



TECHNISCHE UNIVERSITÄT ILMENAU

Fakultät für Elektrotechnik und Informationstechnik

Masterarbeit

Perzeptive Bewertung von Fehlereinflüssen bei der binauralen Auralisation von Kugelarraydaten unter Verwendung des Spatial Audio Quality Inventory - SAQI

vorgelegt von: Mathias Hellmich
eingereicht am: 30.04.2014
geboren am: 31.10.1987 in Dresden
Studiengang: Master Medientechnologie
Anfertigung im Fachgebiet: Elektronische Medientechnik
Fakultät für Elektrotechnik und Informationstechnik
Verantwortlicher Professor: Prof. Dr.-Ing. Dr. rer. nat. h.c. mult. K. Brandenburg
Wissenschaftliche Betreuer: Dipl.-Ing. J. Nowak
Dipl.-Ing. S. Werner

urn:nbn:de:gbv:ilm1-2014200049

Danksagung

An dieser Stelle möchte ich mich bei Dipl. Ing. Johannes Nowak und Dipl. Ing. Stephan Werner für die Betreuung meiner Masterarbeit sowie für die vielen hilfreichen Hinweise in dieser Zeit bedanken.

Ein ganz besonderer Dank geht schließlich an meine Familie und Freunde, die mich stets unterstützt haben.

Kurzfassung

Bei der Aufnahme von Schallfeldern durch Kugelmikrofonarrays treten Messfehler auf. Diese Fehler können zum Beispiel räumliches Aliasing, Mikrofonrauschen und Positionierungsfehler sein. Bemerkbar machen sich diese Fehler als tieffrequentes Rauschen sowie Verzerrungen im hochfrequenten Bereich. Bei der Wiedergabe von räumlichen Schallfeldern, welche mittels Kugelmikrofonarrays aufgenommen werden können, wirken sich diese Fehler ebenso negativ auf die Auralisationsqualität aus. In dieser Arbeit sollen diese Fehler bei der binauralen Wiedergabe über Kopfhörer perzeptiv untersucht werden. Dazu werden zwei Hörtests durchgeführt. Ein Hörtest wurde konzipiert, um den Schwellwert der Wahrnehmung dieser Fehler zu bestimmen. Der zweite Hörtest, welcher auf der Repertory Grid Technik basiert, dient der Zuordnung von einzelnen Fehlern zu akustischen Merkmalen. Zur Beschreibung der Fehler wird das Spatial Audio Quality Inventory (SAQI) herangezogen, welches eine aktuelle Sammlung qualitätsbeschreibender Merkmale darstellt. Die verwendeten Merkmale aus SAQI beschreiben räumliche Eigenschaften und Artefakte wie beispielsweise eine Anhebung des tieffrequenten Bereichs oder eine metallische Klangfarbe. Eine Hauptkomponentenanalyse wurde durchgeführt um die Daten zu ordnen und zu bewerten. Damit ist es möglich, wichtige Merkmalsgruppen zu identifizieren und die perzeptiven Merkmale der Fehlerinflüsse zu bewerten.

Abstract

When recording a sound field with a spherical microphone array, there always occur measurement errors. These errors can be spatial aliasing, microphone noise and positioning errors. These errors are noticeable as low frequency noise as well as distortion in the high frequency range. During playback of spatial sound fields, which can be recorded by spherical microphone arrays, these errors also have a negative effect on the auralisation quality. In this work these errors will be perceptually examined for binaural auralisation via headphones. For this purpose a listening test is designed to determine the perceptual thresholds of these errors. A second listening test, which is based on the repertory grid technique, is used to assign characteristics to the individual errors. To describe these errors, the recently published Spatial Audio Quality Inventory (SAQI) will be used, which is an up to date collection of quality descriptors. The analysis is used to determine whether the descriptive characteristics can be grouped so that recording errors and perceptual features can be linked. The features used in SAQI describe spatial properties and artifacts such as an increase in the low frequency range or a metallic timbre. A principal component analysis was conducted to organize and evaluate the data. Thus it is possible to identify important feature groups and to assess the perceptual characteristics of the error influences.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Zielstellung und Umsetzung	2
1.3	Kapitelübersicht	3
2	Stand der Technik	4
2.1	Qualitätsmerkmale in der Akustik	4
2.2	Hörtests	7
2.2.1	Hörtestmethoden	8
2.2.2	Deskriptive Untersuchungsmethoden	9
2.2.3	Skalierung	11
2.2.4	Befangenheit und Training	12
2.3	Mikrofonarrays	13
2.3.1	Einsatzgebiete	13
2.3.2	Komplexität	15
2.3.3	Beamforming	15
3	Grundlagen	17
3.1	Spatial Audio Quality Inventory	17
3.2	Hörtests	19
3.2.1	Adaptive Prozeduren	19
3.2.2	Repertory Grid Technique	22
3.3	Parameter zur räumlichen Wahrnehmung	23
3.3.1	Azimuth	23
3.3.2	Elevation	26
3.3.3	Distanz	26
3.3.4	Schallquellenbreite	28
3.4	Kugelarraydatenverarbeitung	29
3.4.1	Wellengleichung	29

3.4.2	Schallausbreitung	31
3.4.3	Kugelmikrofonarray	32
3.4.4	Sphärische Harmonische	35
3.4.5	Fehleranalyse	38
3.4.6	Binaurale Auralisation	43
4	Methodik	45
4.1	Konfiguration	45
4.2	Informeller Test	46
4.3	Itemerstellung	47
4.4	Auftretende Probleme	49
5	Hörtest Schwellwertbestimmung	52
5.1	Design	52
5.1.1	Testmethode	52
5.1.2	Testdaten	53
5.2	Auswertung und Ergebnisse	54
5.2.1	Aliasing	54
5.2.2	Rauschen	55
5.2.3	Positionierungsfehler	55
6	Repertory Grid Test	58
6.1	Design	58
6.1.1	Testmethode	58
6.1.2	Testdaten	59
6.2	Auswertung und Ergebnisse	60
6.2.1	Bemerkungen der Probanden	60
6.2.2	Merkmalshäufigkeiten	61
6.2.3	Hauptkomponentenanalyse	63
7	Fazit	72
7.1	Zusammenfassung	72
7.2	Ausblick	73
	Literaturverzeichnis	74
	Abbildungsverzeichnis	82

Tabellenverzeichnis	84
Verzeichnis der verwendeten Abkürzungen	85
Verzeichnis der verwendeten Symbole	86
A Anhang	87
Eigenständigkeitserklärung	94

1 Einleitung

1.1 Motivation

Die virtuelle Akustik beschäftigt sich unter anderem mit der Simulation von Räumen. Ein aktueller Untersuchungstrend der virtuellen Akustik beschäftigt sich dabei mit räumlichen Schallfeldern. Zur Aufnahme dieser räumlichen Schallfelder sowie deren räumlicher Eigenschaften werden Mikrofonarrays verwendet. Um diese Eigenschaften auch korrekt wiedergeben zu können, wird die Wellenfeldsynthese, Ambisonics oder die Binauralisierung über Kopfhörer eingesetzt. Mit diesen Techniken kann beim Hörer eine akustische Illusion des aufgenommenen Raumes erzeugt werden.

Die Aufnahme solcher komplexer Schallfelder begann mit einfachen Stereomikrofonaufnahmen und kann bis hin zu dreidimensionalen Mikrofonanordnungen (*Mikrofonarrays*) erweitert werden, welche mittlerweile ebenfalls kommerzielle Verwendung finden. Die Möglichkeiten zur Wiedergabe solcher Schallfelder sind ebenso vielfältig wie die zur Aufnahme. Moderne Systeme beschäftigen sich mit der Wiedergabe des räumlichen Schallfeldes, wie zum Beispiel die Wellenfeldsynthese ([Berkhout, 1988]), die mehrere Lautsprecher zur Erzeugung von künstlichen Wellenfronten nutzt. Ein anderer Ansatz ist die Wiedergabe durch die binaurale Technik, bei der das räumliche Schallfeld unter der Berücksichtigung der kopfbezogenen Übertragungsfunktionen auf zwei binaurale Kanäle verrechnet wird ([Moller, 1992]). Bei der *Auralisation (Hörbarmachung)* dieser Schallfelder spielt die Qualität eine besondere Rolle. Es entstehen bei der Aufnahme von Schallfeldern durch Mikrofonarrays Fehler, welche sich in Rauschen, Aliasing und Positionierungsfehlern unterteilen lassen. Diese Fehler wurden bereits analytisch beschrieben von [Rafaely, 2005] und haben einen negativen Effekt auf die räumliche Antwort des Mikrofonarrays. Mikrofonrauschen und Positionierungsfehler wirken sich negativ auf die tieffrequenten Anteile des Arraysignals aus, Aliasing dagegen erzeugt Fehler im hochfrequenten Bereich. Wie sich diese Fehler jedoch auf die Wahrnehmung auswirken und wann ein deutlicher Qualitätsverlust festzustellen ist, wurde bisher nur in wenigen Veröffentlichungen untersucht, wie zum Beispiel in [Avni et al., 2013]. Um die Komplexität von Kugelmikrofonarrays zu reduzieren, ist es daher von Interesse, wie-

viele Mikrofone bei gegebener Größe des Mikrofonarrays für eine ausreichende Qualität nötig sind.

Es ist von Interesse, Mikrofonarrays und entstehende Fehler perzeptiv zu untersuchen, um diese Arrays anhand der Ergebnisse für Auralisationsanwendungen gestalten zu können. Dabei spielt nicht nur die Qualitätsbeschreibung der Fehler, sondern auch entsprechende Schwellwerte eine Rolle.

Daher widmet sich diese Arbeit der perzeptiven Evaluation der verschiedenen Fehler, welche bei der Aufnahme von Schallfeldern durch Mikrofonarrays entstehen. Der Grundstein für die auditive Bewertung wurde dabei kurz vor Ende des 19. Jahrhunderts von [Sabine, 1922] mit der bekannten Formel für den Nachhall gelegt, welche in Kapitel [2.1] erläutert wird. Im Laufe der Zeit kamen immer mehr objektive Parameter zur Beschreibung von Qualität hinzu, wie *Intimacy* und *Loudness*. Obwohl bereits etliche Parameter und Merkmale zur Beschreibung entdeckt wurden, bleibt das Forschungsgebiet bis heute aktuell. Eine historische Übersicht gibt dabei [Lacatis et al., 2008].

1.2 Zielstellung und Umsetzung

In der vorliegenden Arbeit werden Fehler perzeptiv untersucht, welche bei der Aufnahme von Schallfeldern durch Kugelmikrofonarrays entstehen. Die Schallfelder werden mit den entsprechenden Fehlern simuliert und anschließend binaural über Kopfhörer wiedergegeben. Die drei Fehlerarten, die in dieser Arbeit untersucht werden, sind Positionierungsfehler, Mikrofonrauschen und räumliches Aliasing, welche in den Grundlagen erläutert werden.

Es wird untersucht, welche Fehlerarten zu welchen hörbaren Qualitätseinbußen führen und wo die Schwellwerte dieser Fehler liegen. Dazu werden zwei Hörtests mit vorwiegend Expertenhörern durchgeführt. Als erstes wurde ein *3 Alternative Forced Choice (3AFC)*-Test genutzt, um eine Schwellwertbestimmung der einzelnen Fehler durchzuführen [Levitt, 1971]. In der zweiten Versuchsreihe wird eine *Repertory Grid Technique (RGT)* verwendet, welche die Zuordnung von Qualitätsmerkmalen zu auftretenden Fehlern ermöglichen soll [Kelly, 1955]. Eine Übersicht über Hörtestmethoden wird gegeben sowie die verwendeten Testmethoden erläutert. Zur Bewertung wird eine neue Sammlung verschiedener Merkmale zur Bestimmung von Audioqualität in dieser Arbeit eingesetzt, genannt *Spatial Audio Quality Inventory (SAQI)* [Lindau et al., 2014].

1.3 Kapitelübersicht

Für die vorliegende Arbeit wurde folgende Gliederung gewählt:

- Kapitel 2 stellt relevante existierende Ansätze zur perceptiven Bewertung von Audiodaten vor. Es wird eine Übersicht über Hörtestmethoden sowie über deskriptive Untersuchungsmethoden gegeben. Desweiteren werden akustische Merkmale zur Beschreibung von Qualität diskutiert.
- Kapitel 3 gibt einen Überblick über die theoretischen Grundlagen. Die räumliche Wiedergabe von Schall wird betrachtet sowie in diesem Zusammenhang stehende Parameter wie beispielsweise *Kopfbezogene Übertragungsfunktionen*. Bei der Verarbeitung räumlicher Signale nimmt die sphärisch harmonische Dekomposition für die Berechnungen einen wichtigen Platz ein. Da sie eine grundlegende Operation für die Zerlegung räumlicher Schallfelder ist, wird auf sie ebenfalls eingegangen.
- Kapitel 4 beschreibt die Vorbereitung und Methodik der Hörtests. Es werden auftretende Probleme sowie die Erarbeitung von Konfigurationsdetails beschrieben.
- Kapitel 5 präsentiert den Aufbau, die Durchführung sowie Untersuchungsdetails und Ergebnisse des *3AFC*-Schwellwerhörtests. Die gefundenen Ergebnisse geben Aufschluss über die Minimalkonfiguration eines Kugelmikrofonarrays für akustische Aufnahmen zur binauralen Wiedergabe über Kopfhörer.
- Kapitel 6 präsentiert den Aufbau, die Durchführung sowie Untersuchungsdetails und Ergebnisse des Repertory Grid Tests. Akustische Merkmale können durch diesen Test den verschiedenen Fehlerarten zugeordnet werden.
- Kapitel 7 fasst die erarbeiteten Ergebnisse kurz zusammen. Weiterhin werden interessante Vorschläge und Ansätze zukunftsweisend formuliert.

2 Stand der Technik

Als Stand der Technik sind die Qualitätsmerkmale und Hörtests zu betrachten, welche den Schwerpunkt dieser Arbeit bilden. Verschiedene Qualitätsmerkmale werden in einem Repertory Grid Test untersucht. Ein ähnliches Konzept schlägt [Avni et al., 2013] vor. Eine Neuerung dazu ist das in dieser Arbeit erstmals verwendete Qualitätsinventar *SAQI*. Ebenso wird eine Übersicht über Mikrofonarrays gegeben.

2.1 Qualitätsmerkmale in der Akustik

In den letzten Jahren wurden Zusammenfassungen von beschreibenden Qualitätsmerkmalen für verschiedene Audiodomänen erstellt. Diese Bereiche beinhalten beispielsweise die Raumakustik, Lautsprecherforschung, Wiedergabesysteme, Audio Coder oder auch VAEs (*Virtual Acoustic Environments*). Dabei basieren solche Zusammenfassungen bisher immer auf Wortschätzen, welche meist aus der Erfahrung der Autoren oder aus literarischen Werken stammen. Die erste analytische Beschreibung in diesem Bereich war die Näherung des Nachhalls durch Sabine [Sabine, 1922], welche in diesem Kapitel beschrieben wird. Erst vor kurzem wurde begonnen, methodisch fortgeschrittenere Prozeduren für die Erstellung solcher Attribute zu entwickeln. Dazu gehören die *Quantitative Descriptive Analysis (QDA)* [Zacharov and Koivuniemi, 2001b], *Free Choice Profiling (FCP)* [Lorho, 2005], oder die *Repertory Grid Technique (RGT)* [Berg and Rumsey, 1999]. Bisher ist noch keiner dieser Wortschätze vollständig oder deckt alle wichtigen Bereiche für die Qualitätsbewertung ab. [Lindau et al., 2014] stellt ein detailliertes Vokabular von perzeptiven Qualitätsmerkmalen vor, welches auf eine große Breite von akustischen Umgebungen, Wiedergabesystemen oder auch Simulationen anwendbar sein soll. Dieses Vokabular wird als *SAQI (spatial audio quality inventory)* bezeichnet und ist das aktuellste im Bereich der auditiven Qualitätsbewertung und findet ebenso in der vorliegenden Arbeit Verwendung. Nur eine weitere Publikation hat sich mit der Entwicklung eines Vokabulars beschäftigt, allerdings nur für *Virtual Acoustic Environments* [Silzle, 2007]. [Bech and Zacharov, 2006] gibt eine Übersicht über Qualitätsvokabulare, welche in Sprachqualität und räumliche Audioqualität sowie

sonstige Qualitätsmerkmale unterteilt sind.

Akustische Beschreibungen finden normalerweise durch verschiedene subjektive Charakteristiken statt. Einige davon finden in der Literatur häufig Erwähnung, jedoch werden diese Merkmale je nach Autor unterschiedlich stark gewichtet. Zu diesen Beschreibungen gehören unter anderem unter anderem *Nachhall*, *Klarheit*, *Wärme* und *Brillanz* sowie *Apparent Source Width* und *Listener Envelopment* [Donovan et al., 2008]. Nachfolgend sollen einige wichtige und häufig in der Literatur auftretende Charakteristiken zur Qualitätsbeschreibung diskutiert werden:

Nachhall: In geschlossenen Räumen wird, wie in Kapitel [3.4.2] beschrieben, der Schall an den Wänden reflektiert, und lässt den Klang einer Schallquelle auch nach ihrem Verstummen einige Zeit weiterklingen. Abhängig ist die Dauer dieses Nachhalls von der Größe des Raumes sowie der Stärke der Absorption durch die Wände. Definiert ist die Nachhallzeit als die Zeit, welche das Signal benötigt, um 60 dB an Schalldruckpegel abzunehmen. [Sabine, 1922] schlug bereits 1920 eine Näherung für die Nachhallzeit vor:

$$T_{60} = 0.161 \frac{V}{\bar{\alpha} S} \quad (2.1)$$

Dabei ist V das Raumvolumen, $\bar{\alpha}$ der durchschnittliche Absorptionskoeffizient und S die Wandfläche des Raumes. Nachhallzeiten können stark variieren, wie Tabelle [2.1] deutlich macht:

Tabelle 2.1: Typische Nachhallzeiten für verschiedene Raumtypen [Boone et al., 1995]

Raumtyp	Nachhallzeit [s]
Wohnungszimmer	0.5
Kino	0.7 – 1.0
Opernhaus	1.2 – 1.6
Kirche	1.5 – 2.5

Eine genauere Bestimmung der Nachhallzeit erfolgt durch Rückwärtsintegration der aufgenommenen Impulsantwort eines Raumes, wobei $h_{SR}(\tau)$ die Impulsantwort zum Zeitpunkt τ darstellt:

$$h_i^2(t) = \int_t^\infty [h_{SR}(\tau)]^2 d\tau \quad (2.2)$$

Aus dem logarithmischen Plot wird diese dann anhand der Gerade von $h_i^2(t)$ berechnet [Schuitman, 2011].

Brillanz: Eine Schallquelle klingt brilliant, wenn sie klar und flächig wahrgenommen wird und viele Oberschwingungen besitzt. Dieser Parameter ist eng verbunden mit der Klangfarbe, und eine Möglichkeit, diese zu beurteilen. Brillanz kann durch die folgende Berechnung des *Treble-Verhältnisses* (*treble ratio*, TR) ausgewertet werden:

$$TR = \frac{RT_{2000} + RT_{4000}}{RT_{500} + RT_{1000}}, \quad (2.3)$$

wobei RT_f die gemessene Nachhallzeit für das Oktav-Frequenzband mit der Mittenfrequenz $f(Hz)$ ist [Beranek, 2010].

Wärme: Wärme ist definiert als die Fülle der Basstöne. Sie wird ähnlich wie die Brillanz ausgewertet und ist ebenfalls stark mit der Klangfarbe verbunden. Berechnet wird sie durch das *Bass-Verhältnis* (*bass ratio*, BR):

$$BR = \frac{RT_{125} + RT_{250}}{RT_{500} + RT_{1000}}, \quad (2.4)$$

Für Musik werden werden Bassverhältnisse von $BR = 1,0 \dots 1,3$ angestrebt. Um die Sprachverständlichkeit beizubehalten, sind niedrigere Werte von maximal $BR = 1,0$ wünschenswert [Weinzierl, 2008] [Beranek, 2010].

Klarheit: Die Klarheit beschreibt, wie gut verschiedene Komponenten eines Signals wahrzunehmen sind. Starker Nachhall kann zum Verschwimmen der Signale führen, was die Hörbarkeit von Details im Signal verschlechtert. Die folgende Formel ist als *Klarheitsindex* (*clarity index*) gebräuchlich:

$$C_t = 10 \log_{10} \left(\frac{\int_0^t h_{SR}^2(t) dt}{\int_t^\infty h_{SR}^2(t) dt} \right) [dB] \quad (2.5)$$

Für Musik wird die Integrationszeit t auf 80 ms (C_{80}) gesetzt. Dagegen wird bei der Betrachtung von Sprache 50 ms (C_{50}) Integrationszeit eingesetzt und eher der Begriff *Deutlichkeit* verwendet. Die Werte des Klarheitsindex liegen in der Regel zwischen -5 und +5 dB [Möser, 2009].

Intimität: Intimität stellt eine Verbundenheit dar, welche vornehmlich in kleinen Räumen herrscht. Sie kann mit dem *Initial Time Delay Gap (ITDG)* bestimmt werden, welche die Zeit zwischen der Ankunft des Direktschalls und der ersten Reflexionen beschreibt. In einer Konzerthalle sind große *ITDGs* vorhanden, was die Intimität senkt. In solchen Räumen liegen die Werte meist zwischen 15 und 30 ms [Beranek, 2010].

Listener Envelopment: Das *Listener Envelopment (LEV)* entspricht der subjektiven Empfindung des "vom Schall eingehüllt sein" und dient damit der räumlichen Interpretation eines Raumes. Beschrieben wird es durch die späten Signalteile der interauralen Korrelation ($IACC_L$), welche in Kapitel [3.3] erläutert werden:

$$LEV = 1 - IACC_L \quad (2.6)$$

Der Wert für die zeitigen Signalanteile korreliert dabei mit der subjektiven Wahrnehmung der "Breite" der Schallquelle, welche als *Apparent Source Width (ASW)* bezeichnet wird [Möser, 2009].

2.2 Hörtests

Als Untersuchungsmethode von auditiven Charakteristiken haben sich Hörtests etabliert, da bei der Wiedergabe von Audioszenen immer die Wahrnehmung des Hörers im Vordergrund steht und nicht einzelne gemessene Parameter.

Hörtests können keine direkte Qualitätsangabe über Audiostreame geben oder ein bestimmtes Problem identifizieren. Allerdings sind sie in der Lage über folgende Klassifizierungen Aufschluss zu geben [Bech and Zacharov, 2006]:

- werden Audiostreame als identisch wahrgenommen oder nicht
- werden Audiostreame gleich, besser oder schlechter wahrgenommen als andere

Derartige Vergleiche sind gängige Praxis bei Hörtests. Um diese Vergleiche vornehmen zu können, wird sehr oft mit akustischen Referenzen und Ankern gearbeitet, welche Qualitätsoptima bzw. Qualitätsminima darstellen. Bei sogenannten *hidden references* (versteckten Referenzen) ist dem Hörer nicht bekannt, welche der gegebenen Stimuli die Referenz darstellt [ITU-R, 2003]. Damit wird festgestellt, ob eine Audioanwendung der Qualität nach verwendbar ist.

Bei Hörtests sollten eine Reihe von Gegebenheiten beachtet werden, um gültige Ergebnisse zu erhalten. Dazu gehört unter anderem die Art des Hörtests sowie Skalenwahl und Skaleneinteilungen. In den folgenden Abschnitten soll ein kurzer Überblick über einschlägige Problematiken gegeben werden.

2.2.1 Hörtestmethoden

Es existiert eine große Anzahl an Hörtestmethoden, die in verschiedenen Standards festgelegt werden. Die in diesem Kapitel vorgestellten Hörtests werden auch als quantitative Hörtests bezeichnet, da diese Tests eine große Anzahl von Probanden und Wiederholungen benötigen, um verwertbare Ergebnisse zu erhalten. Der Stand der Technik für die am häufigsten in der Literatur auftretenden Methoden wird in diesem Abschnitt beschrieben. Die folgende Übersicht ist [Hynninen and Zacharov, 1999] entnommen.

Paarvergleiche

Paarvergleiche werden aufgrund ihrer Einfachheit bereits seit über 100 Jahren eingesetzt. Der Testperson werden zwei Signale präsentiert, von welchem das bessere oder schlechtere ausgewählt werden soll. Dies ist eine weit verbreitete Testmethode, da sie einfach für Hörer zu verstehen und zu erlernen ist, selbst wenn sie nicht mit Hörtests vertraut sind.

Eine Erweiterung dazu sind skalierte Paarvergleiche, welche Intervalskalen zur genaueren Abstufung verwenden. Diese Methode wird häufig im Bereich Telekommunikation und akustische Untersuchungen eingesetzt. Eine Standardisierung dieser Methode erfolgt durch die *ITU-R BT.500-8* [ITU-R, 1998].

ABX Methode

Diese Methode präsentiert der Testperson drei Signale (A,B,X). Es soll festgestellt werden, ob A oder B identisch zu X sind. Der Vorteil liegt in einer eindeutigen Vergleichsbasis, welche das Signal X darstellt. Der Test erfordert dagegen etwas mehr Zeitaufwand als die Paarvergleiche.

Ein ähnliches Konzept stellt die *Triple Stimulus Hidden Reference*-Methode dar, welche die *ITU-R 5-stufige Wertminderungsskala* verwendet. Laut dieser wird nur das sich unterscheidende Signal bewertet, das andere Signal erhält zwangsweise die Maximalbewertung. Diese Methode ist weniger intuitiv und komplexer als die ABX-Methode und wird daher meist für erfahrenere Hörer verwendet. Sie ist in *ITU-R BS.1116-1* [ITU-R, 1994] standardisiert.

Alternative Forced Choice

Alternative Forced Choice (AFC)-Methoden sind weit verbreitet zur Bestimmung von Schwellwerten. Dem Hörer werden zwei Signale präsentiert, von denen er eines anhand eines Merkmals (beispielsweise *Entfernung*) auswählt. Es kann dabei unterschieden werden zwischen dem Vergleich von zwei (*2AFC*), drei (*3AFC*) oder sogar noch mehr Signalen *N-AFC*. Der *AFC*-Test ist eine adaptive Prozedur, bei welcher die aufeinanderfolgenden Stimuli stark von den vorangegangenen Antworten der Testperson abhängen. Bei der klassischen Methode hingegen stehen die aufeinanderfolgenden Stimuli von vornherein fest [Treutwein, 1995]. Verschiedene Formen von *AFC*-Methoden sind bekannt und werden in den Grundlagen näher erläutert, da diese in der vorliegenden Arbeit ebenfalls Verwendung finden [Levitt, 1971].

Rank Order

Diese Methode wird bei mehr als drei Testsignalen angewendet. Die Aufgabe der Testperson besteht darin, die einzelnen Signale nach auf- oder absteigender Stärke des zu untersuchenden Merkmals zu sortieren. Die Methode ist zwar für die Probanden einfach zu verstehen und leicht durchzuführen, allerdings können die Signale nicht einer festen absoluten Qualität zugeordnet werden, sondern nur einer qualitativen Rangfolge [Hynninen and Zacharov, 1999].

MUSHRA

Der *MUSHRA (Multi-Stimulus test with Hidden Reference and Anchor)* dient der subjektiven Beurteilung von Audioqualität, welche bei der *ITU-R BS.1116-1* in der unteren Hälfte der Skala liegen würde. Aus diesem Grund wird eine versteckte akustische Referenz, welche die höchstmögliche Qualität darstellt, sowie ein Anker, welcher die Qualitätsuntergrenze darstellt, verwendet [ITU-R, 2003]. Der Test wird häufig für Audiocodiersysteme verwendet und ist als *ITU-R BS.1534* Empfehlung aufgeführt [Bech and Zacharov, 2006].

2.2.2 Deskriptive Untersuchungsmethoden

Als Gegenstück zu den klassischen Hörtests, wie sie im letzten Abschnitt vorgestellt wurden, kommen deskriptive Untersuchungsformen zum Einsatz. Diese fortgeschritteneren Untersuchungsformen konzentrieren sich auf qualitative Auswertungen von wahrgenommenen Qualitätsaspekten. Probanden werden hierbei nach ihrem Empfinden in Bezug zur wahrgenommenen Qualität befragt. [Strohmeier et al., 2010] schlägt

eine Unterteilung in drei Hauptkategorien vor: *Interview-based Evaluation*, *Consensus Vocabulary Profiling* und *Individual Vocabulary Profiling*. Eine tabellarische Übersicht findet sich in Tabelle [2.2].

Interview-based Evaluation: Testteilnehmer beschreiben bei dieser Variante explizit die Ausprägungen der Testsignale oder deren erfasste Qualität. Ziel dabei ist eine Erstellung von Begriffen, welche in der Lage sind, die Qualität auszudrücken und nachzuprüfen, ob die Teilnehmer die erwarteten Qualitätsaspekte verwenden. Anwendungsbeispiele sind die *Interpretation-Based Quality evaluation* [Radun et al., 2008] und die Methode der *Experienced Quality Factors* [Jumisko-Pyykkö et al., 2007].

Consensus Vocabulary Profiling: Das Ziel des *Consensus Vocabulary Profiling* besteht darin, ein Konsensvokabular zu erstellen, welches dann von trainierten Teilnehmern genutzt wird, um die Qualität in einem separaten Testteil zu bewerten. Die Erstellung des Vokabulars geschieht zumeist in Gruppendiskussionen. Als Beispiel für den Audibereich sollte an dieser Stelle die *Audio Descriptive Analysis and Mapping (ADAM)* Technik [Zacharov and Koivuniemi, 2001a] genannt werden. Weitere Methoden, wie die *Flavor Profile Method*, die *Texture Profile Method* und die *Quantitative Descriptive Analysis* werden von [Bech et al., 1996] vorgestellt.

Individual vocabulary profiling: Diese Methode unterscheidet sich vom *Consensus Vocabulary Profiling* insofern, als dass hier das Vokabular einzelner Probanden aufgenommen und zu Bewertung verwendet wird. Dabei machen sich die Probanden zuerst mit den verwendeten Testsignalen vertraut, um anschließend eine Attributliste zu erstellen. Dazu wird auf die *Repertory Grid Technique* zurückgegriffen, welche in den Grundlagen erläutert wird. Anschließend wird mithilfe dieses Vokabulars ein Test zur Bewertung der Qualität durchgeführt. Anwendung findet diese Variante in *Free-Choice Profiling (FCP)* [Lorho, 2005], *Flash Profiling (FP)* [Delarue and Sieffermann, 2004] sowie die *Repertory Grid Technique (RGT)*, welche als Vokabularerhebung für die erstgenannten Methoden dient und von [Berg and Rumsey, 1999] als eigenständige Testmethode verwendet wird. Neuere Publikationen beschäftigen sich mit einer Kombination aus herkömmlichen quantitativen Hörtestmethoden und qualitativ-deskriptiven Testmethoden. [Strohmeier et al., 2010] stellt eine solche Methode als *Open Profiling of Quality (OPQ)* vor.

Tabelle 2.2: Übersicht über deskriptive Testmethoden

Ansatz	Interview-Based	Consensus vocabulary profiling	Individual vocabulary profiling
Methoden	Interpretation-Based Quality evaluation, Experienced quality factors	Flavor Profile Method, Texture Profile Method, Quantitative Descriptive Analysis, ADAM	Free-Choice Profiling, Flash Profiling, Repertory Grid Technique

2.2.3 Skalierung

Ein wichtiger Betrachtungsgegenstand bei den genannten Hörtests ist die *Skalierung*, bei der eine große Anzahl von Möglichkeiten vorhanden ist. Dabei ist nicht nur die Auflösung der jeweiligen Skala von Interesse, sondern ebenso, ob Anker oder Referenzen Verwendung finden. Bei der Art der Skala kann zwischen vier Hauptkategorien unterscheiden werden:

Nominal: Bei dieser einfachen Skaleneinteilung gibt es nur zwei Möglichkeiten, da ausschließlich unterschieden wird, ob zwei Stimuli sich unterscheiden oder nicht. Diese Skala wird auch als *category scale* bezeichnet.

Ordinal: Die Basisoperation ist hierbei eine Einteilung der Empfindungsstärke. Sie findet zum Beispiel bei *Rank Order Tests* Anwendung, wo mehrere Vergleichsobjekte nach ihrer Qualitätsstufe sortiert werden, wobei alle Qualitätsstufen monoton ansteigen.

Interval: Ähnlich der Ordinalskala wird hier die Empfindungsstärke unterteilt. Jedoch wird gleichzeitig versucht auszudrücken, um wieviel die zu untersuchenden Objekte sich unterscheiden. Es können zum Beispiel Ähnlichkeitsskalen anhand von prozentualen Vergleichen genutzt werden.

Ratio: Diese Skala basiert auf der Intervalskala. Der Unterschied ist, dass eine absolute Wert als Vergleich zur Verfügung steht. Dies kann ein absoluter Nullwert sein, zum Beispiel bei Untersuchungen von Basstönen, wobei der Nullwert keinen Bass enthält.

Exemplarisch sind zwei gebräuchliche Skalen in Tabelle [2.3] zu finden.

Tabelle 2.3: Siebenstufige Vergleichsskala nach der *MUSHRA-Methode* (links) und fünfstufige Wertminderungsskala nach *ITU-R* (rechts) [Bosi, 2004].

Grad	Vergleich	Stufe	Qualität	Wertminderung
3	Viel besser	5	Ausgezeichnet	Nicht wahrnehmbar
2	Besser	4	Gut	Wahrnehmbar, nicht störend
1	Etwas besser	3	Aktzeptabel	Leicht störend
0	Gleich	2	Schlecht	Störend
-1	Etwas schlechter	1	Ungenügend	Sehr störend
-2	Schlechter			
-3	Viel schlechter			

Desweiteren muss beachtet werden, dass bei der Skalierung die Beschreibungen von akustischen Parametern sehr abstrakt sein können. Daher ist es notwendig einen Kompromiss zwischen *Bedeutsamkeit* und *Einfachheit* bei der Erstellung des Hörtests zu wählen, da beide Attribute im Detail in verschiedene Richtungen auseinanderlaufen [Bech and Zacharov, 2006].

Bedeutsamkeit: Der Hörer sollte mit der Skaleneinteilung vertraut sein. Eine eindeutige Erklärung der Skala und der Zielstellung der Messung ist daher äußerst wichtig. Bei Messungen bestimmter akustischer Attribute müssen diese genau erklärt werden. Für die Auswertung ist dies ebenfalls hilfreich, da durch ein vergleichbares Verständnis der Probanden auch die Ergebnisse besser vergleichbar sind.

Einfachheit: Die Einfachheit des Aufbaus ist ausschlaggebend für den Hörer. Gerade bei Benutzeroberflächen können Unklarheiten auftreten, welche den Hörer beeinflussen. Je einfacher der Aufbau ist, desto weniger muss der Hörer auf den Versuch vorbereitet werden. Durch Einfachheit kann ein Test schnell und mit vielen Testsubjekten durchgeführt werden, die Ergebnisse können jedoch bei umfangreicheren Studien an Aussagekraft verlieren.

2.2.4 Befangenheit und Training

Ein großes Problem bei Hörtest ist die *Befangenheit*. Sie sorgt dafür, dass sich der Hörer durch die Umgebung oder seine eigene Mentalität meist unbewusst beeinflussen lässt. Zum Beispiel wird ein Hörer, welcher nicht mit der Skaleneinteilung vertraut ist, anfänglich sehr zögerlich mit der Ausreizung der Skala umgehen. Ebenso beeinflussen diese anfänglich gegebenen Bewertungen die späteren Entscheidungen des Hörers, da

dieser seine Bewertungseinstufung mit vorher gegebenen Werten vergleicht. Lautsprecher, welche für den Hörer sichtbar sind, haben dabei gravierenden Einfluss auf seine Aussagen, weshalb solche Art Tests für gewöhnlich ohne visuellen Kontakt zu den Lautsprechern durchgeführt werden [Bech and Zacharov, 2006].

Aufgrund der *Befangenheit* sollte ein *Training* des Hörers vor der Durchführung des eigentlichen Tests in Betracht gezogen werden. Das Training dient dazu, eine Eingewöhnung mit den experimentellen Konditionen zu schaffen sowie sich mit den neuen Attributen vertraut zu machen. Es kann damit im Vorfeld getestet werden, ob ein experimenteller Aufbau funktioniert oder nicht. Eventuelle Fragen, welche bei den Hörern auftreten, sollten während des Trainings beseitigt werden, da alle Unklarheiten Einfluss auf die erhaltenen Daten des Tests nehmen.

Weitere Einflussgrößen von potentiellen Problemen bei Hörtests werden in [Bech and Zacharov, 2006] ausführlich diskutiert und sind nachfolgend aufgezählt:

Abhörraum: Klassenzimmer, Studio, Konzertsaal

Wiedergabesystem: Surround Sound, Stereowiedergabe, Kopfhörerwiedergabe

Untersuchungsgegenstand: Musik, Sprache, sonstige Geräusche

Hörer: Expertenhörer oder Alltagshörer

2.3 Mikrofonarrays

2.3.1 Einsatzgebiete

Die ersten Mikrofonarrays wurden 1974 von [Billingsley, 1974] eingesetzt und haben sich seitdem erheblich weiterentwickelt [Michel, 2006]. Aktuelle Anwendungsgebiete von Mikrofonarrays sind die Vermessung von räumlichen Schallfeldern und Störgeräuschen in schwer zu vermessenden Umgebungen, wo keine direkten Messungen durchgeführt werden können. Dazu gehören unter anderem folgende Bereiche:

Vermessung von bewegten Objekten: Die Vermessung von bewegten Objekten ist gerade in der Mobilindustrie sehr von Interesse. Störgeräusche sind in diesem Bereich sehr laut und unangenehm. Daher werden Lärmabstrahlungen von Zugfahrzeugen ([King and Bechert., 1979]), von Flugzeugen während des Fluges ([Howell et al., 1986]), Motorgeräuschen von Kraftfahrzeugen ([Michel et al., 2004]) oder auch von Rotorblättern von Helikoptern

([Blacodon et al., 1987]) gemessen. In diesem Zusammenhang wurden ebenso zusammengesetzte Arrays verwendet ([Paschereit and Barsikow, 1994]). Der Abstand der Mikrofone wird dabei von innen nach außen hin größer, was die Auflösung bei tiefen Frequenzen verbessert.

Die Kombination eines Mikrofonarrays mit einer Kamera erlaubt die Überlagerung des aufgenommenen zu untersuchenden Objekts und dem visualisierten Schallfeld und wird als akustische Kamera bezeichnet wird [Heinz et al., 1999].

Quellenlokalisierung: Die Mikrofonarraytechnologie wird ebenfalls zur Quellenlokalisierung eingesetzt. Ebenso ist möglich, Störgeräusche aus anderen Richtungen zu unterdrücken, wenn die Position des aufzunehmenden Objektes bekannt ist. Dies kann in Konferenzsystemen oder auch zur Vermessung in Windtunneln, wo starkes Hintergrundrauschen vorherrscht, eingesetzt werden [Eichler, 2009]. Die akustische Abschattung von Räumen kann ebenso überprüft werden [Möser, 2005].

Wiedergabe: Auch in Film- und Musikproduktionen werden Mikrofonarrays eingesetzt, um ein realistisches Klangbild zu erzeugen. Hierbei steht das räumliche Schallfeld im Vordergrund [Eichler, 2009]. Eine Möglichkeit zur Wiedergabe dieses Schallfeldes ist die *Wellenfeldsynthese (WFS)*. Diese basiert auf dem *Huygenschen* Prinzip, welches besagt, dass eine Schallwelle durch mehrere kleinere Einzelwellen dargestellt werden kann. Die Einzelwellen werden in der *WFS* durch eine Vielzahl von Mikrofonen simuliert. Die Wellenfeldsynthese wird von [Berkhout, 1988] beschrieben. Ein ähnliches Konzept wird mit *Ambisonics* verfolgt, welche erstmals durch [Gerzon, 1973] vorgestellt wurden. Dabei wird ein omnidirektionaler Kanal und weitere dreidimensionale Bi-direktionale (in x-, y- und z-Richtung des kartesischen Koordinatensystems) Komponentenkanäle erstellt. Dieser Ansatz besitzt allerdings nur eine kleine Hörfläche durch die geringe räumliche Auflösung. Weitere Entwicklungen in diesem Bereich lassen höhere Auflösungen zu und werden *Higher Order Ambisonics (HOA)* genannt [Daniel et al., 2003]. Ein anderer Ansatz ist die binaurale Wiedergabe über Kopfhörer, wie sie von [Moller, 1992] beschrieben wird. Das räumliche Schallfeld wird mit kopfbezogenen Impulsantworten gefaltet und somit auf zwei binaurale Kanäle verrechnet.

2.3.2 Komplexität

Es werden verschiedene Formen von Mikrofonarrays verwendet, welche sich wie folgt kategorisieren lassen ([Möser, 2005]):

Eindimensionale Arrays: Bei eindimensionalen Mikrofonarrays die Mikrofone in einer Linie angeordnet. Dies stellt die einfachste Art der Mikrofonanordnung dar, welche zur horizontalen oder vertikalen Lokalisation eingesetzt werden kann.

Zweidimensionale Arrays: Für die zweidimensionale Anordnung sind Kreuzarrays und Ringarrays die geläufigsten Formen, wobei Ringarrays bessere Ergebnisse bei der Lokalisation erzielen als Kreuzarrays [Möser, 2005].

Dreidimensionale Arrays: Für räumliche Aufnahmen werden dreidimensionale Arrays verwendet. Die dreidimensionale Anordnung von Mikrofonen erlaubt eine große Anzahl an Positionierungsmöglichkeiten. Die meistgenutzte ist das Kugelmikrofonarray, bei der die Mikrofone in Form einer Kugel angeordnet sind. Diese Variante bildet in dieser Arbeit die Grundlage und wird daher bevorzugt betrachtet.

2.3.3 Beamforming

Das *Beamforming* erlaubt die Lokalisation von Schallquellen durch mehrere unterschiedlich positionierte Mikrofone. Die Summierung von Signalen dieser Mikrofone ist die einfachste Variante des *Beamforming* und wird als *Sum-Beamforming* bezeichnet. Weitere Methoden beinhalten eine zeitliche Verschiebung (*Delay-and-Sum-Beamforming*) oder Filterung (*Filter-and-Sum-Beamforming*) der Signale [Eichler, 2009]. Verfahren zur Störgeräuschunterdrückung sind beispielsweise *Generalized Sidelobe Canceller*. Ein weit verbreitetes Gegenstück zum *Beamforming* stellt die *Near-Field Acoustic Holography* dar. Deren Hauptvorteil ist, dass die räumliche Auflösung nur vom Mikrofonabstand abhängig ist, nicht aber von der Frequenz [Deblauwe et al., 2007]. [Brandstein and Ward, 2001] gibt einen Überblick über grundlegende Verfahren zur Sprachverbesserung sowie zur Quellenlokalisierung.

Eine Übersicht über wichtige existierende Techniken, sowie deren Vor- und Nachteile, findet sich in Tabelle [2.4].

Tabelle 2.4: Vor- und Nachteile von Beamforming Techniken [McCowan, 2001]

Technik	Vorteile	Nachteile
Delay-Sum	einfach	Schmalbandig, schlechte tief-frequente Auflösung
Sub-array Delay-Sum	Breitbandig	schlechte tieffrequente Auflösung
Superdirectivity	optimale Verstärkung	Annahme von diffusem Rauschen
Nearfield Superdirectivity	Nahfeldquellen, optimale Verstärkung, gute tieffrequente Auflösung	Annahme von diffusem Rauschen, nimmt Rauschen im Fernfeld an
Generalised Sidelobe Canceler	passt sich Rauschbedingungen an, minimiert Rauschenergie, starke Beschränkung auf das Signal	kann Verzerrungen in der Praxis verursachen, schlechte tieffrequente Auflösung
AMNOR	passt sich Rauschbedingungen an, minimiert Rauschenergie, Beschränkung auf das Signal, Verzerrungen können kontrolliert werden	hohe Komplexität, kann leichte Verzerrungen in der Praxis verursachen, schlechte tieffrequente Auflösung, Signal-Ruhe Schwellen werden nicht erkannt
Post-filtering	passt sich Rauschbedingungen an, verbessert Beamformersignale	kann das Signal verzerren

Aktuelle Forschungen im Bereich Mikrofonarrays beschäftigen sich mit der Veränderung der räumlichen Auflösung durch die Ordnung N ([Avni et al., 2013]) oder basieren auf dem *Modal Beamforming*, bei dem richtungsabhängige Radialfilter verwendet werden ([Zotter and Zaunschirm, 2014]). Diese werden errechnet durch die Zerlegung des Wellenfeldes in orthogonale Lösungen der Wellengleichung. Eine Übersicht dazu findet sich in [Teutsch, 2007].

3 Grundlagen

Dieses Kapitel befasst sich mit den Hintergründen zu den verwendeten Qualitätsmerkmalen und den erstellten Hörtests, welche in den Abschnitten [3.1] und [3.2] behandelt werden. Außerdem werden die zum Verständnis notwendigen physikalischen Grundlagen erläutert. Hierbei ist die Aufnahme und Weiterleitung des Schalls ein wichtiges Thema. In Abschnitt [3.3] wird auf die grundlegenden räumlichen Parameter eingegangen. Darauf folgend wird in [3.4] der in dieser Arbeit relevante Prozess der Datenverarbeitung von Kugelarraydaten erläutert.

3.1 Spatial Audio Quality Inventory

Die Einteilung der Qualität ist, wie bereits in der Einleitung ([1.1]) beschrieben, abhängig von unterschiedlichen Faktoren, welche nicht immer eindeutig definiert werden können. Das *Spatial Audio Quality Inventory (SAQI)* stellt eine Sammlung unterschiedlichster Qualitätsmerkmale dar, welche erst kürzlich von [Lindau et al., 2014] zusammengestellt wurden. Ziel war es, ein beschreibendes Vokabular für die detaillierte perzeptuale Bewertung von akustischen Umgebungen zu erstellen. Unter akustischen Umgebungen sind hier verschiedenste Systeme von Wiedergabesystemen bis hin zur Raumakustik zusammengefasst. Dabei wird nicht auf einzelne übergeordnete Qualitätskriterien, wie zum Beispiel *Authentizität* ([Blauert, 1997]), Wert gelegt, sondern auf die Möglichkeit, einen komplexen Gesamteindruck zu gewährleisten. Die wahrnehmbare Geometrie wird ebenso thematisiert wie die Betrachtung gängiger Artefakte oder Dynamikeigenschaften [Lindau et al., 2014]. Im *SAQI* sind derzeit 48 bestehende Merkmale aufgelistet, welche in 8 verschiedene Kategorien unterteilt werden. Sämtliche Qualitätsmerkmale, welche im darin Verwendung finden, sind in Tabelle [3.1] zu finden.

Tabelle 3.1: Tabelle aller Qualitätsmerkmale des Spatial Audio Quality Inventory.

Kategorie	Qualitätsparameter
Klangfarbe	Klangfarbe (hell-dunkel) Klangfarbliche Ausprägung im Höhenbereich Klangfarbliche Ausprägung im Mittenbereich Klangfarbliche Ausprägung im Tiefenbereich Schärfe Rauigkeit Kammfilterartigkeit Metallische Klangfarbe
Tonalität	Tonhaltigkeit Tonhöhe Dopplereffekt
Geometrie	Richtung Azimut Richtung Elevation Vorn-Hinten-Lage Entfernung Tiefenausdehnung Breitenausdehnung Höhenausdehnung Externalisierungsgrad Lokalisierbarkeit Räumliches Zerfallen
Raum	Nachhallstärke Nachhalldauer Nachhallumhüllung
Zeitverhalten	Vorechos Nachechos Zeitliches Zerfallen Knackigkeit Wiedergabegeschwindigkeit Szenenablauf Reaktionsschnelligkeit
Dynamik	Lautheit Dynamik Kompressoreffekte
Artefakte	Tonhaltiges Fremdgeräusch Impulshaftes Fremdgeräusch Rauschhaftes Fremdgeräusch Fremdquelle Geisterquelle Verzerrungen Vibration
Allgemeines	Klarheit Sprachverständlichkeit Natürlichkeit Präsenz Gefallen
	Unterschied allgemein, nicht erfasster Unterschied

Eine Besonderheit bei diesen Merkmalen ist, dass sie ausschließlich auf perzeptiver Vergleichsbasis beruhen und keine abstrakten Werte zur Schätzung von akustischen Eigenschaften aufweisen, wie das bisher bei vielen Qualitätsparametern der Fall war. Damit kann der Vergleich zwischen vergleichbaren Systemen auf perzeptueller Basis einfacher durchgeführt werden. Eine präzisierende Ergänzung ebenso wie eine gebräuchliche Skaleneinteilung findet sich in [Lindau et al., 2014]. Viele der auftretenden Qualitätsmerkmale wurden in der Literatur bereits ausführlich diskutiert und beschrieben, jedoch kommen zur Qualitätsbeurteilung meist nur einige wenige dieser Attribute zum Einsatz, bezogen auf besonders wichtige akustische Eigenschaften der zu beschreibenden Umgebung.

3.2 Hörtests

3.2.1 Adaptive Prozeduren

Adaptive psychophysikalische Prozeduren schätzen Schwellwerte als einzelne Punkte auf der psychometrischen Funktion. Schwellwerte sind keine festgesetzten Grenzwerte, sondern weisen einen fließenden Übergang zwischen wahrnehmbaren und nicht wahrnehmbaren Effekt auf. Dieser fließende Übergang ist in Abbildung [3.1] als psychometrische Funktion beispielhaft abgebildet. Dauf der x-Achse ist die Stimulusstärke abgetragen, die y-Achse zeigt die Häufigkeit der positiven Antworten. $F(x)$ zeigt die Abhängigkeit der positiven Antworten von der Stimulusstärke. Der Schwellwertpunkt wird meist mit X_{50} angenommen, was bedeutet, dass der Unterschied zwischen zu vergleichenden Stimuli nur noch zu 50 % detektiert werden kann [Ciba et al., 2009]. Bei adaptiven Prozeduren werden Stimuli von den vorangegangenen Werten und Antworten des Testsubjektes bestimmt. Da viele psychophysikalische Untersuchungen, insbesondere die Schwellwertbestimmung, nach diesem Muster erstellt werden, liegt die adaptive Prozedur ebenso dieser Arbeit zu Grunde. In diesem Zusammenhang steht auch die *Alternative Forced Choice (AFC)* Methode, welche die adaptive Prozedur für Hörtests bildet. Dabei wird der Hörer gezwungen, sich zwischen zwei (*2AFC*) oder drei (*3AFC*) Stimuli für das qualitativ bessere oder schlechtere zu entscheiden. Eine positive oder negative Antwort des Hörers bestimmt dabei, ob der nächstfolgende Vergleichsstimulus erhöht oder verringert wird. Ein Schwellwert kann anhand der Häufigkeit der jeweiligen Stimulusstärke bestimmt werden.

Es werden dabei drei Paradigmen als Vorgehensweisen für adaptive Prozeduren unterschieden, welche die grundlegende Herangehensweise und die Komplexität bestimmen.

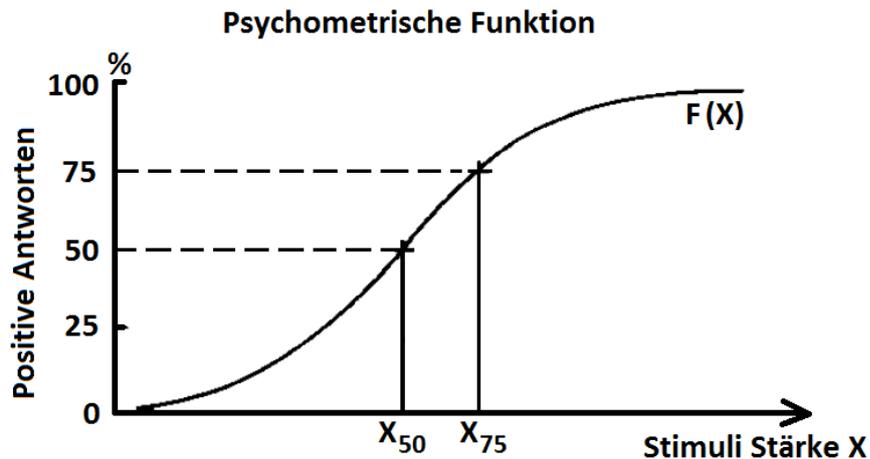


Abbildung 3.1: Psychometrische Funktion: $F(X)$ stellt die erwartete Häufigkeit von positiven Antworten dar. Es ist ein fließender Übergang der Wahrscheinlichkeit von negativen nach positiven Antworten erkennbar [Levitt, 1971]

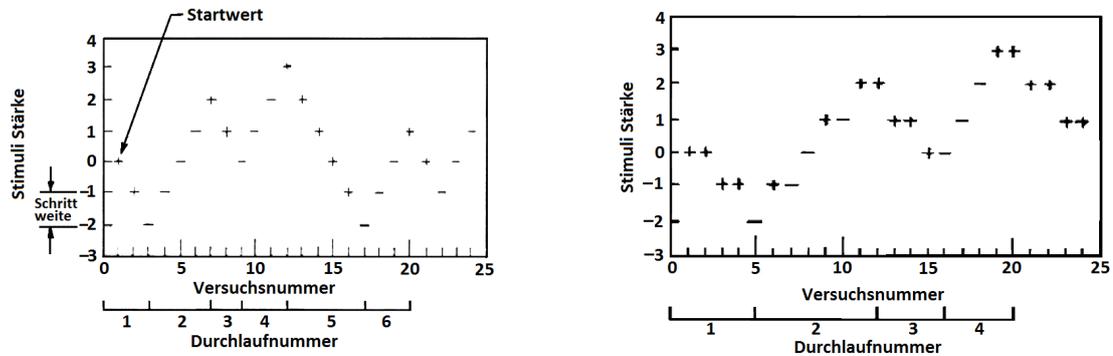
Diese werden in [Leek, 2001] ausführlich diskutiert und sollen hier noch einmal vorgestellt werden.

Staircase: Die einfachste Methode ist die *Simple Up-Down Technique*, auch *Staircase* genannt, bei der ein Durchgang ausschließlich aus positiven oder negativen Antworten des Testsubjektes besteht. Die *Mode* (am häufigsten auftretender Wert) der auf- oder absteigenden Sequenzen deutet darauf hin, an welcher Stelle 50 % positive Antworten zu erwarten sind (X_{50}).

Allerdings können mit dieser einfachen Methode keine anderen Punkte als X_{50} ermittelt werden. Außerdem können bei zu großen Schrittweiten zusätzliche Probleme auftreten, da die Daten schlecht in Bezug zu X_{50} positioniert sein können. Ein weiteres Problem ist die psychophysikalische Auswirkung auf das Testsubjekt, da es möglich ist, den Algorithmus zu durchschauen und Antworten darauf anzupassen, was eine effektive Auswertung der Daten unmöglich macht [Levitt, 1971].

Aus diesem Grund werden *Transformed Staircase* Methoden eingesetzt. Dabei wird ein Set von Antworten benötigt, damit eine Änderung im Stimulus vorgenommen wird. In Abbildung [3.2] sind typische Daten für beide Fälle dargestellt, links die *Simple Up-Down Technique* und auf der rechten Seite die *Transformed Staircase* Methode. Die Stimulusstärke wird als Funktion der Versuchsnummer dargestellt. Ein Pluszeichen deutet auf eine positive Antwort des Probanden hin, ein Minus auf eine negative Antwort. Bei der einfachen Methode wird die Stimuli

Stärke nach jeder Versuchsnummer geändert, bei der *Transformed Staircase* Methode wird bei negativen Werten der Stimulus sofort erhöht, jedoch bei positiven erst der folgende Wert abgewartet, bevor Änderungen vorgenommen werden.



(a) Simple up - down Staircase mit festen Schrittweiten und dem Startwert 0.

(b) Transformed 2up - 1down Staircase für die Wahrscheinlichkeit positiver Antworten von 70,7 %. Ein Set von zwei Stimuli wird benötigt, bevor Änderungen an der Stimulusstärke vorgenommen werden.

Abbildung 3.2: Staircase Methoden (Bilder rekonstruiert nach [Levitt, 1971]).

PEST: Ein weiteres Paradigma, wird *PEST (Parameter Estimation by Sequential Testing)* genannt. Hierbei wird eine spezifische Fehlerwahrscheinlichkeit in Betracht gezogen, um festzustellen, ob erhaltene Antworten in vorgegebenen Grenzen liegen. Wenn dies nicht der Fall ist, kann die Schritttrichtung durch den Algorithmus angepasst werden. Hinzu kommen spezielle Regeln zur Schrittweitenänderung. Ein beispielhafter Durchlauf einer Testserie ist in Abbildung [3.3] dargestellt. Die Grafik ist ähnlich aufgebaut wie die der *Staircase* Methode: Stimulusstärke auf der y-Achse sowie die Versuchsnummer auf der x-Achse. Positive Antworten sind in diesem Fall gefüllte Quadrate und negative Antworten Kreise. Der relevante Unterschied zur *Staircase* Methode ist die Schrittweitenänderung. Dieses Paradigma ist besonders gut geeignet, um Hypothesen zu überprüfen und die Versuchsanzahl so gering wie Möglich zu halten [Levitt, 1971].

Best PEST: Die letzte Methode ist das *Best PEST* Paradigma, welche ebenso in dieser Arbeit seine Anwendung findet. *Best PEST* basiert auf der *Maximum Likelihood* Methode, welche den Startwert von Untersuchungen anhand des Erwartungswertes festsetzt. Dieser kann auch während der Messungen erstellt und beeinflusst werden, indem Informationen aus vorangegangenen Versuchsdaten analysiert werden. Dies macht nachfolgende Messungen effizienter, da Messungen

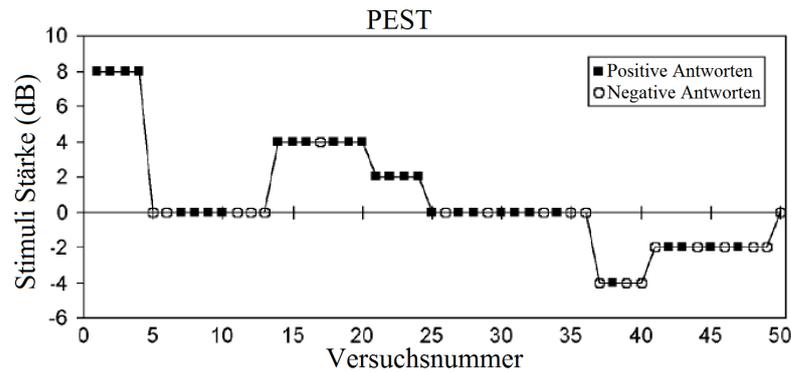


Abbildung 3.3: Beispielhafter Durchlauf der *PEST* Methode [Leek, 2001].

immer in der Nähe des Erwartungswertes durchgeführt werden. Ebenso wird die Schrittweite anhand vorheriger Messwerte festgelegt. Ein beispielhafter Durchlauf einer Testserie ist in Abbildung [3.4] dargestellt, welche ebenso aufgebaut ist, wie der *PEST*-Test in Abbildung [3.3]. Es ist zu erkennen, dass sich die Funktion immer weiter an den durch eine horizontale Linie markierte Schwellwert annähert. Weiterführende Lektüre findet sich unter [Lieberman and Pentland, 1982].

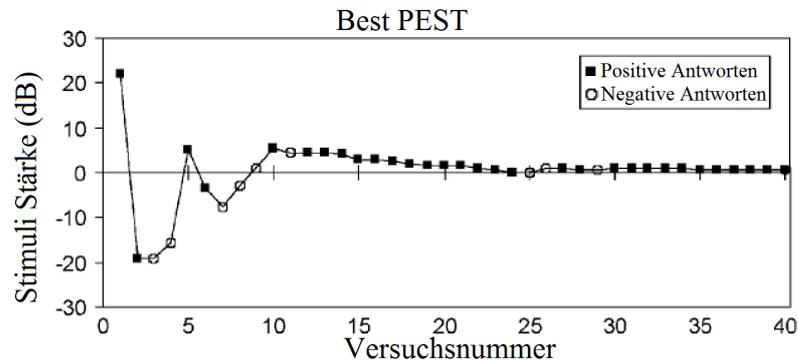


Abbildung 3.4: Beispielhafter Durchlauf der *Best-PEST* Methode [Leek, 2001].

3.2.2 Repertory Grid Technique

Die *Repertory Grid Technik (RGT)* dient der Erfassung von subjektiven Sichtweisen einzelner Personen, welche in diesem Zusammenhang als *persönliche Konstrukte* bezeichnet werden. Sie ist daher kein Test im eigentlichen Sinne, sondern eher eine Methode, um von einem Probanden Informationen zu erhalten und zu strukturieren [Fromm, 1995]. Entwickelt in den 1950er Jahren für psychologische Untersuchungen findet die Technik auch heute noch Anwendung.

Der Test besteht aus drei Testsignalen, aus denen der Proband auswählt, welche beiden

sich ähneln und inwiefern sie sich vom dritten Signal unterscheiden. Somit wird ein individuelles Vokabular an Merkmalen erstellt, mit denen die Qualität beschrieben werden kann. Anschließend werden weitere Testsignale anhand der zuvor erfolgten Beschreibung bewertet. Dazu werden herkömmliche quantitative Hörtests verwendet, wie sie in Kapitel [2.2.1] vorgestellt wurden.

Diese Art von Test wird nicht nur für akustische Untersuchungen verwendet und variiert je nach Einsatzbereich. Da dieser Test sehr flexibel ist, lässt er sich auf viele Erfahrungsbereiche anwenden. Ursprünglich wurde er für wissenschaftliche Untersuchungen im Bereich Psychologie entwickelt.

Es können erhebliche Schwierigkeiten bei diesem Test auftreten, wenn Probanden stark differenzierte Beschreibungen verwenden oder einzelne isolierte Begriffe entstehen, welche ohne Zusammenhang nicht auswertbar sind. Allerdings kann ebenso der Grad der Komplexität durch die Probanden untersucht werden, was neue Möglichkeiten zur Beschreibung der Stimuli ermöglicht.

Die *RGT* für den Audibereich wurde erstmals von [Berg and Rumsey, 1999] angewendet, welche diese Variante detailliert beschreiben. Grundlagen und erste Anwendungen der *RGT* wurden von [Kelly, 1955] vorgeschlagen.

3.3 Parameter zur räumlichen Wahrnehmung

Unser Wahrnehmungsapparat ist mithilfe der Ohren dazu in der Lage, eine Schallquelle präzise im Raum zu lokalisieren. Um die Lokalisation im drei-dimensionalen Raum beschreiben zu können, wird ein Kugelkoordinatensystem genutzt, welches im Kopfbereich des Hörers liegt. Der Azimuth beschreibt den horizontalen Winkel der eintreffenden Reize, der Elevationswinkel dient der vertikalen Beschreibung des Schalls in der Medianebene. Die Distanz ist der dritte Parameter der Richtungsbeschreibung und wird mit r angegeben [Blauert, 1997]. Abbildung [3.5] veranschaulicht das dazugehörige Koordinatensystem.

3.3.1 Azimuth

Die Bestimmung des Azimuthwinkels erfolgt durch zwei Parameter: die interaurale Laufzeitdifferenz (*interaural time difference, ITD*) und die interaurale Pegeldifferenz (*interaural level difference, ILD*) oder auch Intensitätsdifferenz (*interaural intensity difference, IID*). Diese beiden messbaren Differenzen des Schalls werden durch den Kopf als Schallhindernis erzeugt und von den Ohren als Richtungsinformationen aufgenommen.

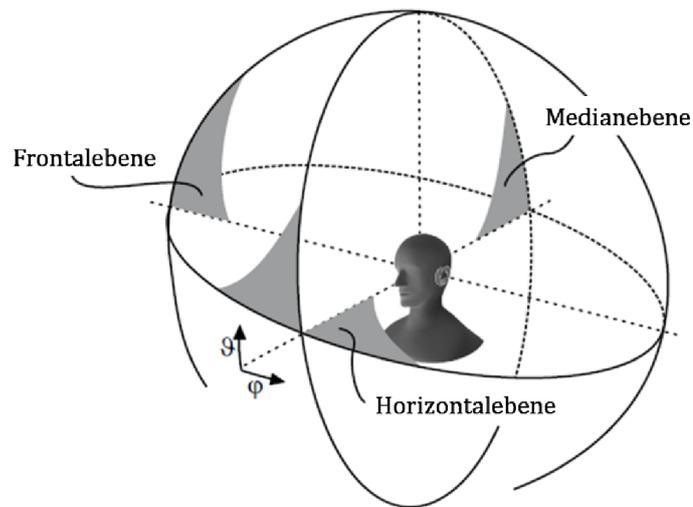


Abbildung 3.5: Kopfkoordinatensystem mit dem Elevationswinkel ϑ und dem Azimutwinkel φ [Fels, 2008].

Die *ITD* entsteht durch das verzögerte Eintreffen des Schalls bei dem schallabgewandten Ohr, und kann zwischen 0 und 800 μs betragen. Mit der Formel

$$ITD = \frac{r(\alpha + \sin \alpha)}{c} \quad (3.1)$$

kann die Laufzeitdifferenz berechnet werden, wobei c die Schallgeschwindigkeit, r den Radius des Kopfes in Meter und α den Schalleinfallswinkel als Radiant darstellt [Rumsey, 2001].

Die *ILD* wird durch die Abschattung des Schalls durch den Kopf verursacht und äußert sich darin, dass ein Schallereignis bis zu 35 dB Schallpegelunterschied bei Aufnahme durch beide Ohren haben kann. Sie wird berechnet durch

$$ILD = 10 \log_{10} \frac{p_L^2}{p_R^2} \quad (3.2)$$

Hierbei ist p_L und p_R der Schalldruckpegel des linken beziehungsweise rechten Ohres [Breebaart and Faller, 2008]. In Abbildung [3.6] ist der Laufzeitunterschied ebenso wie die Pegeldifferenz dargestellt.

Von Rayleigh wurde bereits Anfang des zwanzigsten Jahrhunderts die Duplextheorie entwickelt, nach der beide interaurale Differenzen für die Lokalisation von Schall wichtig sind, jedoch die Laufzeitdifferenz eher bei niedrigen Frequenzen und die Schallpegeldifferenz bei höheren Frequenzen zum Tragen kommt [Macpherson and Middlebrooks, 2002]. Die Relevanz von *ILD* und *ITD* in Abhängigkeit von der Frequenz sind in Abbildung [3.7] verdeutlicht.

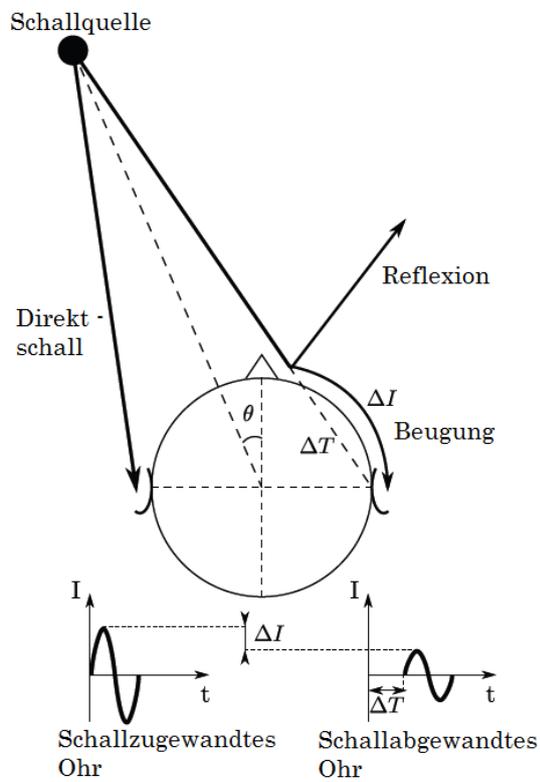


Abbildung 3.6: Signaleinflüsse durch Laufzeit- und Intensitätsdifferenzen [Daniel, 2011].

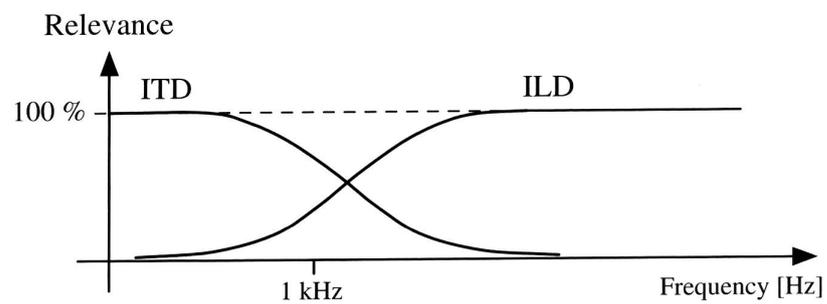


Abbildung 3.7: Frequenzabhängigkeit von ILD und ITD anhand der Duplextheorie [Breebaart and Faller, 2008].

Mit diesem Hintergrundwissen ist es möglich, die wahrgenommene Schallrichtung für den Hörer zu simulieren, indem die genannten Differenzen auf die verwendeten Kanäle (mindestens zwei) angewendet werden. In dem Fall wird von *interaural channel level differences (ICLD)* und *interaural channel time differences (ICTD)* gesprochen [Breebaart and Faller, 2008].

3.3.2 Elevation

Die Möglichkeiten zur räumlichen Beschreibung sind bei den interauralen Differenzen schnell ausgeschöpft, wenn eine Elevationslokalisation durchgeführt werden soll. Das Gehör kann dennoch diese Art von Informationen gewinnen, indem es die spektrale Zusammensetzung der erhaltenen Signale auswertet. Durch Reflexionen und Beugungen des Schalls an Schultern, Kopf und Außenohr wird ein natürlicher Filter erzeugt, welcher das Signal bearbeitet, bevor es vom Trommelfell aufgenommen wird. Diese Filter sind für alle Einfallswinkel verschieden und können auch personenbezogen sehr individuell ausfallen. Die fachspezifische Bezeichnung dieser Filterfunktion lautet *kopfbezogene Übertragungsfunktion*, wobei der englische Begriff *head related transfer function (HRTF)* geläufiger ist. Im Zeitbereich wird von *binaural room impulse responses (BRIR)* gesprochen. Mit binauralen HRTF's (HRTF Paare für beide Ohren) lassen sich über Kopfhörer äußerst realistische akustische Szenen nachbilden [Breebaart and Faller, 2008]. Ein Beispielpaar zweier HRTF findet sich unter Abbildung [3.8].

Blauert erstellte in diesem Zusammenhang richtungsbestimmende Bänder, welche die Wahrscheinlichkeit der Lokalisationsrichtung in der Medianebene (vergleiche Abbildung [3.5]) beschreiben [Weinzierl, 2008]. Grafisch dargestellt sind diese in Abbildung [3.9].

3.3.3 Distanz

Die Distanz ist die letzte Koordinate in der Lokalisation von Schallquellen. Es kann von vier Parametern ausgegangen werden, welche die Distanzwahrnehmung beeinflussen, jedoch werden diese als nicht sehr zuverlässig eingestuft [Loomis et al., 1998] [Rumsey, 2001]:

Schallpegel der Schallquelle: Die wahrgenommene Nähe hängt stark von dem Schalldruckpegel ab, was für bekannte Geräusche ein wichtiger Anhaltspunkt für die Entfernung ist. Bei unbekanntem Geräuschen ist dieser Parameter eher ungenau.

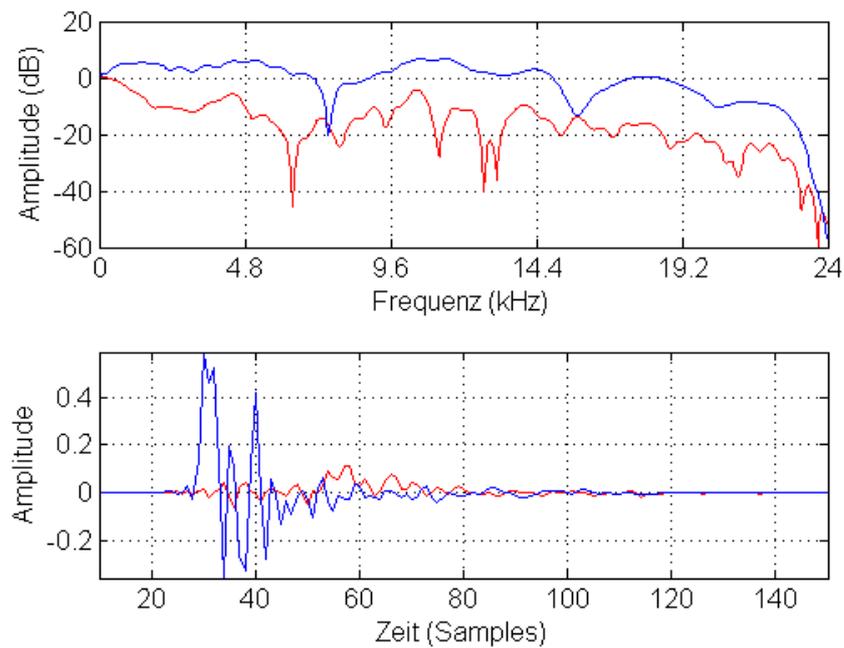


Abbildung 3.8: Beispiel eines HRTF Paares (oben) und das dazugehörige BRIR Paar (unten) bei einer Abtastfrequenz von 48 kHz.

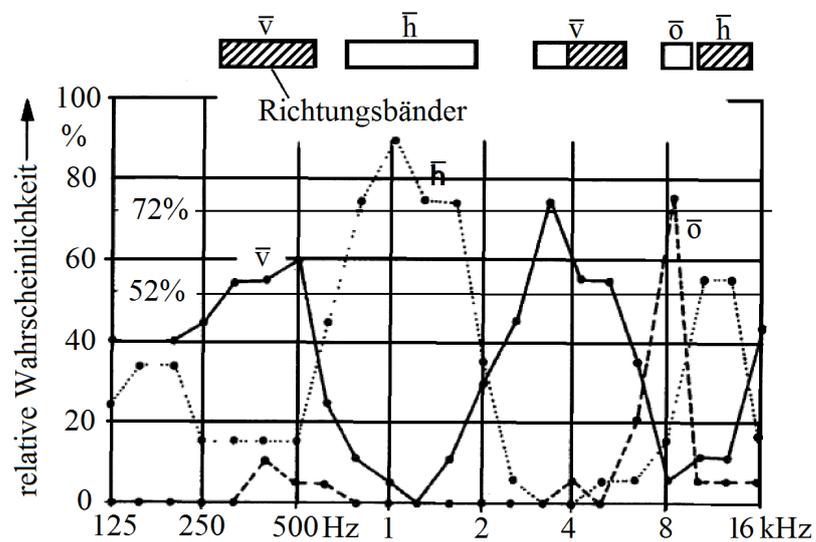


Abbildung 3.9: Richtungsbestimmende Bänder, die Wahrscheinlichkeit der Richtungszuordnung für von vorn (\bar{v}), von oben (\bar{o}) und von hinten (\bar{h}) wird in Abhängigkeit von der Frequenz dargestellt nach [Blauert, 1997].

Hall: Der Direktschall ist bei nahen Quellen sehr viel stärker als das diffuse Schallfeld. Dieser Aspekt spielt gerade bei geschlossenen Räumen eine Rolle, da sie ein stark reflektierendes Umfeld darstellen.

Spektralverteilung: Hohe Frequenzen werden stärker durch den Luftwiderstand abgeschwächt, wodurch weiter entfernte Quellen deutlich weniger hohe Frequenzanteile besitzen.

Reflexionswinkel: Weiter entfernte Quellen besitzen einen sehr geringen Schallwinkel zwischen Direktschall und Reflexionen am Boden. Die ersten Reflexionen kommen daher zeitlich weniger verzögert beim Hörer an, was als Indiz für die Entfernung gilt.

Freifeldversuche haben gezeigt, dass die Distanz ohne reflektierende Umgebung schwer einzuschätzen ist, jedoch eine relative Beurteilung in Bezug zu Vergleichsquellen mit bekanntem Abstand von vielen Hörern sehr gut vollzogen werden kann [Rumsey, 2001].

3.3.4 Schallquellenbreite

Ein weiterer Aspekt in der räumlichen Schallwiedergabe ist die Breite eines Schallerignisses, welche zwischen Punktquelle und diffuser Quelle eingestuft werden kann, bekannt als *apparent source width (ASW)*. Die mathematische Beschreibung davon wird durch die *interaurale Korrelation (interaural correlation, IC)* vorgenommen [Morimoto and Maekawa, 1988] [Hidaka, 1995]. Dieser Wert wird zwischen den Signalen der einzelnen Kanäle durch die normalisierte Kreuzkorrelation berechnet, im binauralen Fall die Signale des rechten (p_R) und linken (p_L) Lautsprechers:

$$IC = \max_{\tau} \frac{\int_{t_1}^{t_2} p_L(t) \cdot p_R(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} p_L^2(t) \cdot \int_{t_1}^{t_2} p_R^2(t) dt}} \quad (3.3)$$

Die IC kann Werte zwischen 0 für diffuse und unabhängige Signale und 1 im Falle einer Punktquelle sowie kohärenten Signalen annehmen [Möser, 2009]. Für die Integrationszeiten t_1 und t_2 wird zwischen sehr zeitigen und späten Signalteilen unterschieden:

$$\begin{aligned} IC_{Early} &: t_1 = 0 \text{ ms} ; t_2 = 80 \text{ ms} \\ IC_{Late} &: t_1 = 80 \text{ ms} ; t_2 = 2000 \text{ ms} \end{aligned} \quad (3.4)$$

In Abbildung [3.10] werden verschiedene Stufen der IC dargestellt, bei denen sich mit steigender Flächennummer die interaurale Korrelation verringert.

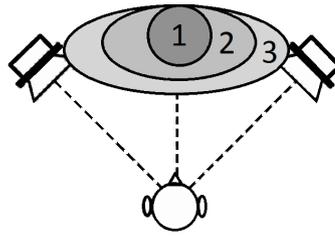


Abbildung 3.10: Die Breite der Schallereignisse steigt (1-3) während die IC zwischen den beiden Ausgangssignalen abnimmt [Faller, 2006].

3.4 Kugelarraydatenverarbeitung

In diesem Kapitel wird eine Übersicht über die Aufnahme und Analyse räumlicher Signale gegeben. Anhand der Wellengleichung wird die Schallausbreitung erklärt, ebenso wie das für solche Aufnahmen geeignete Kugelmikrofonarray. Die mit der dreidimensionalen Darstellung akustischer Daten verbundenen sphärischen Harmonischen dienen als räumliche Fouriertransformation der Signalanalyse. Zuletzt werden die entstehenden Fehler räumlicher Aufnahmen durch Mikrofonarrays betrachtet.

3.4.1 Wellengleichung

Helmholtzgleichung

Die Physik der Schallausbreitung wird mit der Helmholtzgleichung erklärt, welche die Ausbreitung des Schalls als Welle beschreibt und der Wellenfeldsynthese als Basis dient [Melchior, 2011]. Sie kann abgeleitet werden aus Newton's zweitem Gesetz der Bewegung, welche das Verhältnis der Ortsableitung des Drucks p zur zeitlichen Ableitung der Partikelgeschwindigkeit v zu einem Punkt r und zum Zeitpunkt t beschreibt [Williams, 1999]:

$$-\nabla p(r, t) = \rho \frac{\delta v(r, t)}{\delta t} \quad (3.5)$$

Dabei stellt ρ die Massendichte dar und ∇ verdeutlicht die partiellen Ableitungen. Zur Vereinfachung und Rechenbarkeit wird ein ideales Gas als homogenes Ausbreitungsmedium angenommen. Das Hooksche Gesetz geht den entgegengesetzten Weg und verbindet die Ortsableitung der Partikelgeschwindigkeit mit der Zeitlichen Ableitung des Drucks:

$$-\nabla v(r, t) = \frac{1}{K} \frac{\delta p(r, t)}{\delta t} \quad (3.6)$$

K ist dabei das Kompressionsmodul des Mediums. Bildet man die Divergenz der beiden Formeln [3.5] und [3.6] erhält man die Wellengleichung mit der Ausbreitungsgeschwindigkeit (des Schalls) $c = \sqrt{\frac{K}{\rho}}$:

$$\nabla^2 p(r, t) - \frac{1}{c^2} \frac{\delta^2 p(r, t)}{\delta t^2} = 0 \quad (3.7)$$

Durch Fouriertransformation der Druckkomponente erhält man die Helmholtzgleichung:

$$\nabla^2 P(x, \omega) + \left(\frac{\omega}{c}\right)^2 P(x, \omega) = 0, \quad (3.8)$$

wobei $\omega = 2\pi f$ die Kreisfrequenz bezeichnet und $k = \frac{\omega}{c}$ die Wellenzahl. Eine ebene Welle auf Basis der Helmholtzgleichung wird in Abbildung [3.11] dargestellt. Um eine konkrete Lösung für einen raumakustischen Einzelfall lösen zu können, werden noch Informationen über die *Startkonditionen*, die *Schallquellen* und die *Raumgrenzen* benötigt. Werden die Gleichungen [3.7] oder [3.8] nach $p(r, t)$ bzw. $P(r, \omega)$ aufgelöst ergibt sich der Schalldruckpegel eines Wellenfeldes im Raum.

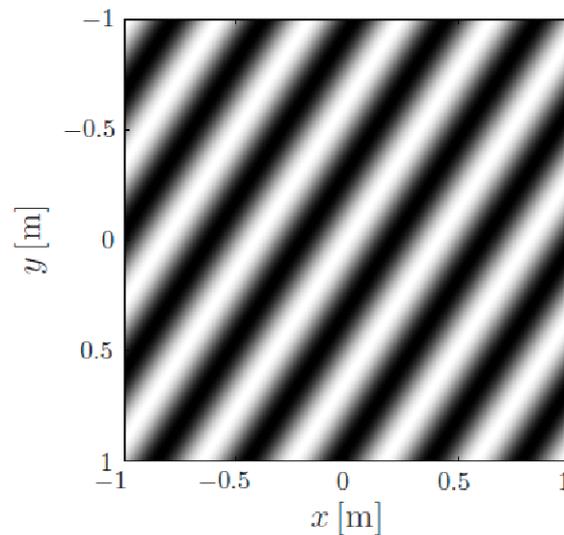


Abbildung 3.11: Schalldruckfeld einer ebenen Welle in der xy-Ebene. Der Grauwertanteil veranschaulicht die Amplitude [Melchior, 2011].

Kugelwellen

Da sich Schall meist in alle Richtungen von einer Punktquelle ausgehend ausbreitet, muss der Spezialfall einer Kugelwelle betrachtet werden. Es wird eine Punktquelle an der Position $r = r_0$ angenommen. Sie wird sphärische Schallwellen (Kugelwellen) wie in Figur [3.12] aussenden.

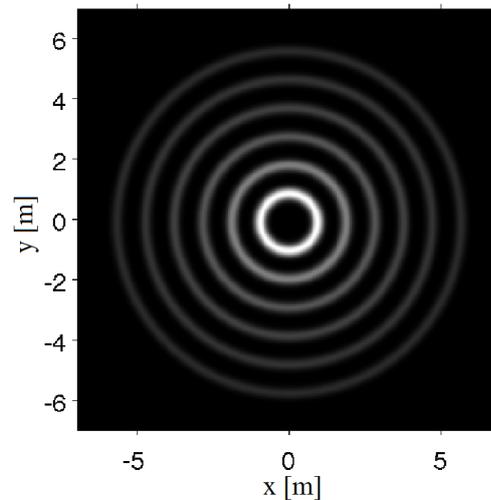


Abbildung 3.12: Sphärische Wellen ausgestrahlt von einer Punktquelle [Schuitman, 2011].

Die Helmholtzgleichung wird sich demnach wie folgt ändern:

$$-\nabla^2 P(r, \omega) + k^2 P(r, \omega) = -4\pi \delta(r - r_0) S(\omega), \quad (3.9)$$

mit dem Spektrum $S(\omega)$ der Quelle. Die Lösung nach $P(r, \omega)$ ergibt:

$$P(r, \omega) = S(\omega) \frac{\exp(-jk |r - r_0|)}{|r - r_0|}, \quad (3.10)$$

Oder für den Zeitbereich:

$$p(r, t) = \frac{s(t - \frac{|r-r_0|}{c})}{|r - r_0|} \quad (3.11)$$

Die Helmholtzgleichung sowie ihre Herleitung wird ebenso in [Williams, 1999] sowie [Melchior, 2011] und [Schuitman, 2011] erläutert.

3.4.2 Schallausbreitung

Bei der Ausbreitung des Schalls in einem geschlossenen Raum wird bei Interaktion des Schalls mit den Wänden ein Teil der eintreffenden Schallenergie reflektiert, ein Teil absorbiert und ein weiterer durch Unebenheiten zerstreut.

Der Schallabsorptionskoeffizient α definiert die Menge an Schallenergie, welche durch die Wand absorbiert wird. Dieser ist abhängig vom Material, von der Frequenz und dem Einfallswinkel. Für die Zerstreung an der Wand wird der *Scattering Coefficient*

s genutzt [Kuttruff, 2009]:

$$s = \frac{E_{streu}}{E_{spiegel} + E_{streu}} \quad (3.12)$$

E_{streu} ist dabei die Gesamtenergie des zerstreuten Schalls, welcher in nichtgespiegelter Richtung reflektiert wird, $E_{spiegel}$ stellt die reflektierte Energie in gespiegelter Richtung dar.

Die verschiedenen Energieanteile werden mit den jeweiligen Koeffizienten in Abbildung [3.13] dargestellt.

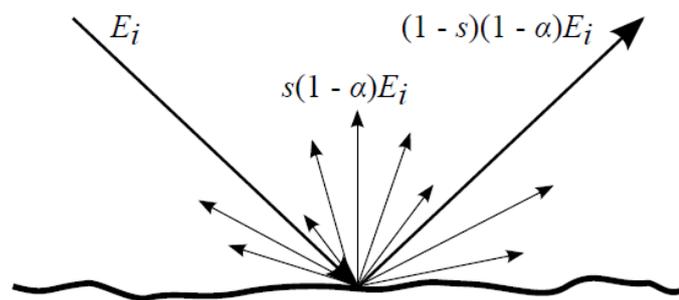


Abbildung 3.13: Auf einer unebenen Oberfläche wird die eintreffende Energie E_i zum Teil reflektiert mit $(1-s)(1-\alpha)E_i$, zerstreut mit $s(1-\alpha)E_i$ und ein Teil αE_i absorbiert [Schuitman, 2011].

Wenn ein Hörereignis vom Hörer oder einem Mikrofon aufgenommen wird, besteht dieses aus dem Direktschall sowie den ersten Reflexionen und dem Nachhall, welche wie zuvor beschrieben durch Reflexion an den Wänden entstehen. Der Direktschall wird direkt von der Quelle zum Hörer ausgesendet, ohne reflektiert zu werden. Erste Reflexionen werden von Wänden reflektiert bevor sie vom Hörer aufgenommen werden und haben eine Verzögerung von etwa 20 bis 100 ms zum Direktschall. Nach 100 ms Verzögerung herrscht in den meisten Räumen nur noch der diffuse Nachhall vor, welcher nicht mehr durch diskrete Reflexionen beschrieben werden kann. Der Schalldruckpegel nimmt dabei mit der Zeit und jeder Reflexion ab, wie in Abbildung [3.14] zu erkennen ist [Schuitman, 2011][Weinzierl, 2008].

3.4.3 Kugelmikrofonarray

Durch ein einzelnes Mikrofon kann keine Information über den Ort einer Schallquelle im Raum entnommen werden (Vergleich Kapitel [3.3]). Mikrofonarrays ermöglichen eine räumliches Abtasten (*Sampling*) eines Schallfeldes und können durch Auswertung der räumlichen Parameter Aufschluss über die Positionierung der Schallquelle

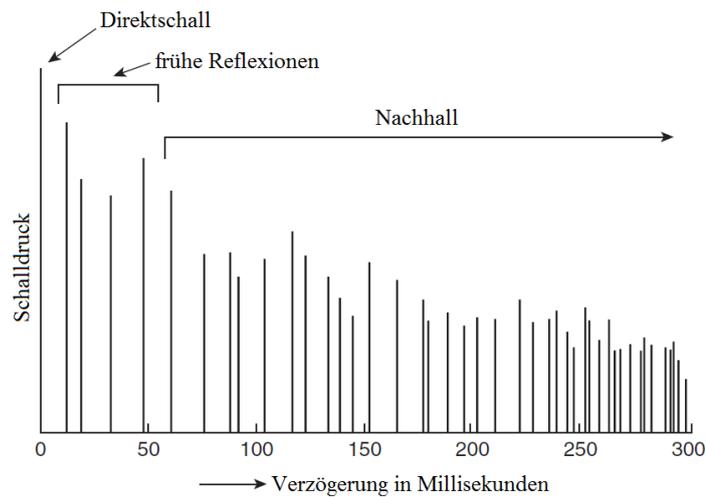


Abbildung 3.14: Die Bestandteile des Schalls: Direktschall, frühe Reflexionen und diffuser Nachhall [Kuttruff, 2009].

im Raum geben. Ein einfaches Mikrofonarray ist das *Linien-Array*, welches in einer Reihe angeordnete Mikrofone verwendet [Möser, 2005]. Allerdings ist damit nur die Quellenlokalisierung auf einer vorgegebenen Linie möglich und es kann nicht unterschieden werden, ob sich das Signal vor oder hinter dem Array befindet. Daher müssen Vorkenntnisse über die ungefähre Position der Schallquelle vorhanden sein, um das *Linien-Array* nutzen zu können. Um dreidimensionale Lokalisation zu gewährleisten, muss auch das Mikrofonarray eine dreidimensionale Anordnung besitzen [Möser, 2009]. Daher wird häufig auf ein *Kugel-Array* zurückgegriffen, wie es beispielhaft in Abbildung [3.15] gezeigt wird. Dabei werden vier grundlegende Möglichkeiten zur Erstellung genutzt [Teutsch, 2007]:

Rigid sphere configuration: Eine geschlossene Kugel wird mit omnidirektionalen Mikrofonen bestückt. Eine weit verbreitete Konfiguration für Mikrofonarrays aufgrund ihrer Einfachheit.

Rigid sphere cardioid configuration: Eine geschlossene Kugel wird mit Mikrofonen mit Nierencharakteristik bestückt. Damit wird die Performance weiter verbessert indem auftretende Fehler minimiert werden.

Open sphere cardioid configuration: Eine offene Kugel wird mit Nierenmikrofonen bestückt. Die offene Kugel besitzt weniger Reflexionen als eine geschlossene Kugel, lässt aber gleichzeitig den Schall durch, der wiederum Mikrofone von der Innenseite der Kugel her beeinflusst. Diesem Problem wird versucht mit Nierenmikrofonen mit Richtcharakteristik entgegenzuwirken.

Open sphere configuration: Eine offene Kugel wird mit Omnidirektionalen Mikrofonen bestückt. Diese Konfiguration ist nicht zu empfehlen, da massive Fehler durch Aliasing auftreten.



Abbildung 3.15: Beispielhaftes offenes Kugelarray mit 120 Mikrofonen und einem Durchmesser von 60 cm [Gfai Tech GmbH, 2013].

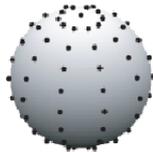
Die Anordnung der Mikrofone (oder rein virtuell der *Sampling Punkte*) bestimmt bei dem Mikrofonarray das *Sampling Schema* (in diesem Zusammenhang als *Quadratur* bezeichnet). Nachfolgend sollen die drei geläufigsten aufgeführt sowie in Abbildung [3.16] grafisch dargestellt werden [Galdo, 2007]:

Chebyshev: Die *Chebyshev Quadratur* besitzt ein einheitliches Sampling für Elevations- und Azimuthwinkel. Durch die sehr einfache Beschreibung aufgrund der gleichbleibenden Winkelabstände wird diese Quadratur häufig eingesetzt, vor allem bei virtuellen Mikrofonarrays. Der Nachteil sind verhältnismäßig viele Mikrofone im Gegensatz zu den anderen beiden Schemata und die dichtere Verteilung an den Polen.

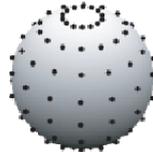
Gauss-Legendre: Mit diesem Schema wird der Azimuthwinkel einheitlich abgetastet. Für den Elevationswinkel werden die Nullstellen der Legendre-Polynome der Ebene $N_{max} + 1$ verwendet. Die Gauss-Legendre Quadratur besitzt die optimale Lösung für die Verteilung in der Elevationsrichtung.

Lebedev: Bei dieser Quadratur wird eine gleichmäßige Verteilung über die Oberfläche der Kugel angestrebt. Dies bedeutet, dass alle Aufnahmepunkte die gleichen

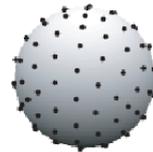
Abstände zu ihren Nachbarpunkten haben, was in der geringsten benötigten Mikrofonanzahl resultiert. Allerdings kann das Gitter somit nicht mehr in Azimuth und Elevationswinkel getrennt werden. Für die Bestimmung der einzelnen Punkte ist keine allgemeine Formel vorhanden und demzufolge wesentlich komplizierter als für die *Chebyshev Quadrature*.



a) Chebyshev Gitter



b) Gauss-Legendre Gitter



c) Lebedev Gitter

Abbildung 3.16: Verschiedene Sampling Schemata auf einer Kugeloberfläche [Melchior, 2011].

3.4.4 Sphärische Harmonische

Für die Aufnahme und Verarbeitung eines Schallfeldes durch Mikrofonarrays ist es notwendig, das gemessene akustische Schallfeld in seine einzelnen Wellenbestandteile zu zerlegen. Dies geschieht mit Hilfe der sphärisch harmonischen Dekomposition, welche sphärische Harmonische verschiedener Ordnungen zerlegt, um die Rekonstruktion eines Feldes durch Überlagerung zu ermöglichen. Es wird sich oft auf einen eigenständigen Berechnungsbereich bezogen, ähnlich dem Zeit- oder Frequenzbereich: Die *sphärisch harmonische Domäne (SHD)*. Abbildung [3.17] veranschaulicht diese Aufspaltung bis zur dritten Ordnung. Im sphärisch harmonischen Bereich können einige Operationen direkt durchgeführt werden, jedoch sind die Ergebnisse nicht intuitiv einsehbar. Für die folgenden Überlegungen gilt das sphärische Koordinatensystem (r, ϑ, φ) , welches in [3.3] eingeführt wurde. Die Funktion $Y_n^m(\vartheta, \varphi)$ der sphärischen Harmonischen mit dem Modus m und der Ordnung n in einem sphärischen Koordinatensystem ist wie folgt definiert:

$$Y_n^m(\vartheta, \varphi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos \vartheta) e^{im\varphi} \quad (3.13)$$

$P_n^m(\cos \vartheta)$ ist dabei die Legendre-Funktion erster Ordnung. Legendre-Funktionen besitzen die Charakteristik, dass sie ein komplettes Set orthogonaler Funktionen für jeden Modus m formen, was eine wichtige Eigenschaft für die sphärischen Harmonischen darstellt. Sie repräsentieren stehende sphärische Wellen in ϑ wohingegen der Term $e^{im\varphi}$

wandernde sphärische Wellen in φ repräsentiert. Eine detaillierte Betrachtung der Legendre Funktionen findet sich in [Williams, 1999].

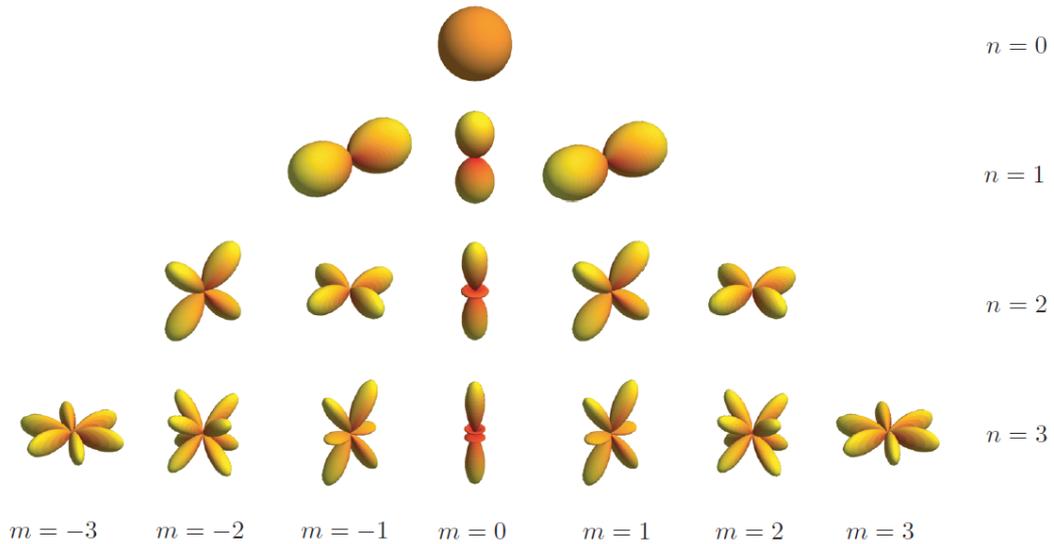


Abbildung 3.17: Komplettes Set für die Sphärischen Harmonischen für $n = 0 \dots 3$ und $m = -3 \dots 3$. Die Plots zeigen die quadrierten Realanteile [Thiergart, 2007].

Die sphärischen Harmonischen sind Lösungen der Helmholtz-Gleichung in sphärischen Koordinaten. Um die gesamte Oberfläche einer Kugel betrachten zu können, wird die sphärische Fouriertransformation und die inverse Fouriertransformation eingesetzt:

$$f_{nm} = \int_0^{2\pi} \int_0^{\pi} f(\vartheta, \varphi) \overline{Y_n^m(\vartheta, \varphi)} \sin\vartheta \, d\vartheta \, d\varphi \quad (3.14)$$

$$f(\vartheta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n f_{nm} \overline{Y_n^m(\vartheta, \varphi)} \sin\vartheta \, d\vartheta \, d\varphi \quad (3.15)$$

Dabei stellt f_{nm} die Koeffizienten der sphärischen Fouriertransformation dar und $f(\vartheta, \varphi)$ die Koeffizienten der inversen Transformation. Für ein Sampling mit endlich vielen Mikrofonen wird eine Gewichtung α_j hinzugefügt, welche vom Sampling-Schema abhängt und das Integral der Fouriertransformation ändert sich zur Summation:

$$p_{nm} = \sum_j \alpha_j f(\Omega_j) \overline{Y_n^m(\Omega_j)} \quad (3.16)$$

Die Variable j reicht über das komplette Set von Sampling Punkten, dabei ist $(\Omega_j =$

ϑ_j, φ_j). Der Schalldruck am Signalausgang kann somit beschrieben werden als:

$$y(kr) = \sum_{n=0}^N \sum_{m=-n}^n p_{nm}(kr) \overline{\omega_{nm}(k)} \quad (3.17)$$

Dabei stellen p_{nm} und ω_{nm} Koeffizienten der sphärischen Fouriertransformation von p und ω dar, k ist die Wellenzahl und r ist der Radius.

Für unterschiedliche Konfigurationen des Kugelarrays, wird ein weiterer Faktor $b_n(kr)$ hinzugefügt, der als *Modenstärke* (*Mode strength*) bezeichnet wird, wobei Θ_0 und Φ_0 auf die Ankunftsrichtung der Welle hinweisen:

$$p_{0nm} = b_n(kr) Y_n^m(\Theta_0, \Phi_0) \quad (3.18)$$

Die Modenstärke beeinflusst die radialen Filterkoeffizienten d_n und damit die Frequenzantwort des Mikrofonarrays. Dabei ändert sich $b_n(kr)$ mit der Konfiguration des Mikrofonarrays wie folgt:

$$b_n(kr) = \begin{cases} 4\pi i^n \left(j_n(kr) - \frac{j'_n(kr)}{h'_n(kr)} h_n(kr) \right) & \text{geschl., Omni} \\ 4\pi i^n j_n(kr) & \text{offen, Omni} \\ 4\pi i^n \frac{1}{2} (j_n(kr) - i j'_n(kr)) & \text{offen, Niere} \\ 4\pi i^n \left(j_n(kr_m) - i j'_n(kr_m) + (i h'_n(kr_m) - h_n(kr_m)) \frac{j'_n(kr_s)}{h'_n(kr_s)} \right) & \text{geschl., Niere} \end{cases} \quad (3.19)$$

wobei j_n und h_n die sphärischen Bessel- und Hankel-Funktionen darstellen, sowie j'_n und h'_n deren Ableitung. Abbildung [3.18] zeigt die Veränderung der Modenstärke bei variabler Ordnung N . Der tieffrequente Bereich fällt stark mit zunehmender Ordnung ab.

Für den allgemeinen Fall der Helmholtzgleichung nach Formel [3.8] wurde von [Melchior, 2011] eine verallgemeinerte Form der Schalldruckberechnung eingeführt durch

$$P(x, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n [C_{nm}(\omega) h_n^{(1)}(kr) + D_{nm}(\omega) h_n^{(2)}(kr)] Y_n^m(\Theta, \Phi). \quad (3.20)$$

C_{nm} und D_{nm} sind definiert durch die Grenzwerte, das bedeutet durch die Positionierung der Quellen. Eine Betrachtung von *Innenraumproblemen* und *Außenraumproblemen*, bei denen die Positionierung der Quellen jeweils innerhalb beziehungsweise außerhalb des Kugelmikrofonarrays stattfindet, findet sich in [Melchior, 2011] und [Williams, 1999]. Die sphärischen Hankelfunktionen erster und zweiter Ordnung

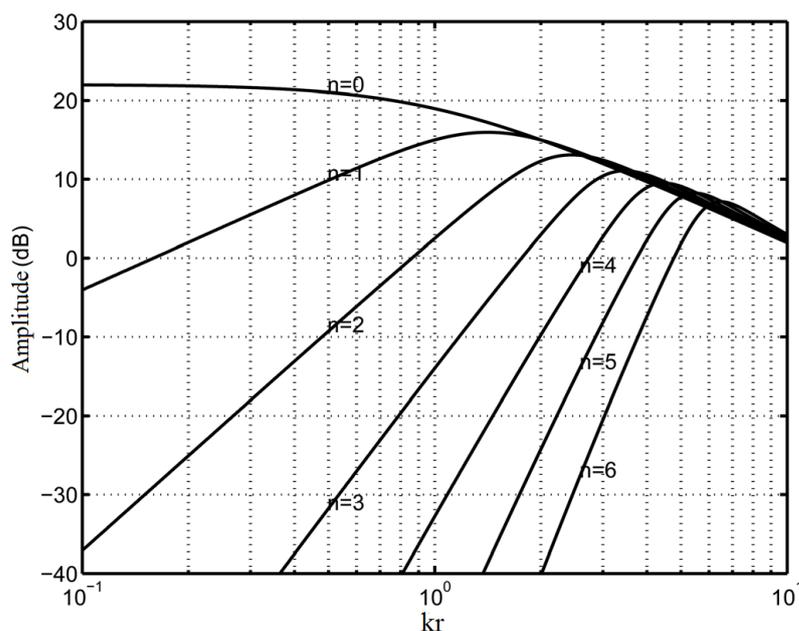


Abbildung 3.18: Modenstärken eines geschlossenen Mikrofonarrays bei unterschiedlicher Ordnung n [Avni et al., 2013].

sind durch die Werte $h_n^{(1)}$ und $h_n^{(2)}$ dargestellt, welche eintreffenden beziehungsweise abgehenden Wellen entsprechen.

Eine ausführlichere Herleitung und weiterführende Betrachtungen der sphärischen Harmonischen geben unter anderem [Melchior, 2011] [Thiergart, 2007] [Rafaely, 2005] [Williams, 1999] [Meyer and Elko, 2002].

3.4.5 Fehleranalyse

Bei der Aufnahme von Schall durch Mikrofone kann es zu Fehlern unterschiedlicher Natur kommen. Drei gravierende Fehlerarten sind bei räumlichen Aufnahmen durch Kugelmikrofonarrays zu betrachten: Spatial Aliasing, Sensorprobleme und Positionierungsfehler.

Spatial Aliasing: Im Anwendungsfall ist die Anzahl der Mikrofone begrenzt, weshalb sowohl im zweidimensionalen als auch im dreidimensionalen Fall Aliasing stattfindet. Aliasing entsteht im dreidimensionalen Fall, wenn die Mikrofone zur Aufnahme zu großen Abstand zueinander besitzen und äußert sich in Gitterkeulen im hochfrequenten Bereich. Diese Gitterkeulen stellen Überhöhungen von Amplitudenwerten in nicht erwünschten Aufnahmerichtungen dar. Um dies zu vermeiden, sollte das *Nyquist Theorem* beachtet werden. Die maximal analysierbare

Frequenz f_{max} bei gegebenem Mikrofonabstand Δx ergibt sich demnach zu:

$$f_{max} = \frac{0.5 c}{\Delta x} \quad (3.21)$$

Die Schallgeschwindigkeit wird mit c ausgedrückt. Da diese Bedingung für den hörbaren Frequenzbereich kaum durchsetzbar ist, muss die Bandbreite nach Formel [3.21] begrenzt oder Gitterkeulen im hochfrequenten Bereich in Kauf genommen werden [Möser, 2009]. Abbildung [3.19] weist derartige Artefakte auf. Entstehende Gitterkeulen werden in Abbildung [3.20] durch zwei verschiedene Ansichten dargestellt. Es kommt neben der Hauptkeule zu weiteren unerwünschten Nebenkeulen, deren Intensität und Anzahl von der Ordnung abhängt.

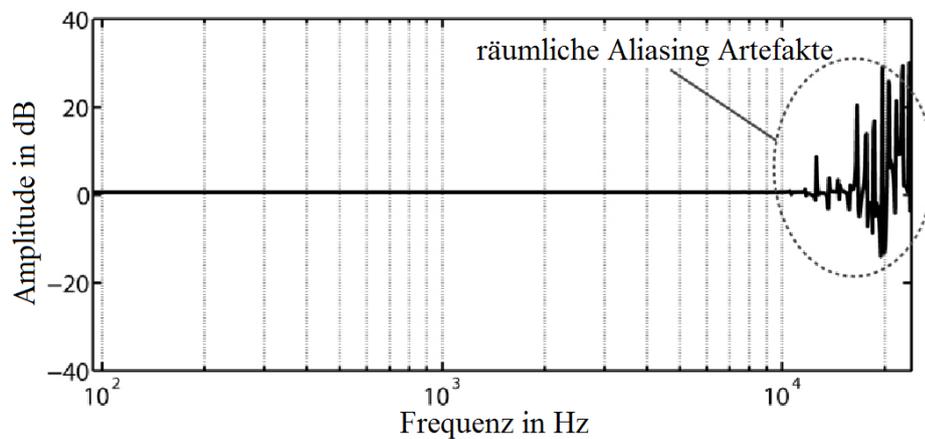


Abbildung 3.19: Artefakte des Spatial Aliasing im hochfrequenten Bereich [Bernschütz et al., 2011].

Rauschen: Bei der Aufnahme durch das Mikrofon treten Fehler auf, die durch gute Sensorqualität entscheidend gemindert werden können. Der Signal-Rausch Abstand (SNR) gibt an, wie groß der Abstand zwischen der Störung durch Rauschen und dem eigentlichen Signal des Mikrofons ist. Dieser Wert fällt bei den heutigen Mikrofonen bereits sehr hoch aus (in der Regel werden ca. 90 dB erreicht). Da ideale Mikrofone ohne Eigenrauschen nicht existieren, muss dennoch mit Störgeräuschen gerechnet werden. Für Richtcharakteristiken, welche zudem Frequenzabhängig sind, ist der ideale Fall ebenso nicht gegeben, was sich negativ auf offene Mikrofonarrays mit Nierenmikrofonen (Vergleich Kapitel [3.4.3]) auswirkt.

Im Falle von räumlichen Aufnahmen, welche durch die sphärisch harmonische Dekomposition mit hoher Auflösung durch großer Ebenenanzahl n zerlegt werden,

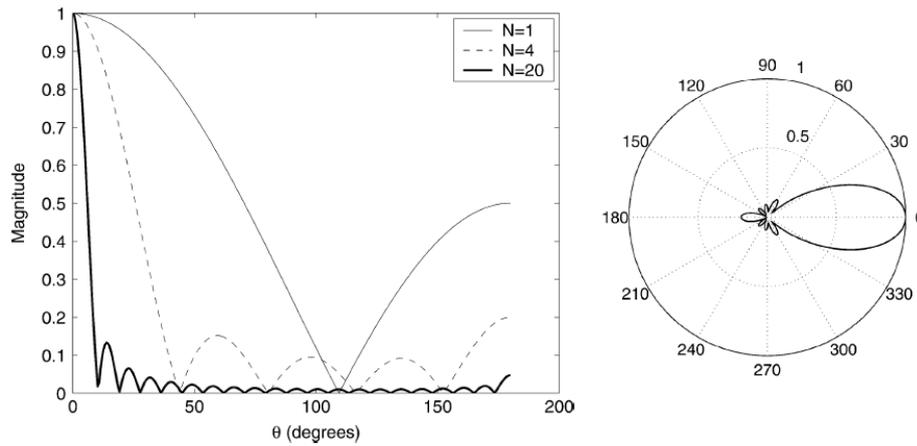


Abbildung 3.20: Gitterkeulen und ihre Richtungsabhängigkeit für drei verschiedene Ordnungen (links) und Richtungsabhängigkeit in der Polardarstellung für Ordnung $N=4$ [Rafaely et al., 2007].

kommt ein weiteres Problem hinzu: Die Robustheit des Arrays lässt für kleine Werte von kr stark nach, weshalb unkorreliertes Rauschen signifikant verstärkt wird. Daher ist ein Minimum des Fehlers festzustellen, der zu den hohen und tiefen Frequenzen hin ansteigt. Dieses Minimum liegt bei

$$kr = n, \quad (3.22)$$

wobei

$$k = \frac{2\pi f}{c} \quad (3.23)$$

entspricht, woraus die optimale Frequenz f berechnet werden kann. Die Schallgeschwindigkeit wird durch c ausgedrückt [Melchior, 2011] [Thiergart, 2007].

Positionierungsfehler: Durch die ungenaue Anordnung der Mikrofone auf dem Kugelarray kann es zu Positionierungsfehlern kommen, welche zu Überhöhungen des tieffrequenten Bereiches führen. Dabei wird unterschieden zwischen zufälligen Positionierungsfehlern und Fehlern aufgrund von konstantem Versatz. Zufällige Fehler sind Normalverteilt mit einer Standardabweichung von etwa $\sigma_\Omega = 1^\circ$ [Albrecht et al., 2007]. Fehler durch konstanten Versatz können entstehen, wenn die Anordnung der Mikrofone nicht korrekt ist, oder durch Überbelastung des Halterungsmaterial nachgibt [Melchior, 2011].

Wenn eine Schallwelle im Raum durch ein Mikrofonarray aufgenommen wird, sollten Mikrofone, welche der Schallwelle zugewandt sind, die größte Amplitude verzeichnen. Beim spatial Aliasing werden nicht nur frequenzbedingte (wie in Abbildung [3.19] verdeutlicht), sondern auch positionsbedingte Fehler wirksam. In Abbildung [3.21(b)] wird ein Kugelarray durch eine Kugel dargestellt. Die aufgenommene Amplitude wird durch eine Farbskala verdeutlicht. Eine Schallwelle wird aus 180° Azimuth und 90° Elevation aufgenommen, wie der Amplitudenanstieg an dieser Stelle zeigt. Bei Fehlern durch spatial Aliasing wird die Hauptamplitude nicht nur in der eigentlichen Detektionsrichtung aufgenommen, sondern auch verstreut von anderen Richtungen kommend auftreten. Diese Fehler sind in den Abbildungen [3.21(a)] für 100 Hz und [3.21(c)] für 5 kHz gezeigt. Aus den Formeln [3.22] und [3.23] lässt sich die Frequenz mit dem geringsten spatial Aliasing für dieses Beispiel errechnen, in welchem die Ordnung der sphärischen Harmonischen $n = 14$ und der Radius $r = 0,25m$ beträgt. Das in Abbildung [3.21(b)] dargestellte Frequenzoptimum liegt bei etwa 3 kHz.

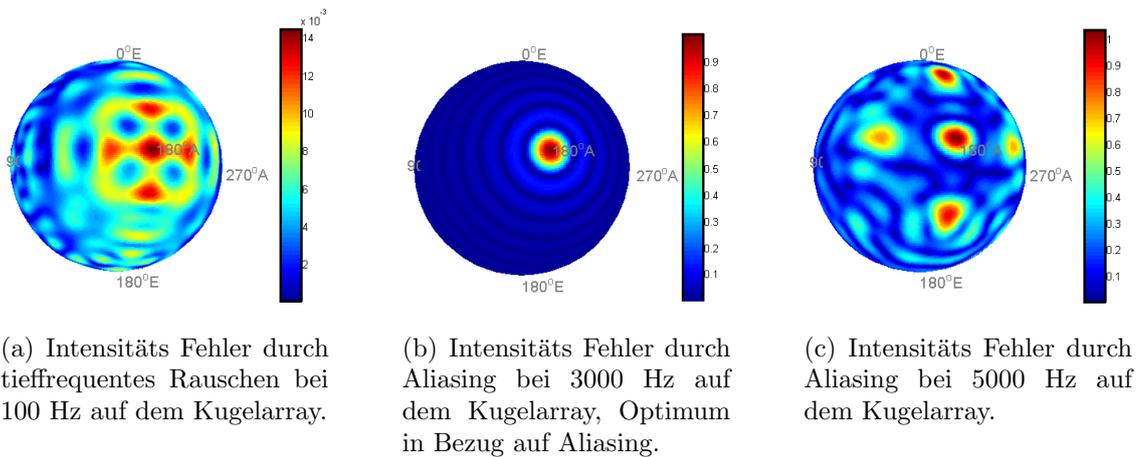


Abbildung 3.21: Rauschen und Aliasing auf einem Kugelarray simuliert durch *SO-FiA*.

Der Schalldruckpegel eines einzelnen Mikrofons bei gegebener Frequenz und unter Einwirkung der vorgestellten Effekte kann wie folgt beschrieben werden:

$$p(kr, \Omega_j) = p(\Omega'_j) + e_j \quad (3.24)$$

wobei

$$p_{nm}(kr) = \alpha_j p(kr, \Omega_j) \quad (3.25)$$

Dabei ist Ω'_j die aktuelle Mikrofonposition und e_j verdeutlicht Rauschen ausgehend vom Mikrofonwandler. Wenn p_{nm} ausgedrückt wird durch [3.16] und in [3.17] eingesetzt wird, erhält man

$$y = \sum_{n=0}^N \sum_{m=-n}^n \sum_{j=1}^M \alpha_j (p(\Omega'_j) + e_j) \overline{Y_n^m(\Omega_j)} \overline{\omega_{nm}} \quad (3.26)$$

Der erste Term wird nun durch die inverse Fouriertransformation ersetzt und umgestellt, um

$$y = \sum_{n=0}^N \sum_{m=-n}^n \overline{\omega_{nm}} \left(\sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} p_{n'm'} \times \left[\sum_{j=1}^M \alpha_j Y_{n'}^{m'}(\Omega'_j) \overline{Y_n^m(\Omega_j)} \right] + \sum_{j=1}^M \alpha_j e_j \overline{Y_n^m(\Omega_j)} \right) \quad (3.27)$$

zu erhalten, worin bereits in eckigen Klammern der Fehlerterm für das Mikrofonrauschen enthalten ist. In den Bereich der eckigen Klammern können weitere Fehlerarten eingesetzt werden, um Aliasing-Fehler

$$\epsilon_{a_{nmn'm'}} = \sum_{j=1}^M \alpha_j Y_{n'}^{m'}(\Omega_j) \overline{Y_n^m(\Omega_j)} \quad n \leq N, n' > N \quad (3.28)$$

und Positionierungsfehler

$$\epsilon_{\Omega_{nmn'm'}} = \sum_{j=1}^M \alpha_j [Y_{n'}^{m'}(\Omega'_j) - Y_{n'}^{m'}(\Omega_j)] \overline{Y_n^m(\Omega_j)} \quad n \geq N, n' > 0 \quad (3.29)$$

zu enthalten. Damit kann der Schalldruckpegel am Signalausgang unterteilt werden in Signal- und Fehlerverteilungen:

$$\begin{aligned} y = & \underbrace{\sum_{n=0}^N \sum_{m=-n}^n \overline{\omega_{nm}} p_{nm}}_{(y_s - \text{Signal})} + \underbrace{\sum_{n=0}^N \sum_{m=-n}^n \overline{\omega_{nm}} \left(\sum_{n'=N+1}^{\infty} \sum_{m'=-n'}^{n'} p_{n'm'} \epsilon_{a_{nmn'm'}} \right)}_{(y_a \text{ Aliasing Fehler})} \\ & + \underbrace{\sum_{n=0}^N \sum_{m=-n}^n \overline{\omega_{nm}} \left(\sum_{n'=N+1}^{\infty} \sum_{m'=-n'}^{n'} p_{n'm'} \epsilon_{\Omega_{nmn'm'}} \right)}_{(y_{\Omega} \text{ Positionierungsfehler})} \\ & + \underbrace{\sum_{n=0}^N \sum_{m=-n}^n \overline{\omega_{nm}} \left(\sum_{j=1}^M \alpha_j e_j \overline{Y_n^m(\Omega_j)} \right)}_{(y_e - \text{Aufnahmerauschen})} \end{aligned} \quad (3.30)$$

Die Formel [3.30] bietet eine Basis zur Fehler- und Leistungsanalyse und kann daher genutzt werden, um die Einflüsse der einzelnen Fehlerarten zu analysieren. Die Abbildung [3.22] stellt eine Analyse der einzelnen Fehler für einen praktischen Anwendungsfall dar. Der gemessene Gesamtfehler wird den einzelnen, durch Berechnung nach Formel [3.30] vorhergesagten Fehlern gegenübergestellt [Rafaely, 2005].

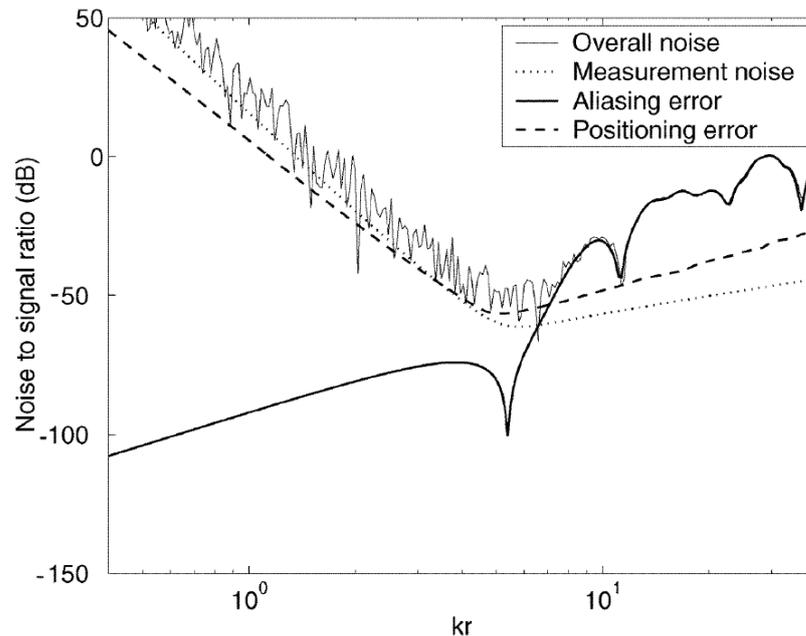


Abbildung 3.22: Gemessener Fehler des Arrayausgangs (Overall Noise) und dazu die theoretisch zu berechnenden Fehlerarten (Measurement noise, Aliasing error, Positioning) [Rafaely, 2005].

3.4.6 Binaurale Auralisation

Nach der Aufnahme eines Schallfeldes durch ein Mikrofonarray soll dieses ebenso wiedergegeben werden. Hierbei sind ebenfalls verschiedene Möglichkeiten gegeben, wie zum Beispiel die Wiedergabe über Lautsprecher oder Kopfhörer. Diese Arbeit konzentriert sich auf die binaurale Auralisation (*Hörbarmachung*) über extraaurale Kopfhörer. Basierend auf der Schallwellendekomposition durch die sphärisch harmonische Zerlegung und einem HRTF-Kugelsatz können die binauralen Impulsantworten für die binaurale Auralisation generiert werden [Melchior et al., 2009].

Der Schalldruck am linken Ohr eines Hörers kann beschrieben werden durch:

$$p^l(k) = \int_0^{2\pi} \int_0^{\pi} a(k, \vartheta, \varphi) H^l(k, \vartheta, \varphi) \sin\vartheta \, d\vartheta \, d\varphi \quad (3.31)$$

Dabei ist $a(k, \vartheta, \varphi)$ die komplexe Amplitude einer ankommenden Schallwelle aus der Richtung (ϑ, φ) und k die Wellenzahl. $H^l(k, \vartheta, \varphi)$ stellt die HRTF des linken Ohres als komplexe Amplitude des Schalldrucks dar, welcher am linken Ohr aufgrund der genannten Schallwelle gemessen wurde. Die Definition ist für das rechte Ohr ebenso gegeben, wenn die entsprechenden HRTF's genutzt werden.

Eine sphärisch harmonische Repräsentation für den Schalldruck des jeweiligen Ohres wird mit

$$p^l(k) = \sum_{n=0}^N \sum_{m=-n}^n \overline{a_{nm}(k)} H_{nm}^l(k) \quad (3.32)$$

sowie

$$p^r(k) = \sum_{n=0}^N \sum_{m=-n}^n \overline{a_{nm}(k)} H_{nm}^r(k) \quad (3.33)$$

ausgedrückt [Avni et al., 2013]. Dabei ist $\overline{a_{nm}(k)}$ die sphärisch harmonische Repräsentation von $a(k, \vartheta, \varphi)$ und $H_{nm}^r(k)$ und $H_{nm}^l(k)$ sind die sphärisch harmonischen Repräsentationen der rechten und linken HRTFs. N bestimmt die Anzahl der Mikrofone auf dem Kugelarray. $H_{nm}^r(k)$ und $H_{nm}^l(k)$ können berechnet werden, indem die sphärisch harmonische Transformation aus Formel [3.14] auf die HRTFs mit Bezug zum jeweiligen Winkel ausgeführt wird. Grafisch lässt sich dieser Sachverhalt grob vereinfacht wie in Abbildung [3.23] darstellen.

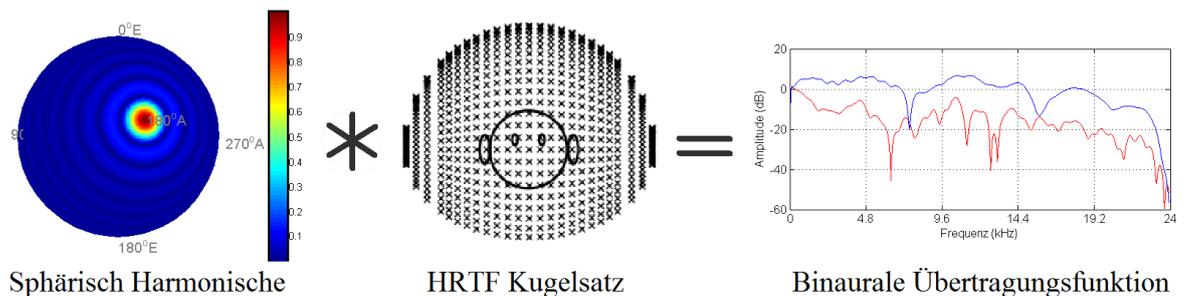


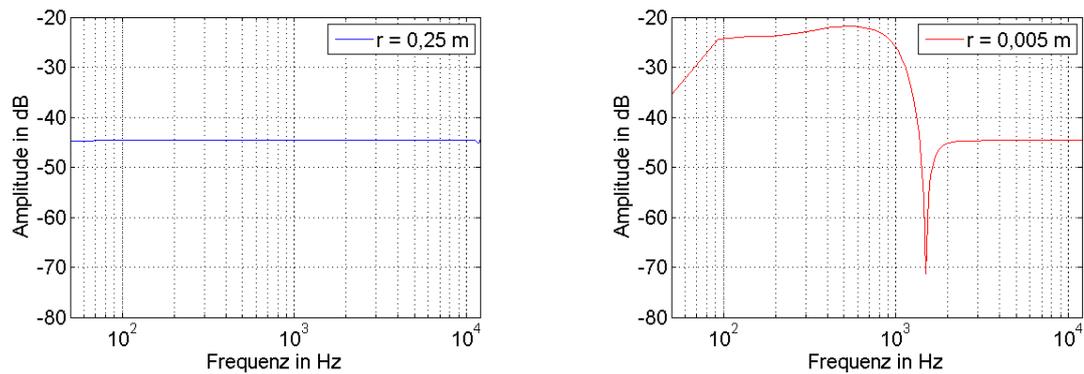
Abbildung 3.23: Visuelle Vereinfachung der binauralen Auralisation von Kugelarraydaten: Faltung und Aufsummierung von Kugelarraydaten mit einem HTRF-Kugelsatz ([Algazi and Duda, 2001]) ergibt eine binaurale Übertragungsfunktion

4 Methodik

In dieser Arbeit werden Schwellwerthörtests für Fehler durchgeführt, welche bei der Simulation oder Aufnahme durch Kugelmikrofonarrays auftreten. Bei der Wiedergabe über Kopfhörer werden diese Fehler hörbar. Die auftretenden Fehler sollen durch Merkmale aus *SAQI* beschrieben werden. In informellen Tests wird das Design und der Aufbau festgelegt und die Durchführbarkeit überprüft. Nachfolgend sind Überlegungen und Feststellungen diesbezüglich aufgeführt.

4.1 Konfiguration

Um die Fehler unter kontrollierten Bedingungen zu untersuchen, werden Simulationen zur Signalgenerierung verwendet. Es wird dabei nach Konfigurationen gesucht, welche externe Fehler minimieren und sich, wenn möglich, nur auf eine Fehlerart auswirken. Dabei muss ein Kompromiss geschlossen werden aus Berechnungszeit und Fehlerminimierung. Es wurde eine Konfiguration für die Wellengeneration gefunden, welche diesen Ansprüchen genügt. Die Ordnung der sphärischen Harmonischen wurde auf $N = 10$ gesetzt, da höhere Ordnungen durch den exponentiellen Einfluss die Berechnungszeit rasch ansteigen lassen. Aufgrund der hohen Ordnung musste der Radius auf $r = 2 \text{ m}$ festgesetzt werden, um tieffrequentes Rauschen zu eliminieren, welches bei höheren Ordnungen und kleinen Radien auftritt, wie Abbildung [4.1] zeigt. Die gewählte *Quadrature* besitzt 4802 virtuelle Mikrofone. Eine höhere Auflösung würde den Aliasingfehler in höhere Frequenzbereiche verschieben, jedoch ebenso die Dauer der Berechnung drastisch erhöhen. Die verwendete Anzahl virtueller Mikrofone bietet bereits ausreichend Audioqualität, da Aliasing erst bei etwa 3 kHz einsetzt und sehr geringe Werte aufweist. Ein Unterschied zu Konfigurationen, welche kein Aliasing aufweisen, wird dabei nicht wahrgenommen. Die Konfiguration und Fehlersimulation wird mithilfe der Schallfeldanalyse Toolbox *SOFiA* durchgeführt. Weitere Informationen zu *SOFiA* finden sich unter [Bernschütz et al., 2011].



(a) Kein Rauschanteil im niederfrequenten Bereich für $r = 0,25\text{m}$.

(b) Rauschanteile im niederfrequenten Bereich für $r = 0,005\text{m}$.

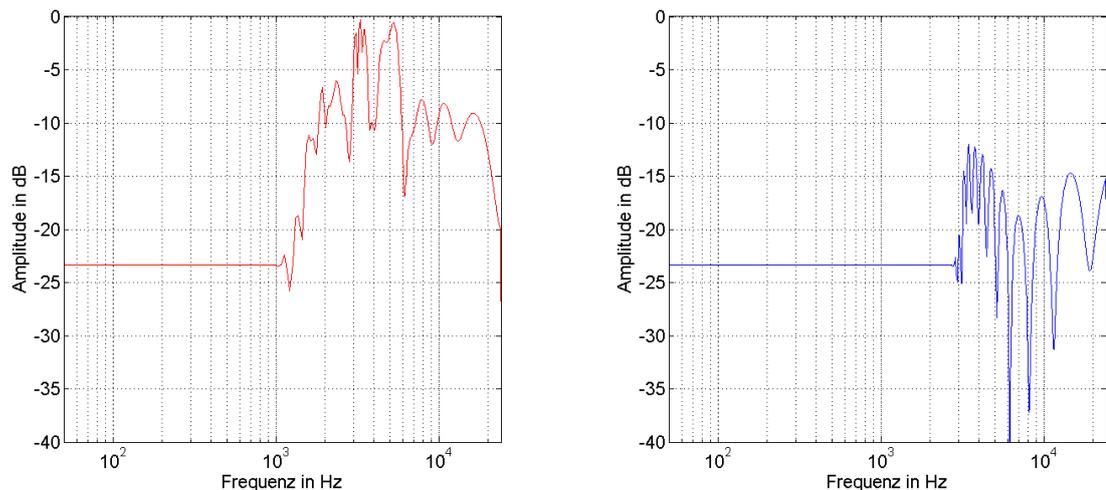
Abbildung 4.1: Abhängigkeit des Rauschanteils im tieffrequenten Bereich vom Radius für $r=0,25\text{m}$ und $r=0,005\text{m}$ bei 1730 Samplingpunkten und der Ordnung $N=5$.

4.2 Informeller Test

Bei der Erstellung eines Hörtests zur Schwellwertermittlung ist es notwendig, einen Überblick über die gesuchten Schwellwerte zu bekommen. Eine Abschätzung der Lage des Schwellwertes sollte bekannt sein, um korrekte Skaleneinteilungen sowie Skalengebiete erstellen zu können. Daher wurde im Vorfeld eine Auswahl von Werten getestet, um einen Überblick über die ungefähren Schwellwerte zu bekommen. Die Ergebnisse dieser Voruntersuchung sind nachfolgend aufgeführt:

Aliasing: Die Simulation von Aliasing Fehlern wird mit der Änderung der Anzahl virtueller Mikrofone vorgenommen. Diese reicht von 1454 bis 5810 in unregelmäßig größer werdenden Abständen. Da die mögliche Mikrofonanzahl in *SOFiA* vorgegeben ist, konnte hierbei auf die Voruntersuchung bezüglich Skaleneinteilungen verzichtet werden. Die höchste verwendete Samplinganzahl ist 5810, da hierbei Aliasing erst bei 20 kHz entsteht und somit keinen Einfluss mehr auf den hörbaren Frequenzbereich ausübt. Weniger als 1000 Sampling Punkte werden allerdings nicht genutzt, da Fehler in diesem Bereich bereits sehr deutlich wahrgenommen werden. Entstehendes Aliasing ist für die gewählten Bereiche in Abbildung [4.2] dargestellt.

Positionierung: Die Positionierung der virtuellen Mikrofone wird wie in Kapitel [3.4.3] beschrieben anhand des *Lebedev Sampling Schema* von *SOFiA* generiert. Ein Positionierungsfehler wird erzeugt, indem auf die ermittelten Positionen ein Zufalls-



(a) Aliasing für 974 Samplingpunkte ist deutlich vorhanden und beginnt bereits bei 1 kHz.

(b) Aliasing für 4802 Samplingpunkte beginnt bei ungefähr 3 kHz und ist wesentlich schwächer ausgeprägt.

Abbildung 4.2: Abhängigkeit des Aliasing im hochfrequenten Bereich von der Anzahl der Samplingpunkte (974, 4802) für die Ordnung $N=10$ und Radius $r=2m$.

wert addiert wird. Es wird ein normalverteilter Winkelversatz generiert, wobei die Standardabweichung σ ausschlaggebend für den Schwellwert ist. Auswirkungen des Positionierungsfehlers sind Abbildung [4.3] zu entnehmen. Es werden Positionierungsfehler für Horizontalen Versatz (Azimuthfehler) und für den Vertikalen Versatz (Elevationsfehler) ausgewertet.

Skalenbereich: $\sigma = 0.001 \dots 0.015^\circ$

Skaleneinteilung: $\Delta\sigma = 0.002^\circ$

Rauschen: Ein Rauschanteil kann hinzugerechnet werden, um Mikrofonrauschen zu simulieren. Der verwendete Signal-Rauschabstand (SNR) ist hierbei der ausschlaggebende Parameter für die Schwellwertermittlung.

Skalenbereich: $SNR = -90 \dots -100 \text{ dB}$

Skaleneinteilung: $SNR_{\Delta} = 1 \text{ dB}$

4.3 Itemerstellung

Nachdem die zu verwendenden Skalenbereiche und die passenden Skaleneinteilungen ermittelt wurden, werden anschließend die Testsignale anhand dieser Vorgaben erstellt.

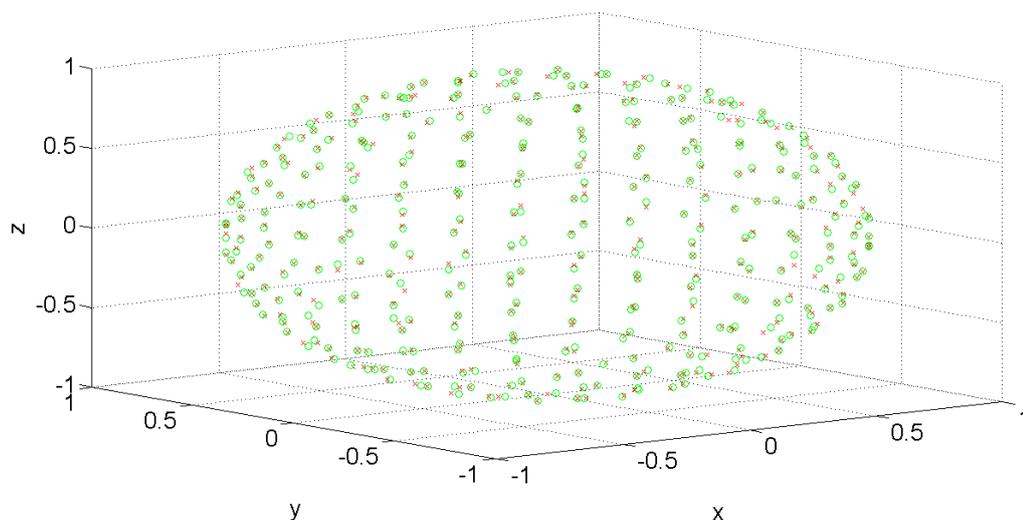


Abbildung 4.3: Mikrofonanordnung für 350 Samplingpunkte im normierten dreidimensionalen Raum. Rote Kreuze stellen die korrekte Position dar, grüne Kreise weisen eine Abweichung von $\sigma = 1^\circ$ auf.

Dazu wird eine Funktion in *SOFiA* genutzt, welche die simulierte Welle in binaurale Impulsantworten umrechnet. Durch Filterung mit diesen Signalen können die Audiosignale gehörrechtig über Kopfhörer mit sämtlichen enthaltenen Richtungsinformationen und Fehlereinflüssen wiedergegeben werden. Für die Pretests wurde jeweils ein Sprach- und ein Musiksignal verwendet, um eine mögliche Signalabhängigkeit der einzelnen Fehler festzustellen. Für alle Fehlerarten wurden Testsignale dieser beiden Signalarten generiert.

Tabelle 4.1: Verwendete Konfigurationen für die Erstellung der Audiosignale mittels *SOFiA*.

Verwendete Konfigurationen	
Kugelarray Radius (m):	$r = 2;$
Kugelarray Konfiguration:	$ac = 2;$ (geschlossene Kugel)
Sampling Frequenz (Hz):	$FS = 48000;$
Azimuthwinkel (rad):	$AZ = \pi/360 \cdot 10;$ (10°)
Elevationswinkel (rad):	$EL = \pi/2;$ (horizontale Ebene)
Ordnung der sphärischen Harmonischen:	$Nsft = 10;$

Die Kugelarraykonfiguration $ac = 2$ symbolisiert ein geschlossenes Kugelarray mit Mikrofonen mit Nierencharakteristik, wodurch das Aliasing bereits im Vorfeld sehr gut unterdrückt werden kann, weshalb sich für diese Konfiguration entschieden wurde. Dabei ist bei offenen Arrays mit omnidirektionalen Mikrofonen deutlich mehr Aliasing zu erkennen als für geschlossene Arrays mit Nierenmikrofonen. Jedoch wird das tieffre-

quente Rauschen für diese Konfiguration erhöht, was sich ebenfalls in der Abbildung zeigt. Eine Rauschunterdrückung ist in *SOFiA* ebenfalls bereits vorhanden. Der Azimutwinkel sämtlicher Signale wurde auf 10° festgelegt, um Symmetrien für die HRFTs zu vermeiden.

4.4 Auftretende Probleme

In Kapitel [2.2] wurden bereits einige Problematiken betrachtet, welche in Hörtests auftreten können. Jedoch kommen in dieser Arbeit einige Neuerungen bezüglich Hörtests vor, welche ebenso einzigartige Problemstellungen aufweisen. Gerade die Verwendung von *SAQI* warf einige Probleme auf.

SAQI bietet eine Vielzahl von Merkmalen für Audiotests, welche sehr umfangreiche Qualitätsbewertungen erlauben. Die unmittelbare Verwendung von *SAQI* hat sich jedoch aus folgenden Gründen als ungeeignet erwiesen:

Merkmalsextraktion: Da *SAQI* sehr viele Qualitätsmerkmale beinhaltet, wurde versucht, zuerst die entsprechenden Merkmale den erzeugten Fehlern passend zuzuordnen. Jedoch ist diese Auswahl sehr subjektiv und kann bei den Probanden schnell zu Unsicherheiten führen, falls diese andere Merkmale für die jeweiligen Fehler verwenden würden. Andererseits würde es die Probanden mitunter überfordern, aus sämtlichen Merkmalen eine Auswahl zu treffen. Ebenso kann die Bewertung und anschließende Auswertung somit einen deutlich größeren Aufwand bedeuten, der den Rahmen dieser Arbeit gesprengt hätte. Es wurde daher entschieden, 17 Merkmale aus dem Inventar auszuwählen, welche für die Aufgabe als passend empfunden wurden. Aus diesen Merkmalen können die Probanden die für sie als passend erachteten Merkmale auswählen.

Schwellwert: Ein weiteres Argument gegen die direkte Verwendung von *SAQI* besteht im Schwellwerttest an sich. Angenommen, einige Merkmale werden den Fehlern im Vorfeld subjektiv zugeordnet und ein Proband wird im Falle von Schwellwerttests befragt, ob ein Unterschied festzustellen ist oder nicht. Selbst wenn der Unterschied nicht mit dem befragten Merkmal zusammenhängt, wird der Proband den Unterschied (wenn wahrnehmbar) erkennen und als solchen beurteilen. Damit kann man zwar einen Schwellwert bestimmen, der allerdings nicht zwangsläufig mit den vorher bestimmten Merkmalen zusammenhängen muss. Es wurde in Betracht gezogen, *SAQI* lediglich für die Merkmalsbestimmung der Schwellwerte zu verwenden. Damit würde eine Schwellwertbestimmung stattfinden und

ein Expertenhörer würde anhand der wahrnehmbaren Unterschiede einige Merkmale zuordnen können. Diese Vorgehensweise wäre jedoch stark subjektiv und von der Expertise des Auswerters abhängig, weshalb sie als unwissenschaftlich erachtet wird.

Daher wurde dieses Vokabular im Laufe der Voruntersuchung als ungeeignet für Hörtests zur Schwellwertbestimmung befunden. Aus diesem Grund werden im Rahmen dieser Arbeit zwei Hörtests erstellt. Ein Test dient wie geplant der Schwellwertbestimmung. Desweiteren wird im zweiten Test die *Repertory Grid Technik (RGT)* verwendet, um eine Analyse über die zutreffenden Merkmale zu erhalten.

Nachfolgend sind die ausgewählte Merkmale aus den Kategorien Klangfarbe, Geometrie, Dynamik und Allgemeines, welche im *RGT* Verwendung finden, aufgelistet:

- Klangfarbe
- Klangfarbliche Ausprägung im Höhenbereich
- Klangfarbliche Ausprägung im Mittenbereich
- Klangfarbliche Ausprägung im Tiefenbereich
- Schärfe
- Rauigkeit
- Kammfilterartigkeit
- Metallische Klangfarbe
- Richtung Azimut
- Richtung Elevation
- Entfernung
- Tiefenausdehnung
- Breitenausdehnung
- Höhenausdehnung
- Lokalisierbarkeit
- Lautheit

- Natürlichkeit
- Klarheit

Es sollte außerdem beachtet werden, dass keine individualisierten *HRTF*'s genutzt wurden, da die Vermessung jedes einzelnen Probanden den Rahmen dieser Arbeit gesprengt hätte. Daher wurden Werte genutzt, welche von einem Kunstkopf stammen. Die Merkmale *Externalisierung* und *Vorn-Hinten-Lage* wurden aus diesem Grund nicht mit einbezogen, da sie als Fehler der binauralen Auralisation durch *HRTF*'s zugeordnet werden und nicht durch die Kugelarraysimulation entstehen.

5 Hörtest Schwellwertbestimmung

5.1 Design

5.1.1 Testmethode

Die in Kapitel [3.4.5] beschriebenen Fehlerquellen wirken sich auf die Auralisationsqualität aus. Um herauszufinden, ab wann der Einfluss eines Fehlers hörbar wird, soll ein Hörtest zur Schwellwertbestimmung durchgeführt werden. Durch Kenntnis des Schwellwertes kann eine Minimalkonfiguration erstellt werden, bei der der Einsatz von Ressourcen (wie zum Beispiel die Anzahl der Mikrofone) so gering wie möglich gehalten werden, aber dennoch Aufnahmen ohne Qualitätseinbußen zu gewährleisten sind. Es wird dabei der perzeptiver Schwellwert von Unterschieden zwischen den Signalen betrachtet. Der Einfluss verschiedener Merkmale von *SAQI* wird nicht berücksichtigt, da die Signalunterscheidung bei diesem Test nicht an Merkmale gebunden ist, sondern lediglich an einen allgemeinen Unterschied. Eine Trennung einzelner Merkmale bei der Betrachtung der Gesamtqualität ist somit nicht möglich. Ein Unterschied zwischen Signalen wird von Probanden nicht einzelnen Merkmalen zugeordnet, aber dennoch bewertet. Eine ausführliche Erläuterung dazu findet sich in Kapitel [4.4] gegeben.

Zur Schwellwertbestimmung haben sich adaptive Prozeduren etabliert, welche in der Literatur in drei verschiedene Verfahrenskategorien unterteilt werden: Die *PEST*-, *Staircase*- und *Best PEST* Methode [Leek, 2001] (siehe Kapitel [3.2.1]). In der vorliegenden Arbeit wird ein *3-AFC* (*three alternative forced choice*)-Hörtest nach dem *Best PEST*-Verfahren durchgeführt ([Lieberman and Pentland, 1982]), da somit eine schnellere Annäherung an den Schwellwert erreicht werden kann und dieses Paradigma weniger vorhersehbar für den Probanden ist. Hierbei werden der Testperson drei Signale A, B und C präsentiert, von denen sich eines von den anderen beiden unterscheidet. Die Testperson muss dieses Signal mithilfe einer grafischen Benutzeroberfläche identifizieren.

Hat die Testperson das richtige Signal markiert, wird die Schwierigkeit erhöht, das bedeutet, dass der Unterschied zwischen den Signalen im nächsten Versuchsteil schwerer

zu erkennen ist. Wenn die Testperson das falsche Signal markiert, wird die Schwierigkeit verringert, die Signalunterschiede sind demnach leichter zu identifizieren. Das Verfahren wird für jede Fehlerart 15 mal wiederholt. Der Schwellwert für jeden Fehler kann anhand der Häufigkeit der durchlaufenen Fehlerstufen bestimmt werden.

Zur Durchführung des Hörtests wurde auf die psychophysikalische Toolbox *WhisPER* der TU Berlin zurückgegriffen. Eine ausführliche Erklärung der in *WhisPER* verwendeten Testmethoden und dem Programm selbst findet sich in [Ciba et al., 2009].

5.1.2 Testdaten

Zur Schwellwertbestimmung werden Kugelarrayschallfelder im Freifeld simuliert und diese entsprechend mit verschiedenen Fehlerausprägungen behaftet. Die Skaleneinteilungen und die zu verwendenden Bereiche wurden bereits in den informellen Vortests, wie in Kapitel [4.2] beschrieben, erstellt und in dieser Form für den Haupttest übernommen. Durch die Abstufungen ergeben sich zwischen 9 und 12 Signale pro Fehlerart und Signalart, welche Tabelle [5.1] zu entnehmen sind. Die Referenzen beinhalten keine Fehleranteile, für Aliasing stellt die höchste Mikrofonanzahl die Referenz dar.

Tabelle 5.1: Abstufungen der einzelnen Fehlersignale.

Aliasing (Sampling Punkte)	Noise (dB)	Azimuth (σ)	Elevation (σ)
1454	-80	30	30
2030	-82	20	20
2354	-84	15	15
2702	-86	13	13
3074	-88	11	11
3470	-90	9	9
3890	-92	7	7
4802	-94	5	5
5810	-96	3	3
	-98	1	1
	-100		

Es werden jeweils ein Sprachsignal und ein Musiksinal ohne Gesang verwendet. Die Länge der Items betragen 6,5 Sekunden (Musik) und 4,5 Sekunden (Sprache).

Der *Schwellwerttest* wurde von 12 Probanden durchgeführt, davon 11 Männer und 1 Frau im Alter zwischen 20 und 40 Jahren. Die Probanden waren sowohl erfahrene Hörer, welche Musikerfahrung aufwiesen oder bereits an Hörtests teilgenommen haben, als auch unerfahrene Hörer ohne musikalische Vorbildung oder Hörtestteilnahmen. Die

Probanden setzten sich aus Studenten (der TU Ilmenau) und Mitarbeitern des *Fraunhofer IDMT* zusammen.

Die Dauer des Tests liegt bei ungefähr 40 Minuten und variiert nur leicht zwischen den Probanden.

5.2 Auswertung und Ergebnisse

Mithilfe eines Schwellwerttests wurden für alle Fehlerarten die Schwellwerte anhand der *Mode*- und *Medianwerte* bestimmt. Die *Mode* bestimmt den am häufigsten auftretenden Wert in einem Histogramm. Dieser wird in dieser Arbeit als der ausschlaggebende Wert zur Schwellwertbestimmung interpretiert. Wenn starke Abweichungen zwischen den Werten vorliegen, kann von einer starken Ungenauigkeit des bestimmten Schwellwertes ausgegangen werden. Wenn Median- und Modewerte übereinstimmen, kann von einem korrekten Schwellwert ausgegangen werden. In den folgenden Abschnitten sind die Schwellwerte dokumentiert. Eine Übersicht der *Mode*- und *Medianwerte* findet sich in Tabelle [5.2].

Tabelle 5.2: Übersicht der erfassten Schwellwerte.

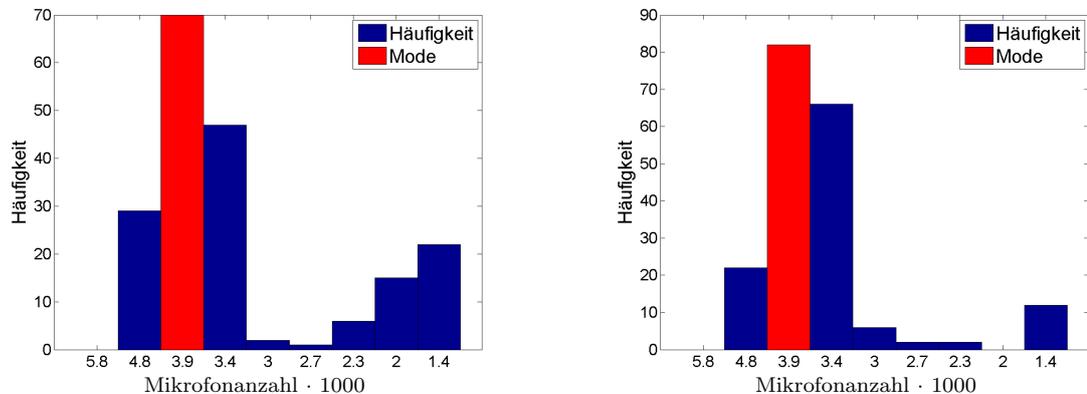
	Musik		Sprache	
	Mode	Median	Mode	Median
Aliasing (Mikrofonanzahl)	3890	3890	3890	3890
Noise (SNR in dB)	-98	-98	-94	-92
Azimuth (σ in $^\circ$)	0,007	0,007	0,005	0,005
Elevation (σ in $^\circ$)	0,030	0,009	0,009	0,009

5.2.1 Aliasing

Für Aliasing haben sich deutliche Grenzen gezeigt. Sowohl bei Musik als auch bei Sprache liegt der Schwellwert bei den durchgeführten Tests bei 3890 Mikrofonen. Durch den deutlich hervortretenden Schwellwert kann die Annahme getroffen werden, dass der Schwellwert noch genauer bestimmt werden kann, wenn nicht die vorgegebene Anzahl von Samplingpunkten verwendet wird, welche durch *SOFiA* vorgegeben ist. *Mode*- und *Medianwerte* stimmen überein, was für einen zuverlässigen Schwellwert spricht.

Histogramme der beiden Signale sind in Abbildung [5.1] zu finden. Die x – Achse zeigt die Fehlerstärke, welche für Aliasing durch die Anzahl der Mikrofone ausgedrückt wird. Die Referenz ist die höchste in *SOFiA* verfügbare Anzahl an Mikrofonen (5810).

Die absolute Häufigkeit wird durch die y – Achse beschrieben. Die Mode ist der am häufigsten auftretende Wert, welcher den Schwellwert darstellt und rot markiert ist.



(a) Schwellwerthistogramm für Aliasing bei Musiksignalen.

(b) Schwellwerthistogramm für Aliasing bei Sprachsignalen.

Abbildung 5.1: Schwellwerthistogramme für Aliasing. Die Mode ist Rot markiert und stellt den Schwellwert dar.

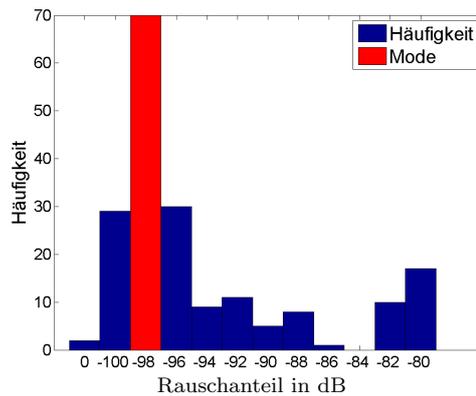
5.2.2 Rauschen

Für die Rauschanteile wurde ein Schwellwert von -98 dB Signal-Rauschabstand bei Musikitems ermittelt. Sprachitems können höhere Werte von -94 dB Rauschanteil aufweisen, bevor ein Unterschied erkannt wird. Der Schwellwert für Musik ist eindeutig definiert, wohingegen Sprachsignale unter den Probanden leichte Unterschiede aufweisen. Für Sprachsignale liegen die *Mode*- und *Medianwerte* nah beieinander, stimmen allerdings nicht überein. Der Medianwert für Sprachsignale liegt mit -92 dB um 2 dB höher als der Modewert. Man kann davon ausgehen, dass der Schwellwert einen weichen Übergang aufweist, welcher zwischen diesen Werten liegt.

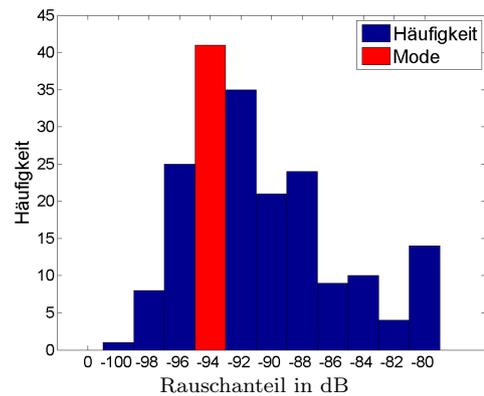
Die Abbildung [5.2] zeigt die Schwellwerte bei verschiedenen Rauschanteilen.

5.2.3 Positionierungsfehler

Azimuth: Positionierungsfehler im Azimuthbereich sind normalverteilt und werden ab einem Sigma von $\sigma = 0,007^\circ$ bei Musiksignalen wahrgenommen. Für Sprachsignale wird der Fehler bei leicht geringeren Fehlerwerten von $\sigma = 0,005^\circ$ wahrgenommen. *Mode*- und *Medianwerte* stimmen hier ebenfalls überein, sowohl für Sprache als auch für Musik, was für einen zuverlässigen Schwellwert spricht. Die



(a) Schwellwerthistogramm für Rauschen bei Musiksignalen.

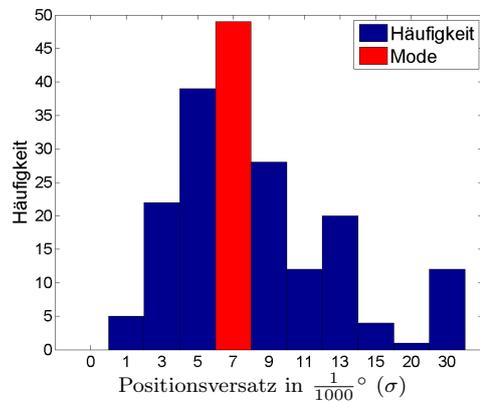


(b) Schwellwerthistogramm für Rauschen bei Sprachsignalen.

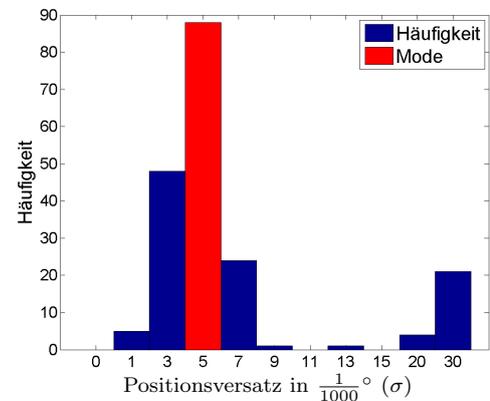
Abbildung 5.2: Schwellwerthistogramme für Mikrofonrauschen. Die Mode ist Rot markiert und stellt den Schwellwert dar.

Abbildung [5.3] zeigt die Schwellwerte bei einem normalverteilten Azimuthversatz als Standardabweichung.

Elevation: Der Schwellwert für die Elevation bei Sprachsignalen liegt mit $\sigma = 0,009^\circ$ bei einem etwas höheren Fehlerwert als beim Azimuth. Musiksignale erzeugen eine Mode von $\sigma = 0,03^\circ$, was unerwartet hoch ist. Der Median liegt jedoch wie beim Sprachsignal bei $\sigma = 0,009^\circ$. Es ist festzustellen, dass mehr als eine Testperson keine Unterschiede feststellen konnte und somit Ausreißer in den Daten entstanden sind. Aus diesem Grund kann der Schwellwert aus den durchgeführten Tests nicht ermittelt werden. Abbildung [5.4] zeigt die Schwellwerte bei einem normalverteilten Elevationsversatz als Standardabweichung.

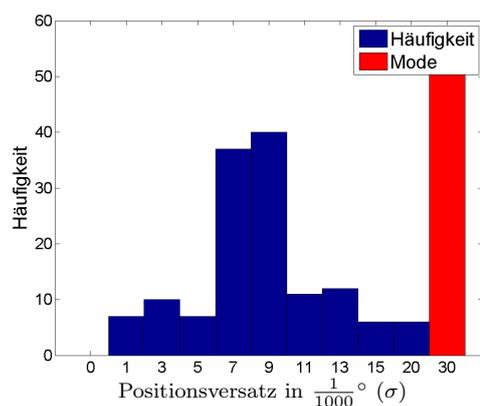


(a) Schwellwerthistogramm für eine Azimuthverschiebung bei Musiksignalen.

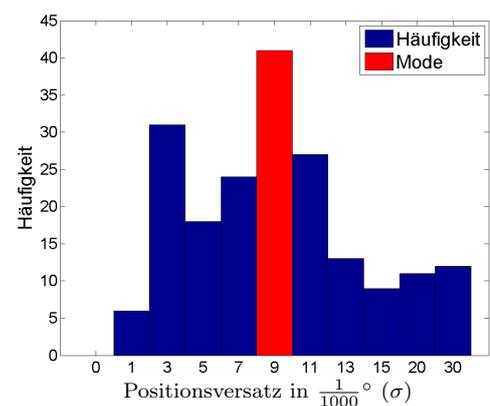


(b) Schwellwerthistogramm für eine Azimuthverschiebung bei Sprachsignalen.

Abbildung 5.3: Schwellwerthistogramme für eine Azimuthverschiebung. Die Mode ist Rot markiert und stellt den Schwellwert dar.



(a) Schwellwerthistogramm für eine Elevationsverschiebung bei Musiksignalen.



(b) Schwellwerthistogramm für eine Elevationsverschiebung bei Sprachsignalen.

Abbildung 5.4: Schwellwerthistogramme für eine Elevationsverschiebung. Die Mode ist Rot markiert und stellt den Schwellwert dar.

6 Repertory Grid Test

6.1 Design

6.1.1 Testmethode

Dieser Test dient der Zuordnung von wahrgenommenen Merkmalen zu den aufgetretenen Fehlern bei der Kugelarraysimulation. Es wird das deskriptive Vokabular *SAQI* als Basis für die Merkmale herangezogen. Um die Anzahl der Merkmale übersichtlich zu halten, werden im Vorfeld einige Merkmale ausgewählt, von welchen, wie bereits in Kapitel [4.4] dargestellt, eine mögliche Zuordnung zu den Fehlern angenommen wird. Die Auswahl geschieht durch wiederholte Vergleiche von Merkmalen und Fehlerstärken sowie Referenzvergleichen. Ebenso wurden Änderungen im Frequenzgang in Betracht gezogen.

Der Test teilt sich in drei Phasen. Zuerst wird ein kurzes Training durchgeführt, um die Probanden mit den Merkmalen vertraut zu machen. Dafür wurden Audiosignale ausgewählt, aus denen die entsprechenden Attribute deutlich erkennbar sind. Dem Probanden wird die Liste von Qualitätsmerkmalen vorgelegt, welche zuvor aus *SAQI* entnommen wurden. Die Aufgabe besteht darin, ein Referenzsignal mit einem Fehler-signal zu vergleichen und Merkmale aus *SAQI* diesen Unterschieden zuzuordnen. Der Versuchsleiter notiert die genannten Merkmale. Anschließend findet eine Bewertung unterschiedlicher Fehlerstufen statt. Die Skala wird dabei von den durch die Probanden gewählten Merkmale bestimmt. Es wird Wert darauf gelegt, die Merkmale getrennt zu betrachten und somit mehrere kurze Tests jeweils für die einzelnen Merkmale zu erstellen.

Der Versuchsleiter befindet sich permanent bei dem Probanden, um bei eventuellen Missverständnissen bezüglich der Merkmalsdefinitionen aus *SAQI* Rücksprache zu halten.

Aus der Untersuchung wird eine Zuordnung von Merkmalen zu Fehlerarten ermöglicht sowie eine Tendenz der Häufigkeit gewählter Merkmale bestimmt.

6.1.2 Testdaten

Es werden für den *RGT*-Test drei Abstufungen der Fehler als Kugelarrayschallfelder im Freifeld simuliert. Für die Abstufung der Fehler wurde die Einteilung nach eigenem Ermessen und Erfahrungswerten erstellt. Sie wurden in *schwach*, *mittel* und *stark* unterteilt. Die Fehlereinflüsse werden für beide Signale separat hinzugefügt, sodass insgesamt 24 Testsignale (4 Fehlerarten in jeweils 3 Abstufungen für 2 Signalarten) vorhanden sind. Die Werte für die Fehlerarten und Abstufungen sind Tabelle [6.1] zu entnehmen.

Tabelle 6.1: Verwendete Abstufungen der Repertory Grid Signale.

	Aliasing (Sampling Punkte)	Noise (dB)	Azimuth (σ in $^\circ$)	Elevation (σ in $^\circ$)
schwach	3470	94	0.005	0.005
mittel	2702	88	0.013	0.013
stark	1454	80	0.030	0.030

Es werden jeweils ein Sprachsignal und ein Musiksinal ohne Gesang verwendet. Die Länge der Signale betragen 6,5 Sekunden (Musik) und 4,5 Sekunden (Sprache).

Der *Repertory Grid Test* wurde von 18 Probanden durchgeführt, davon 14 Männer und 4 Frauen im Alter zwischen 22 und 40 Jahren. Die Probanden waren sowohl erfahrene Hörer, welche Musikerfahrung aufwiesen oder bereits an Hörtests teilgenommen haben, als auch unerfahrene Hörer ohne Musikalische Vorbildung oder Hörtestteilnahmen. Die Probanden waren zu dem Zeitpunkt des Tests am *Fraunhofer IDMT* tätig, als Studenten oder Mitarbeiter.

Die Länge des *RGT* ist stark abhängig von der Erfahrung der Probanden, was eine vorherige Schätzung der Testdauer nicht möglich machte. Erfahrene Teilnehmer sind sehr zügig mit ihren Antworten und der Bewertung der Audioitems und benötigen somit unter Umständen weniger als 60 Minuten für den Test. Unerfahrene Probanden hingegen benötigen zum Teil über 90 Minuten, da sie mehr Zeit für die Identifizierung der Merkmale aufwenden.

6.2 Auswertung und Ergebnisse

6.2.1 Bemerkungen der Probanden

Der Versuchsleiter war während der gesamten Testdauer anwesend, um eventuelle Fragen der Testpersonen beantworten zu können.

Es wurde festgestellt, dass starke Differenzen zwischen erfahrenen Hörern und Probanden ohne Hörtesterfahrung aufgetreten sind. Der Test dauert für unerfahrene Hörer wesentlich länger als bei erfahrenen Hörern. Ebenso versuchen diese, so viele Merkmale aus *SAQI* zu benutzen wie möglich, da sie das Gefühl beschrieben haben, "man muss die [Merkmale] alle mal erwähnen". Erfahrene Probanden hingegen beschränken sich meist auf drei Merkmale, welche sie in den ersten Sekunden als gravierend aufgenommen haben und übertragen ihre Erfahrungen dann auf das Inventar von *SAQI*. Aus diesem Grund musste der Test bei einigen unerfahrenen Probanden verkürzt werden, um den Stressfaktor gering zu halten und keine verfälschenden Ergebnisse aufgrund von Ermüdungserscheinungen zu erhalten.

Das anfängliche Training stellte sich als notwendig heraus, da viele der Probanden äußerten, sie hätten ohne das Training die Merkmale nicht hätten zuordnen können. Die Probanden haben während des *RGT's* weitere Bemerkungen und Hinweise gegeben, welche vom Versuchsleiter zu Untersuchungszwecken aufgenommen wurden. Einige besonders interessante Feststellungen von Probanden und Versuchsleiter werden nachfolgend erläutert.

Kunstkopf HRTF's: Der Azimuth - Versatz wird von einigen Probanden nur sehr schwach wahrgenommen, selbst im Training, wo ein stark richtungsverschobenes Sprachsignal als Beispiel dient. Das deutet darauf hin, dass die verwendeten Kunstkopf *HRTF's* nicht für alle Probanden gleichermaßen einsetzbar sind. Ein ähnlicher Sachverhalt ist bei der Entfernung festzustellen, welche bei einigen Probanden nur als leiser, nicht jedoch als weiter entfernt wahrgenommen werden. Weitere Probanden waren der Auffassung, dass es schwer ist, nur ein Merkmal getrennt zu betrachten.

Begrifflichkeiten: Die *Vorn-Hinten-Vertauschung* wurde aus der Liste der Merkmale entfernt, da sie dem *HRTF*-Problem zuzuordnen ist. Allerdings hat ein Proband genau dieses Phänomen als Elevationsversatz interpretiert. Dies zeigt, dass einige Begrifflichkeiten nicht vollständig verstanden wurden oder eine andere Definition bei dem Probanden vorherrscht als beim Versuchsleiter, was zu einer falschen

Interpretation führt. Es besteht daher die Möglichkeit, dass diese Falschinterpretation auch bei anderen Probanden aufgetreten ist. Ein weiterer Proband hat die Lokalisierbarkeit als Merkmal angegeben, war sich allerdings nicht sicher ob er mit Lokalisierbarkeit nicht eher Externalität meinte.

Trennung: Die Trennung der Begrifflichkeiten war für viele Probanden eine Herausforderung. Wenn diese beispielsweise Schärfe beurteilen sollen, wird die Aufmerksamkeit ebenfalls auf die Natürlichkeit gelenkt. Einige Probanden hatten Bedenken geäußert, dass sie womöglich zwei Merkmale identisch bewerten.

Natürlichkeit: Probanden haben oft bei der Natürlichkeit für die Musikitems angedeutet, dass Musik "an sich ja nichts unbedingt Natürliches ist". Damit war gemeint, dass Musik nicht zwangsläufig natürliche Geräusche erzeugt, sondern ebenso aus nicht gewohnten Klängen bestehen kann. Natürlichkeit ist daher für die meisten eine Plausibilitätsfrage oder richtet sich nach den Gewohnheiten der jeweiligen Person.

Mittbereich: Der Mittbereich der Frequenzen ist in *SAQI* nicht konkretisiert. Daher können die Grenzwerte des Mittbereichs bei den Teilnehmern unterschiedlich ausfallen.

Merkmalsüberschneidung: Einige Merkmale überschneiden sich, je nachdem wie man sie definiert und bei welchen Signalen welche Veränderungen vorgenommen werden. Ein Proband hat dies anhand des Musikstücks erklärt: "Die Hi-Hat ist metallisch. Wenn die Höhen runtergedreht werden, klingt es demnach auch weniger Metallisch." Ein weiterer Teilnehmer hat die Merkmale als nicht eindeutig und sich teilweise überschneidend eingestuft. Diese Bemerkungen basieren auf einer subjektiven Auffassung, dennoch kann damit eine Korrelation einzelner Merkmale in Betracht gezogen werden. Bei vielen Personen tritt das Empfinden auf, dass Höhen abgesenkt werden, obwohl die Tiefen angehoben sind, oder umgekehrt, was wiederum zeigt, dass eher eine Vergleichsbasis als absolute Werte ausgewertet werden. Desweiteren wird von einigen Probanden eine Redundanz beschrieben zwischen der Klangfarbe (hell - dunkel) und den Höhen sowie den Tiefen: "Die Klangfarbe macht doch die Höhen- und Tiefenausprägung aus."

6.2.2 Merkmalshäufigkeiten

Einige Merkmale werden sehr häufig von den Probanden genannt, woran erkennbar ist, welche Merkmale von den Probanden bevorzugt gewählt werden. Nachfolgend ist die

Häufigkeit der Merkmale aufgeführt, deren mögliche Maximalzahl gleich der Anzahl der Probanden (18) ist. Die Gesamthäufigkeit wird in Tabelle [6.2] aufgezählt.

Sehr häufig gewählte Merkmale (13 bis 18 Angaben): Am häufigsten werden die Merkmale *Breitenausdehnung*, *Entfernung*, *Höhen-*, *Mitten-* sowie *Tiefenbereich*, *Klangfarbe*, *Azimuthversatz*, *Natürlichkeit* und *Schärfe* genannt.

Durchschnittlich gewählte Merkmale (7 bis 12 Angaben): 7 bis 12 mal werden die Merkmale *Kammfilterartigkeit*, *Klarheit*, *Lautheit*, *Lokalisierbarkeit* und *Metallische Klangfarbe* angegeben.

Selten gewählte Merkmale (1 bis 6 Angaben): Selten werden die Merkmale *Höhenausdehnung*, *Rauigkeit* und *Elevationsversatz* genannt. Die zu geringe Anzahl an Bewertungen hierbei lässt keine reliable Auswertung zu.

Tabelle 6.2: Häufigkeit gewählter Merkmale.

#	Merkmal	Summe
1	Breitenausdehnung	14
2	Entfernung	15
3	Höhenausdehnung	1
4	Höhenbereich	11
5	Kammfilter	9
6	Klangfarbe	12
7	Klarheit	8
8	Lautheit	10
9	Lokalisierbarkeit	9
10	Metallisch	9
11	Mittenbereich	13
12	Natürlichkeit	12
13	Rauigkeit	3
14	Azimuth Versatz	16
15	Elevations Versatz	6
16	Schärfe	14
17	Tiefenbereich	16
	Gesamtzahl Tests	178

6.2.3 Hauptkomponentenanalyse

Zur Faktorenreduktion wurde eine *Hauptkomponentenanalyse* (*Principal Component Analysis, PCA*) durchgeführt, um die Anzahl von Merkmalen zu reduzieren und eventuell Merkmalsgruppen bilden zu können. Diese Technik wurde erstmals von Pearson vor über 100 Jahren angewendet [Pearson, 1901]. Ausführlich erläutert wird das Thema zum Beispiel in [Jolliffe, 2002]. Die Durchführung der *PCA* und ihre Ergebnisse sind nachfolgend dokumentiert. Eine Beschreibung der durchzuführenden Schritte ist ebenfalls aufgeführt. Ein Flussdiagramm des Ablaufs findet sich unter Abbildung [6.1].

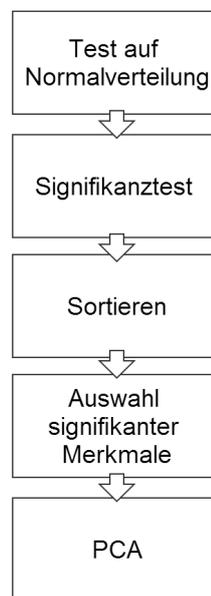


Abbildung 6.1: Flussdiagramm der Arbeitsreihenfolge zur Hauptkomponentenanalyse.

Ablauf

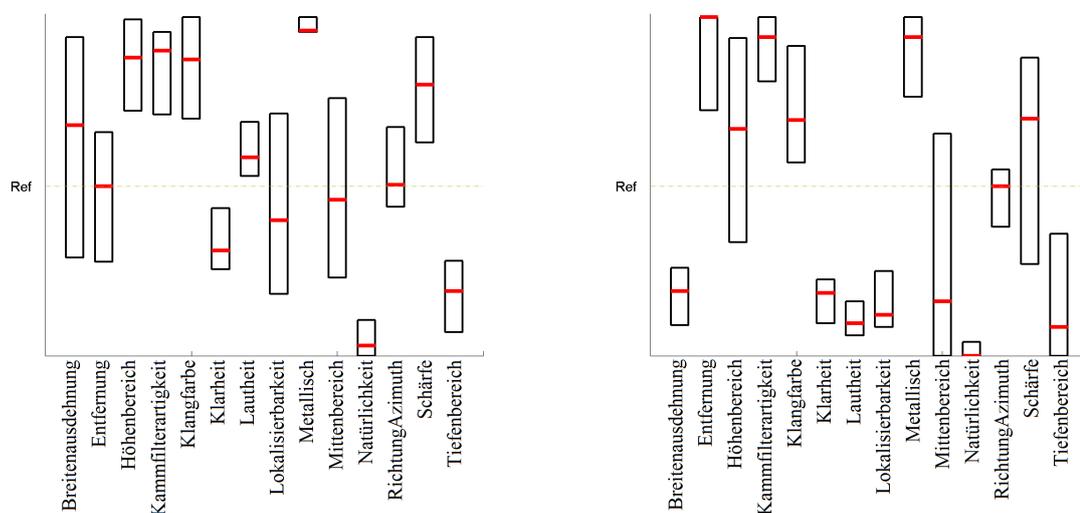
Test auf Normalverteilung

Der erste Schritt zur Auswertung von statistischen Ergebnissen ist die Überprüfung, ob die Daten normalverteilt sind. Dazu wird der *Lilliefors*-Test verwendet. Dieser ist eine Modifikation des *Kolmogorov-Smirnov* Tests. Er ist geeignet, um eine Stichprobe auf Normalverteilung zu überprüfen, ohne dass der Mittelwert oder die Varianz bekannt sind [Lilliefors, 1967]. Der *Lilliefors*-Test ergab, dass von den durchgeführten Tests mehr als 30 % als signifikant abweichend von einer Normalverteilung eingestuft werden. Daher wird die gesamte Untersuchung als nicht normalverteilt angesehen. Aus diesem Grund wird keine Varianzanalyse

durchgeführt, sondern auf parameterfreie Tests wie die *PCA* zurückgegriffen.

Signifikanztest

Nach dem Test auf Normalverteilung wurde ein Signifikanztest durchgeführt, um festzustellen, welche Merkmale deutlich von der Referenz abweichen. Es wurde dabei der nichtparametrische Vorzeichentest mit einem Signifikanzniveau von 5 % gewählt, welches üblicherweise als Grenzwert festgelegt wird ([Chakraborti and Gibbons, 2011]). Signifikant abweichende Merkmale wurden zur Berechnung in die *PCA* aufgenommen. Es wurden zur visuellen Kontrolle Boxplot-Grafiken erstellt, anhand derer die Stärke der Abweichung zur Referenz zu erkennen ist. Für den Aliasingfehler sind die gewählten Merkmale und die dazugehörige Grafik beispielhaft in Tabelle [6.3] und Abbildung [6.2] dargestellt. Die Grafik zeigt sämtliche verwendeten Merkmale auf der *x-Achse*. Die Boxplots stellen die 25%- und 75%-Quartile dar sowie den Median als rot markierte Linie. Die Abweichung zur Referenz (gestrichelte Linie) kann anhand der Lage der Daten nachvollzogen werden. Die restlichen Daten finden sich im Anhang unter [A.1], [A.2], sowie [A.3].



(a) Boxplots der Merkmale für Aliasingfehler bei Musik.

(b) Boxplots der Merkmale für Aliasingfehler bei Sprache.

Abbildung 6.2: Boxplots der Merkmale für Aliasingfehler bei Musik und Sprache. Es wird das obere und untere Quartil sowie der Median dargestellt.

Nachfolgend ist aufgeführt, welche Merkmale aus dieser Überlegung bei den einzelnen Fehlern zutreffen und somit in die *PCA*-Berechnung einfließen. In Tabelle [6.4] ist die komplette Zuordnung aufgelistet.

Tabelle 6.3: Signifikanztest für Aliasing. Die erste Wertezeile stellt die berechnete Signifikanz dar. Wenn diese unter dem Signifikanzniveau von 5% liegt, wird die Nullhypothese (Merkmal wird als Referenz gedeutet) abgewiesen und die zweite Zeile ergibt den Wert 1. Alle Merkmale mit dem Wert 1 finden demnach in der *PCA* Verwendung.

Breite	Entfernung	Höhenbereich	Kammfilter	Klangfarbe	Klarheit	Lautheit	Lokalisierbarkeit	Metallisch	Mittbereich	Natürlichkeit	Azimuthversatz	Schärfe	Tiefbereich
.79	1.0	.00	.00	.04	.07	.11	.51	.00	1.0	.00	.79	.06	.00
0	0	1	1	1	0	0	0	1	0	1	0	0	1

Aliasing: Hier werden für das Musikitem hauptsächlich die Klangfarbe einschließlich Höhen- und Tiefbereich, sowie die Natürlichkeit und eine metallische Klangfarbe angegeben, was ebenfalls deutlich anhand Abbildung [6.3(a)] zu erkennen ist. Aliasing wurde ebenfalls als Kammfilterartig bewertet, was anhand der in hohen Frequenzbereichen auftretenden Überhöhungen nachvollziehbar ist (Vergleich Abbildung [4.2]).

Bei der Sprache wird zudem die Klarheit und die Lautheit und damit verbunden die Entfernung stärker bewertet. Die Lokalisierbarkeit spielt bei der Sprache ebenso eine stärkere Rolle. Man kann daher davon ausgehen, dass die Merkmalszuordnung signalabhängig ist.

Rauschen: Bei Musik wirkt sich Rauschen sehr stark aus. Es wurde eine große Anzahl an Fehlern als signifikant eingestuft, darunter klangfärbliche Ausprägungen, Klarheit, Lautheit oder auch Natürlichkeit.

Im Sprachbereich sind die Änderungen in den tiefen Frequenzbereichen präsent, sowie die Entfernung und die Natürlichkeit. Ein Azimuthversatz wird ebenfalls angegeben.

Positionsfehler: Bei den Positionsfehlern sind sehr viele Übereinstimmungen für Azimuth- und Elevationsversatz zu erkennen. Bei beiden Positionsfehlern werden Merkmale wie Klangfarbe, Klarheit, Lautheit und Natürlichkeit, sowie Schärfe angegeben.

Die Sprachitems weisen hier allerdings verhältnismäßig wenig signifikante

Merkmale auf. Für den Azimuthfehler sind lediglich Entfernung, sowie der Mitten- und Tiefenbereich präsent. Elevationsfehler erzeugen Klangfarbenänderungen, die Schärfe und die Natürlichkeit werden beeinflusst, außerdem kann ein Azimuthversatz nach rechts angegeben werden. Der Azimuthversatz ist für die Elevation ebenfalls bei Musikitems festzustellen. Für das Phänomen, dass ein Azimuthversatz nach rechts für alle Probanden festgestellt wurde, obwohl die Verschiebung in Elevationsrichtung stattfindet, liegt noch keine zufriedenstellende Erklärung vor.

Es ist erkennbar, dass Aliasing als Fehlerart heraussticht, da sie als einzige eine metallische Klangfarbe und eine Kammfilterartigkeit sowohl bei Musik als auch bei Sprache aufweist. Rauschen und Positionsfehler weisen im Musikbereich größtenteils übereinstimmende Fehlermerkmale auf. Bei Sprachitems werden außerdem weniger Merkmale angegeben als bei Musik. Dies ist ebenfalls nur bei Aliasing nicht der Fall, hierbei werden deutlich mehr Merkmale bei Sprache verwendet.

PCA

Da die Reihenfolge der Probanden und der gewählten Merkmale keine Rolle für die Auswertung spielt, werden die Daten in der jeweiligen Testreihe nach ihrem absoluten Abstand zur Referenz geordnet. Mit den Daten aus der entstandenen Matrix wird anschließend die *PCA* berechnet. Diese wurde nur für die starken Fehler durchgeführt (Vergleich Tabelle [6.1]), da bei geringer Fehlerstärke ebenfalls nur geringe Abweichungen zur Referenz von den Probanden bewertet werden. Um eine eindeutige Zuordnung zu ermöglichen, kommen daher nur Testsignale in Frage, bei welchen die Fehler deutlich hörbar sind.

Tabelle 6.4: Zuordnung der Merkmale zu den einzelnen Fehlern durch den Signifikanztest.

Fehlerart	Merkmale Musik	Merkmale Sprache
Aliasing	Höhenbereich Kammfilterartigkeit Klangfarbe Metallisch Natürlichkeit Tiefenbereich	Breitenausdehnung Entfernung Kammfilterartigkeit Klarheit Lautheit Lokalisierbarkeit Metallisch Natürlichkeit Tiefenbereich
Rauschen	Entfernung Höhenbereich Klangfarbe Klarheit Lautheit Lokalisierbarkeit Mittenbereich Natürlichkeit Schärfe Tiefenbereich	Entfernung Klangfarbe Natürlichkeit Azimuthversatz Tiefenbereich
Azimuth	Entfernung Klangfarbe Klarheit Lautheit Mittenbereich Natürlichkeit Schärfe Tiefenbereich	Entfernung Mittenbereich Tiefenbereich
Elevation	Breitenausdehnung Höhenbereich Klangfarbe Klarheit Lautheit Mittenbereich Natürlichkeit Azimuthversatz Schärfe Tiefenbereich	Klangfarbe Natürlichkeit Azimuthversatz Schärfe Tiefenbereich

Ergebnisse

Die Merkmale mit der größten Varianz werden im zweidimensionalen Biplot auf eine Ebene projiziert. Der Anteil der Hauptkomponenten an der Varianz wurde nach Formel [6.1] als Prozentwert ausgedrückt, wobei λ_i und λ_j der i -te beziehungsweise j -te Varianzwert und m die Anzahl an Komponenten darstellt ([Jolliffe, 2002]).

$$G_i = \frac{\lambda_i}{\sum_{j=1}^m \lambda_j} \cdot 100 \quad (6.1)$$

Die Komponenten 1 und 2 besitzen für alle Fehlerarten eine kumulierte Varianz von 80 – 100 %, was darauf hindeutet, dass die ersten beiden Komponenten die Werte bereits sehr gut abbilden. Ebenso ist das Kaiserkriterium für sämtliche Fehlervarianzen erfüllt, welches besagt, dass alle mittelwertfreien Eigenwerte größer als 1 bei der Faktorenreduktion von Bedeutung sind und in die Auswertung mit einfließen sollten [Bacher et al., 2010]. Die Eigenvektoren und Varianzen zur Faktorenreduktion sind beispielhaft für Aliasing in Tabelle [6.5] aufgeführt. Die gesamte Tabelle befindet sich im Anhang [A.1].

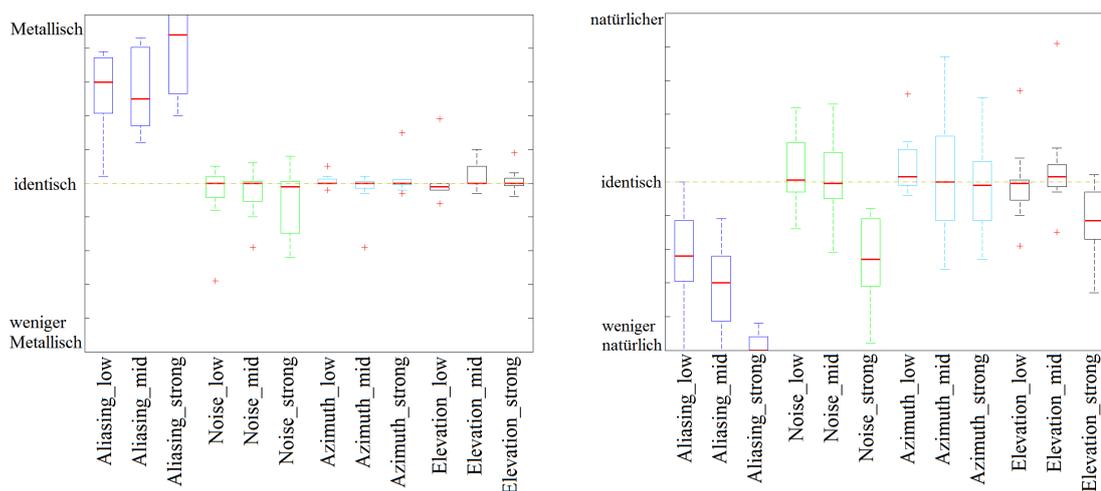
Tabelle 6.5: Eigenwerte und Varianzen für Aliasing.

Komponente	λ	erklärte Varianz (%)	kumulierte Varianz (%)
1	4.69	78.13	78.13
2	0.85	14.12	92.24
3	0.27	4.47	96.71
4	0.12	2.02	98.73
5	0.06	0.99	99.72
6	0.02	0.28	100

Die Analyse der Biplots ergab, dass Aliasing einen Sonderfall bei den Fehlerarten einnimmt. Nur bei Aliasing finden sich die Merkmale kammfilterartig und metallisch wieder, welche zudem eine Gruppe bilden. Natürlichkeit und Tiefenbereich stellen eine weitere Gruppe dar. Eine Überhöhung im tieffrequenten Bereich steht daher für einen natürlicheren Klang. Dies ist jedoch nur bei Aliasing der Fall. Bei allen anderen Fehlerarten wird eine geringere Natürlichkeit für eine Tiefenabsenkung angegeben, wie ebenfalls anhand von Abbildung [6.5] zu erkennen ist. Die Vermutung liegt Nahe, dass die metallische Klangfarbe, welche nur bei Aliasing auftritt und für einen unnatürlichen Klang sorgt, aus den höheren Frequenzbereichen stammt. Boxplots [6.3(a)] und [6.3(b)] verdeutlichen ebenfalls den Zusammenhang zwischen metallischer Klangfarbe und Na-

türlichkeit bei Aliasing. Die Abbildungen zeigen Boxplots der genannten Merkmale über die drei Fehlerstärken schwach (*low*), mittel (*mid*) und stark (*strong*). Boxplots gleicher Fehlerart sind in der gleichen Farbe markiert (Aliasing = Blau, Rauschen = Grün, Azimuthversatz = Cyan, Elevationsversatz = Schwarz). Dabei beschreibt der durchgezogene Kasten die Quartile (25% und 75%), die Mediane sind als roter Strich dargestellt und Ausreißer als rote Kreuze. Whisker stellen Extremwerte dar, welche nicht als Ausreißer interpretiert werden und innerhalb 99,3% der Datenverteilung liegen. Diese sind mit gestrichelten Linien zum Boxplot verbunden.

Eine Auffälligkeit bei den genannten Grafiken ist, dass die Empfindung der Natürlichkeit der Signale bei Rauschen, Azimuth- und Elevationsversatz sogar die Referenz übersteigen kann, was zeigt, dass eine Änderung eines Signals oder sogar ein vermeintlicher Qualitätsverlust nicht zwangswise mit einer Verminderung einer natürlichen Empfindung einhergehen.



(a) Boxplot des Merkmals *metallische Klangfarbe* bei Sprache über den einzelnen Fehlerarten. Nur bei den Aliasingfehlern sind deutliche Abweichungen zu erkennen.

(b) Boxplot des Merkmals *Natürlichkeit* bei Sprache über den einzelnen Fehlerarten. Eine ähnliche Verteilung (umgekehrt) zu metallischer Klangfarbe sind bei Aliasing zu erkennen.

Abbildung 6.3: Boxplots für Natürlichkeit und metallischer Klangfarbe für Sprache über alle Fehler.

Anstelle der Natürlichkeit bildet bei Rausch- und Positionierungsfehlern die Entfernung häufig eine Gruppe mit dem Tiefenbereich. Tieffrequente Signale werden demnach als weiter entfernt wahrgenommen, was den in den Grundlagen ([3.3.3]) erläuterten Faktor der Entfernungswahrnehmung bestätigt. Diese Unterschiede zeigen, dass Merkmale abhängig von der Fehlerart interpretiert werden. Die Natürlichkeit lässt sich bei den genannten Fehlerarten mit der Klangfarbe in Verbindung bringen. Sprachsignale lassen

bei Positionierungs- und Rauschfehlern keine verlässige Clusterbildung zu, da kaum Merkmale für Sprache als signifikant eingestuft wurden.

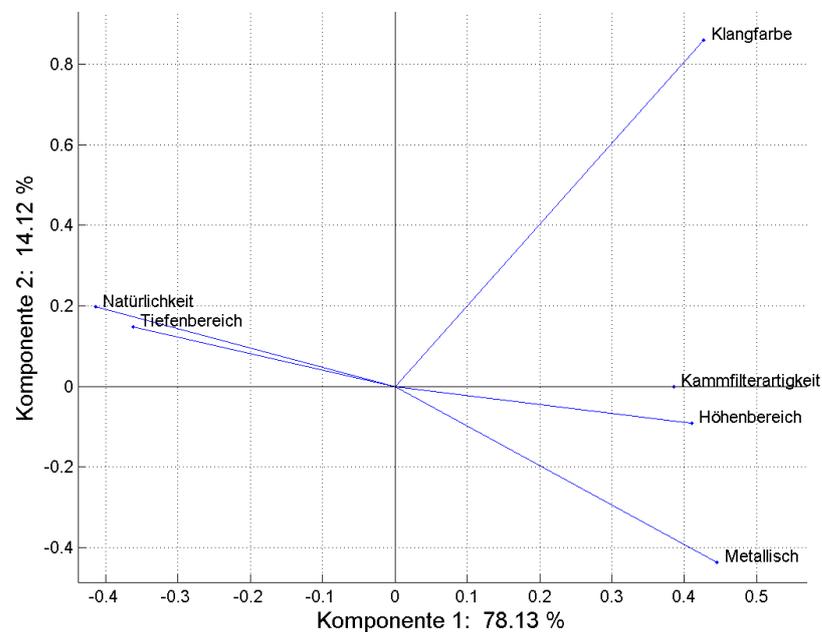


Abbildung 6.4: Biplot für Musik bei Aliasingfehler. Natürlichkeit und Tiefenbereich können einer Gruppe zugeordnet werden, sowie Kammfilterartigkeit, Höhenbereich und metallische Klangfarbe.

In dieser Arbeit wurde eine *PCA* für alle Merkmale und alle Probanden bei einer Fehlerart durchgeführt. In fortführenden Arbeiten könnte eine andere Herangehensweise genutzt werden, in welcher Daten eines Probanden über alle Merkmale und sämtliche Fehlerarten genutzt werden, wie es bereits in [Avni et al., 2013] vorgestellt wurde. Es wäre außerdem möglich, eine unabhängige Komponentenanalyse (*independent component analysis, ICA*) durchzuführen, um den Projektionsraum optimal auf die Achsen abzubilden. Weitere Ausführungen zur *ICA* finden sich unter [Hyvärinen and Oja, 2000].

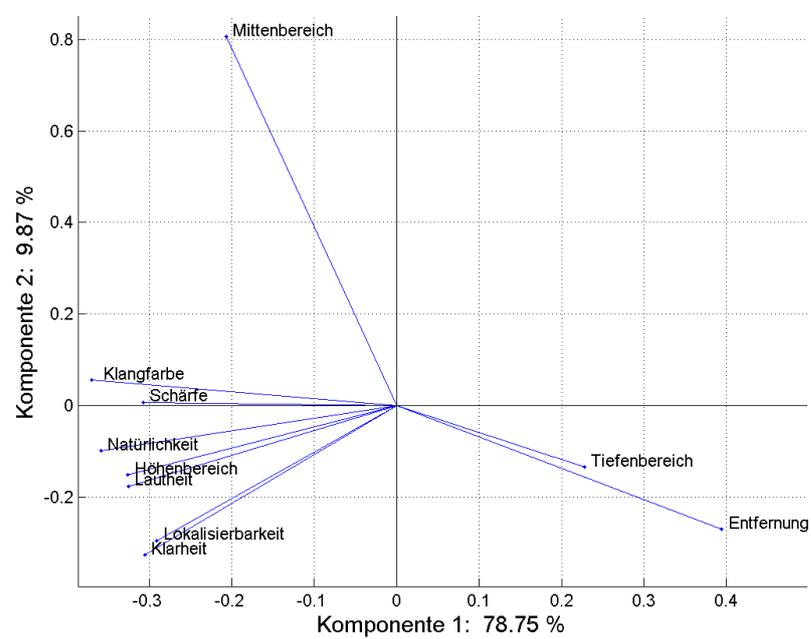


Abbildung 6.5: Biplot für Musik mit Rauschanteil. Tiefenbereich und Entfernung bilden eine Gruppe. Der Mittbereich liegt separat, die restlichen Merkmale könnten ebenfalls gruppiert werden.

7 Fazit

7.1 Zusammenfassung

Ziel dieser Arbeit war die qualitative Bewertung von Kugelarraysimulationen bei binauraler Wiedergabe. Ein besonderer Fokus lag hier auf der perceptiven Bewertung von Messfehlern, wie beispielsweise räumliches Aliasing, Mikrofonrauschen oder Positionierungsfehler. Diese sollten durch Merkmale aus dem Vokabelkatalog *SAQI* beschrieben werden. Desweiteren wurden perzeptive Schwellwerte der einzelnen Fehlerarten gemessen. Beide Aufgaben wurden mithilfe von zwei Hörtests durchgeführt.

Der Schwellwerttest lieferte bei übereinstimmenden *Mode-* und *Medianwerten* nachvollziehbare Ergebnisse. Aliasing weist für Musik- und Sprachsignale gleichermaßen einen Schwellwert von 3890 Mikrofonen auf. Fehler, welche durch Rauschen verursacht werden besitzen einen Schwellwert von -98 dB bei Musiksignalen und -94 dB bei Sprachsignalen. Eine Signalabhängigkeit ist für diese Fehlerart erkennbar. Für Positionierungsfehler wurde ein normalverteilter Winkelversatz in Azimuthrichtung von $\sigma = 0,007^\circ$ für Musik und $\sigma = 0,005^\circ$ für Sprache als Schwellwerte gefunden. Der Schwellwert für den Elevationsversatz konnte nicht eindeutig bestimmt werden, da einige Probanden auch bei hohen Fehlerstärken das fehlerbehaftete Signal nicht identifizieren konnten. Die starken Abweichungen zwischen den Probanden lassen keine verlässlichen Angaben bezüglich eines Schwellwertes zu.

Im zweiten Test wurde eine Variante der *Repertory Grid Technique* durchgeführt, um Merkmale aus *SAQI* den Fehlerarten zuzuordnen. Zur Auswertung wurde eine Hauptkomponentenanalyse durchgeführt, um Merkmale zu gruppieren und den Fehlerarten zuordnen zu können. Dabei nahm Aliasing eine Sonderposition ein, da es als einziger Fehlertyp Merkmale wie wie *metallisch* und *kammfilterartig* aufwies. Fehler durch Mikrofonrauschen und Positionierungsfehler weisen größtenteils ähnliche Merkmale auf. Die Fehlerart wirkt sich ebenso auf die Lage der Merkmale im durch die *Hauptkomponentenanalyse* reduzierten Komponentenraum aus. Ein unnatürlicher Klang wird beispielsweise bei Mikrofonrauschen und Positionierungsfehlern einer Anhebung des tieffrequenten Bereiches zugeordnet, wohingegen die Anhebung des hochfrequenten

Bereiches bei Aliasing zu einem unnatürlicheren Klang beiträgt. Die Hauptkomponentenanalyse wurde mit geordneten Elementen durchgeführt, da die Reihenfolge der Probanden keine Rolle für die Auswertung spielt. Sämtliche Merkmale können somit zu 80 – 100% (je nach Fehlerart) durch zwei Hauptkomponenten abgebildet werden.

7.2 Ausblick

Die Schwellwertmessung könnte in späteren Forschungsarbeiten weitergeführt werden, um Schwellwerte für den Elevationsversatz zu untersuchen oder den Einfluss von veränderten Konfigurationen des Mikrofonarrays auf die Schwellwerte zu überprüfen.

Zur Bestimmung von Qualitätsmerkmalen wäre es möglich, andere in dieser Arbeit vorgestellte Methoden zu verwenden, wie zum Beispiel das *Open Profiling of Quality* oder das *Free Choice Profiling* oder die *Quantitative Descriptive Analysis* geeignet.

Ein weiteres Interessengebiet ist die Aufnahme und Verarbeitung, sowie die Untersuchung der Fehler von realen Messdaten.

Eine Überlegung zur weiteren Bearbeitung ist die Durchführung einer unabhängigen Komponentenanalyse, um den Projektionsraum optimal auf die Hauptachsen abzubilden.

Literaturverzeichnis

- [Albrecht et al., 2007] Albrecht, B., Del Galdo, G., Husung, S., Lotze, J., and Schlessinger, A. (2007). Holographic Sound Field Analysis with a Scalable Spherical Microphone-Array. In *Audio Engineering Society 122th Convention*. Audio Engineering Society.
- [Algazi and Duda, 2001] Algazi, V. and Duda, R. (2001). The cipic hrtf database. *Workshop on the Applications of Signal Processing to Audio and Acoustics, IEEE*, (October):99–102.
- [Avni et al., 2013] Avni, A., Ahrens, J., Geier, M., Spors, S., Wierstorff, H., and Rafaely, B. (2013). Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution. *The Journal of the Acoustical Society of America*, 133(5):2711–21.
- [Bacher et al., 2010] Bacher, J., Pöge, A., and Wenzig, K. (2010). *Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren*. Oldenbourg Wissenschaftsverlag, München.
- [Bech et al., 1996] Bech, S., Hamberg, R., Nijenhuis, M., Teunissen, K., Looren de Jong, H., Houben, P., Pramanik, S. K., and SPIE (1996). Rapid perceptual image description (RaPID) method. In Rogowitz, B. E. and Allebach, J. P., editors, *Human Vision and Electronic Imaging*, volume 2657, pages 317–328. SPIE.
- [Bech and Zacharov, 2006] Bech, S. r. and Zacharov, N. (2006). *Perceptual Audio Evaluation - Theory, Method and Application*. Wiley, West Sussex.
- [Beranek, 2010] Beranek, L. (2010). *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*. Springer.
- [Berg and Rumsey, 1999] Berg, J. and Rumsey, F. (1999). Spatial attribute identification and scaling by repertory grid technique and other methods. *Audio Engineering Society 16th International Conference on Spatial Sound*.

- [Berkhout, 1988] Berkhout, A. J. (1988). A Holographic Approach to Acoustic Control. *Audio Engineering Society*, 36:977–995.
- [Bernschütz et al., 2011] Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (2011). SOFiA Sound Field Analysis Toolbox. In *Proceedings of the ICOSA International Conference on Spatial Audio*, Detmold.
- [Billingsley, 1974] Billingsley, J. (1974). An acoustic telescope. *Aeronautical Research Council*, 35(364).
- [Blacodon et al., 1987] Blacodon, D., Caplot, M., and Élias, G. (1987). A source localization technique for helicopter rotor noise. In *Aeroacoustics 11th Conference*, Sunnyvale, CA.
- [Blauert, 1997] Blauert, J. (1997). *Spatial Hearing: the psychophysics of human sound localization*. MIT Press, Cambridge, revised edition.
- [Boone et al., 1995] Boone, M. M., De Vries, D., and Berkhout, A. J. (1995). Sound control (parts 1 and 2). Technical report, Delft University of Technology, Delft.
- [Bosi, 2004] Bosi, M. (2004). Introduction to Digital Audio Coding and Standards. *Journal of Electronic Imaging*, 13(2):399.
- [Brandstein and Ward, 2001] Brandstein, M. and Ward, D. (2001). *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag Berlin/Heidelberg, New York.
- [Breebaart and Faller, 2008] Breebaart, J. and Faller, C. (2008). *Spatial Audio Processing: MPEG Surround and Other Applications*. Wiley-Interscience.
- [Chakraborti and Gibbons, 2011] Chakraborti, S. and Gibbons, J. D. (2011). *Non-parametric Statistical Inference*. Chapman & Hall/CRC Press, Taylor & Francis Group, Boca Raton, 5th edition.
- [Ciba et al., 2009] Ciba, S., Wlodarski, A., and Maempel, H. (2009). WhisPER, A new tool for performing listening tests. *Audio Engineering Society 126th Convention*.
- [Daniel, 2011] Daniel, A. (2011). *Spatial Auditory Blurring and Applications to Multichannel Audio Coding*. PhD thesis, Université Pierre et Marie Curie.

- [Daniel et al., 2003] Daniel, J., Nicol, R., and Moreau, S. (2003). Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. *New York*, convention(5):2764–2778.
- [Deblauwe et al., 2007] Deblauwe, F., Jansen, K., and Robin, M. (2007). Extending the usability of near-field acoustic holography and beamforming by using focalization. In *14th International Congress on Sound Vibration*, pages 1–8, Cairns, Australien.
- [Delarue and Sieffermann, 2004] Delarue, J. and Sieffermann, J.-M. (2004). Sensory mapping using flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality and Preference*, 15(4):383–392.
- [Donovan et al., 2008] Donovan, A. O., Duraiswami, R., and Zotkin, D. (2008). Imaging Concert Hall Acoustics Using Visual And Audio Cameras. *IEEE International Conference on Acoustics, Speech and Signal Processing*, (ICASSP):5284–5287.
- [Eichler, 2009] Eichler, M. (2009). *Breitbandige Beamforming-Algorithmen zur Erfassung von Audiosignalen mit kompakten Mikrofon-Arrays*. PhD thesis, Goethe-Universität, Frankfurt am Main.
- [Faller, 2006] Faller, C. (2006). Parametric Multichannel Audio Coding : Synthesis of Coherence Cues. *IEEE*, 14(1):299–310.
- [Fels, 2008] Fels, J. (2008). *From children to adults : how binaural cues and ear canal impedances grow*. PhD thesis, TU Aachen.
- [Fromm, 1995] Fromm, M. (1995). *Repertory-Grid-Methodik: ein Lehrbuch*. Deutscher Studien Verlag, Weinheim.
- [Galdo, 2007] Galdo, G. D. (2007). *Geometry Based Channel Modeling for Multi-user MIMO Systems and Applications*. PhD thesis, TU Ilmenau, Ilmenau.
- [Gerzon, 1973] Gerzon, M. (1973). Periphony: With-height sound reproduction. *Journal of the Audio Engineering Society*.
- [Gfai Tech GmbH, 2013] Gfai Tech GmbH (2013). Acoustic Camera, <http://www.acoustic-camera.com/de/product-ac-pro/mikrofon-arrays/sphere-arrays> (Zugriff am 10.11.2013).

- [Heinz et al., 1999] Heinz, G., Döbler, D., and Nguyen, T. (1999). Acoustic Photo- and Cinematography basing on the H-Interference Transformation (HIT). In *Acoustical Society of America 137th Conference*, TU Berlin, Germany.
- [Hidaka, 1995] Hidaka, T. (1995). Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls. *The Journal of the Acoustical Society of America*, 98(2):988.
- [Howell et al., 1986] Howell, G. P., Bradley, A. J., McCormick, M. A., and Brown, J. D. (1986). De-dopplerization and acoustic imaging of aircraft flyover noise measurements. *Journal of Sound and Vibration*, 105:151–167.
- [Hynninen and Zacharov, 1999] Hynninen, J. and Zacharov, N. (1999). GuineaPig - A Generic Subjective Test System for Multichannel Audio. In *Audio Engineering Society 106th Convention*. Audio Engineering Society.
- [Hyvärinen and Oja, 2000] Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks : the official journal of the International Neural Network Society*, 13(4-5):411–30.
- [ITU-R, 1994] ITU-R, R. (1994). BS. 1116, Methods for the subjective assessment of small impairments in audio systems including multi-channel sound systems. *International Telecomm Union, Radiocommunication Assembly*, pages 1–26.
- [ITU-R, 1998] ITU-R, R. (1998). BT. 500-8, Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union, Radio Communication Assembly*.
- [ITU-R, 2003] ITU-R, R. (2003). BS. 1534-1, Method for the subjective assessment of intermediate quality level of coding systems. *International Telecommunications Union Radiocommunication Assembly*.
- [Jolliffe, 2002] Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, New York.
- [Jumisko-Pyykkö et al., 2007] Jumisko-Pyykkö, S., Häkkinen, J., and Nyman, G. (2007). Experienced quality factors: qualitative evaluation approach to audiovisual quality. In *Multimedia on Mobile Devices*, pages 6507–21, San Jose.
- [Kelly, 1955] Kelly, G. A. (1955). *The Psychology of Personal Constructs: Vol 1 and 2*. WW Norton, New York.

- [King and Bechert., 1979] King, W. F. and Bechert., D. (1979). On the sources of way-side noise generated by highspeed trains. *Journal of Sound and Vibration*, 66:311–332.
- [Kuttruff, 2009] Kuttruff, H. (2009). *Room acoustics*. Spon Press, London and New York, 5th edition.
- [Lacatis et al., 2008] Lacatis, R., Gimenez, A., Barba Sevillano, A., Cerda, S., Romero, J., and Cibrian, R. (2008). Historical and chronological evolution of the concert hall acoustics parameters. In *Acoustics'08*, pages 2151–2156, Paris.
- [Leek, 2001] Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & psychophysics*, 63(8):1279–92.
- [Levitt, 1971] Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B):467.
- [Lieberman and Pentland, 1982] Lieberman, H. R. and Pentland, A. P. (1982). Microcomputer-based estimation of psychophysical thresholds: The Best PEST. *Behavior Research Methods & Instrumentation*, 14(1):21–25.
- [Lilliefors, 1967] Lilliefors, H. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62:399–402.
- [Lindau et al., 2014] Lindau, A., Erbes, V., Lepa, S., Maempel, H.-j., Brinkman, F., and Weinzierl, S. (2014). A spatial audio quality inventory for virtual acoustic environments (SAQI). In *Proc. of the EAA Joint Symposium on Auralisation and Ambisonics*, number April, pages 3–5, Berlin. Audio Communication group, TU Berlin.
- [Loomis et al., 1998] Loomis, J. M., Klatzky, R. L., Philbeck, J. W., and Golledge, R. G. (1998). Assessing auditory distance perception using perceptually directed action. In *Perception & Psychophysics*, volume 60, pages 966–980.
- [Lorho, 2005] Lorho, G. (2005). Individual Vocabulary Profiling of Spatial Enhancement Systems for Stereo Headphone Reproduction. In *Audio Engineering Society 119th Convention*, page 19, New York. Audio Engineering Society.

- [Macpherson and Middlebrooks, 2002] Macpherson, E. a. and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, 111(5):2219.
- [McCowan, 2001] McCowan, I. (2001). *Robust Speech Recognition using Microphone Arrays*. PhD thesis, Queensland University of Technology, Australien.
- [Melchior, 2011] Melchior, F. (2011). *Investigations on spatial sound design based on measured room impulse responses*. PhD thesis, Delft.
- [Melchior et al., 2009] Melchior, F., Thiergart, O., and Galdo, G. D. (2009). Dual radius spherical cardioid microphone arrays for binaural auralization. *Audio Engineering Society 127th Convention*.
- [Meyer and Elko, 2002] Meyer, J. and Elko, G. (2002). A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In *IEEE International Conference on Acoustics Speech and Signal Processing*, volume 2, pages II-1781–II-1784. IEEE.
- [Michel, 2006] Michel, U. (2006). History of acoustic beamforming. In *Proceedings of the Berlin Beamforming Conference*, pages 1–17, Berlin, Germany.
- [Michel et al., 2004] Michel, U., Barsikow, B., Böhning, P., and Hellmig, M. (2004). Localisation of moving sound sources with phased microphone arrays. In *Proceedings Inter-Noise*, Prague, Czech Republic.
- [Moller, 1992] Moller, H. (1992). Fundamentals of binaural technology. *Applied Acoustics*, 36(3-4):171–218.
- [Morimoto and Maekawa, 1988] Morimoto, M. and Maekawa, Z. (1988). Effects of Low Frequency Components on Auditory Spaciousness. *Acta Acustica united with Acustica*, 66(4):190 – 196.
- [Möser, 2005] Möser, M. (2005). *Technische Akustik*. VDI-Buch. Springer-Verlag, Berlin/Heidelberg.
- [Möser, 2009] Möser, M. (2009). *Messtechnik der Akustik*. Springer.
- [Paschereit and Barsikow, 1994] Paschereit, C. O. and Barsikow, B. (1994). The microphone array - a tool on the path towards reducing railway noise. In *World Congress on Railway Research*, pages 371–376, Paris. World Congress on Railway Research.

- [Pearson, 1901] Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.
- [Radun et al., 2008] Radun, J., Leisti, T., and Häkkinen, J. (2008). Content and quality: interpretation-based estimation of image quality. *Transactions on Applied Perception*, 4(4):1–5.
- [Rafaely, 2005] Rafaely, B. (2005). Analysis and design of spherical microphone arrays. *IEEE Trans. Speech and Audio Processing*, 13(1):135–143.
- [Rafaely et al., 2007] Rafaely, B., Weiss, B., and Bachmat, E. (2007). Spatial aliasing in spherical microphone arrays. *IEEE Trans. Signal Processing*, 55(3):1003–1010.
- [Rumsey, 2001] Rumsey, F. (2001). *Spatial Audio*. Focal Press, Oxford.
- [Sabine, 1922] Sabine, W. C. (1922). *Collected Papers on Acoustics*. Harvard University Press, Harvard.
- [Schuitman, 2011] Schuitman, J. v. D. (2011). *Auditory modelling for assessing room acoustics*. PhD thesis, TU Delft.
- [Silzle, 2007] Silzle, A. (2007). Quality Taxonomies for Auditory Virtual Environments. In *Audio Engineering Society 122th Convention*, page preprint no. 6993, Vienna.
- [Strohmeier et al., 2010] Strohmeier, D., Jumisko-Pyykkö, S., and Kunze, K. (2010). Open Profiling of Quality: A Mixed Method Approach to Understanding Multimodal Quality Perception. *Advances in Multimedia*, 2010:1–28.
- [Teutsch, 2007] Teutsch, H. (2007). *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*. Springer.
- [Thiergart, 2007] Thiergart, O. (2007). *Sound Field Analysis on the Basis of a Spherical Microphone Array for Auralization Applications*. Diplomarbeit, TU Ilmenau.
- [Tretwein, 1995] Tretwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, 35(17):2503–2522.
- [Weinzierl, 2008] Weinzierl, S. (2008). *Handbuch der Audiotechnik*. Springer, Berlin Heidelberg.

-
- [Williams, 1999] Williams, E. G. (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press.
- [Zacharov and Koivuniemi, 2001a] Zacharov, N. and Koivuniemi, K. (2001a). Audio descriptive analysis & mapping of spatial sound displays. In *International Conference on Auditory Display*, Espoo.
- [Zacharov and Koivuniemi, 2001b] Zacharov, N. and Koivuniemi, K. (2001b). Unravelling the Perception of Spatial Sound Reproduction: Analysis & External Preference Mapping. In *Audio Engineering Society Convention 111*. Audio Engineering Society.
- [Zotter and Zaunschirm, 2014] Zotter, F. and Zaunschirm, M. (2014). Measurement Based modal beamforming using planar circular microphone arrays. In *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, number April, pages 3–5, Berlin, Germany.

Abbildungsverzeichnis

3.1	PsychometricFunktion	20
3.2	Staircase Methoden (Bilder rekonstruiert nach [Levitt, 1971]).	21
3.3	PEST Methode	22
3.4	PEST Methode	22
3.5	Koordinatensystem	24
3.6	Differenzen	25
3.7	Duplextheorie	25
3.8	HRTF	27
3.9	Richtungsbaender	27
3.10	interaurale Korrelation	29
3.11	EbeneWelle	30
3.12	Kugelwellen	31
3.13	Reflexions	32
3.14	Schallbestandteil	33
3.15	Kugelarray	34
3.16	Quadratures	35
3.17	SphericalHarmonics	36
3.18	Modenstärken	38
3.19	Spatial Aliasing	39
3.20	Richtungsabhängigkeit	40
3.21	Rauschen und Aliasing auf einem Kugelarray simuliert durch <i>SOFiA</i>	41
3.22	Fehlerverteilung	43
3.23	BinauralAuralisation	44
4.1	Abhängigkeit des Rauschanteils im tieffrequenten Bereich vom Radius für $r=0,25m$ und $r=0,005 m$ bei 1730 Samplingpunkten und der Ordnung $N=5$	46
4.2	Abhängigkeit des Aliasing im hochfrequenten Bereich von der Anzahl der Samplingpunkte (974, 4802) für die Ordnung $N=10$ und Radius $r=2m$	47

4.3	Mikrofonanordnung für 350 Samplingpunkteim normierten dreidimensionalen Raum. Rote Kreuze stellen die korrekte Position dar, grüne Kreise weisen eine Abweichung von $\sigma = 1^\circ$ auf.	48
5.1	Schwellwerthistogramme für Aliasing. Die Mode ist Rot markiert und stellt den Schwellwert dar.	55
5.2	Schwellwerthistogramme für Mikrofonrauschen. Die Mode ist Rot markiert und stellt den Schwellwert dar.	56
5.3	Schwellwerthistogramme für eine Azimuthverschiebung. Die Mode ist Rot markiert und stellt den Schwellwert dar.	57
5.4	Schwellwerthistogramme für eine Elevationsverschiebung. Die Mode ist Rot markiert und stellt den Schwellwert dar.	57
6.1	Ablauf der PCA	63
6.2	Boxplot für Musik bei Aliasing	64
6.3	Boxplots für Natürlichkeit und metallischer Klangfarbe für Sprache über alle Fehler.	69
6.4	Biplot für Musik bei Aliasing	70
6.5	Biplot für Musik bei Rauschen	71
A.1	Boxplot für Musik bei Rauschen	87
A.2	Boxplot für Musik bei Azimuthfehler	88
A.3	Boxplot für Musik bei Elevationsfehler	88
A.4	Biplot für Musik bei Aliasing	90
A.5	Biplot für Sprache bei Aliasing	90
A.6	Biplot für Musik bei Rauschen	91
A.7	Biplot für Sprache bei Rauschen	91
A.8	Biplot für Musik bei Azimuthversatz	92
A.9	Biplot für Sprache bei Azimuthversatz	92
A.10	Biplot für Musik bei Elevationsversatz	93
A.11	Biplot für Sprache bei Elevationsversatz	93

Tabellenverzeichnis

2.1	Typische Nachhallzeiten für verschiedene Raumtypen [Boone et al., 1995]	5
2.2	Übersicht über deskriptive Testmethoden	11
2.3	Siebenstufige Vergleichsskala nach der <i>MUSHRA-Methode</i> (links) und fünfstufige Wertminderungsskala nach <i>ITU-R</i> (rechts) [Bosi, 2004]. . . .	12
2.4	Vor- und Nachteile von Beamforming Techniken [McCowan, 2001] . . .	16
3.1	Tabelle aller Qualitätsmerkmale des Spatial Audio Quality Inventory. .	18
4.1	Verwendete Konfigurationen für die Erstellung der Audiosignale mittels <i>SOFiA</i>	48
5.1	Abstufungen der einzelnen Fehlersignale.	53
5.2	Übersicht der erfassten Schwellwerte.	54
6.1	Verwendete Abstufungen der Repertory Grid Signale.	59
6.2	Häufigkeit gewählter Merkmale.	62
6.3	Signifikanztest für Aliasing. Die erste Wertezeile stellt die berechnete Signifikanz dar. Wenn diese unter dem Signifikanzniveau von 5% liegt, wird die Nullhypothese (Merkmal wird als Referenz gedeutet) abgewie- sen und die zweite Zeile ergibt den Wert 1. Alle Merkmale mit dem Wert 1 finden demnach in der <i>PCA</i> Verwendung.	65
6.4	Zuordnung der Merkmale zu den einzelnen Fehlern durch den Signifi- kanztest.	67
6.5	Eigenwerte und Varianzen für Aliasing.	68
A.1	Eigenwerte und Varianzen aller Fehler	89

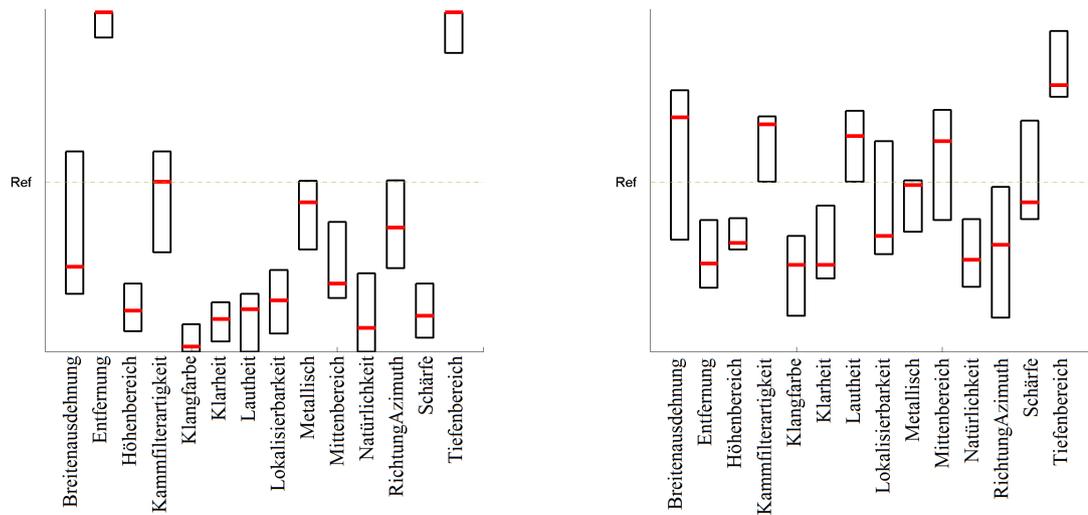
Verzeichnis der verwendeten Abkürzungen

ADAM	Audio Descriptive Analysis and Mapping
AFC	Alternativ Forced Choice
ASW	Apparent Source Width (wahrgenommene Schallquellenbreite)
BR	Bass Ratio (Bassverhältnis)
bzw.	beziehungsweise
ERB	Equivalent Rectangular Bandwidth
FCP	Free Choice Profiling
FP	Flash Profiling
FS	Sampling Frequenz
HRIR	Head Related Impulse Responses (Kopfbezogene Impulsantworten)
HRTF	Head Related Transfer Function (Kopfbezogene Übertragungsfunktion)
HOA	High Order Ambisonics
IACC	Interauraler Korrelationskoeffizient
IC	Interaural Correlation (Interaurale Korrelation)
IID	Interaural Intensity Difference (Interaurale Intensitätsdifferenz)
ILD [ICLD]	Interaural [Channel] Level Difference (Interaurale [Kanal-] Pegeldifferenz)
ITD [ICTD]	Interaural [Channel] Time Difference (Interaurale [Kanal-] Laufzeitdifferenz)
ITDG	Initial Time Delay Gap
LEV	Listener Envelopment
OPQ	Open Profiling of Quality
PCA	Principal Component Analysis
PEST	Parameter Estimation by sequential Testing
QDA	Quantitative Descriptive Analysis
RGT	Repertory Grid Technique
SAQI	Spatial Audio Quality Inventory
SHD	Sphärisch Harmonische Domäne
SNR	Signal-Rausch Abstand
SOFiA	Sound Field Analysis Toolbox
TR	Treble Ratio (Höhenverhältnis)
VAE	Virtual Acoustic Environment
WFS	Wellenfeldsynthese

Verzeichnis der verwendeten Symbole

α	Schallabsorptionskoeffizient
α_j	Mikrofongewichtung
b_n	Modenstärke
c	Schallausbreitungsgeschwindigkeit (343 m/s)
C_{50}, C_{80}	Deutlichkeit, Klarheit
$\epsilon_a, \epsilon_\Omega$	Aliasingfehler, Positionierungsfehler
e_j	Rauschanteil
E	Energie
f (Hz)	Frequenz in Herz
Δf	Frequenzunterschied
f_{nm}	Sphärische Fouriertransformation
h_n	sphärische Hankelfunktion
j_n	sphärische Besselfunktion
H^l	HRTF des linken Ohres
G_i	Varianzanteil der i-ten Hauptkomponente
k	Wellenzahl
K	Kompressionsmodul
λ	Eigenwert (PCA Berechnung)
p	Schalldruckpegel, Druck
ρ	Massendichte
r	Radius des Mikrofonarrays
σ	Standardabweichung (der Fehlerverteilung bei der Positionierung)
s	Scattering Coefficient
S	Wandfläche des Raumes
t	Zeit
T_{60}	Nachhallzeit
RT_f	Nachhallzeit für Frequenzband f
$\Omega=(\varphi, \vartheta)$	Azimuthwinkel, Elevationswinkel
v	Partikelgeschwindigkeit
V	Raumvolumen
Δx	Mikrofonabstand
X_{50}	Schwellwertpunkt für 50 % positive Antworten
Y_m^m	Funktion der sphärisch Harmonischen

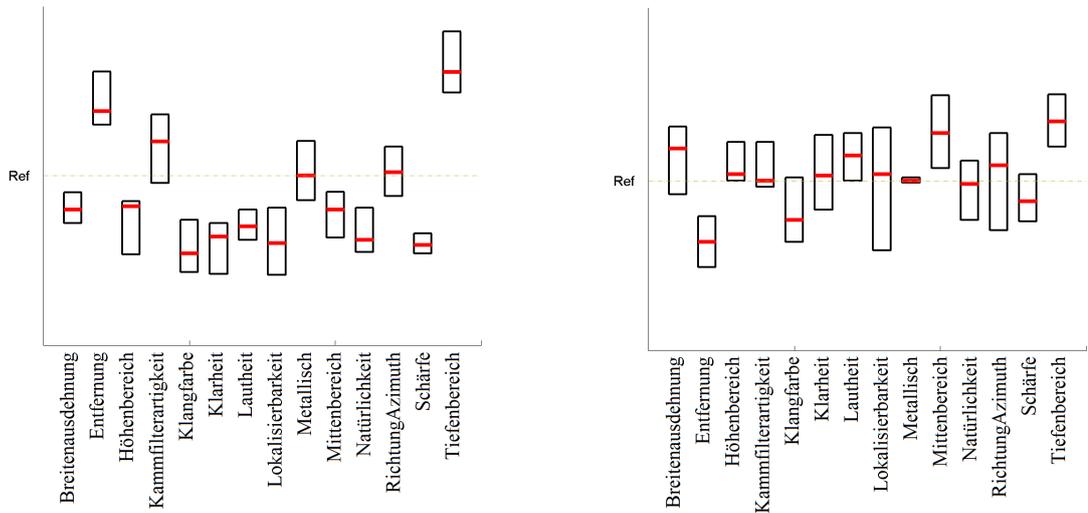
A Anhang



(a) Boxplots der Merkmale mit Rauschanteil für Musik

(b) Boxplots der Merkmale mit Rauschanteil für Sprache

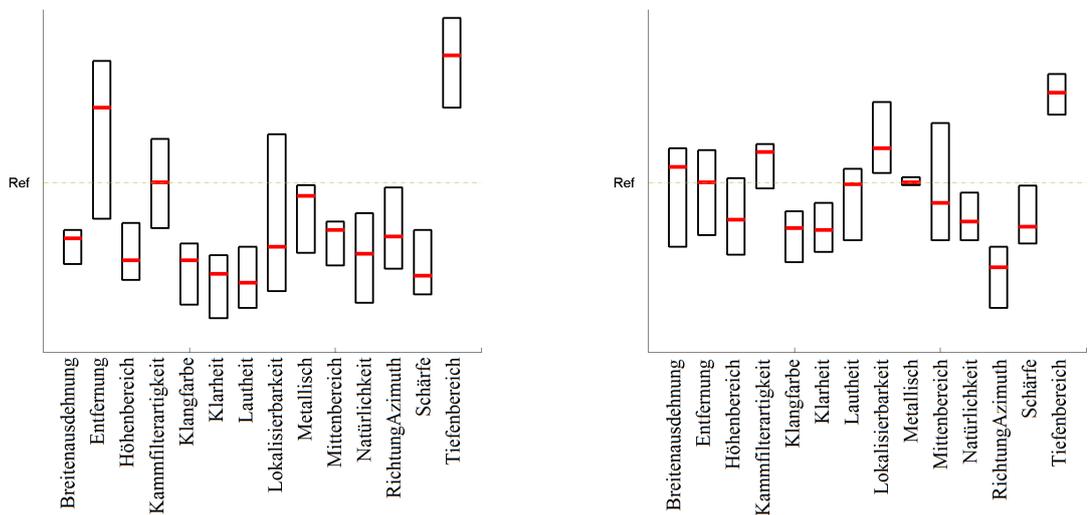
Abbildung A.1: Boxplots der Merkmale mit Rauschanteil für Musik und Sprache. Es werden das obere und untere Quartil sowie der Median dargestellt



(a) Boxplots der Merkmale mit Azimuthversatz für Musik

(b) Boxplots der Merkmale mit Azimuthversatz für Sprache

Abbildung A.2: Boxplots der Merkmale mit Azimuthversatz für Musik und Sprache. Es werden das obere und untere Quartil sowie der Median dargestellt



(a) Boxplots der Merkmale mit Elevationsversatz bei Musik

(b) Boxplots der Merkmale mit Elevationsversatz bei Sprache

Abbildung A.3: Boxplots der Merkmale mit Elevationsversatz für Musik und Sprache. Es werden das obere und untere Quartil sowie der Median dargestellt

Tabelle A.1: Eigenwerte und Varianzen aller Fehler

Fehler	C	λ	erklärte Varianz (%)	kumulierte Varianz (%)	λ	erklärte Varianz (%)	kumulierte Varianz (%)
		Musik			Sprache		
Aliasing	1	4.69	78.13	78.13	6.15	68.34	68.34
	2	0.85	14.12	92.24	1.24	13.78	82.12
	3	0.27	4.47	96.71	0.83	9.27	91.39
	4	0.12	2.02	98.73	0.41	4.5	95.89
	5	0.06	0.99	99.72	0.23	2.51	98.4
	6	0.02	0.28	100	0.12	1.3	99.7
	7				0.02	0.22	99.92
	8				0.01	0.06	99.99
	9				0	0.01	100
Rauschen	1	7.87	78.75	78.75	3.75	74.94	74.94
	2	0.99	9.87	88.62	1.03	20.53	95.47
	3	0.67	6.7	95.33	0.16	3.25	98.71
	4	0.24	2.37	97.7	0.05	1.07	99.78
	5	0.11	1.07	98.77	0.01	0.22	100
	6	0.06	0.59	99.36			
	7	0.04	0.37	99.74			
	8	0.02	0.2	99.94			
	9	0	0.05	99.98			
	10	0	0.02	100			
Azimuth	1	4.72	58.98	58.98	2.86	95.37	95.37
	2	2.09	26.1	85.08	0.11	3.64	99.01
	3	0.55	6.88	91.96	0.03	0.99	100
	4	0.32	3.95	95.91			
	5	0.15	1.89	97.79			
	6	0.15	1.81	99.61			
	7	0.02	0.31	99.92			
	8	0.01	0.08	100			
Elevation	1	7.57	75.67	75.67	4.39	87.84	87.84
	2	1.04	10.36	86.02	0.41	8.28	96.12
	3	0.58	5.79	91.81	0.11	2.29	98.41
	4	0.43	4.32	96.13	0.07	1.41	99.82
	5	0.21	2.05	98.18	0.01	0.18	100
	6	0.1	1.02	99.2			
	7	0.04	0.45	99.65			
	8	0.02	0.17	99.82			
	9	0.01	0.1	99.91			
	10	0.01	0.09	100			

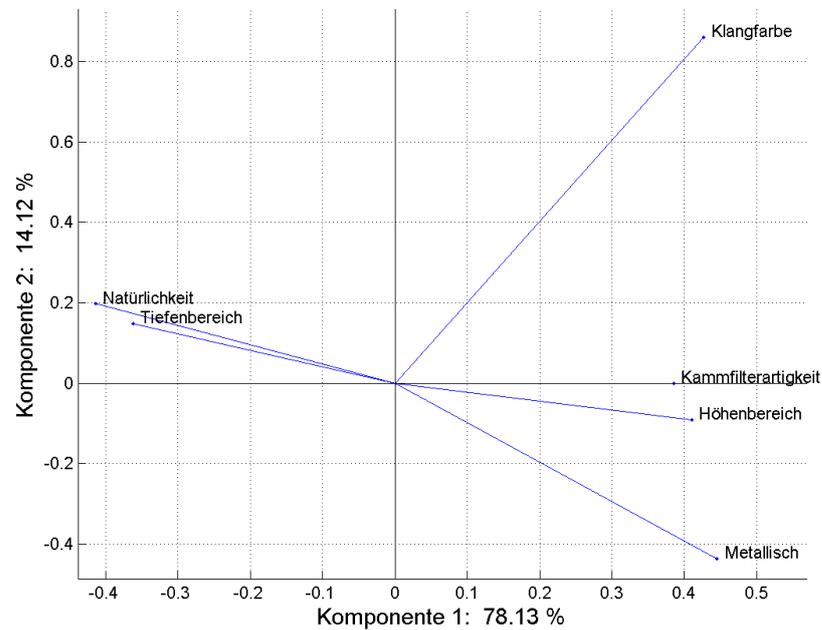


Abbildung A.4: Biplot für Musik mit Aliasingfehler. Natürlichkeit und Tiefenbereich bilden eine Gruppe, sowie Kammfilterartigkeit, Höhenbereich und metallische Klangfarbe

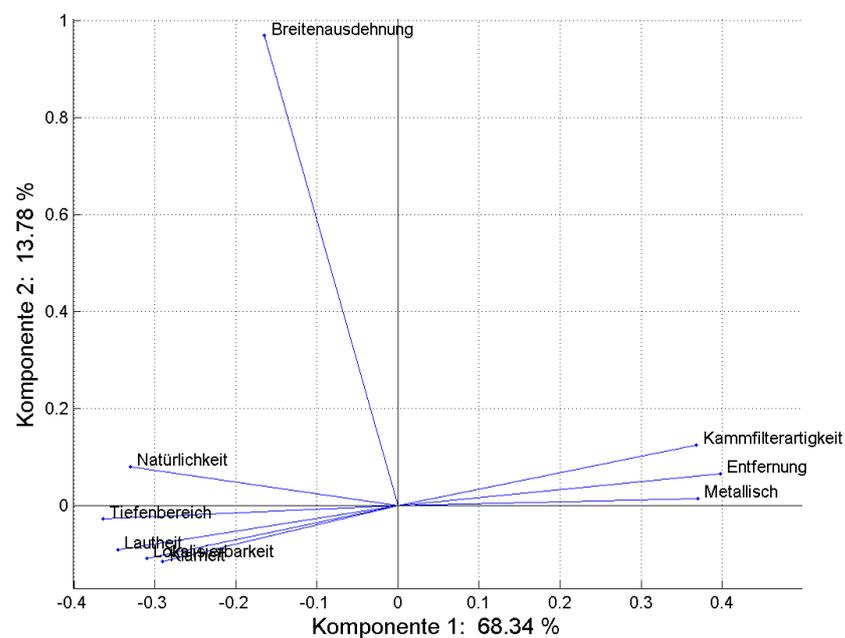


Abbildung A.5: Biplot für Sprache mit Aliasingfehler. Teilung der Merkmale in zwei Gruppen, die Breitenausdehnung liegt als einziges Merkmal separat

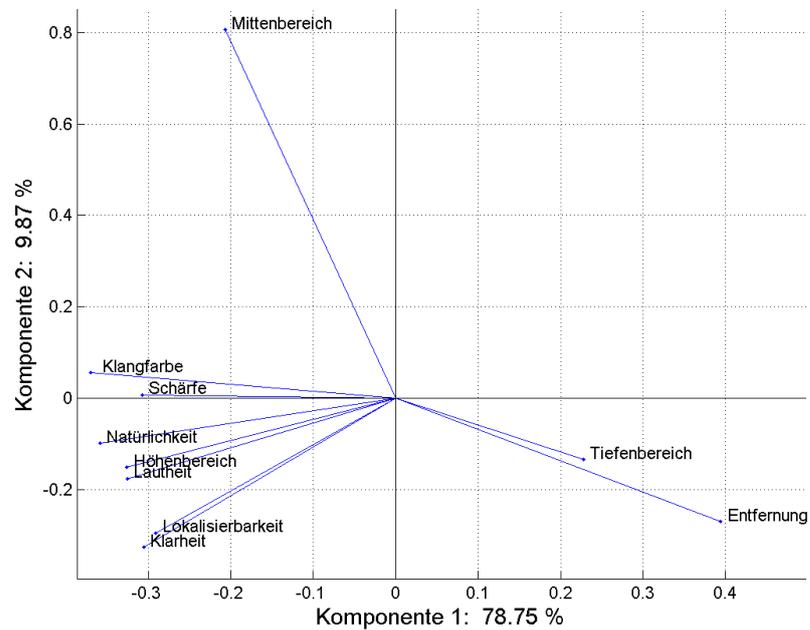


Abbildung A.6: Biplot für Musik mit Rauschanteil. Tiefenbereich und Entfernung bilden eine Gruppe. Der Mittbereich liegt separat. Die linke Merkmalsgruppe beinhaltet den Großteil der Merkmale

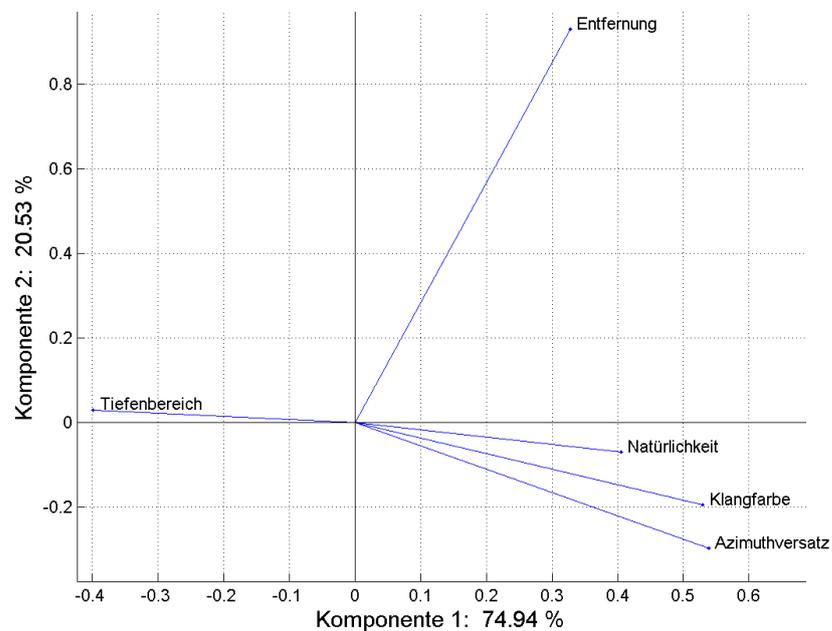


Abbildung A.7: Biplot für Sprache mit Rauschanteil. Natürlichkeit, Klangfarbe und Azimuthversatz bilden eine Gruppe. Die Entfernung und der Tiefenbereich sind separiert

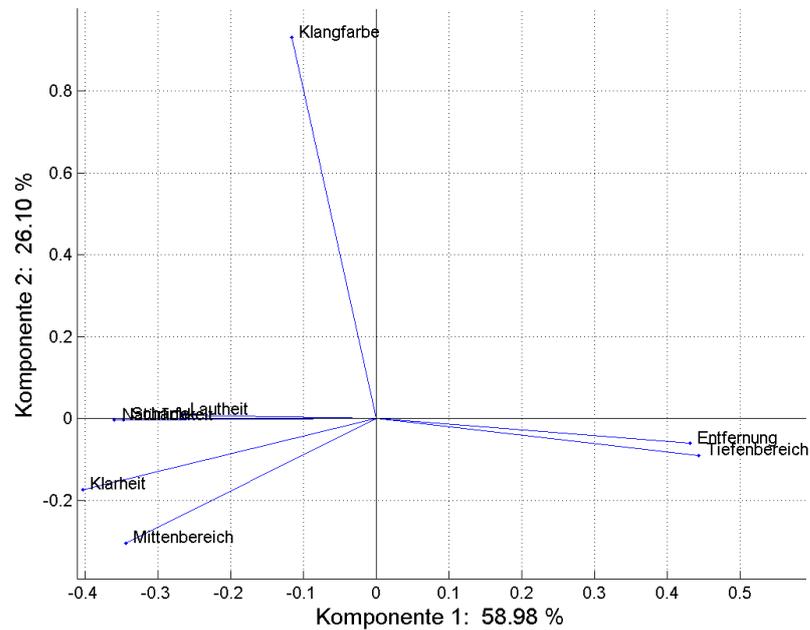


Abbildung A.8: Biplot für Musik mit Azimuthversatz. Wie bei Musik mit Rauschanteil fällt hier ebenfalls die Entfernung mit dem Tiefenbereich zusammen

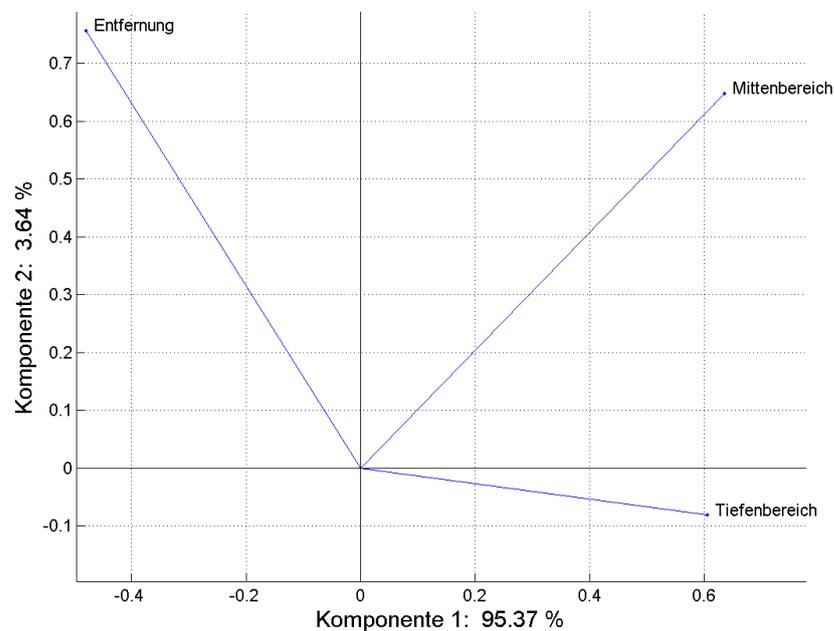


Abbildung A.9: Biplot für Sprache mit Azimuthversatz. Es wurden zu wenig Merkmale als signifikant eingestuft, um eine Gruppenbildung feststellen zu können

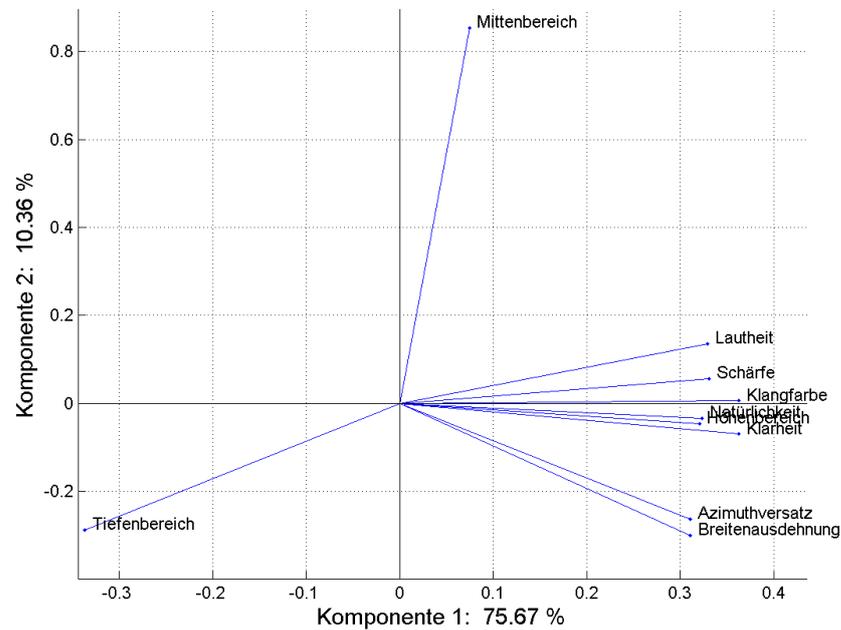


Abbildung A.10: Biplot für Musik mit Elevationsversatz. Azimuthverschiebung und Breitenausdehnung bilden eine Gruppe. Mitten und Tiefen sind separiert. Eine zweite Gruppe beinhaltet weitere Merkmale

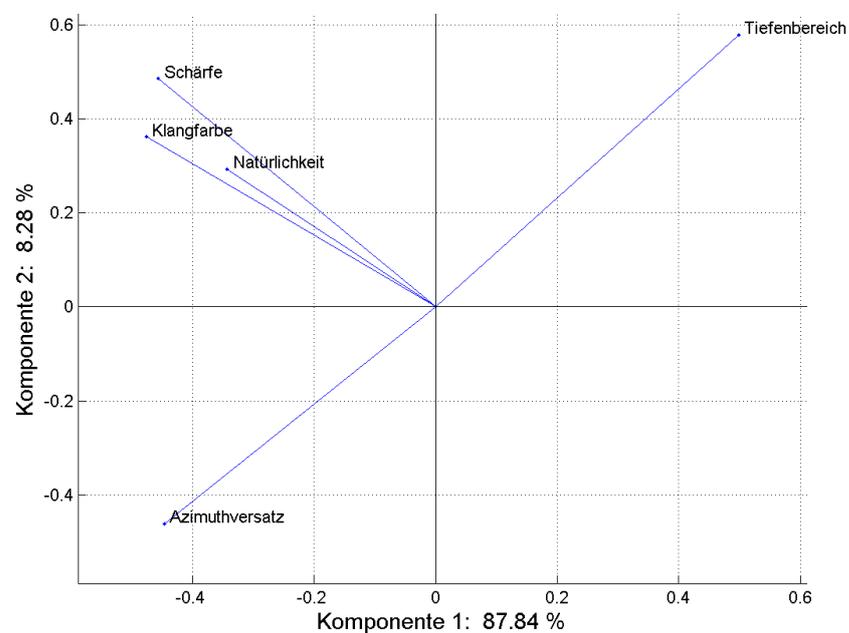


Abbildung A.11: Biplot für Sprache mit Elevationsversatz. Natürlichkeit, Klangfarbe und Schärfe bilden eine Gruppe. Azimuthversatz und Tiefenbereich liegen separat

Eigenständigkeitserklärung

Ich erkläre, dass ich diese Arbeit selbständig durchgeführt und abgefasst habe. Quellen, Literatur und Hilfsmittel, die benutzt wurden, sind als solche gekennzeichnet.

Ilmenau, den 30. 04. 2014

Mathias Hellmich