Related Impulse Response (HRIR), bzw. Head-Related Transfer Function (HRTF) (im Frequenzbereich) als Übertragungsfunktion zusammengefasst [17]. Da das menschliche Ohr als ebenso einzigartig, wie ein Fingerabdruck gilt, sind die HRIR ebenso individuell. Eine Möglichkeit zur Messung dieser Übertragungsfunktionen wird in Abschnitt 2.3.1 vorgestellt.

Die Genauigkeit der Wahrnehmung von Richtungsänderungen ist winkelabhängig. So liegt die höchste Auflösbarkeit des Menschen in der Horizontalebene bei  $\varphi = 0^\circ$ . Dort werden je nach Messart Änderungen von etwa  $1^\circ$  wahrgenommen. Je weiter die zu lokalisierende Quelle nach rechts oder links wandert, desto schlechter wird die Wahrnehmung einer Richtungsänderung. Sie wird zwischen drei bis zehn Mal schlechter als im Frontalbereich [9, S. 41].

Ähnlich verhält sich die Lokalisierbarkeit in der Medianebene ( $\varphi = 0^\circ$  und variablem  $\vartheta$ ). Der Stimulus wird dabei in der Frontalebene, aber auf verschiedenen Elevationsstufen wiedergegeben. Dort zeigt sich für die Frontalebene eine Auflösung von  $9^\circ$ . Mit steigender Elevation steigt diese Ungenauigkeit auf das drei- bis vierfache.

### 2.3 BINAURALE SYNTHESE

Abbildung 3 zeigt das Prinzip des binauralen Hörens einer Schallquelle, welches mithilfe der binauralen Synthese nachgebildet werden kann. Zur Darstellung einer Punktquelle  $x(t)$  in der Horizontalebene



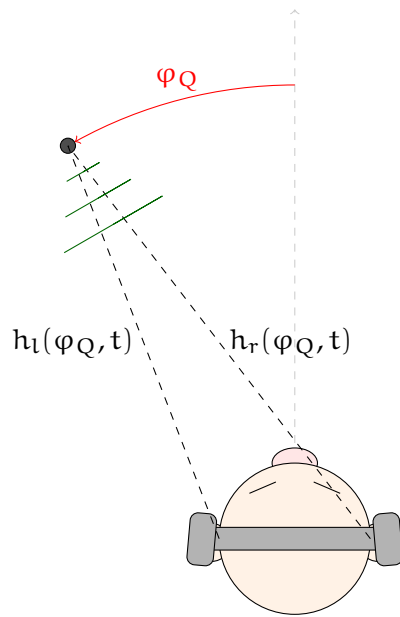


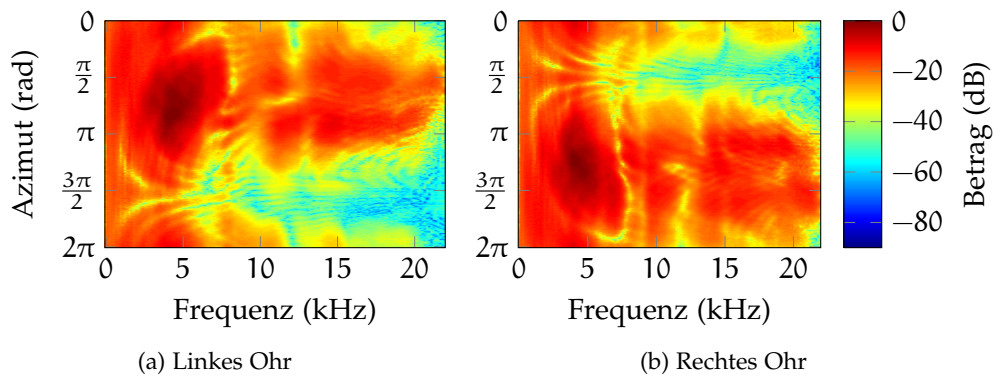
Abbildung 3: Binaurales Hören einer Schallquelle

an der Stelle  $\varphi_Q$ , werden die Ohrsignale durch Faltung (\*) von  $x(t)$  mit den winkelabhängigen HRIR  $h_{\varphi_Q, \{l/r\}}(t)$  generiert. Für die beiden Ohrsignale ergibt sich somit:

$$y_l(t) = x(t) * h_l(\varphi_Q, t), \quad (2)$$

$$y_r(t) = x(t) * h_r(\varphi_Q, t). \quad (3)$$

Zur Berücksichtigung der Kopfbewegungen wird ein Headtracker benutzt. Experimente zeigen, dass ein Headtracker die Vertauschungen, die durch die o.g. „Cones of Confusion“ hervorgerufen werden, signifikant reduziert [8]. Minimale Kopfbewegungen reichen bereits aus, um diese Mehrdeutigkeit der Quellenposition aufzulösen.

Abbildung 4: HRTF für das linke und rechte Ohr für  $\vartheta = 0^\circ$ 

Neben den Betrachtungen von nur einer Quelle im Raum kann diese Anzahl nach belieben vergrößert werden. So ist es möglich z.B.

Lautsprecheranordnungen für Raumklang wie 5.1 oder 7.1 via Kopfhörer wiederzugeben. Dazu werden die Lautsprechersignale mit den kopfbezogenen Übertragungsfunktionen gefaltet und damit im Raum positioniert. In Abbildung 4 sind beispielhaft die Betragsspektren einer HRTF dargestellt. An den Beträgen der Funktionen ist die Abschätzung durch den Kopf zu sehen (z.B. beim linken Ohr beim Azimut-Winkel  $\frac{3\pi}{2}$ ).

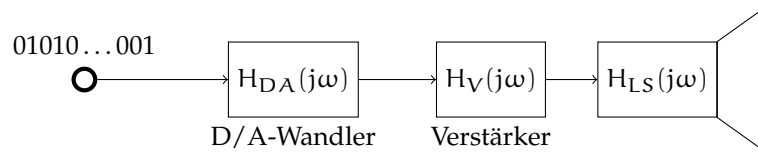


Abbildung 5: Wiedergabe über Lautsprecher

Abbildung 5 zeigt den Signalflussplan der Lautsprecherwiedergabe einer digitalen Bit-Sequenz. Die Bit-Sequenz wird in ein analoges, kontinuierliches Signal gewandelt.  $H_{DA}(j\omega)$  kennzeichnet die Übertragungsfunktion der Digital/Analog Wandler. Die Ausgabe erfolgt über einen Lautsprecher mit Übertragungsfunktion  $H_{LS}(j\omega)$ , der an einem Verstärker mit  $H_V(j\omega)$  betrieben wird. Der Lautsprecher erzeugt eine Druckverteilung, die anschließend vom menschlichen Ohr empfangen wird. Der Druck, der nun am Trommelfell vorherrscht, soll als Referenzdruck  $P_{Ref}(j\omega)$  dienen. Ihn gilt es mittels binauraler Synthese zu reproduzieren [50].

Bei der binauralen Synthese aus Gleichung (2), werden mehrere Faktoren, die das Signal im Vergleich zur Wiedergabe über Lautsprecher zusätzlich beeinflussen, vernachlässigt. Abbildung 6 zeigt den Signalverlauf der binauralen Synthese. Nach Wandlung der Bit-Sequenz in ein analoges Signal, wird die gewünschte Lautsprecher-richtung mittels der passenden HRTF  $H_{HRTF}(\vec{\Omega}_{LS}, j\omega)$  festgelegt. Die Wiedergabe über Kopfhörer verzerrt das Signal mit der Treiberübertragungsfunktion, als auch mit dem Aufbau des Gehäuses und der daraus resultierenden Verfärbungen. Dies wird zusammengefasst als  $H_{KH}(j\omega)$  bezeichnet. Abhängig von der Art, wie die Messung der HRTF erfolgt, beinhaltet sie entweder den Gehörgang oder nicht. Dazu wird das Mikrofon entweder in den Gehörgang eingeführt und blockiert ihn damit oder es wird alternativ ein sehr kleines Sondenmikrofon vor das Trommelfell geschoben. Die Messung des blockierten Gehörgangs wird in den meisten Fällen bevorzugt, da die Ohrkanal-Übertragungsfunktion lediglich ein richtungsunabhängiges Filter ist [36, S. 14].

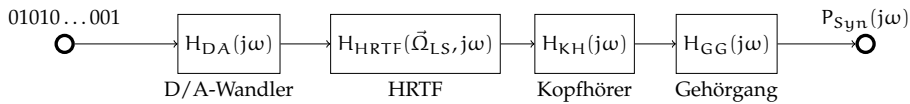


Abbildung 6: Wiedergabe über Kopfhörer

Das vollständig entzerrte Signal erfüllt die Eigenschaft, dass der durch die Lautsprecherwiedergabe erzeugte Druck, dem durch die binaurale Synthese erzeugten Druck entspricht:

$$P_{\text{Ref}}(j\omega) \stackrel{!}{=} P_{\text{Syn}}(j\omega). \quad (4)$$

Die Entzerrung der akustischen Pfade ist schwierig, aber messtechnisch erfass- und korrigierbar.

Ein großer Vorteil der binauralen Synthese ist es, dass sich der Hörer während der Wiedergabe konstant im Sweet-Spot befindet. Es finden keine ungewollten Phasenauslöschungen durch Veränderung der Hörposition statt, die bei der Lautsprecherwiedergabe entstehen können, wenn sich die Hörperson aus dem optimierten Sweet-Spot bewegt.

Weiterhin werden zur Wiedergabe keine Lautsprecher benötigt. Gerade bei zukünftigen Systemen sind eine hohe Anzahl von Lautsprechern gefordert (z.B. 22.2). Weitere Anwendungen finden sich auf Mobilgeräten zur räumlichen Musik- oder Filmwiedergabe.

### 2.3.1 HRTF Messung

Notwendige Bedingung zur Durchführung einer binauralen Synthese ist das Vorhandensein von HRTF für verschiedene Raumwinkel. Diese werden für gewöhnlich in einem schallarmen Raum aufgenommen, um Raumeinflüsse auszuschließen. Im Gegensatz dazu steht die Binaural Room Impulse Response (BRIR), die zusätzlich die Impulsantwort des Raums enthält. Es wird ferner zwischen generischen und individuellen HRTF unterschieden: Generische werden durch messen eines Kunstkopfs (z.B. HEAD acoustics<sup>®</sup> HMS IV, Neumann KU 100 oder KEMAR<sup>®</sup> Manikin) aufgenommen, individuelle dagegen durch messen der Testperson selbst.

Dabei muss die kopfbezogene Übertragungsfunktion aus möglichst jeder Raumrichtung erfasst werden, was je nach Anforderung an die erforderliche Winkelauflösung ein sehr aufwändiger und zeitintensiver Prozess sein kann.

Die hier verwendeten HRTF wurden alle mithilfe eines neuartigen Verfahrens gemessen, das eine Messung der HRTF in acht Elevationsstufen ( $0^\circ, \pm 20^\circ, \pm 40^\circ, \pm 60^\circ, 80^\circ$ ) und mit quasi-kontinuierlicher Azimut-Komponente in einer Dauer von ca. 80 s durchführen kann [21, 53].

Der Messaufbau besteht dabei aus einem Kreisbogen, an dem acht Lautsprecher auf verschiedenen Elevationsstufen angebracht sind. Im gedachten Kreismittelpunkt des Lautsprecherbogens befindet sich ein Stuhl, der auf einem motorisierten Drehteller angebracht ist. Die zu vermessene Person wird dort platziert und mit Messmikrofonen ausgestattet. Es handelt sich hierbei um Messmikrofone, die den Ohrkanal vollständig blockieren, somit nicht den Druck am Trommelfell aufnehmen, sondern am Eingang des Gehörgangs.

Die Messung besteht darin, dass der Drehteller, und damit die Testperson, eine komplette Drehung durchführt, während aus den acht Lautsprechern parallel Testsequenzen abgespielt werden. Bei den Lautsprechersignalen handelt es sich um perfekte Sweeps [48].

Die Erstellung eines perfekten Sweeps erfolgt im Frequenzgang durch Definition von Betrags- und Phasenspektrum. Der Betrag wird bei allen Frequenzen konstant gesetzt. Die Phase  $\Phi(j\omega)$  wird so festgelegt, dass im Zeitbereich ein linearer Anstieg der Frequenz erfolgt. Dies wird erreicht, indem die Gruppenlaufzeit  $\tau_G$  linear über der Frequenz  $f$  ansteigt, woraus durch Integration die Phase gewonnen wird. Anfangs- und Endfrequenz  $f_g$  des Sweeps definieren dabei die entstehenden Integrationskonstanten. Es folgt für die Phase

$$\Phi(f) = - \left( \tau_G(0)f + \frac{\tau_G(f_g) - \tau_G(0)}{2f_g} f^2 \right). \quad (5)$$

Die parallele Messung wird ermöglicht, indem der perfekte Sweep als periodische Folge, phasenversetzt auf den jeweiligen Lautsprechern wiedergegeben wird [10, S. 84].

Die Analyse der Mikrofonensignale erfolgt mittels eines adaptiven Normalize Least Mean Squares (NLMS)-Filters [39, S. 909]. Es kann gezeigt werden, dass perfekte Sequenzen ein optimales Konvergenzverhalten des Algorithmus erzeugen [1].

Die durch den mehrdimensionalen NLMS-Algorithmus geschätzten Impulsantworten lassen sich schreiben als:

$$\hat{h}_n(\varphi_{k+1}) = \hat{h}_n(\varphi_k) + \mu_0 \frac{e(k) \bar{x}_n^T(k)}{\sum_n \|\bar{x}_n(k)\|_2^2}, \quad (6)$$

wobei  $e(k) = y(k) - \hat{y}(k)$  den Adaptionsfehler darstellt und  $n$  einen der in diesem Falle acht Lautsprecher. Nach jedem Eingangssample wird die Impulsantwort angepasst. Dazu wird die  $k$ -te Schätzung genommen und mithilfe des letzten Fehlers korrigiert. Als Eingangsssequenz  $\bar{x}$  werden Referenzaufnahmen ohne den Probanden in der Position der späteren Kopfmittle verwendet, um mögliche Einflüsse des Raumes oder der A/D-Wandler auszuschließen. Vergleiche mit Referenzaufnahmen ergeben, dass die mit diesem Verfahren geschätzten Impulsantworten eine hinreichende Genauigkeit ergeben und somit für die binaurale Synthese verwendet werden können.

Nach einleitenden Beschreibung der binauralen Synthese folgen nun die mathematischen und konzeptionellen Hintergründe zu den hier verwendeten HOA.

#### 2.4 HIGHER-ORDER-AMBISONICS

Die Verwendung von Higher-Order-Ambisonics in der akustischen Signalverarbeitung geht auf Arbeiten von Gerzon und Fellgett zurück [26, 27, 29]. Die ersten Testaufnahmen in der St. Marys Kapelle in Oxford wurden mit einem Mikrofon bestehend aus vier Kapseln mit Nierencharakteristik durchgeführt, die auf den Seiten eines Tetraeders angebracht sind. Diese vier Mikrofon-signale werden als A-Format bezeichnet.

Das A-Format kann durch Umrechnung in vier virtuelle Mikrofon-signale transformiert werden: Ein Kugelmikrofon-signal (s. Abbildung 7a) und drei Mikrofon-signale mit Achter-Charakteristiken, die jeweils in Raumrichtung der Komponenten eines kartesischen Koordinatensystems liegen (s. exemplarisch Abbildung 7b). Diese vier errechneten Signale werden als B-Format bezeichnet.

Das B-Format stellt eine Beschreibung des im Ursprung des Mikro-fons herrschenden Schallfelds dar.

Die zugehörigen theoretischen Überlegungen zum B-Format lassen sich mithilfe der Lösung der Wellendifferentialgleichung

$$\Delta p = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p \quad (7)$$

in sphärischen Koordinaten. Die mathematische Herleitung kann im Anhang A.1 nachgelesen werden.

Die resultierende Lösungsfunktionen zu den Winkelkomponenten  $\vartheta$  und  $\varphi$  (s. Koordinatensystem in Abbildung 2) werden zu den sogenannten sphärischen Harmonischen von Ordnung  $n$  und Grad  $m$  zusammengefasst [54, S. 186]:

$$Y_n^m(\vartheta, \varphi) \equiv \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos \vartheta) e^{jm\varphi}, \quad (8)$$

wobei  $n$  den Grad,  $m$  die Ordnung und  $P_n^m$  die assoziierten Legendre-Funktionen darstellen.

Wichtigste Eigenschaft ist, dass die sphärischen Harmonischen ein orthonormales und komplettes System im  $L^2(\Omega)$  bilden [18, S. 25].

Aufgrund dieser Eigenschaft können diese Funktionen verwendet werden, um jede beliebige Funktion auf einer Kugel mittels Kugel-flächenfunktionen und einer Laplace-Reihe, zu beschreiben [2, S. 761].

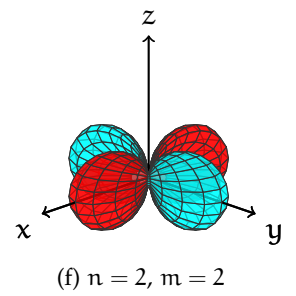
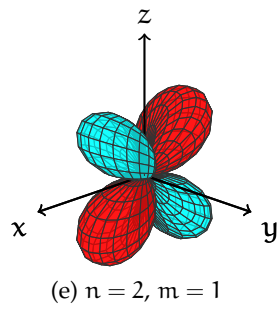
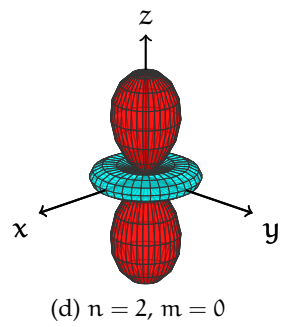
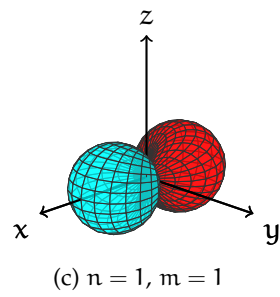
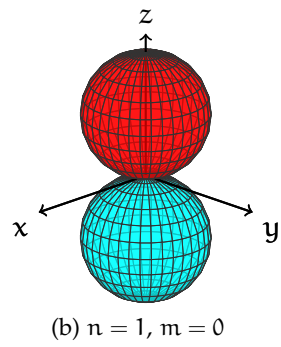
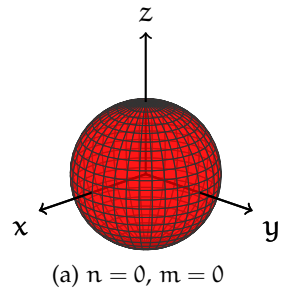


Abbildung 7: Sphärische Harmonische  $\Re\{Y_n^m(\vartheta, \varphi)\}$

Für die Schallfeldsynthese kann der Druck  $p(\vec{\Omega}, j\omega)$  auf einer Kugeloberfläche beschrieben werden als:

$$p(\vec{\Omega}, j\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n (R_{1,nm} j_n(kr) + R_{2,nm} y_n(kr)) Y_n^m(\vartheta, \varphi). \quad (9)$$

Die Lösung teilt sich in die Lösung innerhalb und außerhalb des Lösungsgebietes auf. Für die Beschreibung von quellfreien Gebieten innerhalb der Kugel werden diejenigen Funktionen benutzt, die im Ursprung endlich sind [54, S. 217]. Das sind in diesem Fall die sphärischen Besselfunktionen  $j_n(kr)$ . Das innere Schallfeld lässt sich demnach mit

$$p(\vec{\Omega}, j\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n R_{1,nm} j_n(kr) Y_n^m(\vartheta, \varphi) \quad (10)$$

beschreiben.  $R_{1,nm} j_n(kr)$  wird im Folgenden zu  $\mathring{A}_n^m(kr)$  zusammengefasst und mithilfe der Wellenzahl  $k$  und der Schallgeschwindigkeit  $c$  zu  $\omega = kc$  gesetzt:

$$p(\vec{\Omega}, kc) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \mathring{A}_n^m(kr) Y_n^m(\vartheta, \varphi). \quad (11)$$

Durch die konzeptionelle Analogie zur Fourier-Transformation kann die HOA-Transformation als sphärische Fourier-Transformation bezeichnet werden [42].

Abbildung 7 zeigt die Realteile von  $Y_n^m$  für verschiedene Ordnungen  $n$  und Grade  $m$ . Daraus ergibt sich, dass die vier Signale des B-Formats die sphärischen Harmonischen für  $n = 0$  und  $n = 1$  sind (die negativen Grade sind aus Symmetriegründen nicht dargestellt). Deshalb wird das von Gerzon verwendete Tetraedar-Mikrofon als Schallfeld-Mikrofon<sup>1</sup> erster Ordnung bezeichnet. Es beschreibt das Schallfeld nur approximativ, da es die Reihe der Koeffizienten nach  $n = 1$  abbricht. Je höher die Ordnung  $N$  gewählt wird, desto feiner wird die räumliche Auflösung.

Im Folgenden wird das größte vorkommende  $n$  in einer Schallfeldbeschreibung als Ordnung  $N$  bezeichnet. Aus der Ordnung lassen sich für den dreidimensionalen Fall die Anzahl der verwendeten sphärischen Harmonischen berechnen zu  $O = (N + 1)^2$ .

*Für 2D reduziert sich die Anzahl der Koeffizienten auf  $O = 2N + 1$ .*

#### 2.4.1 Transformationsvorschriften

Zusammenfassend sind hier die Transformationsvorschriften aufgelistet.

<sup>1</sup> Kommerzieller Anbieter dieser Mikrofone ist die Firma TSL Professional Products Ltd. (<http://www.soundfield.com/>).

Die Transformation einer beliebigen Funktion  $f(\vartheta, \varphi)$  in den Bereich der sphärischen Harmonischen ist definiert als eine gewichtete Integration über die Kugeloberfläche  $S^2$  [54, S. 192]:

$$\mathring{A}_n^m = \int_{\vec{\Omega} \in S^2} f(\vartheta, \varphi) \Psi_n^m(\vartheta, \varphi) d\vec{\Omega}, \quad (12)$$

mit  $\Psi_n^m(\vartheta, \varphi) = Y_n^m(\vartheta, \varphi)^*$ , bezeichnet als Mode.

Die Rücktransformation der Koeffizienten  $\mathring{A}_n^m$  ist gegeben durch:

$$f(\vartheta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \mathring{A}_n^m Y_n^m(\vartheta, \varphi). \quad (13)$$

### 2.4.2 Weitere Eigenschaften

**ORTHOGONALITÄT** Die sphärischen Harmonischen  $Y_n^m$  bilden ein Orthonormalsystem, da gilt:

$$\langle Y_n^m(\vartheta, \varphi), Y_{n'}^{m'}(\vartheta, \varphi) \rangle = \int_0^{2\pi} \int_0^\pi Y_n^m(\vartheta, \varphi) Y_{n'}^{m'}(\vartheta, \varphi)^* \sin(\vartheta) d\varphi d\vartheta \quad (14)$$

$$= \delta_{nn'} \delta_{mm'} \quad (15)$$

mit dem Kronecker-Delta

$$\delta_{\alpha\alpha'} = \begin{cases} 1 & \text{für } \alpha = \alpha' \\ 0 & \text{sonst.} \end{cases} \quad (16)$$

**SYMMETRIE** Die sphärischen Harmonischen folgen der Symmetriebeziehung (ohne Beweis) [54, S. 191]

$$Y_n^{-m}(\vartheta, \varphi) = (-1)^m Y_n^m(\vartheta, \varphi)^*, \quad (17)$$

wobei  $(\cdot)^*$  das konjugiert Komplexe darstellt.

### 2.4.3 Darstellung als reelle Koeffizienten

Die sphärischen Harmonischen können reellwertig dargestellt werden. Im Folgenden werden diese als  $R_n^m(\vartheta, \varphi)$  bezeichnet. In der Literatur gibt es diverse Definition [38]. Eine davon lautet:

$$R_n^m(\vartheta, \varphi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos \vartheta) \begin{cases} (-1)^m \sqrt{2} \cos(m\varphi) & m > 0 \\ 1 & m = 0 \\ \sqrt{2} \sin(m\varphi) & m < 0 \end{cases} \quad (18)$$



Die reellen Koeffizienten lassen sich über

$$Y_n^m(\vartheta, \varphi) = \begin{cases} \frac{(-1)^m}{\sqrt{2}} [R_n^m(\vartheta, \varphi) + jR_n^{-m}(\vartheta, \varphi)] & m > 0 \\ R_n^0(\vartheta, \varphi) & m = 0 \\ \frac{1}{j\sqrt{2}} [R_n^m(\vartheta, \varphi) + jR_n^{-m}(\vartheta, \varphi)] & m < 0 \end{cases} \quad (19)$$

in komplexe Koeffizienten überführen.

Beim Schallfeld handelt es sich um ein Skalarfeld. Jeder Raumpunkt wird durch eine skalare, zeitabhängige Größe, den Druck  $p(\vec{\Omega}, t)$ , beschrieben. Zur Abtastung des Schallfelds werden Druckempfänger (Mikrofone) verwendet. Nach der A/D-Wandlung liegt eine reelle Eingangsfolge  $x(t) \in \mathbb{R}$  vor.

Die Darstellung dieses Felds in HOA kann nun mittels komplexer oder reeller sphärischer Harmonischer erfolgen. Bei der Verwendung von reellen Eingangsfolgen (hier:  $x(t)$ ), bieten sich reelle sphärische Harmonische an. Sie nutzen im Gegensatz zum komplexen Pendant lediglich die Hälfte an Speicherplatz und minimieren den Rechenaufwand. Aus diesem Grund werden in den hier vorgestellten Implementierungen durchgehend reelle sphärische Harmonische benutzt. Die Herleitungen werden aus Gründen einer möglichst allgemeinen Darstellung im Komplexen durchgeführt.

#### 2.4.4 Ebene-Wellen-Zerlegung

Das Schallfeld innerhalb und außerhalb einer Kugel kann durch eine unendliche Anzahl ebener Wellen beschrieben werden, die von außen aus allen Raumrichtungen auf die Kugel treffen [43]. So kann ein beliebiges Schallfeld  $P$ , mit den HOA-Koeffizienten  $\mathring{P}_n^m(kr)$ , als gewichtete Überlagerung von ebenen Wellen, dargestellt als Koeffizienten  $\mathring{E}_n^m(\vartheta, \varphi)$ , beschrieben werden.

Dies lässt sich als Integration über alle Raumrichtungen auffassen, die mit der Funktion  $w(\vartheta, \varphi)$  gewichtet werden:

$$\mathring{P}_n^m = \int_{\vec{\Omega} \in S^2} w(\vartheta, \varphi) \mathring{E}_n^m(\vartheta, \varphi) d\vec{\Omega}. \quad (20)$$

Die HOA-Koeffizienten einer ebenen Welle aus der Richtung  $(\vartheta, \varphi)$  können dargestellt werden als [54, S. 227]:

$$\mathring{E}_n^m(\vartheta, \varphi) = 4\pi i^n Y_n^m(\vartheta, \varphi)^*, \quad (21)$$

mit  $i$  als imaginäre Einheit.

Einsetzen von (21) in (20) resultiert in

$$\mathring{P}_n^m = 4\pi i^n \int_{\vec{\Omega} \in S^2} w(\vartheta, \varphi) Y_n^m(\vartheta, \varphi)^* d\vec{\Omega}, \quad (22)$$

*Die weiteren Betrachtungen werden für eine Frequenz  $kr$  und ohne Einheiten vorgestellt.*

wobei der Term unter dem Integral die HOA-Transformation von  $w(\vartheta, \varphi)$  darstellt und somit geschrieben werden kann zu:

$$\mathring{p}_n^m = 4\pi i^n \mathring{w}_n^m. \quad (23)$$

Die Gewichte der ebenen Wellen können mit

$$\mathring{w}_n^m = \frac{\mathring{p}_n^m}{4\pi i^n}. \quad (24)$$

ermittelt werden.

#### 2.4.5 Schallfeldsynthese

Zur Synthese des Schallfelds mit einer begrenzten Anzahl von  $L$  Lautsprechern, kann das Mode-Matching Verfahren verwendet werden. Jeder Lautsprecher wird durch einen Richtungsvektor  $\vec{\Omega}_l$ ,  $l \in [0, L-1]$  beschrieben.

Ausgehend von Gleichung (22) werden nun statt der aus unendlich vielen Richtungen einfallenden ebenen Wellen, nur die Lautsprecherrichtungen benutzt.

Das Integral wird durch eine Summe approximiert:

$$\frac{\mathring{p}_n^m}{4\pi i^n} = \sum_{\vec{\Omega}_l} w_l(\vartheta, \varphi) Y_n^m(\vartheta, \varphi)^*. \quad (25)$$

Dies beschreibt ein lineares Gleichungssystem mit  $L$  Gleichungen, dass für alle  $(N+1)^2$  Koeffizienten gelöst wird. Zur Lösung wird eine Ordnung  $N \geq (\sqrt{L}-1)$  benötigt.

Die Lösung stellen die Gewichtsfunktionen  $w_l(\vartheta, \varphi)$  dar, die als Panning-Funktionen der Lautsprecher bezeichnet werden.

Die Einführung der sphärischen Harmonischen soll hier seinen Abschluss finden. HOA bildet eine aussichtsreiche Plattform für kommende Raumklang-Verfahren. Im Zuge der Standardisierung von MPEG-H wird es als szenenbasiertes Verfahren diskutiert. Im Anschluss an diesen Abschnitt werden statistische Verfahren zur Auswertung von Hörtests vorgestellt.

## 2.5 TESTSTATISTIK

Die Evaluation von Effekten bezüglich der auditiven Wahrnehmungen können oftmals nur mit Hörtests durchgeführt werden. Grund dafür ist der Mangel an Modellen, die den gesamten Prozess der Wahrnehmung simulieren könnten.

Mithilfe mathematischer Modellierung und Abstraktion lassen sich Hörtests planen und auf Signifikanz untersuchen.

Es wird zwischen den parametrischen Tests, die sich auf konkrete Größen wie dem Mittelwert oder die Varianz stützen, sowie den nicht-parametrischen Tests unterschieden.

Ein typischer Testablauf beginnt mit der Aufstellung einer Null-Hypothese  $H_0$ , gefolgt von einer alternativen Hypothese  $H_a$ . Meist stellt die Null-Hypothese den zu widerlegenden Fall dar [7, S. 154].

### 2.5.1 Terminologie

Das Testdesign bedient sich der Terminologie der mathematischen Statistik. Die mathematischen Grundlagen sind u.a. in [11, 35, 37] aufgeführt.

In Vorbereitung auf die nachfolgende Einführung der verschiedenen Testarten seien hier die wichtigsten Termini und Definitionen gegeben. Der Übersicht halber sind die englischen Begriffe in Klammern beigelegt.

**GRUNDGESAMTHEIT (POPULATION)** Die Grundgesamtheit beschreibt eine Menge  $P$  bestehend aus Einheiten mit teilweise übereinstimmenden Eigenschaften, die sie eindeutig zu einer Gruppe zuordnen (z.B. alle existierenden Kopfhörertypen).

**STICHPROBE (SAMPLE)** Eine Stichprobe ist ein (zufällig) gewähltes Element aus der Menge der Grundgesamtheit. Die Menge  $S$  aller Stichproben ist somit eine Teilmenge der Grundgesamtheit  $P$ :  $S \subseteq P$ .

**FREIHEITSGRAD (DEGREES OF FREEDOM)** Der Freiheitsgrad einer Stichprobenverteilung bestimmt die Anzahl derjenigen Werte, welche frei verändert werden können, ohne dass sich das arithmetische Mittel ändert. Bei einer Stichprobenmenge mit  $N$  Stichproben können  $N - 1$  Werte frei variiert werden, der letzte bestimmt sich allerdings durch die vorherige Variation, da sie die Änderungen zur Beibehaltung des Mittelwertes kompensieren muss.

**BEOBSACHTUNG (OBSERVATION)** Eine Beobachtung beschreibt den Wert oder Zustand einer Variable zu einem bestimmten Zeitpunkt.

**STICHPROBENMITTEL** Das Stichprobenmittel ist das arithmetische Mittel:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (26)$$

**STICHPROBENVARIANZ** Die Stichprobenvarianz ist definiert als:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (27)$$

In Abschnitt 2.5.2 wird diese ausführlich hergeleitet.

**SKALENTYPEN** Der Skalentyp ist eine Eigenschaft einer Variable. Diese kann metrisch (z.B. diskreter Fehlerwert) oder nominal sein (kategorial genannt, z.B. Kopfhörer- oder Lautsprecherwiedergabe).

### 2.5.2 Schätzfunktionen

Aufgrund der nur teilweisen Kenntnis der Grundgesamtheit durch die Stichproben sind die statistischen Momente dieser unbekannt. Durch geeignete Schätzer können diese aber näherungsweise errechnet werden.

Sei  $\Theta$  eine Funktion der statistisch unabhängigen, identisch verteilten Zufallsvariablen  $X_1, \dots, X_n$ :

$$\Theta = f(X_1, \dots, X_n), \quad (28)$$

die für jede Stichprobe  $x_1, \dots, x_n$  den Schätzwert

$$\hat{\vartheta} = f(x_1, \dots, x_n) \quad (29)$$

liefert [37, S. 492].

Der optimale Schätzer  $\Theta$  erfüllt folgende Kriterien:

- Erwartungstreue:  $E[\Theta] = \vartheta$
- Konsistenz:  $\lim_{n \rightarrow \infty} \Theta = \vartheta$
- Effizienz: Varianz von  $\Theta$  ist unter allen Schätzfunktionen mit gleichem Stichprobenumfang minimal.

Für die Parameter der meisten Verteilungsfunktionen interessieren insbesondere der Mittelwert- und die Varianzschätzung.

#### 2.5.2.1 Mittelwertschätzer

Als Mittelwertschätzer wird üblicherweise das arithmetische Mittel verwendet:

$$\mu \approx \hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (30)$$

*Beweis der Erwartungstreue.* Werden alle Stichproben  $x_i$  zufällig aus der Grundgesamtheit gezogen, so ist der jeweilige Mittelwert  $\hat{\mu}$ . Daraus folgt für die Erwartungstreue:

$$E[\hat{\mu}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} E\left[\sum_{i=1}^n x_i\right] = \frac{1}{n} [n\hat{\mu}] = \hat{\mu}. \quad (31)$$

□

Die zugehörige Schätzfunktion lautet:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad (32)$$

mit welcher für jede Stichprobe ein Schätzwert prädiziert werden kann.

### 2.5.2.2 Varianzschätzer

$$\sigma \approx \hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (33)$$

mit der zugehörigen Schätzfunktion

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (34)$$

An dieser Stelle wird durch  $(n-1)$  und nicht durch  $n$  dividiert, um einen erwartungstreuen Schätzer zu erhalten. Diese Anpassung ist in der Literatur als Besselsche Korrektur bekannt.

*Beweis der Erwartungstreue.*

$$E[S^2] = E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{1}{n-1} \sum_{i=1}^n E[(X_i - \bar{X})^2] \quad (35)$$

Nebenrechnungen:

$$E[(X_i - \bar{X})^2] = E[((X_i - \mu) - (\bar{X} - \mu))^2] \quad (36)$$

$$= E[(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2] \quad (37)$$

$$= \sigma^2 - \underbrace{2E[(X_i - \mu)(\bar{X} - \mu)]}_{=A} + \underbrace{E[(\bar{X} - \mu)^2]}_{=B} \quad (38)$$

$$A = 2E[(X_i - \mu)(\bar{X} - \mu)] = 2E[(X_i - \mu)]E[(\bar{X} - \mu)] \quad (39)$$

$$= 2E[(X_i - \mu)]E\left[\frac{1}{n} \left(\sum_{j=1}^n X_j - n\mu\right)\right] \quad (40)$$

$$= \frac{2}{n} \sum_{j=1}^n \underbrace{E[(X_i - \mu)(X_j - \mu)]}_{\text{Kovarianz}} = \frac{2}{n} \delta_{ij} E[(X_i - \mu)(X_j - \mu)] \quad (41)$$

$$= \frac{2}{n} E[(X_i - \mu)^2] = \frac{2}{n} \sigma^2. \quad (42)$$

Die Variablen  $X_i$  und  $X_j$  sind statistisch unabhängig, woraus Unkorreliertheit folgt. Unkorrelierte Variablen besitzen die Kovarianz null.

$$B = E[(\bar{X} - \mu)^2] = E[(\bar{X} - E[\bar{X}])^2] = \frac{1}{n} \sum_{i=1}^n (\bar{X} - E[\bar{X}])^2 \quad (43)$$

$$= \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{n} \sum_{j=1}^n X_j - \frac{1}{n} \sum_{k=1}^n \bar{X}_k \right)^2 \quad (44)$$

$$= \frac{1}{n^2} \sum_{i=1}^n \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}. \quad (45)$$

Einsetzen aller Teilergebnisse in (35) ergibt:

$$E[S^2] = \frac{1}{n-1} \sum_{i=1}^n E[(X_i - \bar{X})^2] \quad (46)$$

$$= \frac{1}{n-1} \sum_{i=1}^n \left( \sigma^2 - \frac{2}{n} \sigma^2 + \frac{\sigma^2}{n} \right) = \frac{n\sigma^2}{n-1} \left( 1 - \frac{1}{n} \right) \quad (47)$$

$$= \frac{n\sigma^2}{n-1} \frac{n-1}{n} = \sigma^2. \quad (48)$$

□

### 2.5.2.3 Parameterschätzung mit der Maximum-Likelihood-Methode

Die Maximum-Likelihood-Schätzregel kann zur Parameterschätzung angewendet werden, wenn ausschließlich die Art der Wahrscheinlichkeitsdichte  $p(x; \vec{\alpha})$  bekannt ist.  $\vec{\alpha}$  beschreibt einen  $k$ -dimensionalen Vektor, der die möglichen Parameter für die jeweilige Dichtefunktion beinhaltet.

Die Verbundwahrscheinlichkeit statistisch unabhängiger Stichproben  $x_1, \dots, x_n$  wird in diesem Zusammenhang als Likelihood-Funktion  $L$  betrachtet:

$$L(\vec{\alpha}) = \prod_{i=1}^n p(x_i; \vec{\alpha}). \quad (49)$$

Am Maximum dieser Funktion besteht die höchste Wahrscheinlichkeit, oder anders formuliert die beste Anpassung der Parameter an die gegebene Verteilungsfunktion.

Erreicht wird dieses Maximum über Gradientenbildung der Likelihood-Funktion:

$$\nabla \ln L(\vec{\alpha}) \stackrel{!}{=} 0. \quad (50)$$

Die Bildung des Gradienten führt auf ein lineares Gleichungssystem mit  $k$  unabhängigen Gleichungen.

Exemplarisch sei dies für eine Normalverteilung

$$p(x, \mu, \sigma^2) = N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (51)$$

Das Logarithmieren ermöglicht rechnerische Vorteile bei exponentiellen Dichtefunktionen [35, S. 105].

aufgezeigt. Die zugehörige Likelihood-Funktion ist nach Umformung durch

$$L(\mu, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{(x_1-\mu)^2+\dots+(x_n-\mu)^2}{2\sigma^2}} \quad (52)$$

gegeben. Die partiellen Ableitungen errechnen sich zu

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \stackrel{!}{=} 0 \quad (53)$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 \stackrel{!}{=} 0. \quad (54)$$

Die Anwendung der notwendigen Bedingung ergibt das folgende Gleichungssystem:

$$\frac{1}{\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\mu}) \stackrel{!}{=} 0 \quad (55)$$

$$-\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3} \sum_{i=1}^n (x_i - \hat{\mu})^2 \stackrel{!}{=} 0. \quad (56)$$

Ausführung von Gleichung (55) liefert mit

$$\sum_{i=1}^n (x_i - \hat{\mu}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \hat{\mu} = \sum_{i=1}^n x_i - n\hat{\mu} = 0 \quad (57)$$

$$\Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad (58)$$

den bekannten erwartungstreuen Schätzer aus Abschnitt 2.5.2.1.

Gleichung (56) löst sich zu

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 = s^{*2}. \quad (59)$$

Dieser Schätzer ist nicht erwartungstreu und unterscheidet sich vom erwartungstreuen Varianzschätzer aus Abschnitt 2.5.2.2 durch den konstanten Faktor

$$s^{*2} = \frac{n-1}{n} s^2. \quad (60)$$

Für große  $n$  gilt näherungsweise  $s^{*2} \approx s^2$  [37, S. 507].

### 2.5.3 Regressionsanalyse

Die Regression dient als Basis für eine Vielzahl von statistischen Untersuchungen und Testverfahren.

Das Hauptziel der Regression besteht darin, den Einfluss der abhängigen Variablen  $(x_1, \dots, x_k)$  auf den Mittelwert der Zielgröße genannten unabhängigen Variable  $y$  zu untersuchen. Formal ergibt sich der bedingte Erwartungswert zu

$$E[y|x_1, \dots, x_k] = f(x_1, \dots, x_k). \quad (61)$$

Das allgemeine Regressionsmodell ist definiert als

$$y = E[y|x_1, \dots, x_k] + \epsilon = f(x_1, \dots, x_k) + \epsilon, \quad (62)$$

wobei  $\epsilon$  die nicht von den unabhängigen Variablen erklärte Abweichung vom Erwartungswert ist - die sog. Störgröße [23, S. 19].

Ziel ist es, die Zielvariable möglichst genau schätzbar zu machen, um so eine möglichst optimale Modellierung des gegebenen Sachverhaltes zu erhalten.

Im Folgenden werden die lineare Regression und die Varianzanalyse als Analysewerkzeuge zur Prüfung der Abhängigkeit einer Variablen zu mehreren unabhängigen, vorgestellt.

### 2.5.3.1 Lineare Regressionsmodelle

Bei den linearen Modellen ergibt sich die unbekannte Funktion  $f$  als gewichtete, additive Überlagerung der unabhängigen Variablen, mathematisch fassbar als

$$f(x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad (63)$$

wobei die Konstante  $\beta_0$  in statistischen Programmen als Intercept bezeichnet wird.

Betrachtet man nun  $N$  Beobachtungen der Zielvariable und fasst diese, sowie die restlichen Parameter zusammen, gelangt man zu folgender Darstellung:

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad \vec{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix}, \quad \vec{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{pmatrix} \quad (64)$$

mit der Designmatrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{kN} & \cdots & x_{kN} \end{pmatrix}, \quad (65)$$

wobei die Einsen in der ersten Spalte der Beibehaltung der Konstante  $\beta_0$  dienen.

Damit lässt sich das gesamte lineare Regressionsmodell zu

$$\vec{y} = \mathbf{X}\vec{\beta} + \vec{\epsilon} \quad (66)$$



zusammenfassen.

Zusätzlich werden folgende Annahmen für die Störgröße  $\vec{\epsilon}$  hinzugefügt:

1.  $E[\vec{\epsilon}] = \vec{0}$
2.  $\text{Var}[\vec{\epsilon}] = \sigma^2 \mathbf{I}$  (homoskedastisch)
3.  $\text{Cov}[\vec{\epsilon}] = E[\vec{\epsilon}\vec{\epsilon}^T] = \sigma^2 \mathbf{I}$ , d.h. die Störgrößen sind unkorreliert
4.  $\text{rang}(\mathbf{X}) = K + 1$

Es wird von der klassischen linearen Normalregression gesprochen, wenn zusätzlich  $\vec{\epsilon} \sim N(\vec{0}, \sigma^2 \mathbf{I})$  gilt. Diese Annahme ist für die Konstruktion von Konfidenzintervallen und Hypothesentests notwendig [47, S. 240].

Die benötigten Regressionsparameter  $\vec{\beta}$  werden üblicherweise mit der Methode der kleinsten Quadrate geschätzt und werden als  $\vec{\hat{\beta}}$  bezeichnet. Mithilfe dieser kann ein Schätzwert  $\vec{\hat{y}}$  berechnet werden. Die Differenz der Schätzung  $\vec{\hat{y}}$  und des wahren Wertes  $\vec{y}$  bezeichnet man als Residuum

$$\vec{\hat{\epsilon}} = \vec{y} - \vec{\hat{y}} = \vec{y} - \mathbf{X}\vec{\hat{\beta}}. \quad (67)$$

$\vec{\hat{\epsilon}}$  kann somit als Schätzung der Störgröße angesehen werden.

### 2.5.3.2 Methode der kleinsten Quadrate

Wie anfangs erwähnt basiert die Parameterschätzung zumeist auf der Methode der kleinsten Quadrate. Diese lässt sich als Minimierung der Abweichungsquadrate beschreiben:

$$\text{KQ}(\vec{\beta}) = \min \sum_{i=1}^N (y_i - \vec{x}_i^T \vec{\beta})^2 = \min \sum_{i=1}^N \epsilon^2 = \min \epsilon^T \epsilon \quad (68)$$

$$= \min \vec{y}^T \vec{y} - 2\vec{y}^T \mathbf{X}\vec{\beta} + \vec{\beta}^T \mathbf{X}^T \mathbf{X}\vec{\beta}. \quad (69)$$

Die Minimierung erfolgt über

$$\frac{\partial \text{KQ}(\vec{\beta})}{\partial \vec{\beta}} = -2\mathbf{X}^T \vec{y} + 2\mathbf{X}^T \mathbf{X}\vec{\beta} = 0. \quad (70)$$

Die Lösung des KQ-Schätzers ergibt sich zu

$$\vec{\hat{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y}. \quad (71)$$

Zur Übersicht aller zugrunde liegender mathematischen Sätze sei auf [23] verwiesen.

Der Erwartungswert von  $\vec{y}$  lässt sich nun mit

$$E[\vec{\hat{y}}] = \vec{\hat{y}} = \mathbf{X}\vec{\hat{\beta}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{y} = \mathbf{H}\vec{y} \quad (72)$$

schätzen, wobei  $\mathbf{H}$  als Prädiktionsmatrix bezeichnet wird.

*Die erfolgte Umstellung, sowie die Differentiation sind [23, S. 91] entnommen.*

### 2.5.4 Varianzanalyse

Die Varianzanalyse (engl. Analysis of variance (ANOVA)) wird benutzt, um den Einfluss mehrerer Faktoren (unabhängige Variablen) auf eine abhängige Variable zu untersuchen und kann als Spezialfall einer Regressionsanalyse angesehen werden [24, S. 542]. Dabei ist die unabhängige Variable metrisch, die Faktoren können allerdings nominal, also in verschiedenen Stufen, vorliegen. Bezogen auf die späteren Experimente ist der Lokalisierungsfehler metrisch, die Einflüsse der HRTF liegen in den Stufen generisch und individuell vor.

ANOVA stellt eine Verallgemeinerung des Zweistichproben-t-Test dar [32, S. 91]. Mit der ANOVA ist es möglich mehrere Stichproben zu untersuchen, die jeweils von mehreren Faktoren beeinflusst werden, ohne dass sich der  $\alpha$ -Fehler aufaddiert. Der  $\alpha$ -Fehler ist die Wahrscheinlichkeit dafür, dass die aufgestellte Nullhypothese zurückgewiesen wird, obwohl sie eigentlich vorliegt. Bedingung zur Durchführung einer ANOVA an die Stichproben ist, dass die abhängige Variable Normalverteilung aufweist und die Varianzen der Stichproben gleich ist (Varianzhomogenität).

Im Folgenden werden die Berechnungsschritte der ein- und zweifaktoriellen ANOVA vorgestellt.

#### 2.5.4.1 Einfaktorielle ANOVA

Die einfaktorielle ANOVA geht von einer abhängigen und einer unabhängigen Variable aus. Dabei kann die unabhängige Variable (Faktor) in  $N$  Stufen vorliegen (z.B. generisch oder individuelle HRTF). Jede dieser Faktorstufen definiert aus der Gesamtpopulation des Tests eine Untermenge, die diese Bedingung erfüllt (z.B. Lokalisation wurde mit individueller HRTF vollzogen). Diese Untermengen bilden selbst eine Population mit Mittelwerten  $\mu_n$ , wobei  $n$  die Populationsgruppe bezeichnet.

Die Varianzanalyse geht nun von folgenden Hypothesen aus:

$$H_0 : \mu_0 = \dots = \mu_N \text{ Gleichheit aller Populationsmittelwerte}$$

$$H_1 : \mu_k \neq \mu_l \text{ für mindestens ein Paar } k, l \in [1, N].$$

Den Namen bekommt die Varianzanalyse dadurch, dass sie die Variabilität der Populationsmittelwerte als Indiz dafür nimmt, dass  $H_0$  angenommen oder abgelehnt wird. Je höher dabei die Unterschiede der Mittelwerte ist, desto eher wird  $H_0$  abgelehnt. Als zweites Indiz wird die Variabilität der Population herangezogen. Diese Überlegung wird im F-Bruch zusammengefasst zu:

$$F = \frac{\text{Unterschiede der Populationsmittelwerte}}{\text{Variabilität in der Population}}. \quad (73)$$

Dieser empirisch ermittelbare Wert lässt sich mit

$$F_{\text{empirisch}} = \frac{MS_A}{MS_W} \quad (74)$$

errechnen, wobei  $MS_A$  die mittlere Quadratsumme zwischen und  $MS_W$  die mittlere Quadratsumme innerhalb der Gruppen kennzeichnet. Quadratsummen werden mithilfe des Gruppenmittelwertes

$$\mu_n = \frac{1}{B_n} \sum_{i=1}^{N_n} y_{in} \quad (75)$$

mit  $B_n$  als Stichprobenumfang und  $y_{ij}$  als Beobachtungswert und des Gesamtmittelwertes aller Beobachtungen  $B$

$$\mu = \frac{1}{B} \sum_{j=1}^N \sum_{i=1}^{B_n} y_{ij} \quad (76)$$

zu

$$SS_A = \sum_{j=1}^N B_j (M_j - M)^2 \quad (77)$$

$$SS_W = \sum_{j=1}^N \sum_{i=1}^{B_n} (y_{ij} - M_j)^2 \quad (78)$$

berechnet.  $SS_A + SS_W$  wird als Gesamtstreuung bezeichnet. Die mittleren Quadratsummen ergeben sich durch Normierung auf den Zähler- bzw. Nennerfreiheitsgrad:

$$MS_A = \frac{SS_A}{N-1} \quad (79)$$

$$MS_W = \frac{SS_W}{B-N}. \quad (80)$$

Ähnlich wie beim zentralen Grenzwertsatz der Normalverteilung, kann eine zentrale F-Verteilung bestimmt werden, wenn aus einer normalverteilten Population  $l$  Stichproben entnommen werden und der empirische F-Wert berechnet wird. Die Verteilung die sich einstellt ist in Abbildung 8 dargestellt. Sie hängt ab von den  $s$  Zähler- und  $t$  Nennerfreiheitsgraden.

Die Entscheidung, ob  $H_0$  bestätigt oder abgelehnt wird, wird durch Vergleich des empirischen F-Wertes mit dem der F-Verteilung bestimmt:

$$F_{\text{empirisch}} \geq F_{N-1, B-n; \alpha}, \quad (81)$$

wobei  $\alpha$  dem bereits bekannten Fehler 1. Art entspricht.

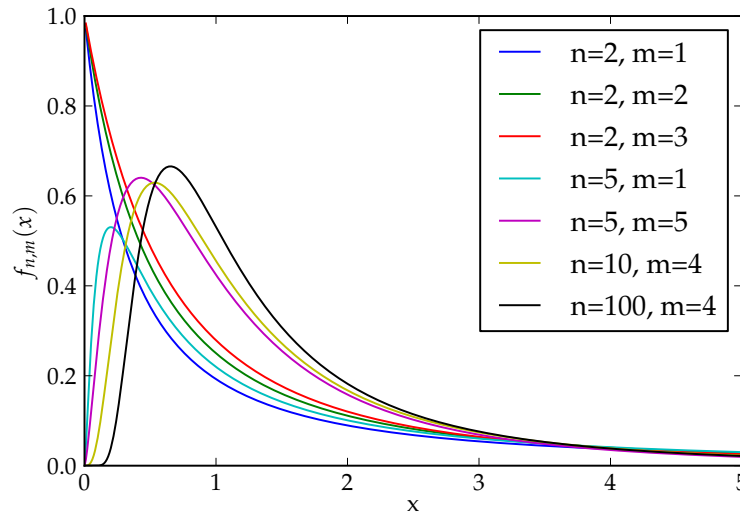


Abbildung 8: Wahrscheinlichkeitsdichtefunktion der F-Verteilung für verschiedene Freiheitsgrade

#### 2.5.4.2 Zweifaktorielle ANOVA

Ausgehend vom Modell der einfaktoriellen ANOVA lässt sich die zweifaktorielle darstellen als [4, S. 82]:

$$y_{ghk} = \mu + a_g + b_h + (ab)_{gh} + \epsilon_{ghk}, \quad (82)$$

wobei  $\mu$  den Mittelwert der Grundgesamtheit,  $a$  und  $b$  die Haupteffekte,  $\epsilon$  den nicht kontrollierbaren Modellfehler und  $g$ ,  $h$  und  $k$  die Stufen selbiger darstellen.  $(ab)_{gh}$  berücksichtigt mögliche Interaktionen zwischen den Haupteffekten.

Zu den genauen Berechnungen der F-Werte, die analog zur einfaktoriellen ANOVA durchgeführt werden, sei auf die im Abschnitt 2.5.4.1 angegebene Literatur verwiesen.

Hiermit endet die Einführung in die Grundlagen dieser Arbeit. Der nun folgende Abschnitt befasst sich mit der Vorstellung der durchgeführten Hörtests.

Mit den Grundlagen aus dem vorhergehenden Kapitel wird nun ein vorhandener Lokalisierungstest untersucht. Dieser wird im Anschluss in modifizierter Version dargestellt und konzeptioniert.

Der Externalisierungstest in Abschnitt 3.3 ist ein neu gestalteter Hörtest. Er untersucht, ob das Hinzufügen von Rauminformationen den Grad der Externalisierung bei binauraler Synthese erhöht.

### 3.1 LOKALISIERUNGSTEST

Dieser Test untersucht die Lokalisierbarkeit eines Stimulus in Abhängigkeit der HOA-Ordnung, des verwendeten Systems zur binauralen Synthese und individueller bzw. generischer HRTF. Letztere wurden mit dem beschriebenen System aus Abschnitt 2.3.1 gemessen.

#### 3.1.1 Testablauf

Der Testablauf ist wie folgt: Der Proband sitzt in einem Abhörraum auf einem Drehstuhl, legt den Kopfhörer an, an dem ein Headtracking-Modul angebracht ist, und steuert den Hörtest über Tastatur und Bildschirm, die sich vor ihm auf einem Tisch befinden.

Das Testsignal ist ein im Frequenzbereich bandbegrenzter Sprachstimulus, bezeichnet als akustischer Zeiger, der bei gegebener Elevation  $\vartheta_p$  zufällig an den Azimut-Winkeln

$$\varphi_p \in \{0^\circ, \pm 30^\circ, \pm 45^\circ, \pm 60^\circ, \pm 90^\circ\}$$

auf den Kopfhörern abgespielt wird. Der Zeiger ist kopffest, d.h. durch Kopffrotation lässt sich dieser in der Ebene im umgebenen Raum positionieren.

Mit kurzem zeitlichen Versatz folgt der identische Stimulus an einer zufälligen Position  $\varphi_a \in [-45^\circ, +45^\circ]$  – bezeichnet als akustischer Anker. Der Anker ist raumfest, verharnt also bei Kopffrotation an der vorgegebenen Stelle im Raum. Abbildung 2 verdeutlicht die Winkelzählung.

Bei dem Stimulus handelt es sich um eine Sprachsequenz, die in Schleife abgespielt wird. Durch die Länge der Stimuli ergibt sich, dass sich Zeiger und Anker in der Wiedergabe überlappen.

Ziel soll sein, durch Kopffrotation den Zeiger auf die Position des Ankers zu bewegen. Die Kopfdrehung wird dabei durch den Headtracker verfolgt. Ein möglich entstehender Winkelfehler wird als Indi-

kator zur Lokalisierungsgenauigkeit der Testperson unter den gegebenen Bedingungen benutzt.

Zur Untersuchung der HRTF-Individualisierung wird zufällig zwischen individueller und generischer HRTF gewechselt.

Der ursprüngliche Test beinhaltet drei Testsysteme zur Generierung der binauralen Kopfhörersignale [6]:

1. HRIR-Referenzsystem: In Abhängigkeit der Kopfausrichtung wird der Stimulus mit den HRIR gefaltet.
2. HOA-System: Der Stimulus wird mithilfe der sphärischen Harmonischen im Raum platziert und über virtuelle Lautsprecher ausgegeben. Die virtuellen Lautsprechersignale werden wie beim HRIR-Referenzsystem mit den richtungsabhängigen HRIR gefaltet und überlagert, um die binauralen Signale zu generieren. Die HRIR sind dabei abhängig von der aktuellen Kopfausrichtung.
3. HOA-Rotation: Basiert auf dem HOA-System, aber anstelle der Rotation der virtuellen Lautsprecher wird hier das Schallfeld im Bereich der sphärischen Harmonischen gedreht woraus folgt, dass die virtuellen Lautsprecher ihre Position relativ zur Kopfausrichtung konstant halten.

Die HOA-basierten Systeme werden mit Ordnung  $N = 6$  betrieben. Im Folgenden werden die drei Testsysteme im Einzelnen näher erläutert.

Das HRIR-Referenzsystem, dargestellt in Abbildung 9, stellt die direkte binaurale Synthese dar [13, S. 141]. Gegeben ist das Quellsignal  $x_s(t)$  und eine gewünschte Richtung  $\vec{\Omega}_s = (\rho, \vartheta, \varphi)^T$ . Da hier die Annahme des Fernfelds getroffen wird, wird  $\rho$  im Folgenden zwar zur Wahrung der allgemeinen Darstellung notiert, aber nicht weiter berücksichtigt. Mithilfe der Kopfausrichtung  $\vec{\Omega}_H(t)$  kann die Richtung der Quelle im Raum konstant gehalten werden. Zur Generierung der Kopfhörersignale wird das Eingangssignal mit den passenden HRIR gefaltet. Letztere sind für alle Raumrichtungen in einer HRIR-Datenbank hinterlegt.

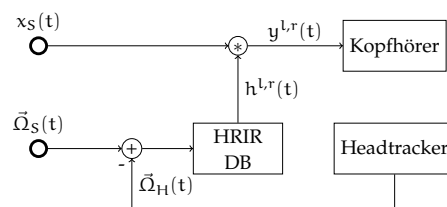


Abbildung 9: System 1, Referenzsystem mit direkter binauralen Synthese

Das HOA-System, dargestellt in Abbildung 10, enkodiert den Stimulus  $x_s(t)$  mit der gewünschten Position  $\vec{\Omega}_s(t)$  in ein HOA-Signal

$\vec{A}(t)$ . Diese wird auf L virtuelle Lautsprecher dekodiert und im Signalvektor  $\vec{a}(t)$  gespeichert. Die L Lautsprecher an den Richtungen  $\vec{\Omega}_{L,i} = (\rho_i, \vartheta_i, \varphi_i)^T$  werden durch 2L HRIR im Raum platziert und müssen bei jeder Kopfbewegung aktualisiert werden, um raumfest zu bleiben. Durch Annahme des Fernfelds wird  $\rho_i$  hier ebenfalls nicht weiter berücksichtigt. Die Ohrsignale  $y^{l,r}(t)$  werden über die ohrrichtige Summation der 2L gefilterten Lautsprechersignale generiert und auf die Kopfhörer ausgegeben.

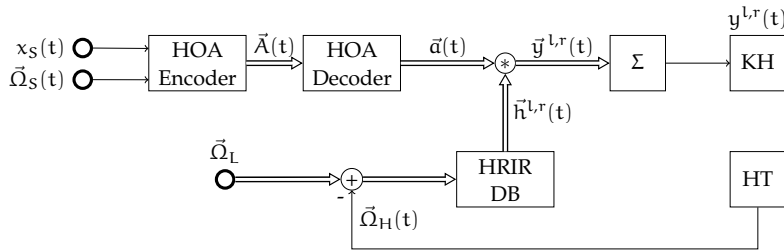


Abbildung 10: System 2, HOA (KH = Kopfhörer, HT = Headtracker)

Das HOA-Rotationssystem, dargestellt in Abbildung 11, hält den Stimulus trotz Kopfbewegung durch Anpassung des akustischen Felds an der gewünschten Stelle. Diese Anpassung geschieht über eine Rotation des akustischen Felds, die in der HOA-Domain effizient durchführbar ist. Es erweitert demnach System 2 durch eine Feld-Rotation und spart somit das Aktualisieren der HRIR ein (Speicherzugriff vs. Rechenoperation).

Die Auswertung in [6] ergibt, dass es wenig Anzeichen für eine Verringerung des Fehlers bei der Verwendung individueller oder generischer HRTF gibt. Zudem ist kein signifikanter Systemunterschied festzustellen. Insbesondere die Unterschiede der zwei HOA-Systeme ist vernachlässigbar gering. Damit ist es möglich das kostengünstige binaurale System 3 für weitere Untersuchungen zu wählen. Die HOA-Systeme spielen einen weiteren Vorteil aus, sobald der wiederzugebene Inhalt enkodiert in HOA vorliegt, womit der Enkodierungsschritt wegfallen würde und lediglich das Rendering durchgeführt werden müsste.

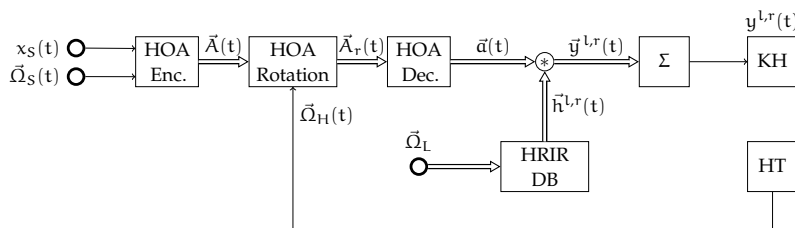


Abbildung 11: System 3, HOA-Rotation (KH = Kopfhörer, HT = Headtracker)

Eine nähere Betrachtung des Testmodus deckt mögliche systematische Fehler auf, die mit den nachstehend formulierten Änderungen minimiert werden sollen.

### 3.2 MODIFIZIERTER LOKALISIERUNGSTEST

Bei der Auswertung des Tests aus Abschnitt 3.1 fällt der relativ geringe Lokalisierungsfehler auf [6]. Durch Modifikation soll der Hörtest nun so gestaltet werden, dass er möglichst isoliert die Effekte einer HOA-Ordnungsvariation untersucht und zusätzlich die Effekte der HRTF-Variation aufzeigt.

Da der vorangegangene Test ergibt, dass es keine relevanten Systemunterschiede gibt, wird im Folgenden nur das Testsystem 3 benutzt (vgl. Abbildung 11). Im Unterschied zum Originaltest werden die Stimuli nur auf der Horizontalebene wiedergegeben (Elevation  $\vartheta = 0^\circ$ ).

Die Modifikationen und erhofften Wirkungsweisen werden im Folgenden kurz erläutert.

#### 3.2.1 *Ordnungsvariation*

Durch Variation der HOA-Ordnung kann die räumliche Auflösung verändert werden. Abbildung 12 zeigt das Winkelspektrum für die Horizontalebene eines HRTF-Datensatzes für das linke Ohr. Es ist folgendermaßen zu interpretieren: Für einen Stimulus mit der Bandbreite  $B$  und zugehöriger Grenzfrequenz  $f_g$ , werden  $M$  Moden benötigt. Die Abbildung zeigt, dass mit steigender Frequenz auch die Anzahl der benötigten Moden zunimmt, da mit höherer akustischer Variabilität eine höhere räumliche Variabilität einhergeht [22].

In der ersten Version der Hörtests wird ein Sprachstimulus verwendet, dessen Bandbreite  $B_S$  bei etwa 6 kHz liegt. Aus diesem Grund wird der nun verwendete Rauschstimulus ebenfalls auf diese Bandbreite begrenzt. Die eingezeichneten Linien in Abbildung 12 zeigen die benötigten Moden bei der Wahl der Bandbreite  $B_S$ . Die äußeren Punkte weisen die Moden  $\pm 16$  auf. Zusammen mit der Null werden  $(2 \times 16 + 1)$  Moden benötigt, was der Berechnung der Anzahl der Koeffizienten  $O = 2N + 1$  für den zweidimensionalen Fall entspricht. Zur Wiedergabe aller räumlichen Informationen, die bei der gewählten Bandbreite  $B_S$  vorhanden sind, würde also Ordnung  $N = 16$  benötigt werden.

Dieser Test untersucht die Lokalisationsgenauigkeit bei einer räumlichen Begrenzung durch die Ordnungen  $N \in \{1, 3, 6\}$ .



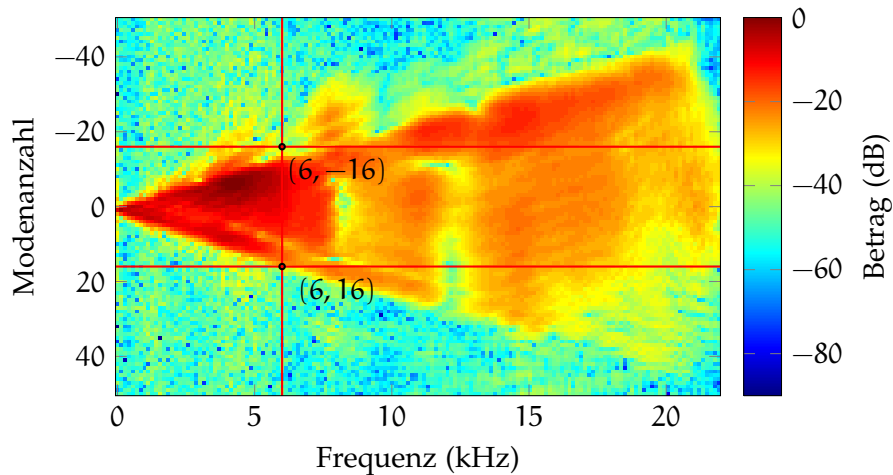


Abbildung 12: Betrag des Winkelspektrums für das linke Ohr

### 3.2.2 Stimuluswahl

Bisher wird eine Sprachsequenz verwendet, dessen Signalenergie über 6 kHz stark abfällt und deshalb als bandbegrenzt angenommen wird. Durch unbewusste oder bewusste Auswertung des Sprachtimbres oder durch gedankliche Synchronisation markanter Stellen im Stimulus ist es möglich, weitere dynamische Informationen aus dem Signal zu erhalten. Diese können herangezogen werden, um kleine räumliche Abweichungen zwischen Zeiger und Anker herauszuarbeiten. Zusammen mit der unbestimmten Anzahl an Stimuluswiederholungen verlängert es zum Einen den Test und zum Anderen können trainierte Testpersonen den Stimulus unter Ausnutzung der o.g. Möglichkeiten sehr genau positionieren.

Um dies zu vermeiden, wird der Sprachstimulus durch 200 ms langes, bandbegrenzttes Rosa-Rauschen ersetzt. Der zeitliche Versatz zwischen Zeiger und Anker beträgt weitere 200 ms. Experimente haben gezeigt, dass dies die minimale Zeit zur Lokalisation ist [51].

### 3.2.3 Stimuluswiederholung

Selbst bei den kurzen Rauschimpulsen ist es möglich, spektrale Verfärbungen wahrzunehmen, allerdings wesentlich schwieriger als beim Sprachstimulus, da die Stimuli zeitlich nicht mehr überlappen. Durch eine Reduktion der Wiederholung der Stimulussequenz wird diese Möglichkeit weiter eingeschränkt. Im Laufe eines Vortests hat sich eine vierfache Wiederholung der Sequenz als ausreichend erwiesen. Danach ist der Hörer gezwungen, eine Eingabe zu tätigen.

### 3.2.4 Zeigerbehandlung

Angelehnt an Hörtests, in denen ein Zeiger in Form eines Stiftes oder Laserpointers benutzt wird, der durch Armbewegung in Richtung Stimulus geführt wird, wird der hier verwendete akustische Zeiger immer im Referenzsystem betrieben. Das hat zur Folge, dass der Zeiger mit der maximalen räumlichen Auflösung arbeitet, auch wenn der zu suchende Anker in einem räumlich bandbegrenzten System betrieben wird.

### 3.2.5 Ergonomie

Zur Steigerung der Ergonomie wird statt der Leertaste der Tastatur eine Infrarotfernbedienung benutzt, um die Zielposition zu bestätigen. Vorteil ist, dass der Hörer sich nicht mehr während des Tests zur Tastatur drehen muss, um die Richtung zu bestätigen.

### 3.2.6 Testablauf

Der Test wird von einem Dritten (typischerweise der Testbetreuer) gestartet. Dieser stellt sicher, dass die HOA-Ordnung korrekt eingestellt ist und von der Testperson nicht gesehen wird. Diese Limitierung auf Einfach-Blindheit ergibt sich aus der momentanen Implementierung des Tests, die die Festlegung der Ordnung des Schallfelds nur beim Start zulässt.

Die Testsoftware erzeugt anschließend eine zufällige Verteilung von 72 Teststimuli. Diese ergeben sich aus der Anzahl der Zeiger-Positionen (neun), der zwei HRTF-Varianten (individuell und generisch), sowie der Forderung, jede Zeiger-Position vier mal abzufragen ( $9 \times 2 \times 4 = 72$ ).

Aufgrund der Begrenzung der Stimuluswiederholungen liegt der zeitliche Aufwand pro Test bei 20 bis 30 min.

### 3.2.7 Statistisches Modell

Die Nullhypothesen für diesen Test können wie folgt definiert werden:

1. Eine Reduktion der Ordnung bewirkt keinerlei Änderung auf die Lokalisierbarkeit.
2. Individuell gemessene HRTF haben keinerlei Einfluss auf die Ortbarkeit eines Stimulus.

Zur statistischen Modellierung wird ein additives Modell gewählt. Die beteiligten Haupteffekte dieses Modells lassen sich aus der Analyse der Testparameter ableiten. In diesem Fall werden die HOA-Ordnung

$\eta$ , die Art der HRTF  $\gamma$ , sowie die  $P$  Testpersonen  $\tau$  variiert. Die zu untersuchende abhängige Variable  $e$ , der Winkelfehler, lässt sich demnach modellieren zu:

$$e_{pti} = \mu + \tau_p + \eta_t + \eta_i + \epsilon_{pti}, \quad (83)$$

wobei  $\mu$  den globalen Mittelwert und  $\epsilon$  einen normalverteilten Fehlerterm beschreiben. Die Indizes  $p \in \{1, \dots, P\}$ ,  $t \in \{1, 3, 6\}$  und  $i \in \{\text{individuell, generisch}\}$  beschreiben die unterschiedlichen Testausprägungen.

### 3.3 EXTERNALISIERUNGSTEST

Zur Verbesserung der Externalisation des Audio-Signals wird der hallfreie Stimulus mit Raumimpulsantworten gefaltet. Diese in der Literatur bekannte Methode soll eine signifikante Erhöhung der Externalisierung hervorrufen [8].

Im ersten Schritt des Tests soll eine möglichst gute Reproduktion der Hörsituation auralisiert werden. Die Echtzeitfähigkeit des Systems ist nicht von oberster Priorität. Das bedeutet, dass die Integration eines Headtrackers in der hier beschriebenen Version nicht möglich ist.

Im Unterschied zum Lokalisierungshörtest findet die Signalprozessierung ausschließlich im HOA-Bereich statt. Der Syntheseschritt (auch bekannt als Rendering) wird in Abschnitt 3.3.5 erläutert.

Der prinzipielle Signalfluss des Tests lässt sich in zwei Abschnitte unterteilen:

1. Eine beliebige Monosequenz wird mithilfe einer Raumimpulsantwort (liegt in HOA-Darstellung vor) in den Bereich der sphärischen Harmonischen transformiert. Die zwei verschiedenen Raumimpulsantworten sind zum Einen gemessenem und zum Anderen synthetischem Ursprungs. Das Verfahren wird im folgenden Abschnitt näher erläutert.

Die Monosequenz ist nun in einem Schallfeld eingebettet und erklingt aus einer definierten Richtung.

2. Unter Verwendung der HRTF werden die Ohrsignale produziert und über Kopfhörer ausgegeben.

Dieser Signalfluss und Generierung der Raumimpulsantworten soll im Folgenden näher beleuchtet werden.

#### 3.3.1 Messung der Raumimpulsantwort

Die Raumimpulsantworten werden mithilfe des Eigenmike<sup>®</sup> im Abhörraum 606 der Firma Technicolor aufgenommen. Der akustisch optimierte Raum hat folgende Maße ( $B \times T \times H$ ): 666,5 cm  $\times$  478,9 cm  $\times$  252 cm, also ein Raumvolumen von ca. 80,5 m<sup>3</sup>.

Dort steht ein Aufbau mit  $L = 32$  Lautsprechern zur Verfügung. Der Aufbau besteht aus 32 Lautsprechern, die in drei Ebenen kreisförmig um einen Mittelpunkt (Sweet-Spot) angeordnet sind. Die Anordnung basiert darauf möglichst viele Standard-Aufbauten (Stereo, 5.1...22.2) abzudecken.

Das Eigenmike<sup>®</sup> ist ein von mh acoustics LLC<sup>1</sup> entwickeltes, sphärisches Mikrofonarray mit  $C = 32$  auf der Kugel verteilten Mikrofonkapseln und einem Radius von  $\rho = 4,2$  cm. Mit diesen 32 Abtastpunkten auf der Kugel ist maximal eine HOA-Ordnung von  $N = 4$  möglich (s. Abschnitt 2.4).

Zur Messung und Auswertung wird ein MATLAB<sup>®</sup>-Skript verwendet. Pro Lautsprecher wird ein logarithmischer Sweep zweimal nacheinander abgespielt. Dabei ist die zeitliche Länge des Sweeps  $\gg$  als die Nachhallzeit des Raumes [25].

Im Anschluss an die Aufnahmephase berechnet das Skript die Impulsantworten, indem es das Anregungssignal durch Division im Spektralbereich wieder herausrechnet. Es verwendet dabei die Aufnahme des zweiten Sweeps, da sich zum Einen der Raum im eingeschwungenen Zustand befindet und es sich durch die zyklische Anregung ebenfalls um ein zyklisches Signal handelt, welches die gesamten Informationen des Raumes in sich vereint.

Nach der Messung liegen nun 32-kanalige Raumimpulsantworten vor, die zu 25 HOA-Koeffizienten gewandelt werden. Dieser Prozess ist in [34] beschrieben. Dazu wird angenommen, dass sich die HOA-Koeffizienten  $\hat{P}_n^m(k\rho)$  aus Gleichung (12) zum Einen in einen Streuteil  $b_n(k\rho)$  und zum Anderen in einen Feldanteil  $\hat{A}_n^m(k)$  zerlegen und sich durch eine Summenapproximation  $\hat{P}_n^m(k\rho)$  errechnen lassen zu:

$$\hat{P}_n^m(k\rho) = b_n(k\rho)\hat{A}_n^m(k) \quad (84)$$

$$\hat{P}_n^m(k\rho) = \sum_{c=0}^{C-1} \alpha \hat{P}_n^m(\vec{\Omega}_c, k\rho) Y_n^m(\vec{\Omega}_c)^*, \quad (85)$$

wobei  $\vec{\Omega}_c = (\vartheta_c, \varphi_c)^T$  die jeweilige Kapselrichtungen und  $\alpha = \frac{4\pi}{C}$  ein Faktor zur Anpassung der Energie beschreibt. Für die Streuung an einer schallharten Kugel gilt zudem

$$b_n(k\rho) = \frac{4\pi j^{n+1}}{(k\rho)^2 \left. \frac{dh_n(x)}{dx} \right|_{x=k\rho}}. \quad (86)$$

Mit der Umstellung

$$\hat{A}_n^m(k) = \frac{\hat{P}_n^m(k\rho)}{b_n(k\rho)} \quad (87)$$

<sup>1</sup> <http://www.mhacoustics.com/>

lassen sich die approximierten HOA-Koeffizienten errechnen. Im weiteren Verlauf der Konvertierung wird noch ein adaptives Wiener-Filter zur frequenzabhängigen Rauschreduktion verwendet.

Am Ende der Raummessung liegen  $L = 32$  HOA-Raumimpulsantworten der Ordnung  $N = 4$  vor. Diese beschreiben die Impulsantwort des Raumes aus  $L$  Anregungsrichtungen. Eine abgeschlossene HOA-Darstellung, die alle Anregungsrichtungen durch Superposition enthält ist nun zwar denkbar, mangels einer geeigneten Referenz, die z.B. aus einer 3D-Simulation stammen könnte, wird darauf allerdings verzichtet und diese Limitierung vorerst hingenommen.

### 3.3.2 Synthetische Raumimpulsantwort

Zur Erzeugung eines künstlichen Pendant zum gemessenen Raum wurde die Bibliothek Spatialisateur (kurz: SPAT)<sup>2</sup> des Institut de Recherche et Coordination Acoustique/Musique (IRCAM) benutzt. Diese steht als Erweiterung im Programm Max/MSP<sup>3</sup> der Firma Cycling '74 zur Verfügung. Max/MSP ist ein Programm, welches es ermöglicht, mittels eines Signalflussplans, echtzeitfähige Systeme zu erstellen. Das Programm SPAT erweitert Max/MSP um räumliche Signalverarbeitungsalgorithmen. Es benutzt HOA, um diese Operationen durchzuführen.

SPAT bietet u.a. die Möglichkeit beliebige Quellenanordnungen auf beliebige Lautsprecheraufstellungen auszugeben. Ebenfalls ist eine Binauralprozessierung integriert, die hier aber nicht verwendet wird.

Für diesen Test interessant ist die Option, unter der Angabe von Raumparametern eine Simulation durchzuführen, die als Ausgabe HOA-Koeffizienten zur Verfügung stellt.

Die Idee ist nun diese Raumimpulsantwort aus dem System zu extrahieren. Im Normalfall erhält SPAT Quellsignale und Quellpositionen und transformiert diese unter Verwendung der gegebenen Raumparameter in die den Bereich der sphärischen Harmonischen. Anschließend werden die Lautsprechersignale für die angegebenen Lautsprecher-Positionen berechnet. Die Raumimpulsantworten werden nicht separat ausgegeben.

Bei Verwendung eines bekannten Anregungssignals können diese allerdings nachträglich extrahiert werden. Dazu wurde ein Max/MSP-Patch erstellt, der das Verhalten des oben beschriebenen MATLAB<sup>®</sup>-Messskripts simuliert. Der virtuelle Raum wird demnach mit einer bekannten Sequenz angeregt und das daraus synthetisierte Schallfeld als HOA-Koeffizienten gespeichert.

Im ersten Schritt wird SPAT mit dem Lautsprecher-Aufbau aus dem Technicolor-Raum 606 initialisiert. Zudem werden die Raumparameter zur Raummodellierung gesetzt (Parameter s. B.1). Das Skript

<sup>2</sup> <http://forumnet.ircam.fr/product/spat/>

<sup>3</sup> <http://cycling74.com/company/>

spielt nun nacheinander das Testsignal auf allen gegebenen virtuellen Lautsprechern ab und speichert die HOA-Koeffizienten des von SPAT generierten Schallfelds in einer Mehrkanal-WAV-Datei ab.

Diese Mehrkanal-WAV-Datei hat nun den gleichen Aufbau wie die echte Messung aus Abschnitt 3.3.1. Da als Testsignal derselbe doppelte Sweep wie beim realen Messen verwendet wurde, kann die Auswerterroutine des MATLAB<sup>®</sup>-Messskripts verwendet werden, um die Raumimpulsantworten zu extrahieren.

### 3.3.3 HOA-Transformation der Quelldaten

Die Transformation des Audiostimulus in den Bereich der sphärischen Harmonischen erfordert zum Einen ein Quellsignal, hier der Monostimulus  $x(t)$ , und die Position  $\vec{\Omega}_0 = (\rho_0, \vartheta_0, \varphi_0)$  der Quelle.

Gleichung (12) folgend, ergeben sich die HOA-Koeffizienten für einen Zeitpunkt  $t_0$  zu:

$$\dot{X}_n^m(t_0) = x(t_0)\Psi_n^m(\vartheta_0, \varphi_0). \quad (88)$$

Die Integration reduziert sich auf ein Winkelpaar  $(\vartheta_0, \varphi_0)$ , da es den einzigen Beitrag zu diesem Schallfeld leistet. Diese Gewichtung wird für jedes Eingangssample durchgeführt. Im Folgenden wird die Frequenzdarstellung  $\dot{X}_n^m(j\omega)$  verwendet, die sich durch die Fouriertransformation der Zeitdarstellung ergibt.

Zur Hinzunahme des Raumeinflusses wird angenommen, dass die HOA-Koeffizienten der Raumübertragungsfunktion  $\dot{R}_n^m(j\omega, \vec{\Omega})$  vorliegen. Das Eingangssignal  $x(t)$  wird mittels Fouriertransformation in den Frequenzbereich transformiert und folglich  $X(j\omega)$  genannt. Der enkodierte Stimulus  $\dot{S}_n^m(j\omega)$  berechnet sich zu:

$$\dot{S}_n^m(j\omega) = X(j\omega)\dot{R}_n^m(j\omega). \quad (89)$$

Die Gleichung kann interpretiert werden als Gewichtung der HOA-Koeffizienten der Raumübertragungsfunktion mit den Frequenzgewichten des Eingangssignals.

### 3.3.4 HOA-Transformation der HRTF

Ebenfalls ausgehend von Gleichung (12) werden die HRIR transformiert. Anders als bei der Stimulustransformation, liegen hier nun Beiträge aus allen Raumrichtungen vor. Da keine analytische Darstellung der HRIR vorliegt, wird dieses Integral in eine Summe überführt und stellt damit ein Quadratur-Problem auf der Kugeloberfläche dar. Dabei finden die sog. Fliege-Punkte Anwendung [28]. Sie liefern für jede HOA-Ordnung  $N$ ,  $O = (N + 1)^2$  Punkte  $\vec{\Omega}_{F,i}$ ,  $i \in [0, O - 1]$  auf der Kugeloberfläche, die den Quadraturfehler minimieren sollen. Für jede Richtung  $\vec{\Omega}_{F,i}$  wird nun die passende HRIR  $h_i(\vec{\Omega}_{F,i})$  aus der Datenbank gelesen und wie in Abschnitt 3.3.3 beschrieben transformiert.

Bei der Auswahl der passenden HRIR entsteht ein Fehler, da diese zwar quasi-kontinuierlich auf dem Azimut vorliegen, die Elevation allerdings nur in acht Stufen. Mittels quadratischem Abstands wird die verfügbare HRIR gesucht, die dem geforderten Fliege-Punkt am nächsten ist. Diese nun vorliegenden  $O$  HOA-Koeffizienten  $\mathring{H}_{n,i}^m(t)$ , können nun zu einem Gesamtfeld  $\mathring{H}_n^m(t)$  additiv überlagert werden. Dieses wird im Folgenden ebenfalls im Frequenzbereich betrachtet.

### 3.3.5 Synthese der Audiodaten

Die Synthese der Audiodaten erfolgt über sog. Beamforming [42]. Die Idee ist, eine Kugelflächenfunktion mithilfe einer zweiten Kugelflächenfunktion – dem Beam – zu gewichten und damit Teile des Schallfelds abzuschwächen oder aber zu verstärken.

Mathematisch drückt sich dies als Integration der Gewichtung über die gesamte Kugeloberfläche aus. Die hier verwendete Analogie beschreibt sich dadurch, das es sich bei der Schallquelle um den Audiosignal  $S(j\omega, \vec{\Omega})$  handelt und die Gewichtung durch die HRTF  $H(j\omega, \vec{\Omega})$  beschrieben wird:

$$Y(j\omega) = \int_{\vec{\Omega} \in S^2} S(j\omega, \vec{\Omega}) H(j\omega, \vec{\Omega}) d\vec{\Omega}. \quad (90)$$

$\vec{\Omega}$  ist der Raumwinkel und  $S^2$  ist die Untermenge von  $\mathbb{R}^3$ , die die Einheitskugel beschreibt.

Zur Deutung als inneres Produkt wird die Substitution  $B(j\omega, \vec{\Omega}) = H^*(j\omega, \vec{\Omega})$  verwendet:

$$Y(j\omega) = \int_{\vec{\Omega} \in S^2} S(j\omega, \vec{\Omega}) B^*(j\omega, \vec{\Omega}) d\vec{\Omega}. \quad (91)$$

Nach dem Einsetzen der Transformationsvorschriften (13) für die Kugelflächenfunktionen ergibt sich:

$$Y(j\omega) = \int_{\vec{\Omega} \in S^2} \sum_{n=0}^{\infty} \sum_{m=-n}^n \mathring{S}_n^m(j\omega) Y_n^m(\vec{\Omega}) \quad (92)$$

$$\sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} \mathring{B}_{n'}^{m'*}(j\omega) Y_{n'}^{m'*}(\vec{\Omega}) d\vec{\Omega} \quad (93)$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^n \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} \mathring{S}_n^m(j\omega) \mathring{B}_{n'}^{m'*}(j\omega) \quad (94)$$

$$\int_{\vec{\Omega} \in S^2} Y_n^m(\vec{\Omega}) Y_{n'}^{m'*}(\vec{\Omega}) d\vec{\Omega}. \quad (95)$$

*Im Zusammenhang mit Dekodierung auf Lautsprecheraufbauten wird  $H^*(j\omega, \vec{\Omega})$  als Panning-Funktion bezeichnet.*

Aufgrund der Orthogonalitätseigenschaft aus Gleichung (14) lässt sich das Integral reduzieren zu:

$$Y(j\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} \mathring{S}_n^m(j\omega) \mathring{B}_{n'}^{m'*}(j\omega) \delta_{nn'} \delta_{mm'}. \quad (96)$$

Diese Funktion hat nur Funktionswerte ungleich Null für den Fall, dass  $n' = n$  und  $m' = m$ . Hieraus folgt, dass sich der Ausdruck für  $Y(j\omega)$  auf

$$Y(j\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \mathring{S}_n^m(j\omega) \mathring{B}_n^{m*}(j\omega) \quad (97)$$

reduzieren lässt.

Durch Rücksubstitution von  $B(j\omega, \vec{\Omega}) = H^*(j\omega, \vec{\Omega})$  bzw.  $\mathring{B}_n^{m'*}(j\omega) = \mathring{H}_n^{m'}(j\omega)$  ergibt sich:

$$Y(j\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \mathring{S}_n^m(j\omega) \mathring{H}_n^m(j\omega). \quad (98)$$

Die einzelnen Ohrsignale ergeben sich dann durch einsetzen der HOA-Koeffizienten  $\mathring{H}_n^m(j\omega)$  aus Abschnitt 3.3.4 für das linke und rechte Ohr.

### 3.3.6 Testablauf

Die Testperson wird im Sweet-Spot des Hörraums platziert, dessen Raumimpulsantworten wie in Abschnitt 3.3.1 beschrieben gemessen werden. Da dieser Test noch keinen Headtracker verwendet, werden die Stimuli vorprozessiert.

Der Test wird mithilfe einer Multi-Stimulus Test with Hidden Reference and Anchor (MUSHRA)-Software durchgeführt [31]. Die angebotenen Stimulus-Richtungen entsprechen den Positionen der realen Lautsprecher. Dies ermöglicht die Wiedergabe des Stimulus aus dem realen Lautsprecher als Referenz.

Aufgabe der Testperson ist es nun, die angebotenen binauralen Signale mit der Lautsprecherreferenz hinsichtlich Externalisierung zu vergleichen und auf einer Skala von null bis zehn zu bewerten. Null entspricht einer Wahrnehmung des Stimulus im Kopf und zehn der Wiedergabe über reale Lautsprecher.

Da es sich bei der wiedergegebenen Referenz um Lautsprecherwiedergabe handelt, wurde auf das Hinzufügen einer versteckten Referenz unter den Testsignalen verzichtet.

Jede Testperson bekommt vor Testbeginn einen Ausdruck mit einem genauen Testablauf, sowie der Möglichkeit weitere Kommentare zu den Stimuli zu hinterlassen, die über eine Bewertung der Externalisierung hinaus gehen. Dieses Informationsblatt kann im Anhang C.2 nachgelesen werden.



### 3.3.7 Implementierung

Anders als beim Lokalisierungstest wird die Implementierung hier in Python<sup>4</sup> durchgeführt. Python ist eine Skriptsprache, die sich durch ihre Einfachheit und Verbreitung auszeichnet. Sie ist zudem quelloffen für alle Plattformen verfügbar.

Mithilfe weiterer Bibliotheken lässt sie sich zu einer MATLAB<sup>®</sup>-ähnlichen Entwicklungsumgebung ausbauen. Wichtige Bestandteile sind hierbei numeric-python (NumPy)<sup>5</sup>, sowie scientific-python (SciPy)<sup>6</sup>. Alle verwendeten Bibliotheken mit Versionsnummer sind im Anhang B aufgeführt.

Das Programm wird in drei Module eingeteilt, die wiederum zusammen das Paket `externalizationTest` bilden. Ein Paket entspricht einer Bibliothek im Sinne anderer Programmiersprachen. Diese Aufteilung ist Standard in Python und gewährleistet eine anschließende Wiederverwendbarkeit des Quelltexts.

Die drei Hauptmodule sind `hoa`, `rooms` und `binaural`. Der Grundaufbau und die Aufgabenverteilung soll im Weiteren kurz umschrieben werden. Implementierungsdetails können der Dokumentation in Anhang B entnommen werden. Das gesamte Paket setzt in allen einzulesenden Daten eine Abtastrate von  $f_s = 48 \text{ kHz}$  voraus.

**MODUL `hoa`:** Dieses Modul liefert die Koeffizienten für die HOA-Transformation. Es nutzt dafür ein weiteres Modul, welches analog zur vorhandenen MATLAB<sup>®</sup>-Implementierung zu Python portiert wurde. Funktions-Rümpfe und Parameter werden identisch zu den MATLAB<sup>®</sup>-Funktionen implementiert. Dies erleichtert zum Einen der Forschungsgruppe die Verwendung der Python-Funktionen und zum Anderen können Ergebnisse der MATLAB<sup>®</sup>-Funktionen als Referenz für die Software-Tests der Python Implementierung genutzt werden.

**MODUL `rooms`:** Das Modul beherbergt die Klasse `room`. Diese lädt Raumimpulsantworten aus Multikanal-WAV-Dateien und fenstert diese mittels eines Hanning-Fensters auf 7200 Samples. Diese Zahl wurde experimentell mithilfe einer Digital Audio Workstation (DAW) und des gemessenen Raumes 606 festgelegt. Bei der gegebenen Abtastrate  $f_s$  ergibt dies eine Länge von  $\frac{7200}{48000} = 150 \text{ ms}$ .

Die Klasse bietet zudem die Ausgabe von Raumimpulsen für mehrere standardisierte Lautsprecheraufstellungen (z.B. Mono, Stereo, 5.1, 7.1 etc.) an. Zur Zuordnung der Lautsprecher zu den jeweiligen Positionen und Impulsantworten wird eine JavaScript Object Notation (JSON)-Datei eingelesen und verarbeitet. Diese beinhaltet die Kanalzuordnungen, sowie alle Lautsprecherpositionen in Kugelkoordination

4 <http://www.python.org/>

5 <http://www.numpy.org/>

6 <http://www.scipy.org/>

vom Sweet-Spot aus gemessen. Damit ist es möglich die gewünschte Standard-Aufstellung aus dem gemessenen Setup zu erhalten. Dies geschieht über eine Nearest-Neighbour Auswertung von gewünschter Position  $\vec{\Omega}_0$  und der Lautsprecherpositionen  $\vec{\Omega}_1$ . Im Allgemeinen kann dies zu Fehlern führen. Durch die Limitierung der Stimulusrichtungen auf reale Lautsprecherrichtungen ist dieser Fehler hier jedoch nicht vorhanden.

**MODUL `binaural`:** Das Kernmodul der Anwendung beherbergt die Klassen `binauralData` und `binauralProcess`. In der momentanen Ausbaustufe lädt `binauralData` die vortransformierten HRIR in einen Datencontainer und stellt sie weiteren Funktionen zur Verfügung.

Im weiteren Verlauf der Entwicklung könnte diese Klasse einer allgemeineren Version weichen, die nicht nur binaurale, sondern ganz allgemeine Panning-Vorschriften enthalten kann und ein binaurales Setup nur ein Spezialfall dessen wäre.

`binauralProcess` vollführt Gleichung (98). Sie übernimmt sowohl das enkodieren mit dem Raumschallfeld, als auch das rendern. Dabei werden momentan Blocklängen von  $B = 8848$  Abtastwerten via schneller Faltung verarbeitet (Overlap-Save-Algorithmus). Diese Länge ist notwendig, da die Raumimpulsantwort  $R = 7200$  Abtastwerte aufweist. Die Blocklänge errechnet sich mit der HRIR-Filterlänge  $H = 318$  Taps und der Fast-Fourier Transform (FFT)-Länge  $F = 2^{14} = 16384$  Taps zu  $B = F - R + H$ .

Damit ist die maximale Blocklänge definiert, mit der das Overlap-Save noch ohne Aliasingartefakte prozessieren kann.

Durch die Definition, dass nur reelle Abtastwerte prozessiert werden, wird in der Implementierung lediglich der positive Teil des Spektrums verwendet, womit sich die Anzahl der Rechenschritte im Vergleich zur komplexen FFT halbiert.

Der zweite Teil des Moduls ist `binauralData`. Diese Klasse lädt die HRIR-Daten und stellt sie anschließend als HOA-Darstellung bereit. Die Transformation ist in Abschnitt 3.3.4 beschrieben.

**MODUL `audiohandling`:** Dieses Modul dient als Sammlung aller allgemeinen Funktionalitäten, die nicht direkt mit der binauralen Synthese verknüpfbar sind. So finden sich hier Funktionen zur Auswahl des Audioadapters oder eine Normalisierung von Audiodaten.

**MODUL `main`:** Gemäß dem Namen ist `main` das Hauptprogramm der Applikation, wobei es eher als Testprogramm während der Entwicklung angesehen werden kann, um schnell Ergebnisse zu produzieren. Nichtsdestotrotz spiegelt es das anvisierte Zusammenspiel der Komponenten wider. Über die Kommandozeile sind einige der Programmparameter setzbar (s. Quelltext 1).

`SKRIPT starttest`: Dieses Skript enthält keine eigene Programmlogik, sondern ruft lediglich das `main`-Modul auf. Die Idee ist, dass sich über diese Skripte die Programmfunktionalität aufrufen lassen, ohne dass weiteres Wissen über die interne Struktur des Paketes vonnöten ist.

`SKRIPT processfiles`: Zur Erzeugung der Testdaten wird dieses Skript verwendet. Es generiert basierend auf einer Stimuli-, Raum- und Testpersonenliste die passenden Kopfhörersignale als WAV-Dateien, die anschließend in der MUSHRA-Software verwendet werden können.

Zur weiteren Automatisierung stehen MATLAB<sup>®</sup>-Skripte zur Verfügung, die beispielsweise die HRIR-Daten vorprozessieren oder proprietäre Extensible Markup Language (XML)-Dateien für die Testsoftware erstellen.

Quelltext 1: Kommandozeilenparameter `main.py`

```
usage: main.py [-h] [-o OUTPUT] [-p] [-H HRTF] [-r]

optional arguments:
  -h, --help            show this help message and exit
  -o OUTPUT, --output OUTPUT
                        Preselect audio output device ID
                        (default: None)
  -p, --processOnly    Only process data - no playback.
                        (default: False)
  -H HRTF, --hrtf HRTF
                        Set HRTF to use. (default: kk_3)
  -r, --rendered       Save encoded HOA-Data as multichannel wav
                        (default: False)
```



## 4.1 MODIFIZIERTER LOKALISIERUNGSTEST

Der Hörtest wird von fünf männlichen Testpersonen wie in Abschnitt 3.2 vorgestellt durchgeführt. Die Evaluation der Messungen soll mithilfe einer ANOVA (s. Abschnitt 2.5.4) vollzogen werden.

Die Auswertung ist zudem in [5] veröffentlicht. Die Auswertung erfolgt mit Python unter der Verwendung der Python-Bibliothek StatsModels [46], sowie der Statistiksoftware R [41].

## 4.1.1 Erster Testdurchlauf mit trainierten Hörern

Die Residuen der ANOVA sind in Abbildung 13 dargestellt. Abbildung 13a zeigt die Wahrscheinlichkeitsverteilung der Residuen. Im Vergleich dazu ist die ideale Normalverteilung mit demselben Mittelwert und derselben Varianz eingezeichnet. Abbildung 13b zeigt den Quantile-Quantile Plot und vergleicht somit ebenfalls die theoretische Normalverteilung mit der Sample-Verteilung.

Beide Abbildungen zeigen Abweichungen zur Normalverteilung. Die Grundcharakteristik der Normalverteilung ist allerdings gegeben, weshalb diese im Folgenden angenommen wird.

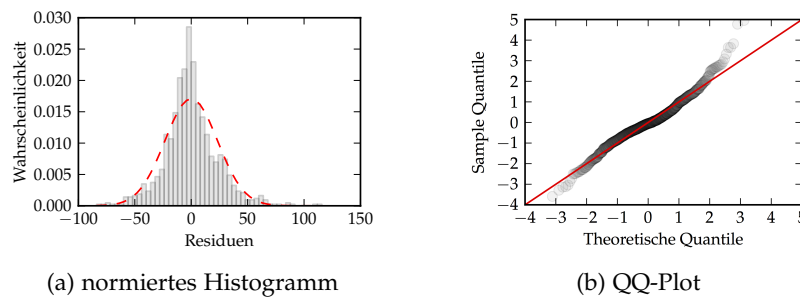


Abbildung 13: Histogramm und QQ-Plot der Residuen

Tabelle 1 gibt einen Überblick über die Testergebnisse. Der Mittelwert ist zudem in Abbildung 14 graphisch aufgetragen. Es ist ersichtlich, dass mit steigender HOA-Ordnung der Winkelfehler sinkt und somit die Lokalisierbarkeit des Stimulus im Mittel schärfer wird. Zudem zeigt die Aufteilung in generische und individuelle HRTF, dass die Verwendung einer individuellen HRTF den Fehler weiter minimiert.

Ein analoges Verhalten zeigt die Standardabweichung, die hier als Unsicherheit der Lokalisierung gedeutet werden kann. Auch hier zeigt

Tabelle 1: Mittlerer Winkelfehler und Standardabweichung

N	HRTF	Mittlerer Fehler	Standardabweichung
1	individuell	22.50°	28.52°
	generisch	30.10°	32.35°
3	individuell	5.27°	17.34°
	generisch	8.13°	23.38°
6	individuell	2.36°	14.01°
	generisch	6.28°	19.48°

sich mit steigender Ordnung und unter Verwendung der individuellen HRTF ein kleiner werdender Winkelfehler.

Diese Ergebnisse entsprechen den Erwartungen, da mit steigender HOA-Ordnung die räumliche Auflösung steigt und somit eine Quelle schärfer dargestellt werden kann. Interessant ist, dass im Gegensatz zu den Ergebnissen des Lokalisierungstest aus Abschnitt 3.1.1 [6] die Wahl der individuellen HRTF den Fehler deutlich verkleinert.

Abbildung 14 zeigt die Interaktionen der beiden Parameter HOA-Ordnung und HRTF. Der parallele Verlauf der beiden Geraden deutet auf keine Interaktion zwischen den beiden Parametern. Zusätzlich ist ein signifikanter Sprung zwischen Ordnung eins und drei. Die Fortführung von dem Verlauf ab Ordnung sechs lässt eine mögliche Sättigung des Fehlers erwarten für höhere Ordnungen. Allerdings enthält diese Darstellung den mittleren Fehler über alle untersuchten Winkel. Eine weiterführende Analyse sollte winkelabhängig durchgeführt werden.

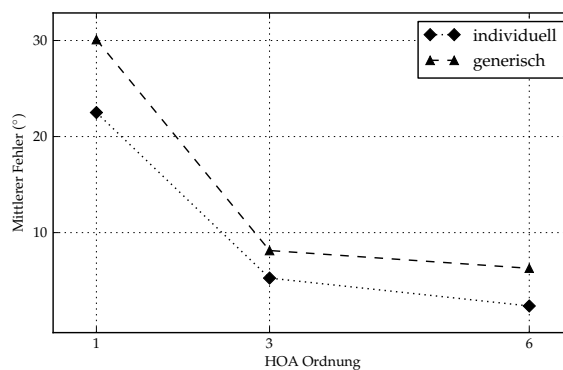


Abbildung 14: Interaktions-Plot der Haupteffekte

Die Boxplots in Abbildung 15 schlüsseln die Ergebnisse auf die jeweiligen Zeigerrichtungen und verwendete HRTF auf. Die Boxplots sind dabei so definiert, dass in den Boxen 50% der mittleren Daten liegen. Die Whisker, gestrichelt dargestellt, sind auf das 1,5-fache des Interquartilabstandes festgesetzt. Der Median der Daten wird durch

einen Strich innerhalb der Box gekennzeichnet. Ausreißer sind mit einem Kreuz dargestellt. Die Winkelzählung des Zeigers wird so festgelegt, dass Linksdrehung einem negativen und Rechtsdrehung einem positiven Winkel entspricht. Die Nullposition entspricht dem Blick des Probanden nach vorn. Die Fehlerberechnung erfolgt im mathematisch korrekten Sinn mit Winkelzählung gegen den Uhrzeigersinn. Daraus erfolgt eine nicht sofort eingängige Interpretation des Winkelfehlers. Es werden zwei Fälle unterschieden: Liegt der Zeiger im Bereich  $[0, -180^\circ]$ , dann entspricht ein positiver Fehler dem Fall, dass sich der Proband zu weit gedreht hat. Befindet sich der Anker auf der anderen Seite, so kehrt sich die Interpretation um. Im mathematischen Sinne hat sich der Proband bei einem positiven Fehler zu weit gedreht.

Zusammengefasst ergibt sich:

- Zeiger  $\in [0, -180^\circ]$ : Positiver Fehler entspricht Überdrehung.
- Zeiger  $\in [0, 180^\circ]$ : Negativer Fehler entspricht Überdrehung.

Auch hier wird im Vergleich der Boxplots deutlich, dass mittlerer Fehler, als auch Streuung des Fehlers mit steigender Ordnung abnimmt. Zusätzlich wird ersichtlich, dass die Lokalisierbarkeit winkelabhängig ist. Die Stimuli aus  $0^\circ$  werden im Mittel gut erkannt, ausgenommen  $0^\circ$  bei Ordnung  $N = 1$  unter Verwendung der generischen HRTF (s. Abbildung 15a). Zu der geringen Ordnung und damit unscharfen Darstellung kommt noch eine schlechtere Externalisierung durch die generische HRTF.

Zu den Seiten nimmt die Unschärfe zu und die Winkelfehler werden größer.

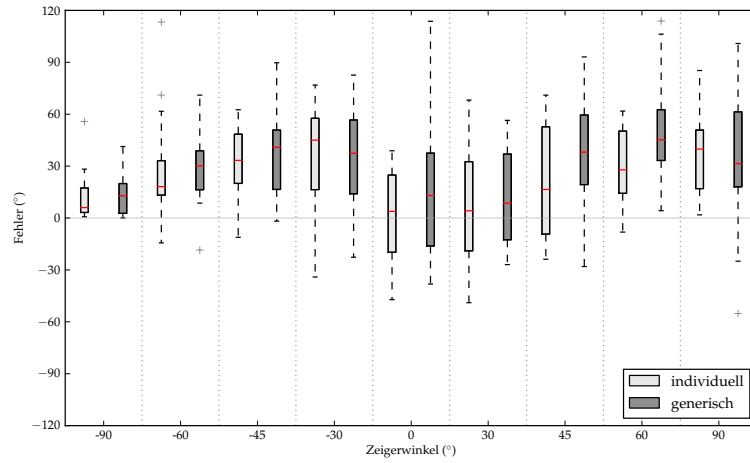
Nach dem deskriptiven folgt der induktive Part. Hier soll untersucht werden, ob die aufgestellten Modelle und Hypothesen aus Abschnitt 3.2.7 durch die Datenlage gestützt werden können.

Dazu wird zuerst das komplette statistische Modell der ANOVA zugrunde gelegt. Tabelle 2 zeigt das Ergebnis. Nun werden die Interaktionsprodukte, die keine signifikanten Effekt aufweisen aus dem Modell ausgeschlossen. In diesem Fall sind das alle Interaktionen mit einem Signifikanzniveau  $> 0,1$ .

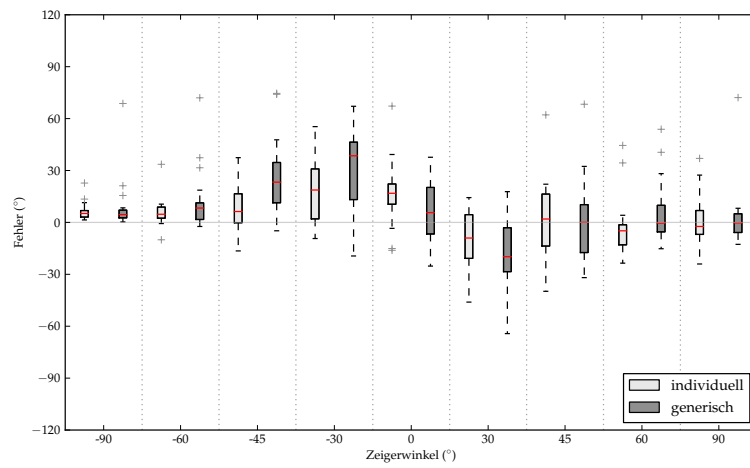
Tabelle 3 zeigt das Ergebnis dieser Modellreduktion. Sie zeigt zum Einen, dass die Wahl des Testhörers signifikant ist und zum Anderen, dass die unterschiedlichen Hörer verschieden stark auf die Ordnungsvariation reagiert haben, da es eine signifikante Interaktion zwischen Ordnung und Hörer gibt.

Dies könnte sich durch den unterschiedlichen Trainingsgrad der Hörer erklären lassen. Trotz des signifikanten Einflusses des Hörers werden diese im Folgenden aus dem Modell ausgeschlossen. Dies hat zur Folge, dass die Hörer als Gesamtmenge betrachtet werden.

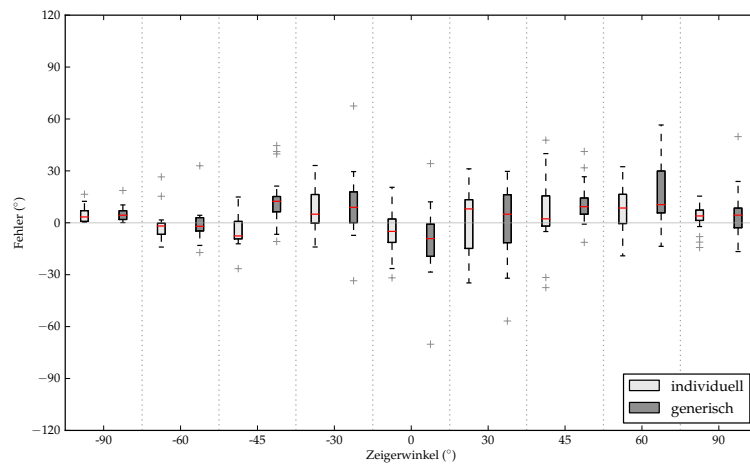
Tabelle 4 zeigt das maximal reduzierte Modell. Es zeigt, dass sowohl der Effekt der HRTF, als auch die HOA-Ordnung signifikant sind.



(a) HOA-Ordnung 1



(b) HOA-Ordnung 3



(c) HOA-Ordnung 6

Abbildung 15: Boxplots mit fünf Testpersonen für die HOA-Ordnungen eins, drei und sechs.



Tabelle 2: ANOVA Analyse I(a): Alle Interaktionen

	df	QS	MS	F	PR(>F)
HRTF	1	6209,33	6209,33	12,72	3,79E-4
Ordnung	2	104719,19	52359,59	107,23	4,21E-43
Hörer	4	29821,59	7455,40	15,27	3,84E-12
HRTF:Ordnung	2	1107,77	553,88	1,13	3,22E-1
HRTF:Hörer	4	3276,93	819,23	1,68	1,53E-1
Ordnung:Hörer	8	41217,49	5152,19	10,55	2,53E-14
HRTF:Ordn:Hörer	8	1223,81	152,98	0,31	9,61E-1
Residuen	1050	512713,89	488,30		

Tabelle 3: ANOVA Analyse I(b): Nur signifikante Interaktionen

	df	QS	MS	F	PR(>F)
HRTF	1	6209,33	6209,33	12,75	3,73E-4
Ordnung	2	104719,19	52359,60	107,48	3,05E-43
Hörer	4	29821,59	7455,40	15,30	3,56E-12
Ordnung:Hörer	8	41217,49	5152,19	10,58	2,28E-14
Residuen	1064	518322,40	487,15		

Die ANOVA zeigt lediglich, dass es einen signifikanten Unterschied in der Wahl der Parameter gibt. Da es aber drei Ordnungen gibt, ist diese Aussage nicht eindeutig. Ein paarweiser Student-T-Test kann verwendet werden, um dieses aufzulösen.

Dieser angewendet auf die Daten deutet an, dass es einen signifikanten Effekt zwischen den HOA-Ordnung eins und drei, sowie eins und sechs gibt. Zwischen drei und sechs ist kein signifikanter Effekt vorhanden. Dies zeigt, dass eine Ordnung größer als drei für diesen Testaufbau keine signifikante Verbesserung in der Lokalisierbarkeit von Quellen aufzeigt.

#### 4.1.2 Ergänzender Testdurchlauf

Im Laufe der Arbeit werden fünf weitere Testperson eingemessen, die anschließend ebenfalls den Test vollziehen.

Die Menge der Probanden besteht nun aus insgesamt neun männlichen Hörern. Die Analyse dazu wird gesondert betrachtet, da sie die vorhergehenden Ergebnisse nicht vollständig bestätigt.

Abbildung 16 zeigt wieder die Verteilungsfunktionen der Residuen, die Gauß-Charakteristik aufweisen.

Tabelle 4: ANOVA Analyse I(c): Ohne Unterscheidung der Hörer

	df	QS	MS	F	PR(>F)
HRTF	1	6209,33	6209,33	11,34	7,87E-4
Ordnung	2	104719,19	52359,60	95,59	6,12E-39
Residuen	1076	589361,48	547,73		

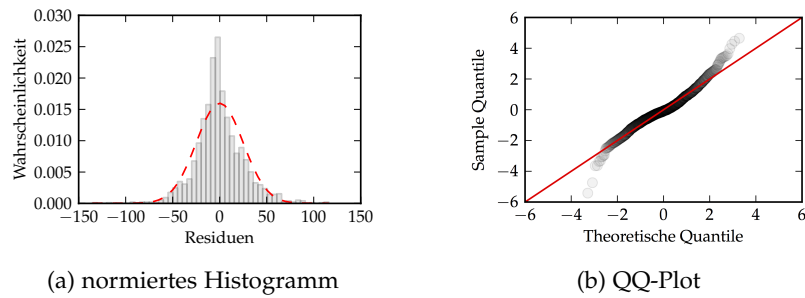


Abbildung 16: Histogramm und QQ-Plot der Residuen

Tabelle 5 zeigt die Mittelwerte und Standardabweichungen. Abbildung 17 führt die Mittelwerte graphisch auf. Dabei ist zu sehen, dass die Geraden im Vergleich zu Abbildung 14 näher zusammen liegen. Bei HOA-Ordnung drei liegen die Mittelwerte sehr nah ( $0,2^\circ$ ) beieinander. Im Vergleich zeigen Tabelle 1 und 5, dass die generischen HRTF ähnliche Fehler erzeugt haben, die individuellen allerdings deutlich größere als bei den trainierten Hörern.

Tabelle 5: Mittlerer Winkelfehler und Standardabweichung

N	HRTF	Mittlerer Fehler	Standardabweichung
1	individuell	$26.45^\circ$	$29.70^\circ$
	generisch	$30.10^\circ$	$32.35^\circ$
3	individuell	$7.93^\circ$	$22.63^\circ$
	generisch	$8.13^\circ$	$23.38^\circ$
6	individuell	$4.63^\circ$	$21.04^\circ$
	generisch	$6.29^\circ$	$19.48^\circ$

Tabelle 6 zeigt die Ergebnisse der ANOVA. Im Gegensatz zu Tabelle 4 weist der Faktor HRTF keine Signifikanz mehr auf. Der Faktor Ordnung ist allerdings wesentlich in der Signifikanz gestiegen.

Eine mögliche Erklärung der Abweichung könnte sein, dass die Testpersonen sehr unerfahren mit dem Testmodus sind. Nach dem Durchlaufen eines Trainings im Referenzsystem (s. Abbildung 9) werden die Tests mit absteigender Reihenfolge durchgeführt, damit ein

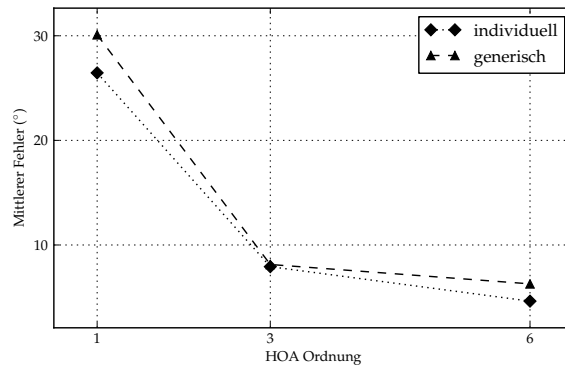


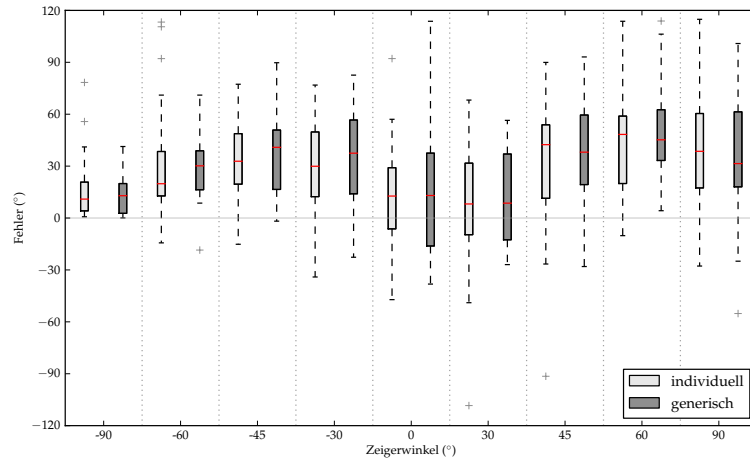
Abbildung 17: Interaktions-Plot der Haupteffekte mit neun Hörern

Sprung in der plötzlich einsetzenden räumlichen Bandbreite durch abrupte Ordnungsreduktion vermieden wird.

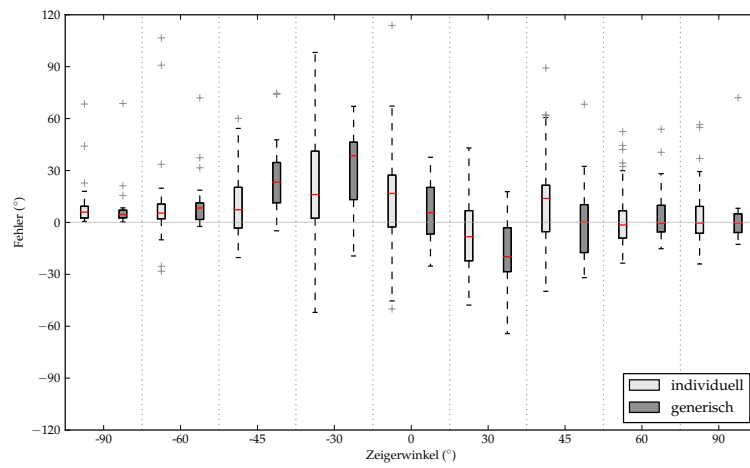
Zudem sollten die Ergebnisse immer im Kontext betrachtet werden, dass die Stimuli alle in der Horizontalebene abgespielt werden. Die individuelle HRTF könnte bei einem Test mit variierenden Elevationen einen größeren Einfluss haben.

Tabelle 6: ANOVA Analyse II: Ohne Unterscheidung der Hörer

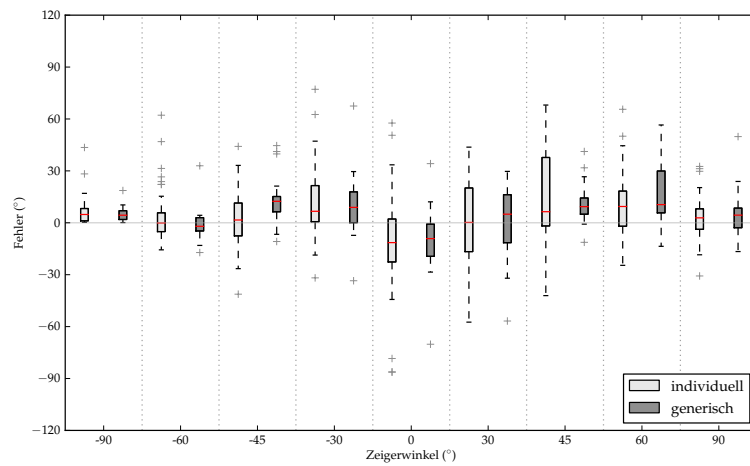
	df	QS	MS	F	PR(>F)
HRTF	1	1313,08	1313,08	2,10	1,47E-1
Ordnung	2	191925,22	95962,61	153,66	1,13E-62
Residuen	1940	1211540,47	624,51		



(a) HOA-Ordnung 1



(b) HOA-Ordnung 3



(c) HOA-Ordnung 6

Abbildung 18: Boxplots mit neun Testpersonen für die HOA-Ordnungen eins, drei und sechs.

## 4.2 EXTERNALISIERUNGSTEST

Dieser Abschnitt beschreibt die Ergebnisse des Externalisierungstests. Der Test wird von neun männlichen Testhörern durchgeführt, die bereits den Lokalisierungstest durchgeführt haben. Das Alter dieser liegt zwischen 25 und 55 Jahren. Die durchschnittliche Testzeit beträgt 26,56 min.

Es werden fünf Systeme gegeneinander getestet:

1. Künstlicher Raum (A) + Generische HRTF,
2. Künstlicher Raum (A) + Individuelle HRTF,
3. Gemessener Raum (B) + Generische HRTF,
4. Gemessener Raum (B) + Individuelle HRTF,
5. Kein Raum (C) + Individuelle HRTF.

Abbildung 19 zeigt die Boxplots der Bewertungen an. Deutlich zu sehen ist die hohe Variabilität der Bewertungen. Alle Konfidenzintervalle der Systeme überschneiden sich. Ferner lässt sich jedoch eine Gruppierung der Systeme erkennen. Raum A und Raum B bilden klare Gruppen. Außer für „A Individ.“ sind die 50%-Boxen der Systeme überschneidungsfrei. Bemerkenswert ist die Gleichheit der Mediane innerhalb der Gruppen. Dies bestätigt, dass die Hörer die in einer Gruppe befindlichen Stimuli ähnlich, wenn nicht sogar gleich bewertet haben.

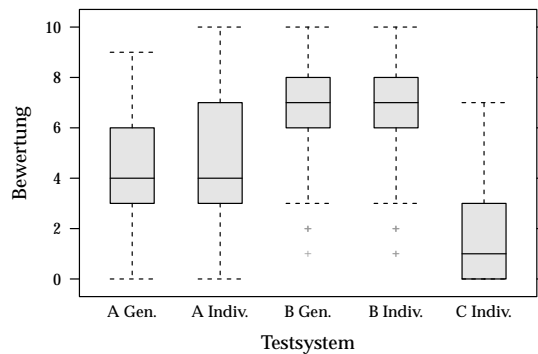


Abbildung 19: Überblick der Testergebnisse

Die ANOVA in Tabelle 7 zeigt signifikante Effekte für die Hauptfaktoren Signal und System, als auch für die Interaktion aus beiden.

Tabelle 8 erweitert diese Analyse durch paarweise T-Tests der Systeme. Zum Einen ist ersichtlich, dass sich Systeme A und B jeweils signifikant von System C unterscheiden. Zusammen mit den Medianen aus Abbildung 19 zeigt dies, dass das Hinzufügen des Raumes prinzipiell eine Verbesserung der Externalisierung ergibt. Es ist ferner

Tabelle 7: ANOVA: Ohne Unterscheidung der Hörer

	df	QS	MS	F	PR(>F)
Signal	8	253,7	31,7	10,16	8,19E-13
System	4	1367,8	342	109,57	< 2E-16
Sig:Sys	32	484,4	15,1	4,85	1,76E-14
Residuen	360	1123,6	3,1		

Tabelle 8: Paarweiser T-Test

	A Gen.	A Individ.	B Gen.	B Individ.
A Individ.	1,00	-	-	-
B Gen.	5,1E-10	5,1E-10	-	-
B Individ.	1,5E-9	1,5E-9	0,86	-
C Individ.	5,5E-16	5,5E-16	< 2E-16	< 2E-16

ersichtlich, dass es keine signifikanten Effekte innerhalb der Raumgruppen A und B gibt. Zur Bewertung, welcher Raum nun am Besten externalisiert werden die Signifikanzstärken von Raum C zu den Räumen A und B untersucht. Hier zeigt sich, dass die Signifikanz von Raum B zu Raum A etwas höher ist als zu Raum A. Folglich externalisiert Raum B (gemessener Raum) in diesem Test am Besten.

Die Ursache der Signifikanz des Signals ist in Abbildung 20 dargestellt. Aufgetragen sind die Boxplots der Bewertungen der verschiedenen Stimuli.

Die freiwilligen Kommentare zu jedem Stimulus sind im Anhang C.3 angegeben. Sie unterstreichen den Eindruck, dass die angebotenen Stimuli in Paaren gruppiert werden. Zudem zeigen sie, dass der gemessene Raum zu sehr externalisiert und halliger scheint als die Referenz selbst. Sie deuten zudem darauf hin, dass es Probleme mit der Lokalisation der Stimuli gibt. Beim Rauschstimulus zeigt sich, dass die Verfärbung durch die Prozesskette sehr deutlich hervortritt und keine Lokalisation möglich ist.

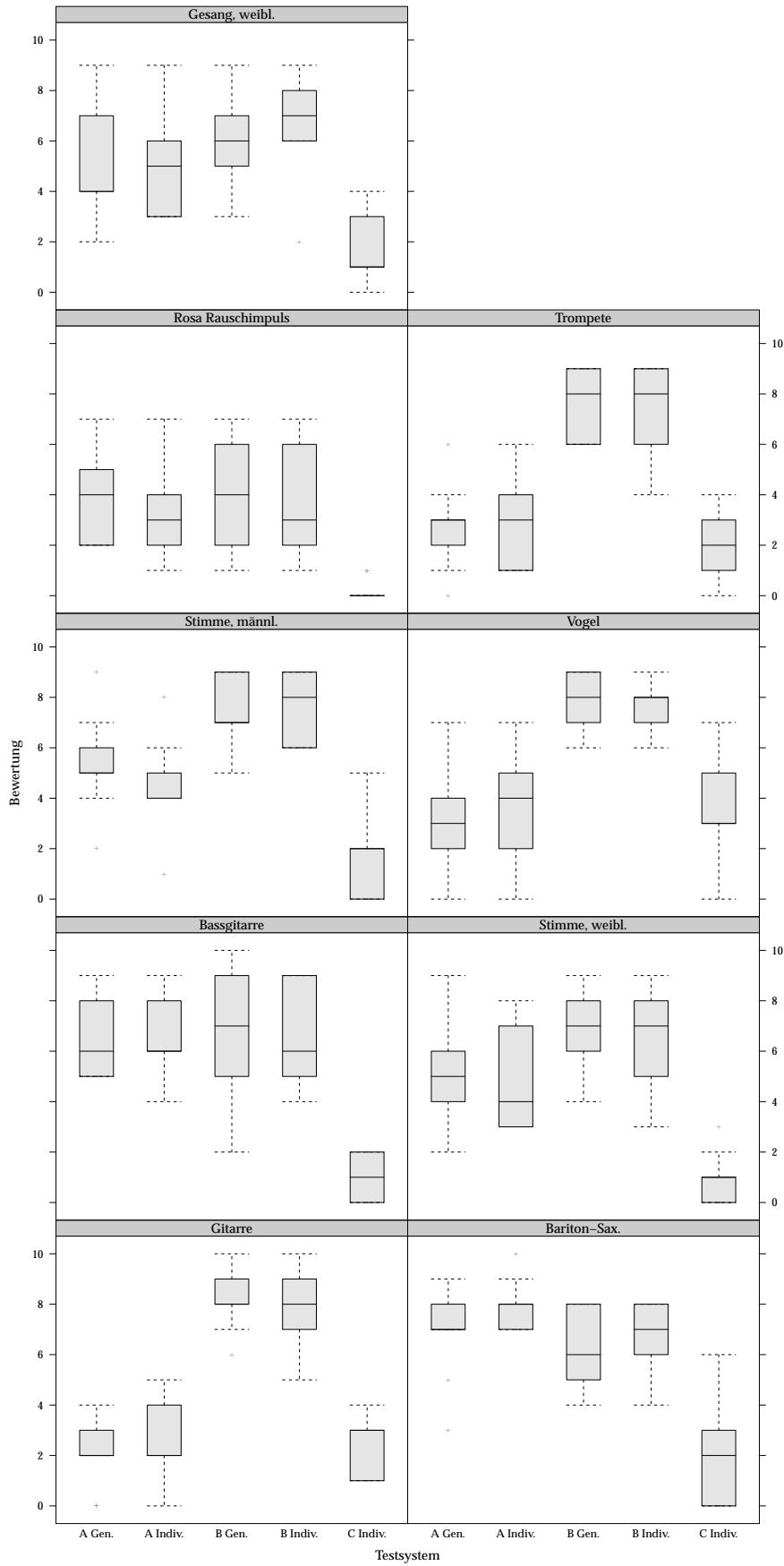


Abbildung 20: Bewertungen aufgeteilt nach Stimulus





## ZUSAMMENFASSUNG

---

Diese Arbeit befasst sich mit der binauralen Synthese unter Verwendung von HOA-Signalen. Neben der Erweiterung eines Lokalisierungstests ist eine komplette Neuimplementierung in Python entstanden, dessen Signale ausschließlich im Bereich der sphärischen Harmonischen prozessiert werden. Es fügt ferner einen Raum hinzu, der die Externalisierung verbessern soll. Der Grad der Verbesserung wird anschließend in einem weiteren Hörtest untersucht.

Im ersten Teil der Arbeit wird ein Überblick über existierende Verfahren zur Generierung von Raumklang gegeben und diese nach ihren unterschiedlichen Ansätzen klassifiziert. Anschließend wird eine Einführung in den Prozess des menschlichen Hörens gegeben. Die Aufgabenstellung der binauralen Synthese und die dabei zu bewältigenden Hürden werden aufgezeigt. Der Binauralteil schließt mit der Vorstellung eines neuartigen Verfahrens zur schnellen Messung von kopfbezogenen Impulsantworten. Nach der Einführung und Herleitung der sphärischen Harmonischen als Schallfeldbeschreibung folgt ein ausführlicher Abschnitt über die Analyse und Bewertung von Testergebnissen mittels statistischer Methoden und statistischer Tests.

Kapitel zwei beginnt mit der Vorstellung des Vorhandenen, dynamischen Lokalisierungstest. Die verschiedenen Testsysteme werden beschrieben und die Resultate des Tests kurz aufgeführt. Es folgt die Modifikation des Lokalisierungstests mithilfe von Stimuluswahl und Ordnungsvariation. Es soll der Einfluss von individuellen und generischen HRTF auf die Lokalisierungsfähigkeit der Probanden bei Variation der räumlichen Auflösung untersucht werden. Im Anschluss daran wird das Konzept des Externalisierungstests vorgestellt. Er hat zum Ziel, eine möglichst gute Reproduktion des Hörraums zu synthetisieren. Dazu wird ein bestehendes Messverfahren mithilfe eines Mikrofonarrays erweitert und zur Beschreibung der Raumimpulsantworten mittels sphärischer Harmonischer benutzt. Zusätzlich werden in einer Raumsimulation dem Messraum ähnliche Raumimpulsantworten zum späteren Vergleich generiert. Die Raumimpulsantworten werden benutzt, um ein Schallfeld zu synthetisieren, welches anschließend mithilfe der HRTF auf den Kopfhörern wiedergegeben wird. Das Kapitel schließt mit einer kurzen Vorstellung der erfolgten Implementierung in Python.

Die Auswertung der durchgeführten Tests schließt sich im vierten Kapitel an. Der modifizierte Lokalisierungstest wird erst mit fünf und im Laufe der Arbeit mit weiteren vier Probanden durchgeführt. Beide Ergebnisse zeigen, dass eine erhöhte räumliche Auflösung eine besse-

re Lokalisation in der Horizontalebene ermöglicht. Die Rolle der Individualisierung der HRTF kann nicht eindeutig gezeigt werden, da sie im ersten Testdurchlauf signifikant, im zweiten hingegen nicht mehr signifikant ist. Alle Hörer erlebten allerdings eine Veränderung in der Wahrnehmung der Höhe des dargebotenen Stimulus bei Verwendung der generischen HRTF. Die Ergebnisse des Externalisierungstests deuten auf eine hohe Stimulusabhängigkeit bei der Wahrnehmung der Externalisierung hin. Zudem weist der gemessene Raum eine bessere Externalisierung auf, als der synthetische. Es wird allerdings auch darauf hingewiesen, dass der gemessene Raum im Vergleich zur realen Hallwahrnehmung im Hörraum deutlich zu viel Hall aufweist.

Die Implementierung des Externalisierungstests, die ausschließlich HOA zur Prozessierung benutzt, ermöglicht eine große Vielfalt an Erweiterungen. Durch das Vorhandensein der Beschreibung des gesamten Schallfelds, kann sehr effizient auf äußere Einflüsse, wie z.B. Raumeinfluss oder Kopfrotation, reagiert werden.

Im Folgenden sollen mögliche weiterführende Aufgaben vorgestellt werden, mit dem Ziel, die Kopfhörerwiedergabe so zu verbessern, dass sie eine realitätsnahe Reproduktion des dargebotenen Raumklangs ermöglicht.

Die Grundintention des Externalisierungstests ist es, die Lokalisierungsfähigkeit zu verbessern. Die neue, flexible Implementierung in Python kann nun dazu genutzt werden, den Lokalisierungstest erneut aufzubauen. Die Schritte dahin sind zum Einen die Sicherung der Echtzeitfähigkeit und die Einbindung eines Headtrackers. Ersteres kann gelöst werden, indem die raumbehafteten Teststimuli vorprozessiert werden. Mithilfe einer Schallfeldrotation im Bereich der sphärischen Harmonischen, kann die Information des Headtrackers genutzt werden, um das Schallfeld an die aktuelle Kopfausrichtung des Probanden anzupassen. Diese Rotation lässt sich effizient auf den HOA-Koeffizienten ausführen.

Die hohe Abhängigkeit der Bewertung vom dargebotenen Stimulus in den Ergebnissen des Externalisierungstests zeigt, dass weitere Untersuchungen und Tests stattfinden sollten. Denkbar ist ein weiterer Test mit zwei bis drei unterschiedlichen Stimuli, die im ersten Durchlauf eine Präferenz für den synthetischen oder den gemessenen Raum aufweisen. Die Untersuchung bestünde darin, die Stimuli mit unterschiedlichen Richtungen im Raum auf ihre Externalisierung zu testen. Als nächstes könnte der Einfluss eines Headtrackers betrachtet werden. Bislang beschränken sich die Tests auf die Horizontalebene. Die Einführung weiterer Elevationsebenen wäre ein konsequenter Schritt zur Nutzung weiterer Vorteile der sphärischen Harmonischen.

Forderung an die Probanden beim Externalisierungstest war, lediglich die Externalisierung zu bewerten. Es stellte sich allerdings im Laufe des Tests heraus, dass eine Trennung von Lokalisierbarkeit und Externalisierung schwierig ist, da die Wahrnehmung der Externalisierung im Allgemeinen mit einer Positionsbestimmung einhergeht. Daher ist ein gewisses Training erforderlich, damit sich der Proband auf den zu beurteilenden Aspekt der Wahrnehmung konzentrieren kann.

Die Kommentare der Probanden zur Externalisierung haben ergeben, dass die Verwendung des gemessenen Raumes zwar besser ex-

ternalisiert, allerdings wesentlich halliger erscheint, als der darzustellende Raum, in dem der Hörtest stattgefunden hat. Die Prozessroutine zur Wandlung der gemessenen Raumimpulse in den Bereich der sphärischen Harmonischen scheint nicht optimal für impulsartige Sequenzen zu sein. Eine Untersuchung und mögliche Optimierung dieser Routine könnte die Raumimpulsantworten näher an den Originalraum bringen.

Mithilfe dieser neuen Erkenntnisse kann ein Vergleichstest mit bestehenden Produkten vollzogen werden. Dazu zählen Softwareprodukte wie das binaurale Rendering des SPAT-Tools oder aber auch Hardwareprodukte wie der SMYTH-Realiser<sup>1</sup>.

---

<sup>1</sup> <http://www.smyth-research.com>

### A.1 LÖSUNG DER SPHÄRISCHEN WELLEN-DIFFERENTIALGLEICHUNG

Die Wellengleichung lässt sich als folgende partielle Differentialgleichung darstellen:

*Als Grundlage der Herleitung dient [12].*

$$\Delta p = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p. \quad (99)$$

Der Laplace Operators  $\Delta$  in Kugelkoordinaten ergibt sich zu

$$\Delta p = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial p}{\partial r} \right) + \frac{1}{r^2 \sin(\vartheta)} \frac{\partial}{\partial \vartheta} \left( \sin(\vartheta) \frac{\partial p}{\partial \vartheta} \right) + \frac{1}{r^2 \sin^2(\vartheta)} \frac{\partial^2 p}{\partial \varphi^2}. \quad (100)$$

Die radiale Komponente lässt sich umformen auf

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial p}{\partial r} \right) = \frac{1}{r^2} \left( \frac{\partial r^2}{\partial r} \frac{\partial p}{\partial r} + r^2 \frac{\partial^2 p}{\partial r^2} \right) \quad (101a)$$

$$= \frac{1}{r^2} \left( 2r \frac{\partial p}{\partial r} + r^2 \frac{\partial^2 p}{\partial r^2} \right) \quad (101b)$$

$$= \frac{2}{r} \frac{\partial p}{\partial r} + \frac{\partial^2 p}{\partial r^2}. \quad (101c)$$

Eingesetzt in (99) ergibt

$$\frac{2}{r} \frac{\partial p}{\partial r} + \frac{\partial^2 p}{\partial r^2} + \frac{1}{r^2 \sin(\vartheta)} \frac{\partial}{\partial \vartheta} \left( \sin(\vartheta) \frac{\partial p}{\partial \vartheta} \right) + \frac{1}{r^2 \sin^2(\vartheta)} \frac{\partial^2 p}{\partial \varphi^2} = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p. \quad (102)$$

Die Lösung dieser partiellen Differentialgleichung erfolgt über den Separationsansatz in der Form

$$p(r, \vartheta, \varphi, t) = R(r)\Theta(\vartheta)\Phi(\varphi)T(t) \quad (103a)$$

$$p_{RR}(r, \vartheta, \varphi, t) = R''(r)\Theta(\vartheta)\Phi(\varphi)T(t) \quad (103b)$$

$$p_{\Theta\Theta}(r, \vartheta, \varphi, t) = R(r)\Theta''(\vartheta)\Phi(\varphi)T(t) \quad (103c)$$

$$p_{\Phi\Phi}(r, \vartheta, \varphi, t) = R(r)\Theta(\vartheta)\Phi''(\varphi)T(t) \quad (103d)$$

$$p_{TT}(r, \vartheta, \varphi, t) = R(r)\Theta(\vartheta)\Phi(\varphi)T''(t) \quad (103e)$$

mit  $p_{FF} = \frac{\partial^2 F(x)}{\partial F^2}$ . Dieser Ansatz nimmt a priori an, dass die partielle Differentialgleichung durch das Produkt der Einzelkomponenten lösbar ist. Gleichzeitig überführt die Teilbetrachtung die Überführung selbiger in gewöhnliche Differentialgleichungen.

*Zugunsten der Übersicht werden die abhängigen Variablen der Separationsfunktionen weggelassen.*

Nach Einsetzen der partiellen Ableitungen in (102) und Division durch  $(R\Theta\Phi T)$  ergibt sich

$$\underbrace{\frac{2}{r} \frac{R'}{R} + \frac{R''}{R} + \frac{1}{r^2 \sin(\vartheta)\Theta} \frac{\partial}{\partial \vartheta} (\sin(\vartheta)\Theta')}_{=L(R,\Theta,\Phi)} + \frac{1}{r^2 \sin^2(\vartheta)} \frac{\Phi''}{\Phi} = \frac{1}{c^2} \frac{T''}{T} p. \quad (104)$$

*Separation der zeitlichen Abhängigkeit t*

$$L(R, \Theta, \Phi) = -k^2 = \frac{1}{c^2} \frac{T''}{T} p \quad (105a)$$

$$\boxed{0 = k^2 T + \frac{1}{c^2} T''} \quad (105b)$$

Aufstellung des charakteristischen Polynoms durch  $e^{\lambda t}$ -Ansatz:

$$\text{Annahme: } T(t) = e^{\lambda t} \quad (106a)$$

$$k^2 e^{\lambda t} + \frac{1}{c^2} \lambda^2 e^{\lambda t} = 0 \rightarrow \lambda = \pm jkc. \quad (106b)$$

Es erfolgt nun die bekannte Substitution  $\omega_k = kc$ , wobei  $\omega_k$  die Kreisfrequenz beschreibt und  $k$  als Wellenzahl interpretiert wird.

Daraus folgt die Lösung der gewöhnlichen Differentialgleichung

$$T(t) = T_1 e^{j\omega_k t} + T_2 e^{-j\omega_k t} \quad (106c)$$

$$(106d)$$

*Separation der radialen Abhängigkeit r*

$$L(R, \Theta, \Phi) = -k^2 \quad (107a)$$

$$r^2 L(R, \Theta, \Phi) = -r^2 k^2 \quad (107b)$$

$$2r \frac{R'}{R} + r^2 \frac{R''}{R} + \frac{1}{\sin(\vartheta)\Theta} \frac{\partial}{\partial \vartheta} (\sin(\vartheta)\Theta') + \frac{1}{\sin^2(\vartheta)} \frac{\Phi''}{\Phi} = -r^2 k^2 \quad (107c)$$

$$2r \frac{R'}{R} + r^2 \frac{R''}{R} + r^2 k^2 = \eta = -\frac{1}{\sin(\vartheta)\Theta} \frac{\partial}{\partial \vartheta} (\sin(\vartheta)\Theta') - \frac{1}{\sin^2(\vartheta)} \frac{\Phi''}{\Phi} \quad (107d)$$

$$\boxed{R'' + \frac{2}{r} R' + \left(k^2 - \frac{\eta}{r^2}\right) R = 0} \quad (107e)$$

Dies ist die Besselsche Differentialgleichung [15, S. 565] deren Lösung die sphärischen Besselfunktionen der ersten  $j_n$  und zweiten Art  $y_n$  sind:

$$R(r) = R_1 j_n(kr) + R_2 y_n(kr). \quad (107f)$$

*Separation des Azimuts  $\varphi$*

$$\eta = -\frac{1}{\sin(\vartheta)\Theta} \frac{\partial}{\partial \vartheta} (\sin(\vartheta)\Theta') - \frac{1}{\sin^2(\vartheta)} \frac{\Phi''}{\Phi} \quad (108a)$$

$$\sin^2(\vartheta)\eta = -\frac{\sin(\vartheta)}{\Theta} \frac{\partial}{\partial \vartheta} (\sin(\vartheta)\Theta') - \frac{\Phi''}{\Phi} \quad (108b)$$

$$\frac{\sin(\vartheta)}{\Theta} \frac{\partial}{\partial \vartheta} (\sin(\vartheta)\Theta') + \sin^2(\vartheta)\eta = \boxed{-\frac{\Phi''}{\Phi} = \xi} \quad (108c)$$

Die Lösung erfolgt analog zur zeitlichen Differentialgleichung in A.1.

$$\Phi(\varphi) = \Phi_1 e^{j\varphi m} + \Phi_2 e^{-j\varphi m} \quad (108d)$$

Da  $\varphi$  eine angulare Komponente ist, kann eine  $2\pi$ -Periodizität angenommen werden.

$$\Phi(\varphi) = \Phi(\varphi + 2\pi) \rightarrow e^{j2\pi m} \rightarrow m \in \mathbb{Z}. \quad (108e)$$

*Separation der Inklination  $\vartheta$*

$$\frac{1}{\sin(\vartheta)} \frac{\partial}{\partial \vartheta} (\sin(\vartheta)\Theta') + \eta \Theta - \frac{m^2}{\sin^2(\vartheta)} \Theta = 0 \quad (109a)$$

$$\frac{1}{\sin(\vartheta)} \frac{\partial}{\partial \vartheta} (\sin(\vartheta)\Theta') + \Theta \left( \eta - \frac{m^2}{\sin^2(\vartheta)} \right) = 0 \quad (109b)$$

Zur weiteren Lösung wird die Variablensubstitution  $\psi = \cos(\vartheta)$  vorbereitet.

$$\frac{d}{d\psi} = \frac{d\vartheta}{d\psi} \frac{d}{d\vartheta} = \left( \frac{d\psi}{d\vartheta} \right)^{-1} \frac{d}{d\vartheta} = -\frac{1}{\sin(\vartheta)} \frac{d}{d\vartheta} \quad (110a)$$

$$(1 - \psi^2) \frac{d}{d\psi} = (1 - \sin^2(\vartheta)) \frac{d}{d\psi} = -\sin(\vartheta) \frac{d}{d\vartheta} \quad (110b)$$

$$\frac{d}{d\psi} (1 - \psi^2) \frac{d}{d\psi} = -\frac{d}{d\psi} \left( \sin(\vartheta) \frac{d}{d\vartheta} \right) = \frac{1}{\sin(\vartheta)} \frac{d}{d\vartheta} \left( \sin(\vartheta) \frac{d}{d\vartheta} \right) \quad (110c)$$

Die Durchführung der Substitution ergibt:

$$\Theta(\varphi) = \Theta(\psi) \quad (111a)$$

$$\frac{1}{\sin(\vartheta)} \frac{\partial}{\partial \vartheta} \left( \sin(\vartheta) \frac{\partial \Theta(\psi)}{\partial \vartheta} \right) + \Theta(\psi) \left( \eta - \frac{m^2}{\sin^2(\vartheta)} \right) = 0 \quad (111b)$$

$$\Theta(\psi) \left[ \underbrace{\frac{1}{\sin(\vartheta)} \frac{\partial}{\partial \vartheta} \left( \sin(\vartheta) \frac{\partial}{\partial \vartheta} \right)}_{\text{aus (110c)}} + \left( \underbrace{\eta - \frac{m^2}{\sin^2(\vartheta)}}_{\text{aus (110a)}} \right) \right] = 0 \quad (111c)$$

$$\Theta(\psi) \left[ \frac{d}{d\psi} (1 - \psi^2) \frac{d}{d\psi} \left( \eta - \frac{m^2}{1 - \psi^2} \right) \right] = 0 \quad (111d)$$

Gleichung (111d) wird durch die sogenannten assoziierten Legendre Funktionen der ersten  $P_n^m$  und zweiten Art  $Q_n^m$  gelöst:

$$\Theta(\vartheta) = \Theta_1 P_n^m(\cos(\vartheta)) + \Theta_2 Q_n^m(\cos(\vartheta)). \quad (112)$$

Die Funktionen der zweiten Art  $Q_n^m$  divergieren an den Polen  $\psi = \pm 1$ , woraus direkt  $\Theta_2 = 0$  folgt.

Die Lösung der winkelabhängigen Komponenten lässt sich zusammenfassend als

$$\Theta(\vartheta) \Phi(\varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \Theta_{nm} P_n^m(\cos(\vartheta)) e^{j\varphi m} \quad (113)$$

darstellen. Diese werden als sphärische Harmonische  $Y_n^m(\vartheta, \varphi)$  definiert als:

$$Y_n^m(\vartheta, \varphi) \equiv \Theta_{nm} P_n^m(\cos(\vartheta)) e^{j\varphi m}. \quad (114)$$



## IMPLEMENTIERUNGSDetails

---

Der Externalisierungstest aus Kapitel 3.3 wurde mithilfe folgender Frameworks implementiert:

- Python 2.7
- Numpy 1.7.1
- Scipy 0.12.0
- Scikits.audiolab 0.11.0

Die statistische Analyse erfolgt mit folgenden Bibliotheken:

- Statsmodels 0.4.3
- Pandas 0.11.0
- R 2.15.0

### B.1 SPAT-PARAMETER

Die verwendeten Versionen sind:

- Max/MSP 6.0.8
- Spat~ 4.5.4

Der synthetische Raum wird im SPAT mit diesen Parametern erzeugt:

- size: 63 m<sup>3</sup>
- reverberance: 46,5
- heaviness: 27,4
- liveness: 35
- early:
  - start: 5 ms
  - stop: 6,3 ms
  - distr: 0,4
  - shape: 0,5
- cluster:
  - start: 6,3 ms

- stop: 13,5 ms
- distr: 0,4
- reverb:
  - frequency-low: 225 Hz
  - frequency-high: 6000 Hz
  - start: 12,7
  - modal density: 1

---

# **Externalization Test Documentation**

*Release 1.0*

**Stefan Balke, Stefan Abeling, Jan-Mark Batke**

November 02, 2013



# CONTENTS

<b>1</b>	<b>Modules</b>	<b>3</b>
1.1	hoa Module . . . . .	3
1.2	rooms Module . . . . .	4
1.3	binaural Module . . . . .	6
1.4	audioHandling Module . . . . .	8
1.5	Executables . . . . .	9
1.6	Todo-List . . . . .	10
<b>2</b>	<b>References</b>	<b>11</b>
	<b>Bibliography</b>	<b>13</b>
	<b>Python Module Index</b>	<b>15</b>
	<b>Python Module Index</b>	<b>17</b>
	<b>Index</b>	<b>19</b>



This Python package is meant to provide a listening test for binaural synthesis using Higher-Order-Ambisonics signals.

It was developed in 2013 in the proceedings of Stefan Balke's diploma thesis [\[balke\]](#).

The test explores the effect of externalization when adding a real room (in the best case: the room in which the test is performed) to the scene.





# MODULES

## 1.1 `hoa` Module

*Module author: Stefan Balke <stefan.balke@technicolor.com>*

`externalizationTest.hoa.encode` (*timeVal, theta0, phi0, order*)  
 Encode mono time signal as ambisonics signal

**Parameters** `timeVal` : array\_like

sample of time signal

**theta0** : float

elevation (degree)

**phi0** : float

azimuth (degree)

**order** : int

ambisonics order

**Returns** `ambiSignal` : array\_like

encoded time signal

`externalizationTest.hoa.getAmbiVec` (*order, theta0, phi0, sNorm='norm', sType='real'*)

Calculates ambisonics coefficients.

Calculates the real spherical harmonics and saves them into the return vector. The interleaved output follows the shape:

( $c_{nm}$  = coefficients,  $n$  = order,  $m$  = mode)

$$A_n^m = \vec{A} = (c_{00}, c_{1-1}, c_{10}, c_{11}, c_{2-2}, c_{2-1}, c_{20}, c_{21}, c_{22}, \dots)^T$$

**Parameters** `order` : int

ambisonics order

**theta0** : float

elevation (degree, 'ear-level' = 0? elevation)

**phi0** : float

azimuth (degree, counting counterclockwise)

**sNorm** : { 'norm', 'N3D', 'sch0', 'sch1', 'sconv' }, optional

**type of normalization:**

- `norm` - fully normalized

- N3D- normalized
- sch0 or N3D- seminormalized
- sch1 - Schmidt-Semi-normalized (Matlab)
- sconv - norm. over all spherical convolutions

**sType** : { 'real', 'real\_signed', 'real\_unsigned', 'complex' }, optional

**type of the spherical harmonic**: real, real\_signed, real\_unsigned or complex

**Returns** **ambiVec\_O** : ndarray  
ambisonics vector

## 1.2 rooms Module

Module author: Stefan Balke <stefan.balke@technicolor.com>

**class** externalizationTest.rooms.**room**(*setup*, *pathToJSONConf*, *pathToMeas*, *nfft*, *theta=0*, *phi=0*)

Bases: object

Room settings

### Methods

**`__room__calcCutMarks`** ()

Get cut marks for truncation of HOA coefficients

**Parameters** **hoaCoeffs** : array\_like

HOA coefficients

**Returns** **hoaCoeffs** : array\_like

List of HOA Coefficients

**.. todo::**

implement window size with envelope detection perhaps?

**`__room__get22point2Setup`** ()

Get loudspeaker positions for a 22.2 setup.

**Returns** **availablePositions** : array\_like

Loudspeaker positions that fit best

**`__room__get5point1Setup`** ()

Get loudspeaker positions for a 5.1 setup.

$p \in \{30^\circ, 330^\circ, 110^\circ, 250^\circ\}$

**Returns** **availablePositions** : array\_like

Loudspeaker positions that fit best

**`__room__get7point1Setup`** ()

Get loudspeaker positions for a 7.1 setup.

Source: Dolby

$p \in \{0^\circ, 30^\circ, 330^\circ, 110^\circ, 250^\circ, 150^\circ, 210^\circ\}$

**Returns availablePositions** : array\_like

Loudspeaker positions that fit best

**\_\_room\_\_getFliegePoints** (*noOfPoints*)

**\_\_room\_\_getFreeMonoSetup** (*theta, phi*)

Get loudspeaker positions for a mono setup.

**Returns availablePositions** : array\_like

Loudspeaker positions that fit best

**\_\_room\_\_getMonoSetup** ()

Get loudspeaker positions for a mono setup.

**Returns availablePositions** : array\_like

Loudspeaker positions that fit best

**\_\_room\_\_getNearestLoudspeakers** (*neededPositions*)

Get nearest loudspeakers for given positions in space.

The neededPositions  $\vec{n}_i$  are compared to the available positions give by the discrete loudspeaker positions  $\vec{p}_j$ .

The comparison is performed by finding the minimum vector distance of all possible combinations:

$$\min \|\vec{p}_j - \vec{n}_i\|$$

**Returns availablePositions** : array\_like

Loudspeaker positions that fit best

**\_\_room\_\_getStereoSetup** ()

Get loudspeaker positions for a stereo setup.

Searches for best matching loudspeaker positions available. The positions  $p_n$  for stereo are:

$$p \in \{30^\circ, 330^\circ\}$$

**Returns availablePositions** : array\_like

Loudspeaker positions that fit best

**\_\_room\_\_loadHoaCoeffs** ()

Loads multi-channel wave-files with HOA coefficients

---

### Todo

avoid lists for hoaCoeffs specify window, is hanning window correct?

---

**\_\_room\_\_setLspPositions** ()

Get loudspeaker positions from JSON file

**getCutMarks** ()

Get cut marks for trunaction of HOA coefficients

**Parameters hoaCoeffs** : array\_like

HOA coefficients

**Returns hoaCoeffs** : array\_like

List of HOA Coefficients

.. todo:: :

implement window size with envelope detection perhaps?

**getFilterLength** ()

Getter for the private variable filterLength

**Returns** filterLength : int

Number of filter taps

**getHoaCoeffs** ()

Getter for the private variable \_\_hoaCoeffs

**Returns** hoaCoeffs : array\_like

List of HOA Coefficients

**getHoaCoeffsFD** ()

Getter for the private variable \_\_hoaCoeffsFD

Getting HOA-Coeffs in the frequency-domain

**Returns** hoaCoeffsFD : array\_like

List of HOA Coefficients

**getHoaCoeffsFDForChannel** (*channel*)

Getter for the private variable hoaCoeffsFD

**Parameters** channel : int

Loudspeaker channel

**Returns** hoaCoeffsFD : array\_like

HOA coefficients for *channel* in the frequency-domain

**getHoaOrder** ()

Getter for the private variable hoaOrder

**Returns** hoaOrder : int

HOA-order

**getSetupChannels** ()

Getter for the private variable setupChannels

**Returns** setupChannels : array\_like

List of chosen setup channels

## 1.3 binaural Module

*Module author: Stefan Balke <stefan.balke@technicolor.com>*

**class** externalizationTest.binaural.**binauralData** (*pathToHrir, nfft, targetOrder=4, test-Set=False*)

Bases: object

This class loads binaural impulse responses from a MATLAB file and transforms them to frequency domain with a given FFT-length. The Ambisonics transform was done in a separate MATLAB script using Fliege points.

---

### Todo

Add openDAFF support, Add encoding to Ambisonics, Use pseudo-inverse to encode

---

**Parameters** pathToHrir : str

path to the measured HRIR-Data

**nfft** : int  
number of FFT bins

**targetOrder** : int, optional  
Ambisonics order, default=4

**testSet** : bool, optional  
Generate test set (HRIR is just a delta impulse) default=False

## Methods

**`_binauralData__loadHRIR()`**  
Load HRIR in hoa from stereo wave file

**`_binauralData__transformHRIR()`**  
Transform HRIR to HRTF

**`getHRIR()`**  
Get function for HRTF data

**Returns** **hrir** : array\_like  
HOA coefficients for left and right HRIR

**`getHRTF()`**  
Get function for HRTF data

**Returns** **hrtf** : array\_like  
HOA coefficients for left and right HRTF

**`getHRTFOrder()`**  
Get order of the data

**Returns** **order** : int  
Order of data

**class** `externalizationTest.binaural.binauralProcess` (*room, binauralData, noOfChannels, blockSize, nfft*)

Bases: `object`

This class blockwise processes audio using the overlap-save. It encodes multichannel audio-input to Ambisonics and renders it with the use of an Ambisonics representation of the HRTFs.

**Parameters** **room** : `room`  
room object with room transfer functions

**binauralData** : `binauralData`  
binauralData object with HRTF information

**noOfChannels** : int  
number of channels of the input file

**blockSize** : int  
specifies the block length of the input blocks, needed for memory allocation

**nfft** : int  
number of FFT bins

### Methods

**\_binauralProcess\_\_processEar** (*ear*)

Render encoded audio input with HRTF.

**Parameters** **ear** : int

ear to process

**Returns** **yEar** : array\_like

Frequency bins of processed audio signal

**encodeAudio** (*roomChannel*, *audioChunk*)

Encode the audioChunk to HOA using the room measurements. The audio input is encoded by using the HOA-representation of the room response and weight it by the input samples.

**Parameters** **audioChunk** : array\_like

audio block with samples in the time domain

**Returns** **encodedAudio** : array\_like

Frequency-HOA representation of audio input

**getEncAudio** (*audioChunk*)

Returns encoded audio block.

**Parameters** **audioChunk** : array\_like

PCM audio samples

**Returns** **array\_like** :

Encoded audio in HOA-frequency domain.

**getEncAudioTD** ()

Return the encoded audio signals in the time-domain (TD) for saving the audio data and process it later.

**Returns** **array\_like** :

Encoded audio in HOA-time domain.

**process** (*audioChunk*)

Process hoa input for each input channel and render it to the ears using fast-convolution (overlap-save).

**Parameters** **audioChunk** : array\_like

audio block with samples in the time domain

**Returns** **y** : array\_like

Time signal of processed audio signal

## 1.4 audioHandling Module

The module is a loose collection for all audio handling concerns. *Module author: Stefan Balke <stefan.balke@technicolor.com>*

`externalizationTest.audioHandling.listAudioDevices` (*interface*, *listType='io'*)

Print all available audio devices.

**Parameters** **interface** : pyaudio

Object under scan

**listType**: {'io', 'in', 'out'} :

Show Inputs, Outputs or both

`externalizationTest.audioHandling.normalizeWavFile` (*path*, *refLevel=0*)

Normalize audio data.

Normalizes audio data to a reference Level by finding the highest level in all channels and normalize all other samples accordingly.

Find float2int conversion here: <http://percival-music.ca/teaching/python/labs/lab3.html>

**Parameters** **path** : char

Path to the (multi-channel) wav-file

**refLevel** : int, optional

Reference level in dB, the default value is 0. The audio signal is understood as an amplitude signal, so sticking to ~0.5 means to enter -6 here.

## 1.5 Executables

The executables are meant to provide an user-interface without the need or further knowledge of the used classes. The *main* module can be seen as glue-code for the implemented classes. Most of its functionality is controllable by command-line parameters.

For preparing a listening test, *processFiles* can be used to generate binaural renderings for different rooms, signals and test participants. *Module author: Stefan Balke <stefan.balke@technicolor.com>*

`externalizationTest.main.main` (*argv*)

Main function. Uses command line parameters to render audio input.

**Parameters** **argv** : array\_like

command line arguments

*Module author: Stefan Balke <stefan.balke@technicolor.com>*

`externalizationTest.processFiles.generateReferenceLspFile` (*room*, *refFile*, *out-File*)

Generate multichannel wav files as reference track.

**Parameters** **room** : room

room object

**refFile** : char

path to reference file

**outFile** : char

path to output file

`externalizationTest.processFiles.processRooms` (*roomList*, *tracks*, *subjects*, *hrtfOrder*, *blockSize*, *dataPath*)

Generate files for listening test.

Iterates over all rooms, tracks and subjects to generate a complete Testset for the MUSHRA test.

**Parameters** **roomList** : list

list with rooms to be processed

**tracks** : list

list with tracks to be processed

**subjects** : list

list with subjects to be processed

**hrtfOrder** : int

specifies the HRTF order

**blockSize** : int

Blocksize of the process blocks

**dataPath** : char

path to the data

## 1.6 Todo-List

---

### Todo

Add openDAFF support, Add encoding to Ambisonics, Use pseudo-inverse to encode

---

### Todo

avoid lists for hoaCoeffs specify window, is hanning window correct?

---



# REFERENCES



# BIBLIOGRAPHY

[balke] Stefan Balke, *Perception-based analysis of Higher-Order-Ambisonics signals using binaural synthesis*,  
Institut für Kommunikationstechnik, Leibniz-Universität Hannover, 11.2013



# PYTHON MODULE INDEX

## a

audioHandling (*Unix, Windows*), 8

## b

binaural (*OSX, Unix, Windows*), 6

## e

externalizationTest.audioHandling, 8

externalizationTest.binaural, 6

externalizationTest.hoa, 3

externalizationTest.main, 9

externalizationTest.processFiles, 9

externalizationTest.rooms, 4

## h

hoa (*OSX, Unix, Windows*), 3

## m

main (*OSX, Unix, Windows*), 9

## p

processFiles (*OSX, Unix, Windows*), 9

## r

rooms (*OSX, Unix, Windows*), 4



# PYTHON MODULE INDEX

## a

audioHandling (*Unix, Windows*), 8

## b

binaural (*OSX, Unix, Windows*), 6

## e

externalizationTest.audioHandling, 8

externalizationTest.binaural, 6

externalizationTest.hoa, 3

externalizationTest.main, 9

externalizationTest.processFiles, 9

externalizationTest.rooms, 4

## h

hoa (*OSX, Unix, Windows*), 3

## m

main (*OSX, Unix, Windows*), 9

## p

processFiles (*OSX, Unix, Windows*), 9

## r

rooms (*OSX, Unix, Windows*), 4





# INDEX

- `_binauralData__loadHRIR()` (externalization-Test.binaural.binauralData method), 7
- `_binauralData__transformHRIR()` (externalization-Test.binaural.binauralData method), 7
- `_binauralProcess__processEar()` (externalization-Test.binaural.binauralProcess method), 8
- `_room__calcCutMarks()` (externalization-Test.rooms.room method), 4
- `_room__get22point2Setup()` (externalization-Test.rooms.room method), 4
- `_room__get5point1Setup()` (externalization-Test.rooms.room method), 4
- `_room__get7point1Setup()` (externalization-Test.rooms.room method), 4
- `_room__getFliegePoints()` (externalization-Test.rooms.room method), 5
- `_room__getFreeMonoSetup()` (externalization-Test.rooms.room method), 5
- `_room__getMonoSetup()` (externalization-Test.rooms.room method), 5
- `_room__getNearestLoudspeakers()` (externalization-Test.rooms.room method), 5
- `_room__getStereoSetup()` (externalization-Test.rooms.room method), 5
- `_room__loadHoaCoeffs()` (externalization-Test.rooms.room method), 5
- `_room__setLspPositions()` (externalization-Test.rooms.room method), 5
- audioHandling (module), 8
- binaural (module), 6
- binauralData (class in externalizationTest.binaural), 6
- binauralProcess (class in externalizationTest.binaural), 7
- encode() (in module externalizationTest.hoa), 3
- encodeAudio() (externalization-Test.binaural.binauralProcess method), 8
- externalizationTest.audioHandling (module), 8
- externalizationTest.binaural (module), 6
- externalizationTest.hoa (module), 3
- externalizationTest.main (module), 9
- externalizationTest.processFiles (module), 9
- externalizationTest.rooms (module), 4
- generateReferenceLspFile() (in module externalization-Test.processFiles), 9
- getAmbiVec() (in module externalizationTest.hoa), 3
- getCutMarks() (externalizationTest.rooms.room method), 5
- getEncAudio() (externalization-Test.binaural.binauralProcess method), 8
- getEncAudioTD() (externalization-Test.binaural.binauralProcess method), 8
- getFilterLength() (externalizationTest.rooms.room method), 6
- getHoaCoeffs() (externalizationTest.rooms.room method), 6
- getHoaCoeffsFD() (externalizationTest.rooms.room method), 6
- getHoaCoeffsFDForChannel() (externalization-Test.rooms.room method), 6
- getHoaOrder() (externalizationTest.rooms.room method), 6
- getHRIR() (externalizationTest.binaural.binauralData method), 7
- getHRTF() (externalizationTest.binaural.binauralData method), 7
- getHRTFOrder() (externalization-Test.binaural.binauralData method), 7
- getSetupChannels() (externalizationTest.rooms.room method), 6
- hoa (module), 3
- listAudioDevices() (in module externalization-Test.audioHandling), 8
- main (module), 9
- main() (in module externalizationTest.main), 9
- normalizeWavFile() (in module externalization-Test.audioHandling), 9
- process() (externalizationTest.binaural.binauralProcess method), 8
- processFiles (module), 9
- processRooms() (in module externalization-Test.processFiles), 9

room (class in externalizationTest.rooms), 4  
rooms (module), 4

## HÖRTEST-HANDOUTS

---

### C.1 LOKALISIERUNGSTEST

Herzlich willkommen zum Lokalisierungstest!

In diesem Test soll die Lokalisierungsfähigkeit in Abhängigkeit von individueller- und generischer- (Kunstkopf) HRTF untersucht werden. Zudem wird mittels wechselnder HOA-Ordnung eine räumliche Bandbegrenzung vorgenommen.

#### C.1.1 Testablauf

- Es werden nacheinander zwei identische Stimuli abgespielt:
  1. Zeiger:  $\varphi_p \in \{0^\circ, \pm 30^\circ, \pm 45^\circ, \pm 60^\circ, \pm 90^\circ\}$   
Kopffest, lässt sich also durch Kopffrotation bewegen
  2. Anker:  $\varphi_a \in [-45^\circ, +45^\circ]$   
Raumfest
- Für diesen Test befinden sich beide Stimuli in der Ohr-Ebene (Elevation =  $0^\circ$ ).
- Das Stimuli-Paar wird viermal wiederholt.
- Der Zeiger wird immer im Referenzsystem gerendert (indiv. HRTF + quasi-unendliche HOA-Ordnung).
- Die Anker-Ordnung ist für einen Testdurchgang konstant.
- Der Anker wird zufällig mit individueller oder generischer HRTF gerendert.

Ziel des Testes soll sein, Anker und Zeiger möglichst deckungsgleich durch Kopf- bzw. Körperdrehung zu positionieren.

Es werden insgesamt 72 Positionen abgefragt. Durch Vortests wurde eine durchschnittliche Zeit von 25 Minuten für den gesamten Test festgestellt. Nach Ablauf der Zeit sollte eine Pause eingelegt werden und der Test später fortgesetzt werden. Dazu das Testprogramm bitte **nicht** schließen.

*Unterschiedliches Rendering kann Quellen erhöht erscheinen lassen – nicht verwirren lassen sondern versuchen zu positionieren.*

## C.2 EXTERNALISIERUNGSTEST

Herzlich willkommen zum Externalisierungstest!

Ziel dieses Hörversuches ist es, die Externalisierungseigenschaften der binauralen Synthese unter Verwendung verschiedener Raummodelle und Kopfübertragungsfunktionen zu untersuchen.

Im Test werden verschiedene, möglichst raumfreie Mono-Stimuli verwendet. Diese werden mit HOA enkodiert und anschließend über eine Renderingmatrix auf die beiden Ohrkanäle dekodiert.

## C.2.1 Testablauf

Insgesamt werden zehn Stimuli in einem MUSHRA-Vergleichstest (ohne Referenz) angeboten. Diese werden jeweils auf fünf Arten prozessiert:

*Während des Tests ist die Reihenfolge der Stimuli zufällig.*

1. Individuelle HRTF + gemessener Raum 606
2. Generische HRTF + gemessener Raum 606
3. Individuelle HRTF + künstlicher Raum
4. Generische HRTF + künstlicher Raum
5. Individuelle HRTF + kein Raum.

Jeder Stimulus wird an einer unterschiedlichen Raumposition in der Ohr-Ebene platziert (Elevation =  $0^\circ$ ).

Während des Testes sollte der Kopf nach Möglichkeit nicht bewegt werden.

## C.2.2 Bewertung

Es sollte versucht werden nur den Grad der Externalisierung der einzelnen Stimuli zu bewerten. Anomalien wie Rauschen oder Positionsänderung während eines Durchganges sind in den Anmerkungen auf der Rückseite dieses Dokuments einzutragen.

Die Bewertungsskala ist in zehn Stufen unterteilt, wobei die Wertung null einer Lokalisierung im Kopf und zehn einer lautsprecher-gleichen Wiedergabe entspricht.

c.2.3 *Optionale Anmerkungen*

STIMULUS: Rauschen

---

---

STIMULUS: Gesang, weiblich

---

---

STIMULUS: Sprache, weiblich:

---

---

STIMULUS: Sprache, männlich

---

---

STIMULUS: Gitarre

---

---

STIMULUS: Bassgitarre

---

---

STIMULUS: Baritonsaxophon

---

---

STIMULUS: Trompete

---

---

STIMULUS: Vogel

---

---

## C.3 KOMMENTARE EXTERNALISIERUNGSTEST

STIMULUS: Rauschen

- totaler Verlust Lokalisierung
- Bei keinem Signal Lokalisierung möglich
- Lokalisation nicht möglich, Frequenzen haben unterschiedliche Positionen
- krasse Verfärbungen
- keine Lokalisierung, Rauschen kam von überall

STIMULUS: Gesang, weiblich

- 2x sehr viel Hall (blechern) → Bahnhofshalle
- Teilweise Richtung vertauscht (vorne rechts, statt hinten links)
- 4: Quelle sehr breit, Ortung nicht möglich, 1 und 4: starker Nachhall

STIMULUS: Sprache, weiblich:

- 1 und 2 sehr hallig

STIMULUS: Sprache, männlich

- 2x zu viel Raum
- 1 und 2: gute Externalisierung, klingt aber sehr künstlich; 4 und 5: besserer Klang aber kaum Externalisierung, Richtungsfehler in allen Signalen (gilt ebenso für Gesang und weibliche Sprache)
- Starker Nachhall → anderer Raum

STIMULUS: Gitarre

- 1 und 3: Raum wirkt deutlich größer
- Referenz sehr trocken (zu), d.h. Quelle unterschiedlich klein

STIMULUS: Bassgitarre

- 2x viel Raum (mehr als Referenz)
- Entweder gute Externalisierung und schlechte Lokalisierung oder gute Lokalisierung und schlechte Externalisierung (4)
- Richtungsfehler, springt nach hinten (1 und 2)
- Richtungen immer eher hinten (statt vorne links)

STIMULUS: Baritonsaxophon

- Lokalisierung hinten links statt vorne, sehr hallig
- 2 und 5: Starker Nachhall → Quelle wirkt weiter
- 1 und 4: Quelle sehr breit, Ortung kaum möglich

STIMULUS: Trompete

- 2x klingt wie 2 Trompeten, 1 vorne links statt rechts
- 1 und 2: Sehr hallig

STIMULUS: Vogel

- 2x verstärktes Rauschen + lauter + Hall
- Eigentlich zu starke Externalisierung.
- Kopfhörer verfälscht Richtung der Referenz.
- 3 und 4: Raum wirkt deutlich größer.
- Stimulus hat nur höhere Frequenzen. Hier scheint mehr Reverb zu externalisieren.

ALLGEMEINE KOMMENTARE:

- Externalisierung ist meist sehr gut, bin überrascht. Nur die Richtung und die Breite der Quelle stimmt nicht. Als Feature nicht schlecht.
- 3 Kategorien bei fall allen, außer z.B. Trompete
  1. Trocken (1 Signal)
  2. Raum, relativ hallig (2 Signale)
  3. Raum, wenig hallig, könnte passen (2 Signale)



---

# Audio Engineering Society Convention Paper

Presented at the 135th Convention  
2013 October 17–20 New York, USA

*This Convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42<sup>nd</sup> Street, New York, New York 10165-2520, USA; also see [www.aes.org](http://www.aes.org). All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.*

---

## Investigation of HRTF Sets Using Content with Limited Spatial Resolution

Johann-Markus Batke<sup>1</sup>, Stefan Abeling<sup>1</sup>, Stefan Balke<sup>2</sup>, Gerald Enzner<sup>3</sup>

<sup>1</sup>Audio & Acoustics, Technicolor Research & Innovation, Karl-Wiechert-Allee 74, 30625 Hannover, Germany

<sup>2</sup>Institute of Communications Technology, Leibniz Universität Hannover, Appelstraße 9a, 30167 Hannover, Germany

<sup>3</sup>Institute of Communication Acoustics, Ruhr-Universität Bochum, ID/2/227, 44780 Bochum, Germany

Correspondence should be addressed to Johann-Markus Batke ([Jan-Mark.Batke@technicolor.com](mailto:Jan-Mark.Batke@technicolor.com))

### ABSTRACT

Headphone rendering of sound fields represented by Higher Order Ambisonics (HOA) is greatly facilitated by the binaural synthesis of virtual loudspeakers. Individualized head related transfer function (HRTF) sets corresponding to the spatial positions of the virtual loudspeakers are used in conjunction with head-tracking to achieve the externalization of the sound event. We investigate the localization accuracy for HOA representations of limited spatial resolution using individualized and non-individualized HRTF sets.

### 1. INTRODUCTION

Binaural synthesis relies on the existence of head related impulse response (HRIR) sets that enable the perception of an auditory event outside the listener's head when using headphones [11]. The externalization and localization of the auditory event depends on the properties of the HRIR sets among other parameters. The individual measurement of the HRIR set for the listener is often claimed to be a key factor for a successful binaural synthesis.

The HRIR is measured using artificial heads or human subjects [4]. The conventional spatial sampling of usu-

ally 5–10 degrees azimuth spacing and 10–20 degrees elevation spacing on the auditory sphere leads to restrictions in the directional resolution of the HRIR [16]. In our investigations, we use 3D-continuous-azimuth HRIR functions that are obtained with a recently proposed dynamic measurement apparatus [8, 15]. However, in prior investigations it turned out that good localization results can also be achieved using non-individual HRIR sets [1].

In this work we address the question of how much the localization accuracy is affected when content of limited spatial resolution is presented to the listener. The auditory event is rendered using a set of virtual loudspeakers.



The content is provided by means of Higher Order Ambisonics (HOA) [17, 7]. The spatial resolution is reduced by lowering the order of the HOA representation. The effect on the localization accuracy is examined while using individual or non-individual HRIR sets.

Following this introduction, the principles of binaural synthesis and Higher Order Ambisonics are briefly outlined. After that, the properties of head related impulse responses are investigated in more detail in section 4. The listening experiment in section 5 tests the relation between spatial resolution and localization accuracy.

## 2. BINAURAL SYNTHESIS

### 2.1. Basics

Binaural synthesis virtualizes acoustic sources for headphone playback using the head related impulse responses or head related transfer functions (HRTF), if expressed in the frequency domain [11]. Today this technology receives growing attention as headphones are often used for audio playback on mobile devices. One example application of binaural synthesis in this case is the creation of virtual loudspeakers placed in a standard surround sound loudspeaker layout which enables the playback of surround sound content on mobile devices [5].

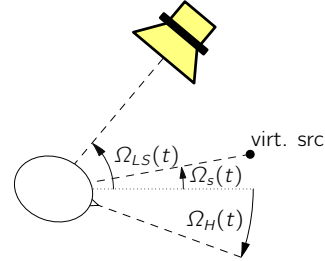
A sound source signal  $x_S(t)$  may appear at source direction  $\Omega_S(t)$  as indicated in Fig. 1. Throughout the paper positions are expressed in the standard spherical coordinate system by  $(r, \theta, \phi)$ . For convenience, the pair of angles  $\theta, \phi$  is summarized as direction  $\Omega = (\theta, \phi)$ . The relative direction  $\Delta\Omega(t)$  according to the listener's head direction  $\Omega_H(t)$  can be written then as

$$\Delta\Omega_S(t) = \Omega_H(t) - \Omega_S(t). \quad (1)$$

Accordingly, if a number of virtual loudspeakers is generated, the relative direction is calculated by  $\Delta\Omega_{LS}(t) = \Omega_H(t) - \Omega_{LS}(t)$ .

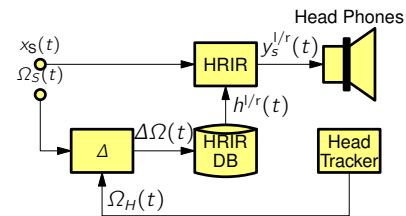
### 2.2. HRIR Rendering System

The binaural ear signals  $y_s^{l/r}(t)$  (l/r: left/right) are computed by convolving the signal  $x_S(t)$  with HRIR functions  $h^{l/r}(t)$  as shown in Fig. 2. These two functions are taken from the HRIR database (HRIR DB) depending on the actual listening direction  $\Delta\Omega(t)$ . If  $L$  virtual loudspeakers are used to reconstruct the sound source, the respective number of HRIR functions needs to be



**Fig. 1:** 2D-projection showing the head orientation  $\Omega_H(t)$ , the virtual source direction  $\Omega_S(t)$ , and the virtual loudspeaker direction  $\Omega_{LS}(t)$ .

updated depending on the directions of the loudspeakers  $\Delta\Omega_{LS}(t)$ . This obviously implies a massive memory-access demand. It may further come along with parallel convolution and cross-fading operations to avoid audio artefacts related to time-varying HRIR rendering [14]. Online interpolation of sampled HRIRs is a conventional alternative to reduce the memory-access demand at the cost of additional computational load, but it was not applied here to clearly avoid detrimental interpolation effects in our results [1].



**Fig. 2:** A rendering system for binaural synthesis using HRIR data.

## 3. HIGHER ORDER AMBISONICS

### 3.1. Basics

Higher Order Ambisonics is an approach to describe audio content independently from any playback system [7, 18]. HOA is mathematically based on a spherical harmonic expansion of the sound field information that is truncated at some order  $N$  [17]. The sound field pressure  $p(\vec{r}, k)$  at position  $\vec{r} = (r, \theta, \phi)$  with wave number  $k$  is approximated by

$$p(\vec{r}, k) = \sum_{n=0}^N \sum_{m=-n}^n A_n^m(k) J_n(kr) Y_n^m(\theta, \phi) \quad (2)$$

where  $A_n^m(k)$  denotes a HOA coefficient,  $j_n(kr)$  is the spherical Bessel function, and  $Y_n^m(\theta, \phi)$  is the spherical harmonic function [17]. The  $(N + 1)^2$  coefficients  $A_n^m(k)$  of this expansion are used for signal transmission and need further processing at the receiver side. Depending on the targeted loudspeaker layout different rendering approaches exist [7, 18, 2].

For binaural synthesis virtual loudspeakers are rendered if HOA content is played back. As a side effect, the listener is always positioned in the sweet spot. In principle, arbitrary numbers of loudspeakers can be used. However, the computational cost for the HRIR updates in the rendering system is of specific interest as discussed in section 2.1.

### 3.2. HOA Rendering System

In our contribution, the input signal  $x_s(t)$  in the HOA transmission chain is encoded using the plane wave representation given by  $A_n^m(k) = 4\pi i Y_n^m(\Omega_s)^*$  [17]. The rendering system following the encoding stage avoids the aforementioned HRIR update [12]. While head movements require new HRIR functions because the directions of the virtual loudspeakers change, the system described here rotates the entire sound field and the virtual loudspeakers remain in static positions in relation to the head. Thus the HRIR data stays unchanged and no further memory access is performed. The block diagram in Fig. 3 uses the head orientation  $\Omega_H(t)$  solely in the rotation block of the processing chain, all other directions remain static. The additional sound field rotation can be realized using efficient algorithms [6].

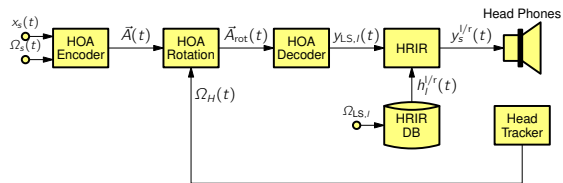
The virtual loudspeakers of the HOA rendering system are positioned according to Fliege et al. [10]. The HRIR data that are required for these directions are assigned using the nearest neighbor method as provided by openDAFF [15].

In a prior work [1] we compared the presented HOA rendering system to a second HOA-based rendering system utilizing a frequently update of the HRIR functions instead of the HOA rotation. With the result that both systems performed quite similar considering the localization accuracy we can now focus on one system.

## 4. HEAD RELATED IMPULSE RESPONSES

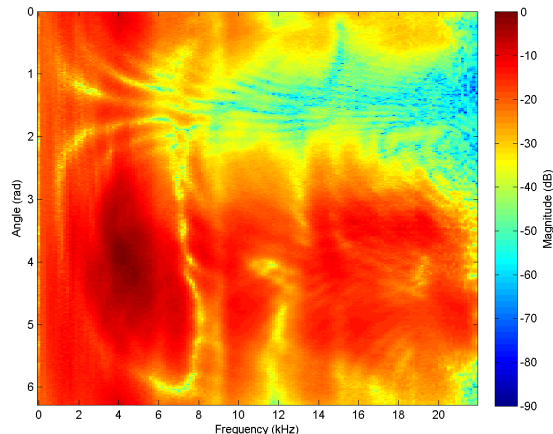
### 4.1. Measurements

The 3D-continuous-azimuth HRIR functions used in this work are obtained with a dynamic measurement appa-



**Fig. 3:** HOA rendering system using HOA processing for the content played back. The entire sound field is rotated depending on the listener’s head direction whereas the virtual loudspeakers remain at static positions.

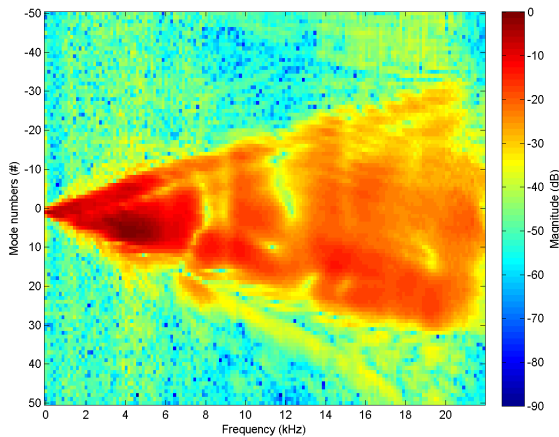
atus [8, 15]. The measurement data has a spatial resolution of 0.1 degrees in azimuthal direction and eight inclination angles of 30, 50, . . . , 170 degrees. The HRIR functions have a length of 308 samples with a sample rate of 44.1 kHz. We investigate the properties of these HRIR sets by computing the magnitude spectrum along the equator. As an example, Fig. 4 shows the resulting spectrum for the left ear. At lower frequencies only little variations in amplitude occur for the different directions. Going to higher frequencies it is visible that the directional dependency of the magnitude spectrum becomes higher, i.e., the magnitude along the equator changes more rapidly.



**Fig. 4:** Spectrum of the HRIR of the left ear of one subject on an inclination level of 90 degrees along the equator.

To obtain more insight to this dependency, the spatial Fourier analysis of the measured HRTF data along the az-

imuthal direction is computed and shown in Fig. 5. The resulting angular spectrum has a triangular shape, i.e., the spatial bandwidth that indicates the necessary order for a HOA representation [9, 13].



**Fig. 5:** The angular spectrum of the measured HRTF data at inclination level of 90 degrees along the equator. The spectrum indicates the required mode numbers for a given frequency bandwidth.

#### 4.2. Representation in HOA

The content played back by the HOA rendering system described in section 3.2 has limited the spatial resolution depending on the order  $N$ . As the convolution of the HRIR functions and the HOA decoder result could be carried out in the HOA coefficients domain, a spatial resolution of the selected HOA order would be sufficient for the HRIR sets as well. In our implementation it is done in the spatial domain at full spatial resolution, though.

In the following the band limited representations of HRIR sets are discussed only to illustrate the effect of the spatial bandwidth limitation caused by the use of the HOA representation for the content.

In the listening test the used stimuli have a spectral bandwidth of 6 kHz. All stimuli are placed somewhere on the equator ( $\theta = 90^\circ$ ). The inspection of Fig. 5 leads to an order of  $N = 15$  which is necessary to represent the full spatial information of the 6 kHz stimuli at the equator. Fig. 10a shows the result.

The spatial structure of the magnitude spectrum sufficiently matches the original data as shown in Fig. 4. Fig-

ures 10b, 10c, and 10d show the spatial spectra for orders  $N = 6, 3, 1$ . Matching regions are increasingly limited to lower frequencies.

## 5. EXPERIMENT

The concept of *localization blur* is widely discussed in the literature [4]. Listeners are not able to localize auditory events in the auditory space as exactly as they are measured in the physical space.

The goal of our experiment is to investigate the localization accuracy depending on the spatial resolution of the HOA content. In particular, we are interested in the impact of individual HRIR sets in contrast to generalized sets.

### 5.1. Setup

For the experiment we use a closed headphone model (Beyerdynamic DT770M) combined with a head tracking system (Polhemus Fasttrack).

To enable the user to indicate the perceived direction, we use an acoustical pointer and an acoustical anchor. The acoustical pointer is implemented by a stimulus being fixed in the head coordinate system. So the listener is able to control this direction by head movements to point to a certain direction. The presented azimuthal directions used in the test are  $0, \pm 30, \pm 45, \pm 60, \pm 90$  degrees, with inclination angle of 90 degrees. To achieve best localization accuracy, the acoustical pointer is rendered by the HRIR based system, see section 2.2.

The acoustical anchor is the same stimulus as used for the pointer to ensure the same spectral properties. The anchor is randomly positioned with azimuthal angles of  $-45 \dots 45$  degrees in the world coordinate system. The anchor is played back by the HOA based rendering system described in section 3.2.

Low-pass filtered pink noise bursts with a cut-off frequency of 6 kHz are used as a stimulus. The first noise burst is played as an anchor for 200 ms followed by a break of 1.8 s, then the pointer is played back for 200 ms using the same stimulus. This sequence is repeated four times.

The subjects in the experiment are asked to bring the directions of pointer and anchor to coincidence by head rotation. Four measurements are taken per direction. No visual cues are given during the experiment although the

	df	Sum of squares	Mean square	F	PR(>F)
HRIR	1	6209.333	6209.333	11.336	$7.869 \cdot 10^{-04}$
HOA order	2	104719.186	52359.593	95.593	$6.121 \cdot 10^{-39}$
Residual	1076	589361.478	547.734		

**Table 1:** Results of factorial ANOVA analysis of the reduced model with independent variables HRIR and the HOA order.

front position is calibrated by the subject before performing the test. Individual and non-individual HRIR (taken from the dummy head *HMS II.3 by Head Acoustics*) are used in the experiment. They are switched randomly during the test without knowledge of the subject.

According to the spectral bandwidth measurement that was performed using the dummy head, an order of  $N = 15$  is required for a spectral bandwidth of 6 kHz. In a former experiment it turned out that HOA content at order  $N = 6$  is undistinguishable from our reference system using no HOA processing. Therefore the only decreasing HOA orders from  $N = 6$  are used in the experiment.

## 5.2. Measurements and Results

The available number of test subjects was five. The subjects are all related to the project and experienced listeners. All are male and of different age.

Each subject performs three blind tests: HOA order  $N \in \{1, 3, 6\}$ . Due to the described setup in section 5.1, each test consists of 72 measurements (angles  $\times$  repetitions  $\times$  HRIRs), resulting in a total sample size of 1080. Each test takes about 20 minutes.

The test results are shown in Figs. 7-9. No front-back confusions were considered.

To evaluate the test statistics an additive ANOVA model is chosen with full interactions of the different factors. The common requirements for ANOVA are fulfilled [3]. However, it turns out that there are no significant interactions between the factors for a type I error  $\alpha = 0.05$ .

Fig. 6 shows the mean localization error over all tested angles for individual and non-individual HRIRs versus the different HOA orders. The almost parallel lines in this interaction plot already suggest the lack of interactions between the HOA order and the type of HRIR.

Thus, the reduced model for the localization error  $e_{ti}$  is

$$e_{ti} = \mu + \eta_t + \gamma_i + \varepsilon_{ti} \quad (3)$$

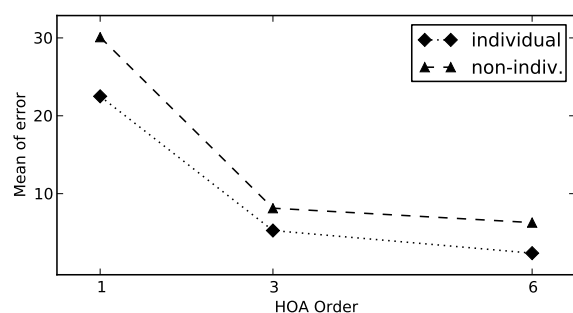
where  $\mu$  is the overall mean,  $\eta_t$  is the HOA order effect with  $t \in \{1, 3, 6\}$ ,  $\gamma_i$  the HRIR effect with  $i \in \{1 \hat{=} \text{indiv.}, 2 \hat{=} \text{non-indiv.}\}$  and  $\varepsilon_{ti}$  the normally distributed error.

Table 1 shows the results of the ANOVA test using the reduced model. It can be seen that both effects - HOA order and HRIR are significant for  $\alpha = 0.05$ .

### 5.2.1. Impact of Tested Factors

It can be seen from the plots that the localization error decreases with increasing HOA order. Fig. 6 points out that there is a high difference in error between order order  $N = 1$  and order  $N = 3$ . It also shows that the use of individualized HRIR leads to a lower average localization error throughout all orders.

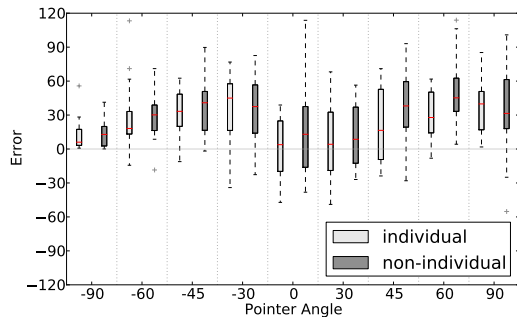
The box plots in Figs. 7-9 reflect the subject's reports about the uncertainty especially with order  $N = 1$ . The dispersion of the data is higher compared to higher HOA orders which was to be expected.



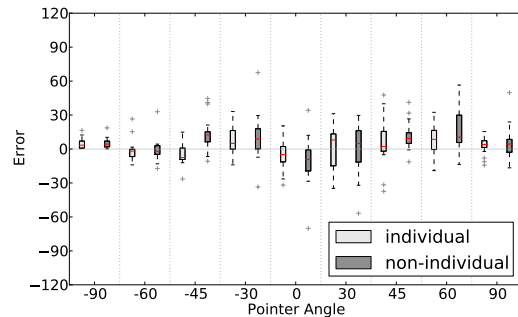
**Fig. 6:** Interaction plot showing the average error vs. HOA order for individual and non-individual HRIRs.

## 6. SUMMARY AND CONCLUSION

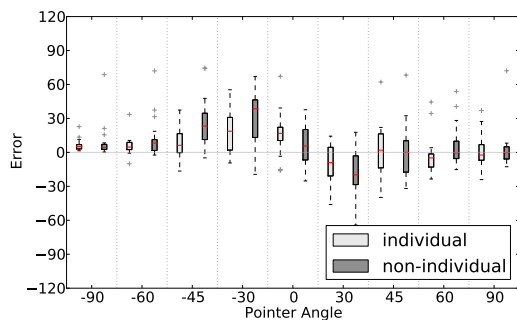
Reducing the performance requirements of an HOA based binaural synthesis system can greatly be achieved



**Fig. 7:** Error average vs. pointer angles for order  $N = 1$



**Fig. 9:** Error average vs. pointer angles for order  $N = 6$



**Fig. 8:** Error average vs. pointer angles for order  $N = 3$

by reducing the HOA order. In this work, the impact of the HOA order reduction of  $N \in \{1, 3, 6\}$  on the localization accuracy of a binaural synthesis system is evaluated in a perceptive listening test using individual and non-individual HRIRs.

Therefore, the concepts of binaural synthesis and HOA is described, and the design of the applied HRIR and HOA rendering systems are presented.

The influence of the decreasing HOA order is discussed on the HOA representation of the continuous-azimuth HRIR data. This clearly shows the mismatch between the different HOA orders.

The measurement results of our experiment verify the theoretical considerations and point out the significant impact of the HOA order with a huge localization accuracy improvement from  $\bar{e}_{11} = 22.5$  to  $\bar{e}_{31} = 5.3$  in contrast to  $\bar{e}_{61} = 2.36$ . The measurements also indicate that HRIR individualization increases the localization accuracy but with a smaller impact. An ANOVA analysis con-

firms the significance of these two factors. However, our test results do not show significant interactions between them.

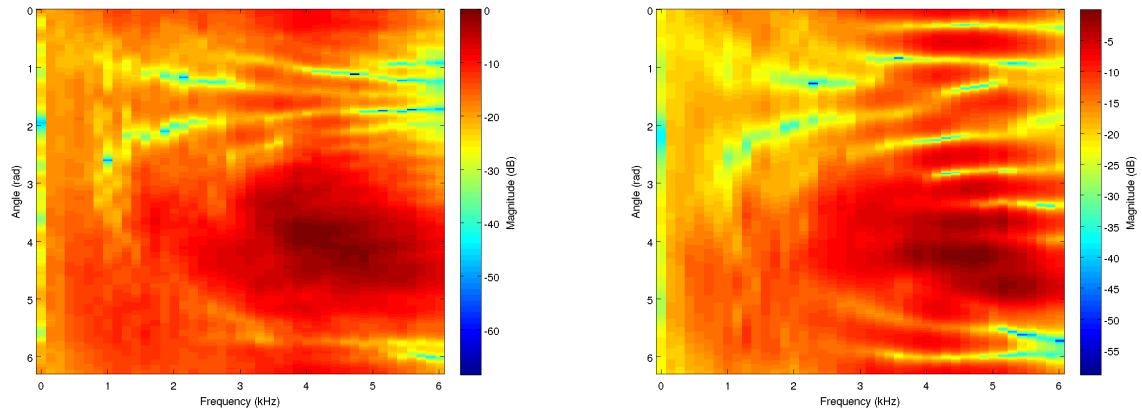
Compared to our prior study [1], noise bursts and a limited number of repetitions allowed corrections of the perceived direction only to a certain extent. The subjects were no longer able to use changes in timbre due to spectral coloration as another cue.

In a next evaluation step, the number of subjects will be increased to be able to perform a more detailed statistical analysis. As an additional factor the externalization of the stimulus perceived by the subject is aimed to be evaluated.

## 7. REFERENCES

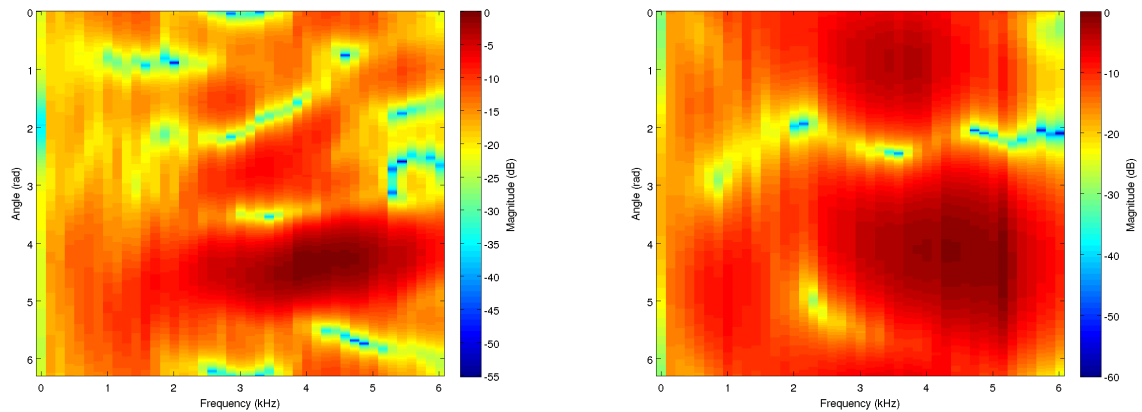
- [1] Johann-Markus Batke, Stefan Abeling, Peter Jax, Michael Weinert, and Gerald Enzner. On the use of continuous-azimuth hrir in higher-order ambisonics rendering. In *AIA-DAGA including the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA)*, 2013.
- [2] Johann-Markus Batke and Florian Keiler. Assessment of spatial audio signals using a higher order ambisonics representation. In *Deutsche Jahrestagung für Akustik (DAGA 2011)*, Düsseldorf, March 2011.
- [3] Søren Bech and Nick Zacharov. *Perceptual audio evaluation-Theory, method and application*. Wiley.com, 2007.

- [4] Jens Blauert. *Spatial Hearing - The Psychophysics of Human Sound Localization*. MA: MIT press., Cambridge, 1997.
- [5] Jeroen Breebaart, Jürgen Herre, Lars Villemoes, Craig Jin, Kristofer Kjörling, Jan Plogsties, and Jeroen Koppens. Multi-channel goes mobile: Mpeg surround binaural rendering. In *AES 29th International Conference*, 2006.
- [6] Cheol Ho Choi, Joseph Ivanic, Mark S. Gordon, and Klaus Ruedenberg. Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion. *Journal of Chemical Physics*, 111(19), November 1999.
- [7] Jérôme Daniel. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, Université Paris 6, 2001.
- [8] Gerald Enzner. 3d-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009.
- [9] Gerald Enzner, Martin Krawczyk, Falk-Martin Hoffmann, and Michael Weinert. 3d reconstruction of hrtf-fields from 1d continuous measurements. 2011.
- [10] Jörg Fliege and Ulrike Maier. A two-stage approach for computing cubature formulae for the sphere. Technical report, Fachbereich Mathematik, Universität Dortmund, 1999. Node numbers are found at <http://www.mathematik.uni-dortmund.de/lx/research/projects/fliege/nodes/nodes.html>.
- [11] Henrik Moller. Fundamentals of binaural technology. *Applied Acoustics*, 1992.
- [12] M. Noisternig, T. Musil, and A. Sontacchi R. Holdrich. 3d binaural sound reproduction using a virtual ambisonic approach. In *IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003*, pages 174–178, July 2003.
- [13] M. A. Poletti. A unified theory of horizontal holographic sound systems. *J. Audio Eng. Soc.*, 48(12):1155–1182, December 2000.
- [14] Frank Wefers and Michael Vorländer. Frequency domain filter exchange for dft-based fast convolution. In *AIA-DAGA including the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA)*, 2013.
- [15] Micheal Weinert, Gerald Enzner, Jan-Mark Batke, Peter Jax, and Christiane Antweiler. Komfortable messung und bereitstellung individueller kopfbezogener impulsantworten als opendaff. In *DAGA*, 2012.
- [16] Fred Wightman and Doris Kistler. Measurement and validation of human hrtfs for use in hearing research. *Acta Acustica united with Acustica*, 91:429–439, 2005.
- [17] Earl G. Williams. *Fourier Acoustics*, volume 93 of *Applied Mathematical Sciences*. Academic Press, 1999.
- [18] Franz Zotter. *Analysis and Synthesis of Sound Radiation with Spherical Arrays*. PhD thesis, Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, 2009.



(a) Order  $N = 15$  results in a good approximation of original HRTF function as shown in Figure 4.

(b) Order  $N = 6$  shows a fair approximation of original HRTF function, the structure of the magnitude spectrum is maintained up to 3 kHz.



(c) Order  $N = 3$  only gives a limited approximation of the original HRTF function.

(d) Order  $N = 1$  yields a poor approximation of the original HRTF function.

**Fig. 10:** The HRTF representations visually indicate the limited spatial resolution caused by the HOA transformation. The shown frequency range is limited to 0–6 kHz according to the stimulus of the listening test.





## LITERATUR

---

- [1] C. Antweiler und M. Antweiler. "System identification with perfect sequences based on the NLMS algorithm: Sequences and sets of sequences with low crosscorrelation and impulse-like autocorrelation and their applications". In: *AEU. Archiv für Elektronik und Übertragungstechnik* 49.3 (1995), S. 129–134.
- [2] G. B. Arfken, H. J. Weber und F. E. Harris. *Mathematical methods for Physicists - A comprehensive guide*. 7. Elsevier, 2013.
- [3] Auro Technologies. "Productions in AURO-3D". In: *White Paper* (2011).
- [4] K. Backhaus, B. Erichson, W. Plinke und R. Weiber. *Multivariate Analysemethoden*. 9. Aufl. Springer-Verlag Berlin Heidelberg, 2000.
- [5] J.-M. Batke, S. Abeling, S. Balke und G. Enzner. "Investigation of HRTF Sets Using Content with Limited Spatial Resolution". In: *135th AES Convention, New York* (Okt. 2013).
- [6] J.-M. Batke, S. Abeling, P. Jax, M. Weinert und G. Enzner. "On the Use of Continuous-Azimuth HRIR in Higher-Order Ambisonics Rendering". In: *40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA)* (2013).
- [7] S. Bech und N. Zacharov. *Perceptual Audio Evaluation - Theory, Method and Application*. Wiley, 2007.
- [8] D. Begault, E. Wenzel, A. Lee und M. Anderson. "Direct comparison of the impact of Head Tracking, Reverberation, and Individualized HRTFs on the Spatial Perception of a virtual Speech Source". In: *J. Audio Eng. Soc* 49.10 (2001), S. 904–916.
- [9] J. Blauert. *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*. The MIT Press, 1996.
- [10] J. Blauert. *The Technology of Binaural Listening*. Springer, 2013.
- [11] G. Box, J. Hunter und W. Hunter. *Statistics for experimenters: design, innovation, and discovery*. Wiley series in probability and statistics. Wiley-Interscience, 2005.
- [12] P. Boyle. *Physical Mathematics*. Lecture Notes. 2012. URL: [http://www.ph.ed.ac.uk/~paboyle/Teaching/PhysicalMaths/2012/notes\\_2012.pdf](http://www.ph.ed.ac.uk/~paboyle/Teaching/PhysicalMaths/2012/notes_2012.pdf).
- [13] J. Breebaart und C. Faller. *Spatial Audio Processing: MPEG Surround and Other Applications*. John Wiley & Sons, Ltd, 2007.

- [14] A. S. Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.
- [15] I. Bronštejn, K. A. Semendjajew, G. Musiol und H. Mühlig. *Taschenbuch der Mathematik*. Bd. 7. Harri Deutsch Verlag, 2008.
- [16] Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. *Neuer Rekord bei Smartphones*. Sep. 2013.
- [17] C. I. Cheng und G. H. Wakefield. "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space". In: *Audio Engineering Society Convention 107*. Audio Engineering Society. 1999.
- [18] D. Colton und R. Kress. *Inverse Acoustic an Electromagnetic Scattering Theory*. Hrsg. von J. E. Marsden und L. Sirovich. Springer-Verlag Berlin Heidelberg New York, 1998.
- [19] comScore, Inc. *48 Prozent der deutschen Mobilebesitzer nutzen Smartphones*. Okt. 2012. URL: [http://www.comscore.com/ger/Insights/Press\\_Releases/2012/12/48\\_Prozent\\_der\\_deutschen\\_Mobilebesitzer\\_nutzen\\_Smartphones](http://www.comscore.com/ger/Insights/Press_Releases/2012/12/48_Prozent_der_deutschen_Mobilebesitzer_nutzen_Smartphones).
- [20] Dolby Laboratories, Inc. "Dolby Atmos - Next-Generation Audio for Cinema". In: *White Paper* (2013).
- [21] G. Enzner. "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering". In: *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*. IEEE. 2009, S. 325–328.
- [22] G. Enzner, M. Krawczyk, F.-M. Hoffmann und M. Weinert. "3D reconstruction of HRTF-fields from 1D continuous measurements". In: *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE. 2011, S. 157–160.
- [23] L. Fahrmeir, T. Kneib und S. Lang. *Regression*. 2. Aufl. Springer Heidelberg Dordrecht London New York, 2009.
- [24] L. Fahrmeir, R. Künstler, I. Pigeot und G. Tutz. *Statistik*. 5. Aufl. Springer Berlin Heidelberg New York, 2007.
- [25] A. Farina. "Simultaneous measurement of impulse response and distortion with a swept-sine technique". In: *Audio Engineering Society Convention 108*. Audio Engineering Society. 2000.
- [26] P. Fellgett. "Ambisonics part one: General system description". In: *Studio Sound* 17.8 (1975), S. 20–22.
- [27] P. B. Fellgett. "Ambisonic reproduction of directionality in surround-sound systems". In: *Nature* 252.5484 (1974), S. 534–538.
- [28] J. Fliege und U. Maier. *Integration nodes for the sphere*. aufgerufen am 29.10.2013. 2010. URL: <http://www.mathematik.uni-dortmund.de/lxx/research/projects/fliege/nodes/nodes.html>.

- [29] M. A. Gerzon. "Ambisonics. Part Two: Studio Techniques". In: *Studio Sound* 17.8 (1975), S. 24–26.
- [30] R. Huch und K. Jürgens. *DMensch, Körper, Krankheit: Anatomie, Physiologie, Krankheitsbilder; Lehrbuch und Atlas für die Berufe im Gesundheitswesen*, 6. Elsevier GmbH, München, 2011.
- [31] ITU. "Method for the subjective assessment of intermediate quality level of coding systems". In: *Recommendation BS.1534-1 (01/03)* (2003).
- [32] M. Janczyk und R. Pfister. *Inferenzstatistik verstehen: von A wie Signifikanztest bis Z wie Konfidenzintervall*. Springer Berlin Heidelberg New York, 2013.
- [33] R. Koenen. *Overview of the MPEG-4 Standard*. Techn. Ber. MPEG, 2002.
- [34] S. Kordon, H. Kropp, A. Krueger und J.-M. Batke. "Optimization of Spherical Microphone Array Recordings". In: *International Conference on Spatial Audio*. 2011.
- [35] H. G. Musmann und J. Ostermann. *Skript zur Vorlesung Statistische Methoden der Nachrichtentechnik*. Leibniz Universität Hannover, 2005.
- [36] R. Nicol. "Binaural Technology". In: Audio Engineering Society. 2010.
- [37] L. Papula. *Mathematik für Ingenieure und Naturwissenschaftler 3: Vektoranalysis, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Fehler- und Ausgleichsrechnung*. Viewegs Fachbücher der Technik. Vieweg + Teubner, 2008.
- [38] M. Poletti. "Unified description of ambisonics using real and complex spherical harmonics". In: *Proc. Ambisonics Symposium*. 2009.
- [39] J. G. Proakis. *Digital Signal Processing: Principles, Algorithms, and Applications, 4/e*. Pearson Education India, 2007.
- [40] V. Pulkki. "Virtual sound source positioning using vector base amplitude panning". In: *Journal of the Audio Engineering Society* 45.6 (1997), S. 456–466.
- [41] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: <http://www.R-project.org/>.
- [42] B. Rafaely. "Analysis and design of spherical microphone arrays". In: *Speech and Audio Processing, IEEE Transactions on* 13.1 (2005), S. 135–143.
- [43] B. Rafaely. "Plane-wave decomposition of the sound field on a sphere by spherical convolution". In: *The Journal of the Acoustical Society of America* 116 (2004), S. 2149.

- [44] L. Rayleigh. "XII. On our perception of sound direction". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13.74 (1907), S. 214–232.
- [45] S. Spors, R. Rabenstein und J. Ahrens. "The theory of wave field synthesis revisited". In: *124th AES Convention*. 2008, S. 17–20.
- [46] Statsmodels Developers. *StatsModels - Statistics in Python*. 2013. URL: <http://statsmodels.sourceforge.net/>.
- [47] W. Stier. *Empirische Forschungsmethoden*. 2. Aufl. Springer Berlin Heidelberg New York, 1999.
- [48] A. Telle, C. Antweiler und P. Vary. "Der perfekte Sweep-Ein neues Anregungssignal zur adaptiven Systemidentifikation zeitvarianter akustischer Systeme". In: *Proc. of Deutsche Jahrestagung für Akustik (DAGA)* (2010), S. 341–342.
- [49] TOMORROW FOCUS Media. *Mobile Effects 2013-1 - A part of our lifes*. 2013.
- [50] F. Völk. "System theory of binaural synthesis". In: *Audio Engineering Society Convention 131*. Audio Engineering Society. 2011.
- [51] F. Völk, F. Heinemann und H. Fastl. "Externalization in binaural synthesis: effects of recording environment and measurement procedure". In: *Journal of the Acoustical Society of America* 123.5 (2008), S. 3935.
- [52] D. Wang und G. Brown. *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley, 2006.
- [53] M. Weinert, G. Enzner, J.-M. Batke, P. Jax und C. Antweiler. "Komfortable Messung und Bereitstellung individueller kopfbezogener Impulsantworten als OpenDAFF". In: *Proceedings of German Annual Conference on Acoustics (DAGA)*. DEGA. 2012, S. 703–704.
- [54] E. G. Williams. *Fourier Acoustics*. ACADEMIC PRESS, 1999.