

Binaural 3D Audio - Die Geschichte und die Zukunft

Bachelorarbeit aus Aufnahmetechnik 1, SE

Luca Langgartner

Betreuung: Univ.Prof. DI Dr. Alois Sontacchi

Graz, 26. Juni 2018



institut für elektronische musik und akustik



EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am

.....

(Unterschrift)

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz,

.....

(signature)

Abstract

This thesis was authored with the intent of presenting the theoretical background, the history and the current state of research in the field of binaural 3D Audio. In addition, in chapter 7, a headphone prototype which was developed at the Institut für Elektronische Musik und Akustik of the University of Music and Performing Arts Graz and the evaluation of a test made with this headphone is presented.

Zusammenfassung

In dieser Arbeit soll die Thematik der dreidimensionalen binauralen Wiedergabe zusammengefasst vorgestellt werden. Dabei wird versucht, neben dem theoretischen Grundwissen, neue wissenschaftliche Erkenntnisse vorzustellen um somit die Gedanken für zukünftige Entwicklungen anzuspornen. Im vorletzten Kapitel wird auf den am Institut für Elektronische Musik und Akustik der Universität für Musik und Darstellende Kunst Graz entwickelten Kopfhörerprototypen eingegangen und einen mit diesem Kopfhörern durchgeführter Test evaluiert.

Inhaltsverzeichnis

1	Einleitung	6
2	Räumliches Hören	7
2.1	Das menschliche Gehör	7
2.2	Lokalisation	8
2.2.1	Das kopfbezogenen Koordinatensystem	8
2.2.2	Ipsilaterales und kontralaterales Ohr	9
2.2.3	Minimal Audible Angle (MMA)	9
2.2.4	Lokalisation und Lateralisation	9
2.3	Faktoren zur Bestimmung der Schallquellenposition	9
2.3.1	Monaurale Faktoren: Lokalisation in der vertikalen Ebene	9
2.3.2	Binaurale Faktoren	10
3	Die Außenohrübertragungsfunktion (HRTF) und die binaurale Impulsantwort (HRIR)	12
3.1	Theoretischer Hintergrund	12
3.2	Messung von Außenohrübertragungsfunktionen	14
3.3	Berechnung von Außenohrübertragungsfunktionen	15
3.4	Lokalisationsqualität verschiedener HRTFs	16
4	Dreidimensionale Audiotechniken	16
4.1	Stereo	16
4.2	Surround Sound	17
4.3	Hemisphärisches Surround	17
4.4	Ambisonics	18
4.5	Objektbasiertes Audio	18
5	Binaurales 3D Audio über Kopfhörer	18
5.1	Externalisation	19
5.2	Kopfpositionsverfolgung	19
5.3	Verbesserung der Lokalisationsqualität durch Zufügung von Hall	20
5.4	Nicht individualisierte Binauralsynthese	20
6	Binaurales 3D Audio über Lautsprecher	21

	5
6.1 Übersprechkompensation	21
6.2 Transaural Audio	22
7 Evaluierung der Testergebnisse des am IEM entwickelten Kopfhörer Proto- typ	23
7.1 Qualitative Auswertung der Testergebnisse	24
7.2 Quantitative Auswertung der Testergebnisse	25
8 Schlussfolgerung und Ausblick	28

1 Einleitung

Klang kann durch das menschliche Gehör gleichzeitig von allen Richtungen wahrgenommen werden und somit kann, im Gegensatz zu anderen Sinnesorganen, leicht ein immersives Wahrnehmungserlebnis erzeugt werden. In diesem Zusammenhang ist die Binauraltechnik eines der wichtigsten Mittel um ein solches Erlebnis zu schaffen. In dieser Arbeit wird die Thematik der binauralen Aufnahme- und Wiedergabe- und Synthetisierungstechnik und alle diesbezüglichen relevanten Problemstellungen erfasst. Dabei wurde versucht neueste wissenschaftliche Erkenntnisse einzubeziehen und diesen einen gerechten theoretischen Hintergrund zu verleihen. In Kapitel 2 wird der theoretische Hintergrund über das räumliche Hören vorgestellt und jene Faktoren die zur Schallquellenlokalisierung beitragen behandelt. Im dritten Kapitel wird die Außenohrübertragungsfunktion -jene Übertragungsfunktion welche die binaurale Enkodierung ermöglicht-, deren Messung und Berechnung vorgestellt. Im darauf folgenden Kapitel werden 3D Audio Techniken und deren Einsatz bei binauraler Wiedergabe erfasst. Die Wiedergabe über Kopfhörer und Lautsprecher und die diesbezüglichen Problematiken werden in den nächsten zwei Kapiteln vorgestellt. In Kapitel 7 erfolgt die Evaluierung eines mit dem am Institut für Elektronische Musik und Akustik entwickelten Kopfhörerprototypen durchgeführten Tests. Kapitel 8 stellt eine Zusammenfassung der Arbeit, samt Schlussfolgerung und Ausblick auf mögliche zukünftige Entwicklungen, dar.

2 Räumliches Hören

2.1 Das menschliche Gehör

Das menschliche Ohr kann nach [Xie13] in drei Teile aufgeteilt werden: Außenohr, Mittelohr und Innenohr. Das Außenohr setzt sich aus der Ohrmuschel, oft auch Pinna genannt, und dem äußerem Gehörgang zusammen. Die Form und die Dimension der Ohrmuschel sind bei jedem Menschen verschieden, dabei beträgt die Durchschnittslänge 65 mm. Der Durchmesser des äußeren Gehörgangs beträgt durchschnittlich 7 mm und die Länge 27 mm [MS04]. Die Form der Pinna hat einen sehr starken Einfluss auf eintretende Schalleignisse, überhaupt im hochfrequenten Bereich, und mitsamt des Gehörgangs kommt es zu Resonanzmoden in diesem Bereich. Diese Effekte spielen eine sehr wichtige Rolle in der Lokalisation von hohen Frequenzen. Durch diese Resonanzmoden ist außerdem ersichtlich, dass der Mensch sensibler im Bezug zu Frequenzen im Bereich zwischen 2 und 5 kHz ist.

Das Mittelohr besteht aus Trommelfell, Paukenhöhle, Gehörknöchelchen und Eustachischer Röhre. Das Mittelohr hat die Aufgabe eines Impedanzwandlers: Die Luftschwingungen, die am Trommelfell anliegen, werden über die Gehörknöchelchen, welche diesen mit dem Ovalen Fenster der Cochlea verbinden, in mechanische Schwingungen umgewandelt. Grund für diese Impedanzwandlung ist die mit Flüssigkeit gefüllte Cochlea. Durch die Flächenunterschiede zwischen Trommelfell und Ovalen Fenster und durch die Hebelwirkung der Gehörknöchelchen kommt es zu einer äquivalenten akustischen Impedanz in Luft und in Flüssigkeit.

Das Innenohr besteht hauptsächlich aus der Cochlea, wegen seiner Form auch Gehörschnecke genannt, welche die Aufgabe hat die mechanischen Schwingungen in elektrische Impulse für das Gehirn umzuwandeln. Zwei Membranen, die Basilarmembran und die Reissner-Membran, teilen die Cochlea in drei flüssigkeitsgefüllte Gänge auf, wobei die zwei äußeren durch einen kleinen Fenster, dem Helicotrema, verbunden sind. Die akustischen Impulse werden in mechanische Impulse in Form von ebenen Wellen entlang der Basilarmembran umgewandelt. Auf der Basilarmembran befindet sich das Corti-Organ mit vier Reihen Haarzellen. Durch diese Haarzellen werden die mechanischen Schwingungen in Nervenimpulse umgewandelt. Die Cochlea wirkt als Frequenzanalysator, da die mechanischen Eigenschaften der Basilarmembran entlang dessen Länge, durch den Einfluss der äußeren Haarzellen, sehr unterschiedlich ist und deswegen die Antwort der Basilarmembran auf eintretende Impulse

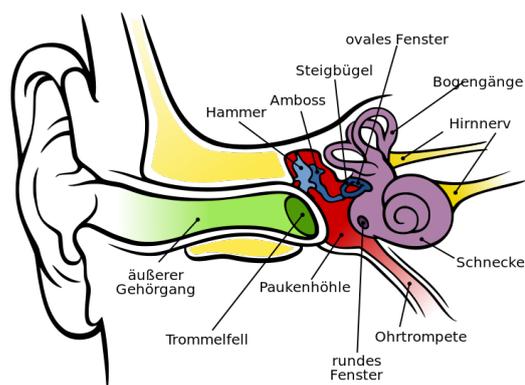


Abbildung 1 – Der Aufbau des menschlichen Ohres. [Han16]

stark frequenzabhängig ist. Somit kommt es z.B. bei einem Ton zu einer starken Aussteuerung der Membran an einem bestimmten Punkt. Tiefe Frequenzen werden in der Nähe des Apex der Cochlea dargestellt, hohe Frequenzen hingegen in Basisnähe, wobei Recherchen zeigen, dass bei Frequenzen über 500 Hz die lineare Distanz zwischen dem Apex und der maximalen Auslenkung proportional der logarithmischen Frequenz wird. Dadurch ergibt sich eine bessere Frequenzauflösung im Bereich bis 500 Hz [Xie13].

2.2 Lokalisation

Neben der Wahrnehmung von Tonhöhen, Lautstärken und Klangfarben, kann durch das menschliche Gehör auch die Position einer Schallquelle im Raum erfasst werden. Um dieses Phänomen begreifen zu können, werden zuvor wichtige Begriffe und Theorien klargestellt:

2.2.1 Das kopfbezogenen Koordinatensystem

Mit *Horizontalebene* wird die waagrechte Ebene, welche durch beide Ohren verläuft, verstanden. Hingegen wird mit *Frontalebene* jene vertikale Ebene beschrieben welche ebenfalls durch beide Ohren verläuft. Die *Medianebene* ist jene senkrechte Ebene im Raum, welche durch die Kopfmittle verläuft und somit gleichen Abstand zwischen linkem und rechtem Ohr vorweist.

Der *Elevationswinkel* ϕ beschreibt den Winkel zwischen wahrgenommener Schallquellenposition und Horizontalebene in der Vertikalebene. Der *Azimuthwinkel* δ hingegen, beschreibt den Winkel zwischen Quellenposition und Medianebene, in der Horizontalebene.

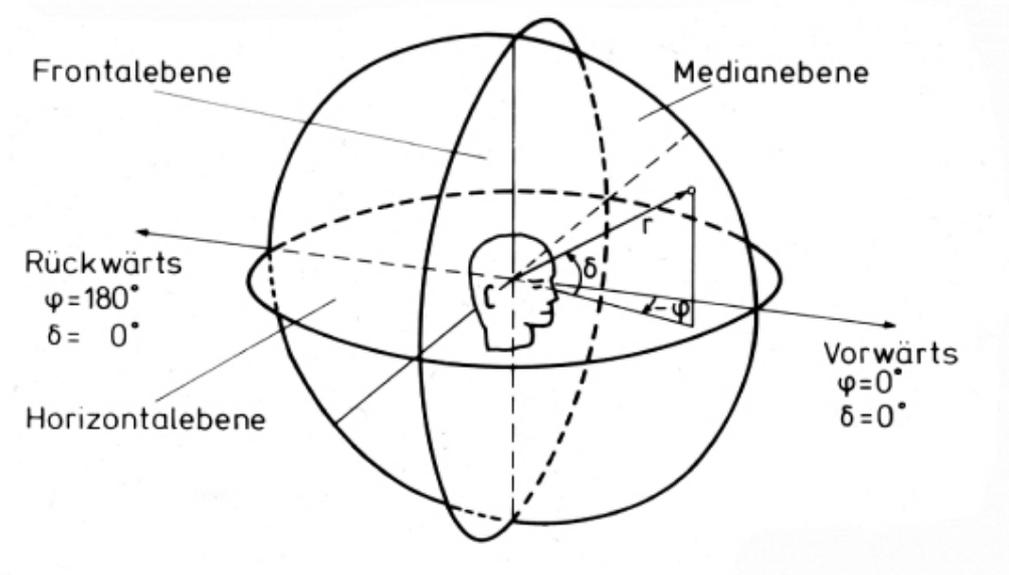


Abbildung 2 – Das kopfbezogene Koordinatensystem.

2.2.2 Ipsilaterales und kontralaterales Ohr

Das *ipsilaterale Ohr* ist das dem Schallereignis zugewandte Ohr, zu dem der Direktschall einen kürzeren Weg zurücklegen muss. Das *kontralaterale Ohr* ist das dem Schallereignis abgewandte Ohr.

2.2.3 Minimal Audible Angle (MMA)

Die menschliche Lokalisationsschärfe ist vor der horizontalen Ebene am höchsten, wobei der *minimal audible angle* (MMA), der kleinst-wahrnehmbare Winkel, $\Delta\phi$ zwischen 1° und 3° in der Horizontalebene liegt. In anderen Regionen nimmt der MMA ab: $\Delta\phi$ ist in seitlichen Richtungen bis zu dreimal und in Rückwärtsrichtung bis zu zweimal so groß. In der Medianebene ist die Lokalisationsfähigkeit des Menschen äußerst gering. Der MMA $\Delta\delta$ liegt hier bei 4° bei weißem Rauschen und bei Sprachsignalen beträgt dieser 17° .

2.2.4 Lokalisation und Lateralisation

Mit Lokalisation versteht man die Zuordnung einer Schallquellenposition im dreidimensionalen Raum. Mit Lateralisation versteht man die Zuordnung der Position auf der Horizontalebene, also um rund um den Hörer.

2.3 Faktoren zur Bestimmung der Schallquellenposition

Bei der Beschreibung der akustischen und psychoakustischen Faktoren zur Bestimmung der Schallquellenposition, wird meist nur eine Schallquelle betrachtet. Dies kann auf die Linearität der Schallfeldgleichung rückgeschlossen werden, wodurch die Überlagerung der durch mehrere Schallquellen entstehenden Signale, sofern diese dekorreliert sind, ermöglicht wird. Die Positionswahrnehmung beruht auf die wahrgenommenen Unterschiede zwischen den Signalen an ipsi- und kontralateralem Ohr (binaurale Faktoren), wodurch hauptsächlich Positionen in der horizontalen Ebene detektiert werden können, und jenen Qualitätsmerkmalen, die an beiden Ohren wahrgenommen werden können (monaurale Faktoren). Die Außenohrübertragungsfunktion beschreibt die Filterwirkung der Ohrmuschel und weiterer Komponente wie Oberkörper und Kopf. Mit HRTF (*Head Related Transfer Function*) wird die Außenohrübertragungsfunktion in der Frequenzdomäne bezeichnet, mit HRIR (*Head Related Impulse Response*) jene in der Zeitdomäne.

2.3.1 Monaurale Faktoren: Lokalisation in der vertikalen Ebene

Bei der Bestimmung der Schallquellenposition in der vertikalen Ebene spielt das monaurale Spektrum, welches durch Interaktion mit der Pinna zustande kommt, eine wichtige Rolle. Es kann tatsächlich ein spektraler Notch, welcher sich von 5 kHz bis 10 kHz bewegt, wenn die Quelle sich von 0° Elevation (direkt vor dem Hörer) nach 90° (über dem Kopf des Hörers) bewegt, beobachtet werden [MOD89]. Wenn sich ein Signal von der

Quelle zum Ohr bewegt, wird es durch Reflektions- und Beugungseffekten verändert. In Bezug auf die Position der Quelle im Raum, kann es zu Filterungen in bestimmten Frequenzbereichen kommen. Der Hörer nutzt dies um eine Richtung der anderen vorzuziehen. Andere monaurale Faktoren werden durch die Beziehung zwischen direktem und reflektiertem Signal gewährleistet. Monaurale Faktoren sind generell unklarer als binaurale Faktoren, da das Hörsystem zuvor die akustischen Merkmale des Signals schätzen muss um diese dann mit dem gefilterten Signalen, welche am Ohr ankommen, vergleichen zu können.

Spektrale Faktoren

Wie schon erwähnt, haben die spektralen Verfärbungen eines Signals, welche durch Reflexionen und Beugungen am Ohr, am Brustkorb und am Kopf zustande kommen, einen großen Einfluss auf die vertikale Lokalisationsfähigkeit und ebenfalls um vorne-hinten Ambiguität zu lösen. Ein frühes vereinfachtes Modell, vgl. [Bat67], zeigt wie direkter und reflektierter Schall zum Eingang des Ohrkanals kommen. Die relative Verzögerung zwischen den beiden Signalen ist somit richtungsabhängig, da andere Einfallsrichtungen die Reflexion von anderen Teilen der Ohrmuscheln implizieren. Durch Interaktion zwischen direktem und reflektiertem Schall kommt es zu spektralen Verfärbungen, welche somit ebenfalls als richtungsabhängig gelten. Dieses Modell kann jedoch nur für Frequenzen oberhalb von 2-3 kHz gelten, da nur in diesem Bereich die Wellenlänge des eintreffenden Schalls klein genug relativ zur Dimension der Ohrmuschel (ca. 65 mm) ist. Ebenso kann die Ohrmuschel, wessen Form sehr unregelmäßig ist, nicht gänzlich als einheitlich reflektierende Fläche aus der Perspektive der geometrischen Akustik gesehen werden und deswegen kann angenommen werden, dass die Reflexionen, die an der Pinna zustande kommen, komplexe Eigenschaften aufweisen. Durch die stark individualisierte Form jeder Ohrmuschel kann ein solches Modell nur sehr schwer verallgemeinert werden. Ein späteres Modell [TS68] basiert auf Resonanzen welche in den Hohlräumen der Ohrmuschel zustande kommen. So können Resonanzen bei 3, 5, 9, 11 und 13 kHz erkannt werden, wobei die Amplitude dieser Resonanzen stark richtungsabhängig ist. Im Laufe der Jahre wurden weitere Modelle über die Lokalisationsfähigkeit über spektrale Faktoren formuliert, jedoch konnte man bis heute noch keinen klaren und quantitativen Zusammenhang zwischen Einfallsrichtung und spektraler Veränderung des Signals finden. Es ist jedoch klar, dass jeder in der Lage ist diese spektrale Faktoren zur Quellenlokalisierung zu nutzen [Xie13].

2.3.2 Binaurale Faktoren

Binaurale Faktoren sind für die Lokalisation in der horizontalen Eben sehr wichtig und beruhen auf die wahrgenommenen Unterschiede zwischen den beiden Ohren.

Interaurale Zeitdifferenzen (ITD)

Durch die Beugung niederfrequenter Schallwellen am Kopf kommt es zu Laufzeitunterschieden zwischen ipsilateralem und kontralateralem Ohr. Die ITD (*Interaural Time*

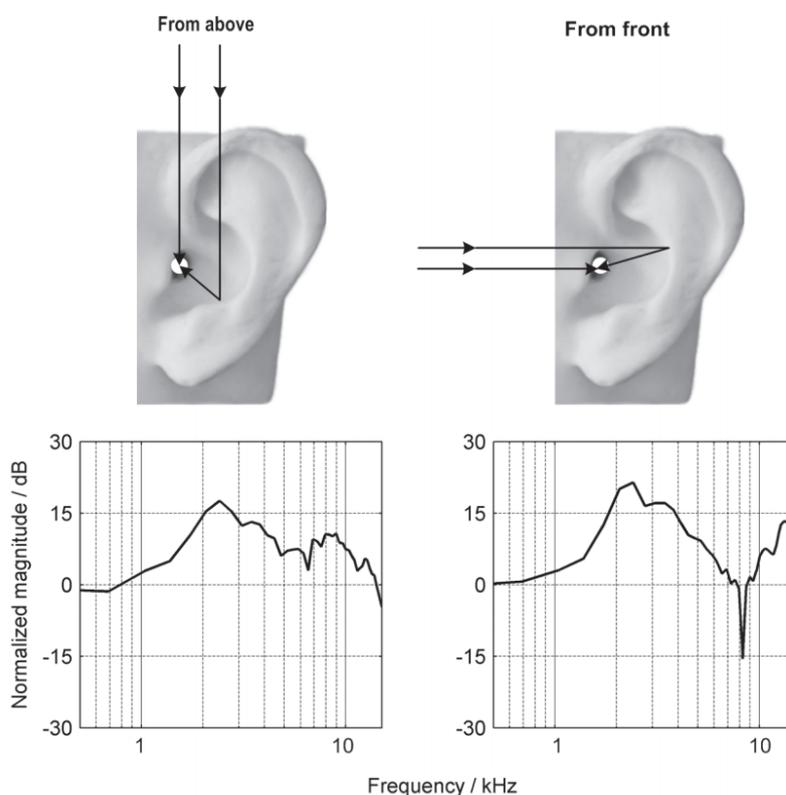


Abbildung 3 – Interaktion der Ohrmuschel mit eintreffendem Schall aus zwei typischen Richtungen und die daraus folgende spektrale Verfärbung. Übernommen von [Xie13]

Difference) ist hauptsächlich bei Frequenzen $< 1,5$ kHz wirksam, da die Wellenlänge bei solchen Frequenzen größer als der durchschnittliche Kopfdurchmesser ist ($\lambda > d_{Kopf}$). Die Laufzeitunterschiede werden vom Menschen als Phasenunterschiede wahrgenommen, wodurch die Lokalisation der Quelle gewährleistet wird.

Interaurale Pegeldifferenzen (ILD)

Falls die Wellenlänge der eintretenden Frequenzen kleiner als der Kopfdurchmesser ($\lambda < d_{Kopf}$) ist, wirkt die ILD (*Interaural Level Difference*). Durch Abschattung der Schallwelle am Kopf kommt es zwischen ipsi- und kontralateralem Ohr zu Pegeldifferenzen. Diese ermöglichen die Lokalisation einer Quelle im Raum. Ist diese Differenz > 25 dB, wird die Schallquelle ganz aus ipsilateraler Richtung wahrgenommen.

Cone of Confusion

Bei der Lokalisation der am Cone of Confusion liegenden Quellen kommt es oftmals zu Fehlern. Quellen welche auf der Oberfläche des Cone of Confusion liegen erzeugen gleiche ILD und ITD und dadurch ist eine Lokalisation dieser sehr schwierig. Oftmals wird deshalb eine Quelle, die sich in der vorderen Hälfte befindet, von hinten wahrgenommen oder umgekehrt (Vorne-Hinten-Ambiguität). Der Mensch kompensiert dieses Phänomen

auf natürliche Weise mit leichten Kopfbewegungen, wodurch ILD und ITD, gegenläufig für beide Definitionsrichtungen verlaufend, variiert werden.

3 Die Außenohrübertragungsfunktion (HRTF) und die binaurale Impulsantwort (HRIR)

3.1 Theoretischer Hintergrund

Um die Lokalisation einer Quelle im virtuellen Raum zu gewährleisten, ist es nötig das Signal dieser mit einer sogenannten Außenohrübertragungsfunktion zu filtern. Mit HRTF (*Head Related Transfer Function*) wird die Außenohrübertragungsfunktion in der Frequenzdomäne bezeichnet, mit HRIR (*Head Related Impulse Response*) die in der Zeitdomäne. Die Außenohrübertragungsfunktion beinhaltet die Filterwirkung von Ohrmuschel, Kopf und Oberkörper. Mit der Annahme einer beliebigen Schallquellenposition, wie in Abbildung 4 gezeigt, sind zwei HRTFs, H_L und H_R , für linkes und rechtes Ohr, mit folgender Gleichung definiert:

$$H_L = H_L(r, \delta, \phi, f, a) = \frac{P_L(r, \delta, \phi, f, a)}{P_0(r, f)} \quad (1)$$

$$H_R = H_R(r, \delta, \phi, f, a) = \frac{P_R(r, \delta, \phi, f, a)}{P_0(r, f)}.$$

P_L und P_R stellen jeweils den komplexen Schalldruck an den beiden Ohren dar. P_0 stellt den komplexen Schalldruck in der Frequenzdomäne im Koordinatenursprung, also in der Mitte des abwesenden Kopfes, und nimmt folgende Form an:

$$P_0(r, f) = j \frac{k \rho_0 c Q_0}{4\pi r} \exp(-jkr), \quad (2)$$

wobei ρ_0 die Mediumsdichte ist, c die Schallgeschwindigkeit, Q_0 die Schallintensität der Punktquelle und $k = 2\pi f/c$ die Wellenzahl ist.

HRTFs sind also Funktionen der Frequenz und der Schallquellenposition bezüglich Distanz r , Azimuth Winkel δ und Elevationswinkel ϕ . Außerdem wird jede HRTF individuell durch die anatomischen Eigenschaften des Menschen geformt. Dies wird in obigen Gleichungen durch die Variable a ausgedrückt, welche Parameter über die Dimensionen relevanter anatomischer Strukturen beinhaltet [Xie13]. Die Messposition für P_L und P_R variiert in der Literatur, jedoch gilt das Trommelfell als die natürlichste Wahl obwohl dies bei einem Menschen, im Gegensatz zur Nutzung eines Kunstkopfes, umständlich erscheinen mag. Deswegen wird der Einfluss des Ohrkanals oft vernachlässigt. Es wurde gezeigt, dass wenn der Ohrkanal als 8 mm dickes Rohr approximiert wird, kann dieses als eindimensionale Übertragungslinie für Frequenzen deren Wellenlänge viel größer als

diesen Durchmesser repräsentiert werden. Eine Wellenlänge von 8 mm entspricht einer Frequenz von 42.5 kHz, ein Vierfaches dieser Wellenlänge entspricht ca. 10 kHz. Deswegen gilt diese Annahme für Frequenzen kleiner als 10 kHz. In verschiedenen Punkten im Ohrkanal kann die räumliche Information korrekt erfasst werden, sei es bei binauralen Aufnahmen als auch bei der Messung von HRTFs.

Unser Gehörssystem kann mit einem Schallübertragungsmodell dargestellt werden welches zur Kalibrierung der Wiedergabekette und zur Einstellung individueller Variationen genutzt werden kann. Dieses Modell wird in Abbildung 5 dargestellt. Das Schallfeld außerhalb des Ohres wird mit einem Thévenin Äquivalent, bestehend aus dem (open-circuit) Schalldruck P_1 und einer Generator Impedanz, dargestellt. Die Generator Impedanz ist gleich der Radiationsimpedanz Z_1 . P_1 existiert nicht bei normalen Hörsituationen, wenn der Ohrkanal jedoch blockiert wird, z.B. mit einem Mikrophon, und die Volumenschnelle somit Null wird, kann P_1 außerhalb des Hindernisses gemessen werden. Der Schalldruck am Trommelfell wird mit P_3 repräsentiert, mit P_2 der am offenen Ende des Ohrkanals welcher mit einem Zweitor dargestellt wird. Die Last dieses Zweitores wird durch die Impedanz Z_3 ausgedrückt und die Eingangsimpedanz mit Z_2 . Der natürliche Schalldruck am Eingang des Ohrkanals wird mit

$$\frac{P_2}{P_1} = \frac{Z_2}{Z_1 + Z_2} \quad (3)$$

bestimmt. Der Schalldruck am Trommelfell kann wie folgt dargestellt werden:

$$\frac{P_3}{P_2} = \frac{Z_3}{Z_2}; \quad \frac{P_3}{P_1} = \frac{Z_3}{Z_1 + Z_2} \quad (4)$$

Die Schallübertragung von Quelle zu Ohr kann in zwei Teile unterteilt werden: einen richtungsabhängigen Teil, von der Quelle bis zum Ohrkanaleingang, und einen richtungsunabhängigen Teil im Ohrkanal bis hin zum Trommelfell. Daraus folgt, dass die Schalldrücke P_1 , P_2 und P_3 , oder auch an anderen Positionen im Ohrkanal gemessenen Schalldrücke den gleiche räumlichen Informationsanteil besitzen. Sie können somit alle als binaurale Schalldrücke in der Definition der Außenohrübertragungsfunktionen benutzt werden.

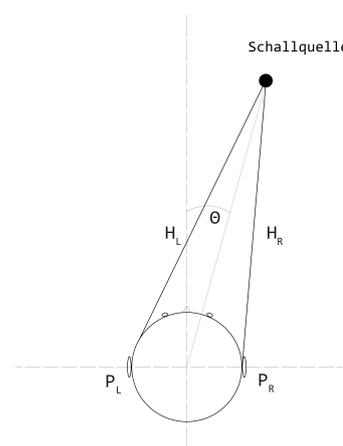


Abbildung 4 – Punkt-schallquelle im Raum

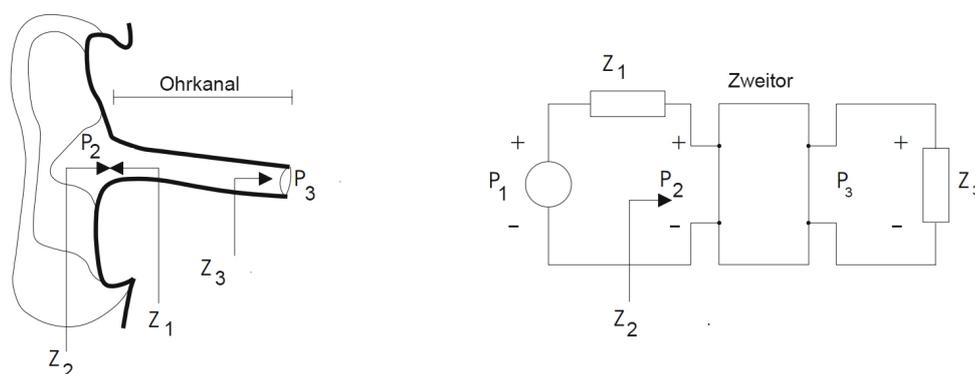


Abbildung 5 – Schallübertragungsmodell des menschlichen Außenohrs, anatomisch und analog - adaptiert aus [Mol92]

3.2 Messung von Außenohrübertragungsfunktionen

Die ideale Außenohrübertragungsfunktion beschreibt die Änderung von einer sich frei verbreitenden ebenen Welle zu den Schalldrücken die sich ergeben, wenn ein Hörer das Schallfeld betritt und die Verbreitung der Welle behindert. Bei Messungen im Labor, ergeben sich ebene Welle nur als Approximation, da sich eine Quelle in der reflexionsfreien Kammer immer auf einer begrenzten Distanz zum Hörer befinden wird und deswegen immer Kugelcharakteristik vorweisen wird. Auch die reflexionsfreie Kammer an sich stellt nur eine Approximation dar, da meist bei tiefen Frequenzen trotzdem Reflexionen zustande kommen, ebenfalls durch Stühle, Aufhängungen, usw..

Zur Messung von individualisierten Außenohrübertragungsfunktionen, wird in einer reflexionsfreien Kammer ein impulsartiges oder längeres breitbandiges Signal abgespielt um diesen dann mit kleinen Mikrofonen am Eingang des Ohrkanals einer Testperson aufzunehmen, wobei der Einfluss der Ohrmuschel und anderer anatomischer Strukturen in die Messung einfließt. Wird in beiden Ohren gleichzeitig gemessen, werden in den Filtern ebenfalls interaurale Differenzen miteinbezogen. Somit können alle relevanten räumlichen Faktoren für eine gewisse Quellenposition, einem gewissen Hörer und einer gewissen räumlichen Umgebung gemessen werden. Die Filter die sich aus diesen Messungen ergeben sind Beispiele für Filter mit endlicher Impulsantwort (FIR-Filter, *Finite Impulse Response Filter*) und können wie schon in 3.1 erläutert, sei es im Frequenzbereich (HRTF) als auch im Zeitbereich (HRIR) dargestellt werden. Um von einer Darstellungsweise in die andere zu wechseln, wird eine Fourier-Transformation durchgeführt. Filterung im Frequenzbereich geschieht mit einer Punkt für Punkt Multiplikation mit dem Frequenzspektrum des Signals, im Zeitbereich geschieht dies mit einer Faltungsoperation. Durch die Filterung eines beliebigen Signals kann diesem somit eine räumliche Eigenschaft verlieht werden und falls diese Filterung in Echtzeit geschieht, können zusätzlich auch die Einflüsse von Quellpositionsänderungen und Kopfbewegungen miteinbezogen werden. Es gibt verschiedene Möglichkeiten die Außenohrübertragungsfunktionen zu erfassen: Es können Kunstköpfe bzw. Kunsttorsos zum Erfassen von generellen HRTFs genutzt werden, die Mikrofonposition im Inneren des Ohrkanals kann variiert werden, sowohl sich im

Laufe der Zeit das Blockieren des Ohrkanals mit einem kleinen Mikrofon als Standard etabliert hat, und es kann ebenfalls das Stimulationssignal in Länge und Art verändert werden. Generell soll besonders Wert auf die Messumgebung gesetzt werden, um Reflexionen jeglicher Art und somit auch Messfehler zu vermeiden. Auch die Kopfpositionen und die Mikrofonpositionen sollten sehr empfindlich behandelt werden, da jede kleine Änderung dieser zu erheblichen Messfehlern führen kann.

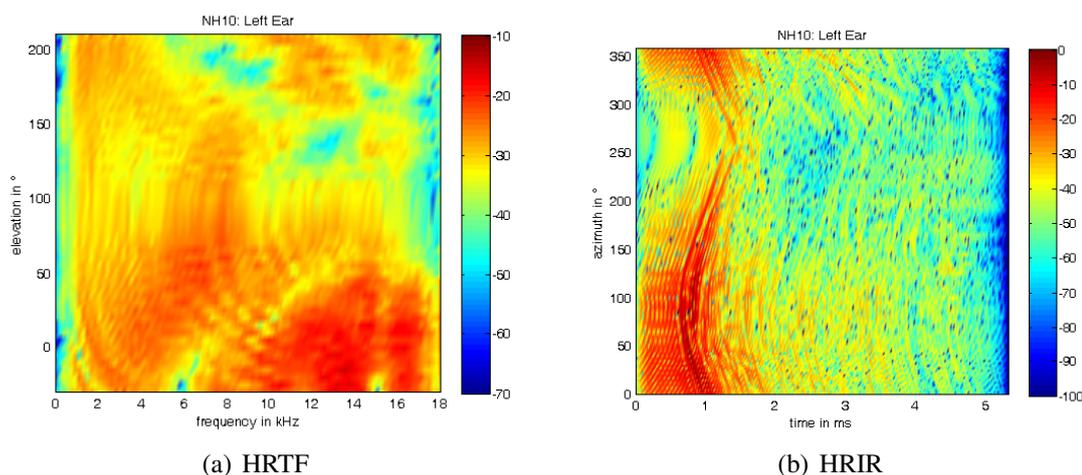


Abbildung 6 – Messung von Außenohrübertragungsfunktion (HRTF) und binauraler Impulsantwort (HRIR). Entnommen aus dem ARI HRTF Database des Acoustic Research Institute der Österreichischen Akademie der Wissenschaften. Gemessen mit einem KE-4-211-2 Mikrofon von Sennheiser im Inneren des Ohrkanales

3.3 Berechnung von Außenohrübertragungsfunktionen

Ein weiterer Zugang zur Erstellung von Außenohrübertragungsfunktionen ist die mathematische Berechnung dieser. Da die Messung individualisierter HRTFs sehr langwierig und fehleranfällig ist, wird über die Lösung der Wellengleichung oder der Helmholtzgleichung, unter Beachtung der Dimensionen von Kopf, Ohrmuschel und Torso, versucht eine individualisierte Übertragungsfunktion zu berechnen.

Ein sphärischer Kopf (ohne Ohrmuscheln und Torso) ist die einfachste Herangehensweise zur Berechnung von HRTFs, ist meistens für Frequenzen unter 3 kHz grob gültig und kann außerdem fundamentale Eigenschaften von HRTFs, wie z.B. ITD, ILD, den Effekt der Ohrposition und die Distanzabhängigkeit von Nahfeld HRTFs beinhalten.

Ein weiteres einfaches Modell, ist das sogenannte Schneemann-Modell, wobei Kopf und Oberkörper mit zwei Kugeln verschiedener Durchmesser simuliert werden. Dadurch kann der Einfluss des Brustkorbes auf qualitative Weise mitberechnet werden. Obwohl dieses Modell eine leichte Verbesserung im Gegensatz zum ersten vorgestellten Modell darstellt, wird dadurch immer noch eine wesentlich niedrige Lokalisationsqualität erreicht. Wie in Sektion 2.3.1 erläutert wurde, spielt bei der Lokalisation von hohen Frequenzen die Form der Ohrmuschel eine substantielle Rolle und auch die genauen Dimensionen von Kopf und Brustkorb sollten bei der Berechnung von HRTFs berücksichtigt werden. Genaue Lö-

sungen für die Übertragungsfunktion eines Kopfmodells mit präzisen geometrischen Eigenschaften zu berechnen ist sehr rechenaufwendig und deswegen wurden Berechnungsmethoden entwickelt, unter anderem die herkömmlich benutzte BEM (*Boundary Element Method*), welche diesen Prozess erleichtern. Beim BEM wird einst das Randwertproblem der Wellen- oder Helmholtzgleichung in ein Oberflächenintegral umgewandelt. Die Grenzoberfläche wird diskretisiert und in ein Netz kleiner Elemente umgewandelt, woraus sich schließlich eine Reihe linearer Gleichungen ergibt. Die Kopfgeometrie wird für diesen Zweck meist mit einem Laserscanner oder möglicherweise auch über Fotografie erfasst. Ein wesentlicher Nachteil dieser Methode ist die benötigte hohe Rechenleistung und die daraus folgenden hohen Kosten. Durch die Anwendung des akustischen Prinzip der Reziprozität, kann der Rechenaufwand nochmals verringert werden, es können jedoch dadurch Fehler auftreten. Es werden derzeit Algorithmen zur Verbesserung der BEM-basierten HRTF Berechnung entwickelt. Jegliche Vereinfachung der geometrischen Eigenschaften des modellierten Kopfes können ebenfalls zu effizienterer aber ungenauerer Berechnung führen. [AR17]

3.4 Lokalisationsqualität verschiedener HRTFs

Wie schon in Kapitel 3.2 erwähnt, können für die Messung von HRTFs sei es Testpersonen als auch Kunstköpfe bzw. Kunsttorsos in Verwendung genommen werden. Der wesentliche Nachteil der Benutzung von Testpersonen ist, dass jegliche kleine Bewegung dieser die Messung stark beeinträchtigen kann. Dies ist bei Benutzung von Kunstköpfen natürlich nicht der Fall, jedoch können die Übertragungsfunktionen, die sich von derartigem Messverfahren ergeben nie individualisiert sein und sind wahrscheinlich nicht für jeden Menschen geeignet, wodurch die Lokalisationsqualität beeinträchtigt wird. Die Erstellung von Kunstköpfen, wie z.B. KEMAR oder Neumann KU-100, basiert auf gemittelte Messungen von Kopfgröße und Ohrmuschelform und versuchen den durchschnittlichen Menschen nachzubilden.

4 Dreidimensionale Audiotechniken

In diesem Unterpunkt wird die Geschichte verschiedener Beschallungstechniken erläutert und erklärt in welchem Zusammenhang diese über binaurale Synthese wiedergegeben werden können.

4.1 Stereo

Die einfachste Wiedergabemethode ist, nach der Monowiedergabe, die Stereowiedergabe. Historisch gesehen wurde Musik anfangs im meist über kopfhörerähnliche Systeme übertragen und abgespielt. Das Theatrophon, welches 1881 von Clément Ader entwickelt wurde, erlaubte einem Hörer über die Telefonlinie Konzerte der Pariser Oper mitzuverfolgen. Dieses System basierte auf die Wiedergabe von jeweils zwei Mikrofonen, welche auf

dem Bühnenrand platziert waren, auf zwei Telefonmuscheln. Diese Erfindung wird heutzutage als frühes Beispiel von binauraler Aufnahme und Wiedergabe gesehen. Zu dieser Zeit hatte jedoch der Begriff *binaural* noch die Bedeutung von 'Hören mit zwei Ohren'. Erst in den 1930er Jahren wurde erstmals die moderne Unterscheidung zwischen Stereo und Binaural vorgeschlagen. Sowohl anfangs meistens, man denke an den militärischen Bereich (Sonar, Radar), über Kopfhörer abgehört wurde, gab es gleichzeitig auch einen Aufschwung in der Lautsprecherentwicklung.

Beim Mixen und Mastern der meisten Musikstücke geschieht die Abhöre mittels einem Lautsprecherpaar, obwohl heutzutage der Verbraucher meistens eine Kopfhörerwiedergabe wählt. Das Klangbild welches über Kopfhörerwiedergabe zustande kommt, entspricht nie dem vom Toningenieur bevorzugten Klangbild welches über Lautsprecherwiedergabe entstand, da dabei kein Übersprechen zwischen den beiden Treibern entstehen kann. Durch Kopfhörer-Verbesserungsalgorithmen wurde versucht dieses Problem zu beheben und in einer Studie, vgl. [MMMC15], wurde tatsächlich bewiesen, dass über *Nahfeld Übersprechsimulation* bei den meisten Testpersonen eine Verbesserung des klanglichen Eindruckes im Gegensatz zu herkömmlicher Kopfhörerwiedergabe gewährleistet wird. Durch binaurale Simulation zweier Schallquellen vor dem Hörer kann ebenfalls ein derartiges Ergebnis erreicht werden. Durch die Simulation zweier Lautsprecher kann eine Phantomschallquelle zwischen den beiden Schallquellen wahrgenommen werden, wie sie bei üblicher Lautsprecherwiedergabe wahrgenommen wird.

4.2 Surround Sound

Mit der Weiterentwicklung von Stereo zu 5.1 Surround - Wiedergabe konnte dem Verbraucher erstmals das Gefühl von räumlichen Hören (wenn auch künstlich) vermittelt werden. Derartige Systeme kamen meist in der Spielfilmbranche zur Verwendung und bestehen meist aus drei Front-Lautsprechern, zwei hinteren Lautsprechern sowie einem Subwoofer. Mit den 90er Jahren und der Entwicklung von neuen verbraucherorientierten Formaten wie die DVD, wurden solche Systeme in Form von Heimkinos immer beliebter. Es gibt eine Vielzahl an diversen Surround Formaten und Anordnungen, die herkömmlichste ist dabei sicherlich die 5.1 Anordnung, wobei dieses heutzutage beliebig mit mehreren Lautsprechern aufgerüstet werden kann. Es besteht die Möglichkeit diese Wiedergabemethode für Kopfhörerwiedergabe über Binauralsynthese zu simulieren.

4.3 Hemisphärisches Surround

Bei Hemisphärischem Surround kommen zusätzlich zu den Lautsprechern auf der Horizontalebene auch Lautsprecher entlang der oberen Hemisphäre zum Einsatz. Eine solche Lautsprecheranordnung kann oftmals in Kinos oder Konzertsälen beobachtet werden und kommt für die Wiedergabe eines dreidimensionalen Schallfeldes zum Einsatz. Bekannte für eine solche Wiedergabe sind z.B. *Auro-3D* ® oder *Dolby Atmos* ®. Bei *Auro-3D* kommen bis zu 27 Lautsprecher zum Einsatz (*AuroMax 26.1*), aufgeteilt in drei Zonen: "Surround Sound"-Lautsprecher in der horizontalen Ebene, "Height"-Lautsprecher welche im 30°-Winkel zum Kopf angebracht werden und "Overhead Ceiling"-Lautsprecher

welche an der Raumdecke angebracht werden. Durch die Lautsprecher in der Decke können sowohl lokalisierbare Geräusche als auch Reflexionen die Natürlichkeit des erzeugten Klangbildes verbessern. Beide Formate unterstützen eine objektbasierte Wiedergabe wie sie in Kapitel 4.5 erläutert wird.

4.4 Ambisonics

Ambisonics wurde 1973 vom Mathematiker und Audiotechnik-Begeisterten Michael Gerzon entwickelt. Seine Kodierungsmethode stach im Gegensatz zu den herkömmlichen matrixbasierten Methoden heraus, weil sie Kugelflächenfunktionen nutzte um die einzelnen Teile des Schallfeldes um den Hörer zu encodieren. Die Basis der Kugelflächenfunktionen kann beliebig gewählt werden (Ambisonics N-ter Ordnung) jedoch kommt es mit steigender Ordnung auch zu mehr Rechenleistung. Bei endlicher Ambisonicsordnung kommt es zu einer Schallfeldapproximation mit einer definierten Anzahl an Kugelflächen. Je niedriger die Ordnung, desto geringer die räumliche Präzision der Nachbildung. Gerzon arbeitete bei der Herstellung des ersten Ambisonics Mikrofon erster Ordnung, dem *Soundfield-Microphone*, mit. Heutzutage gibt es ambisonische Mikrofone mit bis zu 64 Kanälen und neue Software-basierte Methoden zur Verräumlichung mit Ambisonics höherer Ordnung. Ambisonics hat den Vorteil, dass Aufnahme und Wiedergabe von einander getrennt behandelt werden können und auch deswegen wird es heute oftmals als flexibles Produktions- und Vertriebsformat für dreidimensionales Audiomaterial eingesetzt. [FZS15] [KZ14]

4.5 Objektbasiertes Audio

Bei objektbasierter 3D Audiowiedergabe wird ein Audio-Element definiert, dazu werden diesem Element Parameter (Metadaten) zugeordnet, welche Informationen über Lautstärke, Position, usw. beinhalten und wodurch die Audio-Elemente in Lautsprecher übersetzt werden. Audio-Elemente sind meist Mono-Signale welche aufgenommen oder synthetisiert wurden. Objektbasiertes dreidimensionales Audio kommt bei interaktiven Simulationen oder Videospielen zum Einsatz, wobei die Audio-Elemente meist als Punktschallquellen definiert werden. Das Mixen von 5.1 Surround Sound nutzt einen objektbasierten Zugang. Vorteil der objektbasierten Audiowiedergabe und Mischung ist die Wiedergabesystemunabhängigkeit, die intuitiven Personalisierungsmöglichkeiten, z.B. bei der Erstellung von immersiven Soundscapes und effiziente Produktionsabläufe.

5 Binaurales 3D Audio über Kopfhörer

Binaurale Aufnahmen bzw. Aufnahmen die zur Binauralabhöre aufbereitet wurden, werden allgemein über Kopfhörer abgehört. In diesem Zusammenhang gibt es viele Faktoren welche die Wiedergabe von binauralen Aufnahmen beeinflussen können. Im folgenden Abschnitt werden diese erläutert.

5.1 Externalisation

Der Begriff *Externalisation* weist auf die Wahrnehmung einer Schallquelle außerhalb des Kopfes (*Außer-Kopf-Lokalisation*) bei Kopfhörerwiedergabe hin. Bei gewöhnlicher Kopfhörerwiedergabe wird die Quellposition meist im Kopf wahrgenommen (*Im-Kopf-Lokalisation*). Ziel der binauralen Synthese und Wiedergabe ist hingegen dieses Gefühl zu unterdrücken und eine Wahrnehmung der Quelle außerhalb des Kopfes zu gewährleisten, da die wahrgenommene Natürlichkeit des Hörereignisses somit zunimmt. Ein Beispiel von Wahrnehmung einer Schallquelle im Kopf ist die eigene Stimme bei zugehaltenen Ohren. Es wurden viele Gründe für eine solche Wahrnehmung bei Kopfhörerwiedergabe vermutet: Kopfhörer könnten eine unnatürliche Knochenübertragung und Druck am Kopf erzeugen; Invarianz des Signals bei nicht-statischen Kopfdimensionen; Natürliche Resonanzen zwischen Kopfhörer und Mikrofonen. Falls die Signale an den beiden Ohren sehr ähnlich sind werden diese vom Gehirn in ein einziges Hörereignis zusammengefügt und auch die wahrgenommene Nähe der Schallquellen kann zur Wahrnehmung im inneren des Kopfes führen. Auch die Unterdrückung oder das Alterieren der räumlichen Attribute die von der Ohrmuschel gewährleistet sind kann zu Im-Kopf-Lokalisation führen. [Bla83] Kopfhörer weisen außerdem immer eine gewisse Filtercharakteristik vor und beim Platzieren der Kopfhörer auf den Ohren kommt es zu einem "akustischen Hohlraum" welches eine eigene Übertragungsfunktion besitzt. All diese Effekte müssen kompensiert werden um die Externalisation von Schallquellen bei Kopfhörerwiedergabe zu ermöglichen.

Drei Faktoren tragen zur Externalisation über Kopfhörer bei. Das Einfügen von räumlichen Attributen, d.h. angemessene (künstliche oder natürliche) Verhallung der Signale, wird als wichtigster Faktor angesehen, zudem trägt auch der Einsatz von Außenohrübertragungsfunktionen zur Externalisation bei. Der dritte Faktor ist vom Einsatz eines Kopfpositionssensors gewährleistet, wobei eine interaktive akustische Umgebung erzeugt wird, wie in der Sektion 5.2 ausführlich erklärt wird. Fluktuationen der interauralen Pegeldifferenzen, welche sich durch Kombination von kopfbezogener und raumbezogener binauraler Synthese ergeben, haben ebenfalls Einfluss auf die wahrgenommene Externalisation. Es wurde erforscht, dass eine Unterdrückung dieser Fluktuationen, durch den Einsatz einer Gammaton Filterbank für Analyse und Resynthesierung, die Externalisation von Signalen mit Frequenzanteilen über 1 kHz verringern [CSB⁺13]. Eine weitere Methode zur Verbesserung der Externalisation schlägt den Einsatz von Dekorrelation und spektraler Notch-Filterung vor. Neben dem herkömmlichem Einsatz von Außenohrübertragungsfunktionen und Verhallung, werden die Signale dekorreliert und bei gewissen Frequenzen gefiltert. Diese Frequenzen lassen sich von Messungen der Außenohrübertragungsfunktionen für Schallquellen vor dem Hörer ablesen. Die zusätzliche Filterung und Dekorrelation tragen zu einer wesentlichen Verbesserung der Externalisation, überhaupt in frontaler Richtung bei [KCK⁺10].

5.2 Kopfpositionsverfolgung

Wie schon in Sektion 2.3.2 kurz erläutert wurde, nutzt der Mensch kleine Kopfbewegungen aus um vorne-hinten Ambiguitäten zu lösen und generell um die Lokalisation einer

Schallquelle zu verbessern. Es gibt viele Techniken zur Kopfpositionsverfolgung: Es gibt intertiale, magnetische, akustische und optische Sensoren. Es kann sich diesbezüglich als sinnvoll erweisen, diese Kopfbewegungen in die binaurale Simulation einzubinden. In einem Kopfhörer kann ein Bewegungssensor eingebaut werden, welcher Kopfbewegungen verfolgt und dementsprechend können der dreidimensionale Raum bzw. die virtuellen Schallquellen im Raum um den Hörer verschoben werden. Dies kann nicht nur für die Lokalisationsverbesserung ein wichtiger Punkt sein, sondern auch für interaktive Spiele oder Aufführungen, in denen sich der Hörer in einem künstlichen Raum bewegen soll. Durch den Einbau eines solchen Sensors wird auch der subjektiv wahrgenommene Realismus/Wahrheitssinn verbessert [SHLV99]. In einem kürzlich erschienenen Paper [HSM⁺17] wird der Einfluss von Head-Tracking auf die Externalisation von Sprachsignalen für nicht-individualisierte binaurale Synthese untersucht. Es konnte gezeigt werden, dass die Verfolgung der Kopfbewegungen tatsächlich zur Externalisation mit nicht-individualisierten HRTFs beiträgt.

5.3 Verbesserung der Lokalisationsqualität durch Zufügung von Hall

In Sektion 2.3.1 wurde erläutert, dass das Verhältnis zwischen direktem und reflektiertem Schall vom Gehirn als monauraler Faktor zur Klangquellenpositionsbestimmung genutzt werden kann. Dieses Verhältnis ist insbesondere für die Distanzwahrnehmung sehr wichtig und es kann darum bedeutsam sein bei Binauralsynthese oder -aufnahmen einen natürlichen oder künstlichen Hall beizufügen wodurch die Lokalisationsqualität verbessert werden kann, wie auch in Sektion 7 ersichtlich gemacht wird. Unser Gehirn kann bei geschlossenen Augen durch den wahrgenommenen Hall bestimmen ob wir uns in einem kleinen oder in einem relativ großen Raum befinden, deswegen kann eine geeignete Wahl des eingesetzten Halleffekts dem dreidimensionalen Hörereignis einen verbesserten räumlichen Eindruck verleihen.

5.4 Nicht individualisierte Binauralsynthese

Wie schon in oberen Sektionen beschrieben wurde, wäre für eine ideale binaurale Wiedergabe eine individualisierte Außenohrübertragungsfunktion nötig. Da relativ wenige Menschen die Möglichkeit der Berechnung einer individualisierten HRTF besitzen, werden meistens standardisierte Übertragungsfunktionen aus den vielen verfügbaren Datenbanken eingesetzt. Diese haben jedoch den großen Nachteil, dass sie nicht für jeden Menschen das gewünschte Ergebnis liefern. Im Laufe der Jahre wurden verschiedene Versuche durchgeführt, bei denen verschiedene standardisierte HRTFs, unter anderem auf Lokalisationsqualität, untereinander verglichen wurden. Das Problem mit solchen Vergleichen ist jedoch, dass eine Übertragungsfunktion nie für jeden Menschen geeignet sein wird. Um dieses Problem zu lösen, könnte ein Auswahlverfahren zur Individualisierung der Übertragungsfunktionen angewendet werden. In der Literatur kann man viele Beispiele solcher Auswahlverfahren finden. Es kann also vermutet werden, dass dem Verbraucher

in Zukunft die Möglichkeit der Individualisierung der HRFTs geboten wird. Ein weitere Herangehensweise zur Lösung diese Problems ist die physische Aufrüstung der Kopfhörer. Durch den Einbau von mehreren Treibern in der Kopfhörermuschel können ebenfalls binaural Aufnahmen abgespielt werden. Die Anzahl der Lautsprecher ist variabel und die Anordnung dieser ahmt meistens die Anordnung der Lautsprecher bei Surround Wiedergabe nach. Neben Kopfhörern mit mehreren Treibern wurden auch Kopfhörer entwickelt bei denen der Treiber auf der vorderen Hälfte der Kopfhörermuschel platziert wird, wobei der Schall direkt auf die Ohrmuschel projiziert wird. Dies spiegelt eine herkömmliche Lautsprecherwiedergabe. Es wurde erforscht, dass derartige Kopfhörer zu Verminderung von vorne-hinten Ambiguitäten, besserer frontaler Lokalisation und eine Klangverfärbung verhindern können [STG13]. Ein weiteres Beispiel für die physische Modifizierung von Kopfhörern sind die Kopfhörerprototypen welche derzeit am Institut für Elektronische Musik und Akustik der Universität für Musik und darstellende Kunst Graz entwickelt werden und welche in Kapitel 7 näher beschrieben werden. Die Besonderheit dieses Prototypen ist die Kombination von Wiedergabe über mehrere kleine MEMS-Lautsprecher und HRTF-basierter Wiedergabe über den Haupttreiber.

6 Binaurales 3D Audio über Lautsprecher

Eine weitere Wiedergabemöglichkeit von binauralem, dreidimensionalem Audio ist jene über Lautsprecher. Ziel dieser Wiedergabemethode ist es, an jeder Ohrmuschel jeweils nur die Schalldrucksignale des ipsilateralen Kanals einer Stereospur wiederzugeben. Falls das Stereosignal mit einer Außenohrübertragungsfunktion des Hörers encodiert wurde und ebenfalls angebrachte ITD und ILD in Verwendung kam, würde für eine ideale Wiedergabe jeweils der rechte Kanal am rechten Ohr und der linke Kanal am linken Ohr angebracht werden, wie es bei Kopfhörerwiedergabe der Fall ist. Dies kann bei Lautsprecherwiedergabe nicht der Fall sein, da es beim Weg vom Lautsprecher zum Ohr zwischen den beiden Kanälen zu Übersprechen kommt. Um eine binaurale Lautsprecherwiedergabe zu ermöglichen ist es deswegen nötig dieses Übersprechen auszulöschen (*Cross Talk Cancellation*). Ohne eine derartige Auslöschung würden die interauralen Faktoren zwangsläufig unbrauchbar gemacht werden.

6.1 Übersprechkompensation

Für eine vereinfachte Beschreibung der Übersprechkompensation kann behauptet werden, dass das Signal vom rechten Kanal phasenverdrehet dem linken Kanal zugefügt und umgekehrt. Das Signal wird zusätzlich gefiltert und im Bezug auf die Übertragungscharakteristik des "Übersprechpfades" verzögert, wie in Abbildung 7 schematisch dargestellt wurde. Eine perfekte Übersprechkompensation kann relativ leicht entworfen werden, jedoch kann diese zu signifikanten Verzerrungen im abgespielten Signal führen. Dies kommt wegen der schlecht konditionierten Inversion der Übertragungsfunktion des Systems zustande. Die Verfärbung des Spektrums zeigt Spitzen über 30 dB bei gewissen Frequenzen und dies kann die Lautsprecher stark belasten wodurch es zu einer starken Verminderung der

Wellenfeldsynthese ist jene die von G. Theile vorgestellt wurde und den Namen 'Binaural Sky' trägt [MWTF05]. Ein Lautsprecherarray welcher über den Kopf des Hörers befestigt wird erzeugt an den Ohren fokussierte Quellen, die mit geeigneter Filterung eine hochwertige Wiedergabe binauraler Signale ermöglicht. Es kann somit eine dreidimensionale raumakustische Abbildung ohne den Einsatz von Kopfhörern oder Surround-Lautsprechern erreicht werden. Die Wellenfeldsynthese beruht darauf, dass mit Hilfe von Lautsprecherarrays ein dreidimensionales Schallfeld in der Hörzone generiert wird, welches mit einem realen Schallfeld übereinstimmt. Der Hörer kann sich somit frei bewegen ohne das dabei die Lokalisation der Schallquellen vernachlässigt wird. Der 'Binaural Sky' besteht aus einem kreisförmigen Lautsprecherarray welche über den Hörer platziert wird. Tests haben gezeigt, dass durch diese Technik eine hohe Lokalisationsqualität gewährleistet wird. Im Gegensatz zu binaurale Wiedergabe über zwei Lautsprecher haben Reflektionen in diesem Fall keine destruktive Wirkung, da von oben beschallt wird und nur Reflektionen am Boden möglich sind. Die Anwendung einer solchen Methode ist bei einem herkömmlichen Verbraucher jedoch durch hohe Einbaukosten schwer realisierbar.

7 Evaluierung der Testergebnisse des am IEM entwickelten Kopfhörer Prototyp

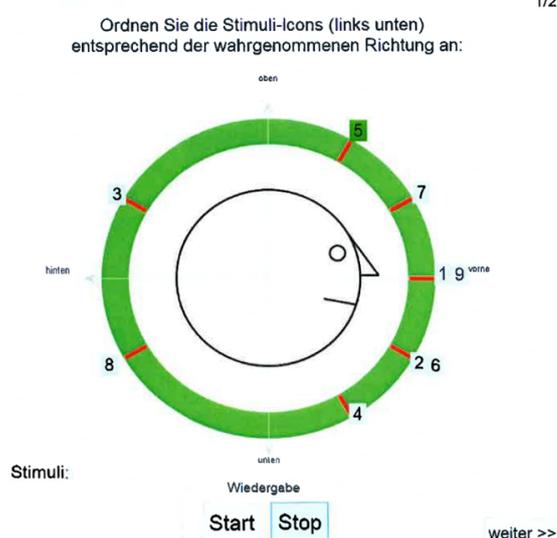


Abbildung 8 – Im Test benutzte grafische Oberfläche für die Zuordnung der Stimuli

1/2

Im Rahmen der Entwicklung eines Kopfhörerprototypen am Institut für Elektronische Musik und Akustik (IEM), wurden Untersuchungen mit verschiedenen Testpersonen durchgeführt. Bei der Evaluierung dieser Testergebnisse wurde zwischen Mitarbeitern des Institutes und Externen Testpersonen unterschieden. Der in dieser Arbeit evaluierte Teil bestand in der Zuordnung diverser Schallquellen im Zusammenhang mit der wahrgenommenen Position dieser auf der Medianebene. Es konnten sieben kurze Musikstückauszüge angehört werden, zwei dieser wurden doppelt ausgelegt, um diese dann schließlich auf einer grafischen Oberfläche (Abbildung 8) der wahrgenommenen Position zuzuordnen. Der entwickelte Kopfhörer basiert auf die Zusammensetzung von "klassischen" 3D Kopfhörern, mit der Nutzung

von HRTFs über den Hauptlautsprecher, und dem zusätzlichen Einbau von sieben MEMS-Lautsprecher auf dem Perimeter der Kopfhörermuscheln, entsprechend den sieben Positionen wie in Abbildung 8. Im niederfrequenten Bereich werden die Kopfhörer mit HRTFs über die Haupttreiber betrieben, die MEMS Lautsprecher kommen erst im

hochfrequenten Bereich zum Einsatz. Dies hängt mit der Bauweise der kleinen MEMS Lautsprecher zusammen, welche nur hohe Frequenzen abstrahlen können. In diesem Sinne wurde der Test vier Mal durchgeführt: einmal rein mit HRTFs, einmal mit der Kombination von HRTFs im tieffrequenten Bereich und MEMS Lautsprechern im hochfrequenten Bereich und jeweils nochmals mit einem zusätzlichen Raumanteil.

Zur Auswertung der Testergebnisse wurde ein MATLAB® Skript entwickelt welches die Testergebnisse grafisch darstellt. Wie in den folgenden Abbildungen erkennbar ist, wird jeder Richtung eine Farbe zugeordnet: Je schmaler die Linie in eine Richtung, desto richtiger die Antworten. In diesem Skript werden außerdem die Testpersonen (auf "Plausibilität") gefiltert, indem kontrolliert wird ob die doppelt ausgelegten Stimuli an der gleichen Position positioniert wurden. Testpersonen bei denen dies nicht der Fall war, wurde für die Auswertung ausgeschlossen. Außerdem wurde kontrolliert ob für jede Richtung maximal zwei Stimuli zugeteilt wurden. Testpersonen bei denen dieser Fehler auftrat wurden ebenfalls von der Auswertung ausgeschlossen.

7.1 Qualitative Auswertung der Testergebnisse

Stellt man die "korrigierten" Auswertungen den ursprünglichen Auswertungen gegenüber, ist eine deutliche Verbesserung in Genauigkeit der Zuordnung sichtbar, wie in Abbildung 9 veranschaulicht wird. Außerdem ist ersichtlich, dass die dem Institut internen Testteilnehmer im Gegensatz zu den externen Testteilnehmern deutlich besser abschneiden (anschaulich beim Vergleich zwischen Abbildungen 10 und 11). Dies kann verschiedene Gründe haben: Einerseits ist es wahrscheinlich, dass diese Testpersonen diesen oder einen ähnlichen Test mit dem Prototypen durchgeführt haben könnten und durch diese Erfahrung besser abschneiden; Andererseits kann vermutet werden, dass diese Testpersonen durch Erfahrung und möglichem Feedback auf Richtigkeit im Laufe der Zeit "trainiert" wurden. In [MCD⁺12] wurde gezeigt, dass kurze Trainingseinheiten und Feedback auf Richtigkeit zu einer Verbesserung der Lokalisationsfähigkeit bei Nutzung von nicht-individualisierten HRTFs führt .

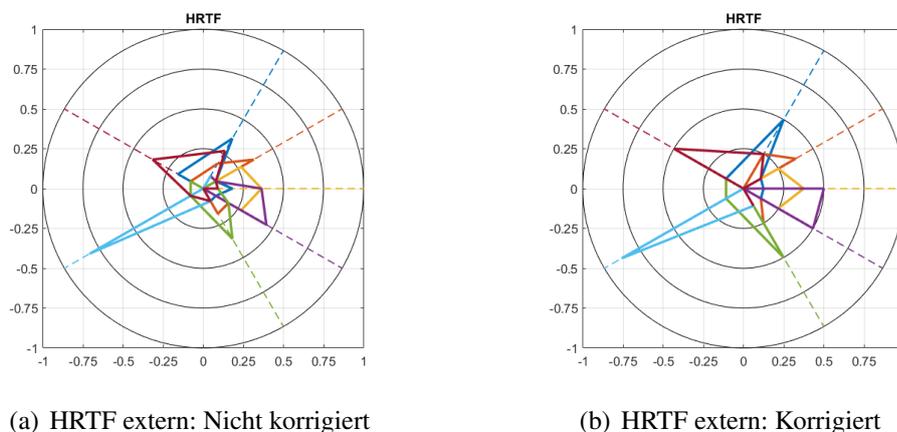


Abbildung 9 – Vergleich zwischen korrigierter Auswertung und nicht korrigierter

Aus den grafischen Auswertungen kann ausgelesen werden, dass die Lokalisation in der

hinteren Hälfte der Medianebene leichter fällt als die in der vorderen Hälfte. Dies kann auf die geringere Distanz zwischen den Stimuli in der vorderen Hälfte der Medianebene, wie in Abbildung 8 dargestellt, zurückgeführt werden, wodurch Ambiguitäten zustande kommen können. Nach Betrachtung der Auswertungsergebnisse, kann ausgesagt werden, dass die Konstellation mit MEMS Lautsprechern und räumlichen Anteil die deutlich besseren Resultate liefert, wie in den Abbildungen 10(d) und 11(d) anschaulich ist.

7.2 Quantitative Auswertung der Testergebnisse

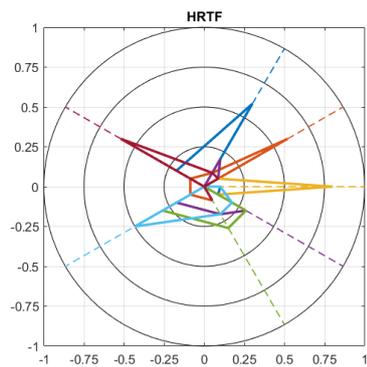
Um neben der qualitativen Auswertung der grafischen Darstellungen der Testergebnisse, welche in Sektion 7.1 durchgeführt wurde, auch eine quantitative Auswertung durchzuführen, wurde das MATLAB® Skript erweitert, sodass die durchschnittliche prozentuale Richtigkeit der Antworten der einzelnen Betriebsmethoden numerisch dargestellt werden kann. Wie schon durch die qualitative Auswertung aus Kapitel 7.1 ersichtlich gemacht werden kann, zeigt auch die quantitative Auswertung, wie sie in Tabelle 1 dargestellt wird, dass institutsinterne Teilnehmer deutlich besser abschneiden als die externen Teilnehmer. Durch die angegebenen Prozentsätze wird klargemacht, dass die Anordnung mit Beschallung über MEMS Lautsprecher und zusätzlichem räumlichem Anteil bei den internen Teilnehmern die beste Lokalisationsqualität liefert. Auch bei den externen Teilnehmern ist diese Anordnung vorne dabei, jedoch wird sie um zwei Prozentpunkte von der Anordnung mit HRTF und Raumanteil übertroffen. Es ist, wie schon bei der qualitativen Auswertung vermutet, eine wesentliche Verbesserung der Lokalisationsqualität bei Beifügung von räumlichen Informationen beobachtbar. In Tabelle 2 wird ein Vergleich der Antwortrichtigkeit externer Teilnehmer vor und nach der obig beschriebenen Teilnehmerselektion dargestellt. Auch in diesem Fall kann die Vermutung einer wesentlichen Verbesserung bestätigt werden: Die Antwortsicherheit steigt um bis zu 10 Prozent an.

	Exterene Teilnehmer	Interne Teilnehmer
Antwortrichtigkeit für HRTF :	51.78 %	52.86 %
Antwortrichtigkeit für MEMS :	32.86 %	75.32 %
Antwortrichtigkeit für HRTF+room :	63.49 %	50.65%
Antwortrichtigkeit für MEMS+room :	61.90 %	79.22%

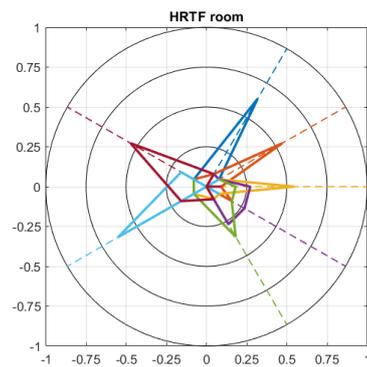
Tabelle 1 – Vergleich der Antwortrichtigkeit interner und externer Testteilnehmer

	Selektierte exterene Teilnehmer	Alle externen Teilnehmer
Antwortrichtigkeit für HRTF :	51.78 %	44.16 %
Antwortrichtigkeit für MEMS :	32.86 %	33.76 %
Antwortrichtigkeit für HRTF+room :	63.49 %	53.24%
Antwortrichtigkeit für MEMS+room :	61.90 %	57.14%

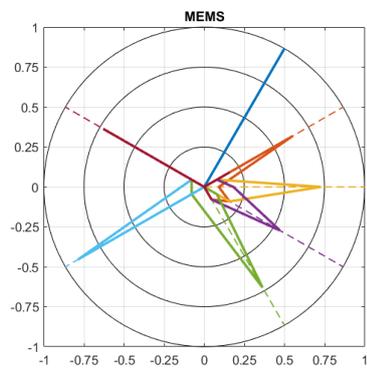
Tabelle 2 – Vergleich der Antwortrichtigkeit externer Testteilnehmer vor und nach der Teilnehmerselektion



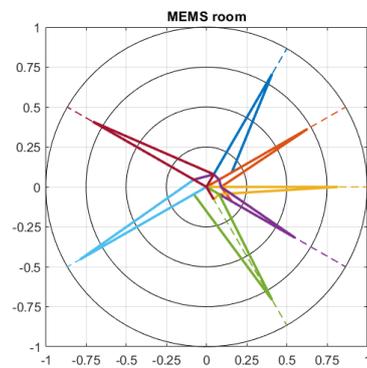
(a) Auswertung HRTF intern



(b) Auswertung HRTF Room intern

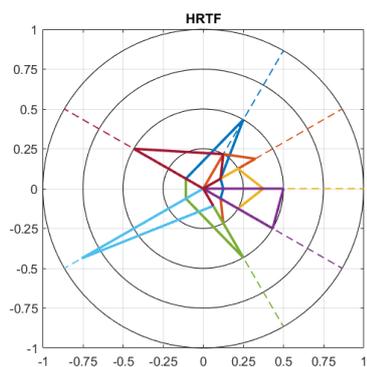


(c) Auswertung MEMS intern

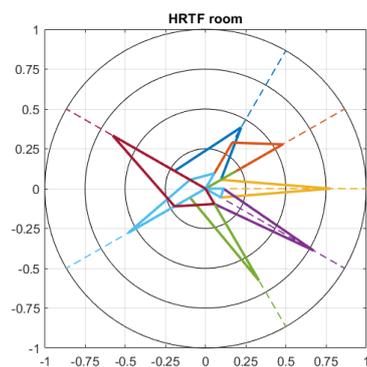


(d) Auswertung MEMS Room intern

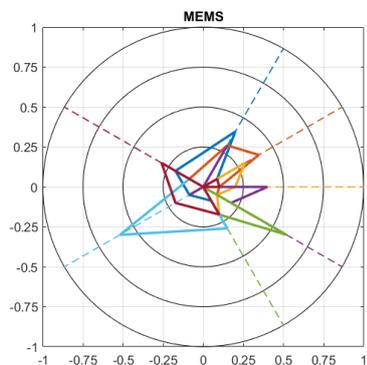
Abbildung 10 – Grafische Auswertung der Testergebnisse interner Teilnehmer



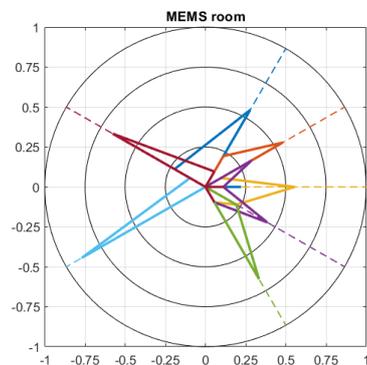
(a) Auswertung HRTF extern



(b) Auswertung HRTF+Room extern



(c) Auswertung MEMS extern



(d) Auswertung MEMS+Room extern

Abbildung 11 – Grafische Auswertung der Testergebnisse externer Teilnehmer

8 Schlussfolgerung und Ausblick

Es wird vermutet, dass Binauraltechnik in Zukunft eine immer zentralere Rolle in der verbraucherorientierten Klangwiedergabe spielen wird. Schon jetzt geschieht die Musikwiedergabe meist über Kopfhörer, wodurch aber das vom Mischer erzeugte Klangbild, welches meist über Lautsprecherabhöre erstellt wurde, verfälscht werden kann. Es ist somit denkbar, dass in Zukunft diesem Phänomen durch den Einsatz von Binauraltechnik entgegengearbeitet wird, wie in Sektion 4.1 erläutert wurde. Durch die immer zunehmende Rechenleistung auch bei mobilen Geräten, kann vermutet werden, dass zukünftig sehr viele dieser Geräte für Binauralwiedergabe ausgestattet sein werden und sich dadurch auch die Unterhaltungsmusikbranche immer stärker mit dreidimensionaler Aufbereitung interessieren wird. Binaurale Systeme kommen heute schon sehr häufig bei Spielen, wie z.B. bei Ego-Shootern in Verwendung. Auch der Aufschwung von Spielen und Simulationen in der virtuellen Realität bringt einen immer stärkeren Einsatz von Binauraltechnik mit sich. Auch bei künstlerischen Installationen kann ein Einsatz von Binauraltechnik zu einem immersiven Erlebnis führen. Die Personalisierung von den Außenohrübertragungsfunktionen wird in Zukunft ein wichtiges Forschungsfeld darstellen, da es immer wichtiger sein wird binaurale Aufnahmen für ein weites Publikum verfügbar zu machen. Es könnten dafür auch selbstkalibrierende Systeme vermutet werden wodurch z.B. durch einen kurzen Test die Außenohrübertragungsfunktion auf den Nutzer abgestimmt wird.

Literatur

- [AR17] P. G. Agnieszka Roginska, *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio (Audio Engineering Society Presents)*. Routledge, 2017.
- [Bat67] D. W. Batteau, “The role of the pinna in human localization,” *Proc. R. Soc. Lond. B*, vol. 168, no. 1011, pp. 158–180, 1967.
- [Bla83] J. Blauert, “Spatial hearing,” 1983.
- [Cho17] E. Choueiri, “Binaural audio through loudspeakers,” in *Immersive Sound*. Focal Press, 2017, pp. 138–193.
- [CSB⁺13] J. Catic, S. Santurette, J. M. Buchholz, F. Gran, and T. Dau, “The effect of interaural-level-difference fluctuations on the externalization of sound,” *The Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. 1232–1241, 2013.
- [FZS15] M. Frank, F. Zotter, and A. Sontacchi, “Producing 3d audio in ambisonics,” in *Audio Engineering Society Conference: 57th International Conference: The Future of Audio Entertainment Technology—Cinema, Television and the Internet*. Audio Engineering Society, 2015.
- [Han16] F. Handrick, Susann und Diercks. (2016, Nov) Das menschliche Ohr. <http://de.serlo.org/biologie/biologie-menschen/sinne-organe-menschen-ohr>.
- [HSM⁺17] E. Hendrickx, P. Stitt, J.-C. Messonnier, J.-M. Lyzwa, B. F. Katz, and C. De Boishéraud, “Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis,” *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 2011–2023, 2017.
- [KCK⁺10] Y. G. Kim, C. J. Chun, H. K. Kim, Y. J. Lee, D. Y. Jang, and K. Kang, “An integrated approach of 3d sound rendering techniques for sound externalization,” in *Pacific-Rim Conference on Multimedia*. Springer, 2010, pp. 682–693.
- [KZ14] M. Kronlachner and F. Zotter, “Spatial transformations for the enhancement of ambisonic recordings,” in *Proceedings of the 2nd International Conference on Spatial Audio, Erlangen*, 2014.
- [MCD⁺12] C. Mendonça, G. Campos, P. Dias, J. Vieira, J. P. Ferreira, and J. A. Santos, “On the improvement of localization accuracy with non-individualized hrtf-based sounds,” *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 821–830, 2012.
- [MMMC15] E. Manor, W. Martens, A. Marui, and D. Cabrera, “Nearfield crosstalk increases listener preferences for headphone-reproduced stereophonic imagery,” *Journal of the Audio Engineering Society*, vol. 63, no. 5, pp. 324–335, 2015.
- [MOD89] B. C. Moore, S. R. Oldfield, and G. J. Dooley, “Detection and discrimination of spectral peaks and notches at 1 and 8 khz,” *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 820–836, 1989.

- [Mol92] H. Moller, “Fundamentals of binaural hearing,” *Appl. Acoust*, vol. 36, pp. 171–218, 1992.
- [MS04] D.-Y. Ma and H. Shen, “Handbook of acoustics,” *Beijing, PRC: Science*, 2004.
- [MWTF05] D. Menzel, H. Wittek, G. Theile, and H. Fastl, “The binaural sky: A virtual headphone for binaural room synthesis,” in *Proc. of intern. tonmeister Symposium, Hohenkammer*, 2005.
- [Rum01] F. Rumsey, *Spatial Audio*. Focal Press, 2001.
- [SHLV99] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, “Creating interactive virtual acoustic environments,” *Journal of the Audio Engineering Society*, vol. 47, no. 9, pp. 675–705, 1999.
- [STG13] K. Sunder, E.-L. Tan, and W.-S. Gan, “Individualization of binaural synthesis using frontal projection headphones,” *Journal of the Audio Engineering Society*, vol. 61, no. 12, pp. 989–1000, 2013.
- [TS68] R. Teranishi and E. Shaw, “External-ear acoustic models with simple geometry,” *The Journal of the Acoustical Society of America*, vol. 44, no. 1, pp. 257–263, 1968.
- [Xie13] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display (A Title in J. Ross Publishing’s Acoustic)*. J. Ross Publishing, 2013.