

Entwicklung einer nutzerfreundlichen Individualisierung von Binauraltechnik

Bachelorarbeit

Im Studiengang
Audiovisuelle Medien

Vorgelegt von

Sibylle Neumeier

Matrikelnummer: 24417

am 03. April 2016

an der Hochschule der Medien Stuttgart

Erstprüfer: Prof. Oliver Curdt

Zweitprüfer: Prof. Dr. Simon Wiest

Kurzfassung

Die vorliegende Bachelorarbeit beschäftigt sich mit der Problematik der konsumentenfreundlichen Individualisierung von HRTFs für Virtuelle Realität auf mobilen Endgeräten. Das Ziel der Arbeit ist es, eine Lösung für die nutzerfreundliche Einflussnahme auf ein binaurales Klangbild zu schaffen. Durch nicht angepasste Ohrübertragungsfunktionen kann es zu gravierenden Fehllokalisationen kommen.

Es werden zunächst die Grundlagen des Hörens vorgestellt, ein Überblick über bisherige Annäherungsversuche an das räumliche Hören gegeben und ein Individualisierungsfaktor für das Vorhaben der nutzerfreundlichen Anpassung an das persönliche Hörbild ermittelt. Im Anschluss wird das Konzept zu einer Methode erstellt, in der der Nutzer die interaurale Laufzeitdifferenz seinen anatomischen Daten entsprechend manipulieren kann.

Abstract

This bachelor thesis deals with the problem of consumer-friendly individualisation of HRTFs for virtual reality on mobile devices. The aim of this study is to provide a solution for a user-friendly influence on binaural sound. Through unmatched ear transfer functions, serious localization mistakes can occur. This thesis is trying to develop a user-friendly method to influence the personal binaural experience.

First, the basics of hearing are presented, followed by an overview of previous overtures to spatial hearing. Then an individualisation factor for the project of the user-friendly adaptation to the personal soundscape is determined.

In the last chapter a concept for a method is created, in which the user can manipulate the interaural time difference in terms of their anatomical data.

Eidesstattliche Versicherung

Name: Neumeier

Vorname: Sibylle

Matrikel-Nummer: 24417

Studiengang: AM7

Hiermit versichere ich, Sibylle Neumeier, an Eides statt, dass ich die vorliegende Bachelorarbeit mit dem Titel „Entwicklung einer nutzerfreundlichen Individualisierung von Binauraltechnik“ selbständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinne nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der eidesstattlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 23 Abs. 2 Bachelor-SPO (7 Semester) bzw. § 19 Abs. 2 Master-SPO der HdM) sowie die strafrechtlichen Folgen (gem. § 156 StGB) einer unrichtigen oder unvollständigen eidesstattlichen Versicherung zur Kenntnis genommen.

Unterschrift

Abkürzungsverzeichnis

AR	Augmented Reality
BRIR	Binaural Room Impulse Response
BRS	Binaural Room Scanning
CD	Compact Disc
dB	Dezibel
DS	Dynamic Sources
ER	Early Reflections
FC	Filter Crossfading
FIF	Filter Interpolation Frequency Domain
FIR	Finite Impulse Response
FIT	Filter Interpolation Time Domain
FUR	Full Update Rate
GUI	Graphic User Interface
HCI	Human-Computer-Interaction
HEQ	Headphone Equalization
HRTF	Head Related Transfer Function
Hz	Hertz
IKL	Im-Kopf-Lokalisation
ILD	Interaural Level Difference
IR	Impulse Response
ITD	Interaural Time Difference
JND	Just Notable Difference
LTI	Linear-Time-Invariant
MCI	Mensch-Computer-Interaktion
NOS	Nederlanse Omroep Stichting (Niederländische Rundfunkstiftung)
ORTF	Office de Radiodiffusion Télévision Française
OSS	Optimales Stereo-Signal
PRIR	Personal Room Impulse Response
SAD	Spatial Audio Designer
SC	Spectral Cues

SDK Support Development Kit
VR Virtual Reality

Abbildungsverzeichnis

Abbildung 1: Das menschliche Ohr

Abbildung 2: Ebenen des Hörens

Abbildung 3: HRTF-Kurven von 10 Personen bei frontalem Schalleinfall

Abbildung 4: Fehllokalisation bei reinen Tönen

Abbildung 5: Zweikanal-Stereo-Abhörordnung

Abbildung 6: Trennkörper- und Kugelflächen-Stereophonie

Abbildung 7: Kunstkopf KU 100 der Firma Neumann

Abbildung 8: Darstellung der schnellen Faltung

Abbildung 9: Prognostizierter Anstieg von VR-Nutzern bis 2018

Abbildung 10: Anzahl heruntergeladener VR-Apps 2014/15

Abbildung 11: Lokalisationsunschärfe bei nicht individualisierten HRTFs

Abbildung 12: Lokalisation bei manipulierter ITD

Abbildung 13: Qualitätsniveaus der implementierten Anwendung durch Sander u.a.

Abbildung 14: Extraktion der ITD mittels Onset Detektion

Abbildung 15: Grafische Darstellung der Anwendung

Inhaltsverzeichnis

Kurzfassung	3
Abstract	3
Eidesstattliche Versicherung	4
Abkürzungsverzeichnis	5
Abbildungsverzeichnis.....	6
1. Einleitung: Überblick 3D-Anwendungen	9
2. Grundlagen: Hören	10
2.1. Das menschliche Ohr	10
2.2. Hörgewohnheiten	11
2.2.1. Entfernungshören	11
2.2.2. Im-Kopf-Lokalisation.....	13
2.3. Lokalisation durch die Head-Related-Transfer-Function (HRTF).....	14
2.3.1. Die Ebenen des Hörens	15
2.3.2. Monaurale Cues.....	16
2.3.3. Binaurale Cues	18
3. Annäherung an ein realistisches Klangbild	24
3.1. Stereophonie	24
3.2. Surround	26
3.3. Auro-3D.....	27
3.4. Binauraltechnik	28
3.4.1. Kunstkopfstereophonie.....	30
3.4.2. Binauralsynthese	32
3.3.3. Probleme der Binauraltechnik	36
4. Userfreundliche Binauraltechnik	38
4.1. Stand der Technik	38
4.1.1. Virtual Reality (VR)	38
4.1.2. Virtualisierung von Räumen	40
4.2. Binauraltechnik für mobile VR	43
4.2.1. Individualisierung	43
4.2.2. Wahl des Individualisierungsfaktors.....	45

4.2.3. Modell der Anwendung	48
4.2.4. Anwendungsmöglichkeiten	49
4.3. Technische Umsetzung	50
4.3.1. Anforderung an die Technik.....	50
4.3.2. ITD-Errechnung.....	52
4.3.3. Aufbau der Anwendung.....	57
5. Fazit und Ausblick.....	59
Literatur- und Quellenverzeichnis:	61
Bücher:.....	61
Fachartikel und Paper:.....	62
Homepages und Blogs:	64
Zeitungsartikel:	65
Bildquellen	65

1. Einleitung: Überblick 3D-Anwendungen

Virtuelle Realität wird immer beliebter. Auf der FMX 2015 waren die Einsatzmöglichkeiten dieser Technologie in aller Munde. Gerade was interaktive und immersive Formate im Bereich Gaming oder in Filmen angeht, waren sich die Entwickler einig: Virtual Reality wird populär und das Format der Zukunft sein. Das Ziel dieses Formats ist eine möglichst detailgetreue Darstellung der Realität. In diversen Techniken wird dabei auch bereits mit binauralem Audio gearbeitet.

Virtual Reality ist zudem auch für einen durchschnittlichen Konsumenten immer einfacher verfügbar. Zwar sind Headsets wie die Oculus Rift oder Samsung Gear noch teuer, aber bereits mit dem Smartphone kompatibel. Das letztens erschienene Samsung Galaxy S6 ist beispielsweise in der Lage, sich mit einer VR-Gear zu verbinden. So können Kamerabilder auf dem Smartphone zu VR-fähigen Bildern verarbeitet werden. Mit der steigenden Zugänglichkeit und Erschwinglichkeit von Virtual Reality muss auch überlegt werden, wie dieses für den Nutzer optimiert werden kann. Da gerade bei der akustischen Wiedergabe die Immersion durch mangelnde Individualisierung gestört werden kann, liegt der Fokus dieser Arbeit auf einer nutzerfreundlichen Anpassung des Höreindrucks.

Dabei soll geprüft werden, welche Faktoren sich für die Individualisierung der akustischen Simulation eignen und wie der Nutzer auf diese zugreifen kann. Im Anschluss wird das Konzept für eine Methode entwickelt, die es dem Nutzer ermöglicht, durch eine einfache Individualisierung der interauralen Laufzeitdifferenzen auf die Güte der Realitätsdarstellung Einfluss zu nehmen. Die entstandene Anwendung ist nicht für den professionellen Studioeinsatz gedacht, sondern für Konsumenten von VR, die nur schwer die Möglichkeit haben, durch Messungen ihre Hörerlebnisse zu individualisieren.

Dazu soll zuerst ein Überblick über das menschliche Hören gegeben werden, im Anschluss werden bisherige Methoden einer räumlichen und realistischen Darstellung aufgezeigt. Zuletzt wird der Faktor zur Individualisierung bestimmt und die Funktionsweise der Anwendung erklärt. Im Ausblick werden weitere Individualisierungsmöglichkeiten diskutiert.

2. Grundlagen: Hören

Um die Möglichkeiten der Individualisierung von Binauralsignalen zu prüfen, muss zuerst überlegt werden, welche Faktoren der auditiven Wahrnehmung sich hierfür eignen. Im Folgenden sollen die Bestandteile und Funktionen des menschlichen Hörens und der Lokalisation von Schallereignissen erklärt werden. Dazu wird die Terminologie nach Blauert verwendet.¹ Die physikalisch messbaren Werte von akustischen Signalen werden mit dem Präfix „Schall-“, die subjektive Wahrnehmung dieser mit dem Präfix „Hör-“ gekennzeichnet.

2.1. Das menschliche Ohr

In Abbildung 1 wird schematisch das menschliche Ohr mit den für das Hören relevanten Bereichen dargestellt. Es besteht aus drei verschiedenen Segmenten: Dem Außen-, dem Mittel- und dem Innenohr. Das Außenohr hat die generelle Funktion eines Antransportorgans, also dem Erfassen von möglichst viel Schallenergie aus den interessierenden Richtungen.² Durch ihre individualspezifischen Merkmale ist die Ohrmuschel für die Lokalisation besonders wichtig.³ Auf diese wird in Kapitel 2.3. eingegangen. Das Mittelohr besteht aus Trommelfell und Gehörknöchelchen. Im Innenohr befinden sich die Endorgane des Gehörsinns sowie die des Gleichgewichtssinns, die Vestibularorgane.

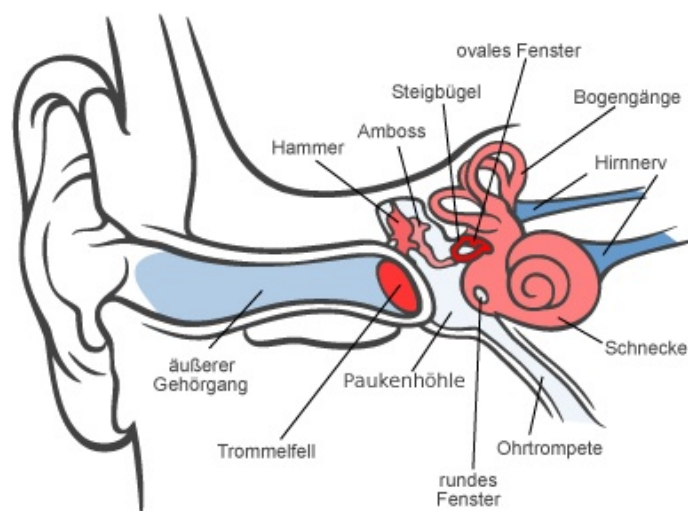


Abbildung 1: Das menschliche Ohr

¹ (Vgl. Blauert, 1974, S. 5)

² (Vgl. Zwicker, 1982, S. 21)

³ (Vgl. Blauert, 1974, S. 4)

Trifft Schall auf das Außenohr, wird er durch den Gehörgang zum Trommelfell weitergeleitet. Dieses erfüllt mit der Paukenhöhle im Mittelohr die Rolle eines Schallempfängers. Die Schwingungen, denen das Trommelfell ausgesetzt wird, werden anschließend weiter auf die Gehörknöchelchen übertragen.⁴ Hammer, Amboss und Steigeisen leiten die Schwingungen in das eigentliche Hörorgan, die Cochlea (Schnecke) weiter, die mit Flüssigkeit gefüllt ist. Durch die Schwingungen werden die sich auf der Basilarmembran befindenden Flimmerhärchen gebeugt. Diese Bewegungen werden von den inneren Haarzellen in elektrische Impulse umgewandelt und an den Hörnerv weitergeleitet.⁵

Eine weitere Möglichkeit für den Transport von Schallereignissen zum Hörnerv stellt der Knochenschall dar. Hier trifft der Schall auf das Schläfenbein, dessen Vibration den Gehörgang anregt. Er ist angesichts der deutlich höheren Intensität des Schalls, der direkt auf das Trommelfell trifft, allerdings meistens zu vernachlässigen.⁶

2.2. Hörgewohnheiten

Das Hören und Auswerten von Schallsignalen ist ein Prozess, der von Gelerntem geprägt ist. Durch zuvor Erlebtes ordnet der Mensch, je nach Assoziation, Schallereignisse unterschiedlich ein. Dabei können Erfahrungen oder Erwartungen als Referenz dienen oder sogar das Hörereignis verändern.⁷

2.2.1. Entfernungshören

Die Hörereignisentfernung bezeichnet die Entfernung zwischen der Schallquelle und dem Mittelpunkt der Achse zwischen den Ohren. Der Mensch ist in der Lage, die Entfernung von Schallereignissen aufgrund seiner Erfahrung, aber auch anhand von verschiedenen Parametern richtig einzuordnen. Eine große Rolle spielt hier das Wiedererkennen von Signalen. Geräusche, die das Gehirn kennt und mit bestimmten Situationen verknüpft, werden unabhängig von der tatsächlichen Entfernung eher dort lokalisiert, wo der Hörer sie durch seine Vorkenntnisse einordnet. So wird

⁴ (Vgl. Blauert, 1974, S. 42f.)

⁵ (Vgl. Görne, 2008, S. 105 ff.)

⁶ (Vgl. Blauert, 1974, S. 43)

⁷ (Vgl. Begault, 2000, S. 29)

beispielsweise geflüsterte Sprache oder das Klappern von Scheren meistens nah am Kopf verortet.⁸⁹ Der gegenteilige Effekt findet sich bei übertrieben lautem Sprechen oder Schreien. Hier wird die Entfernung aufgrund der Vorkenntnisse eher als zu groß eingeschätzt. Handelt es sich um neutralere und noch vertrautere Signale, entspricht die Schätzung durch die Versuchsperson dagegen meist sehr genau der Schallquellenentfernung. Ein Beispiel für solche Signale ist menschliche Sprache bei normaler Lautstärke, die der Mensch bereits in unterschiedlichen Entfernungen wahrgenommen hat.¹⁰

Im freien Schallfeld gilt das 1/r-Gesetz:

$$p \sim \frac{1}{r}$$

p = Schalldruck

r = Abstand zur Schallquelle

Das Gesetz besagt, dass der Schalldruck umgekehrt proportional mit der Entfernung zur Schallquelle abnimmt. Dies bedeutet eine Halbierung der Lautstärke, also einem Absinken um 6 dB, die das Ohr registrieren kann und bei zeitlich stationären Signalen zum Entfernungshören nutzt.¹¹ Für die richtige Entfernungsortung ist aber vorausgesetzt, dass die Lautstärke der Schallquelle dem auditiven System in bestimmten Entfernungen bekannt ist und eingeordnet werden kann.¹²

Ein weiterer Parameter ist die Dämpfung des Signals durch die Luft. Bei Entfernungen von über 15 Metern wird das Signal entfernungs- und frequenzabhängig verzerrt. Dabei werden höhere Frequenzen absorbiert, sodass sich die Klangfarbe des Signals verändert.¹³ Eine andere Klangfarbenveränderung stellt sich bei besonders kleinen Entfernungen zur Schallquelle ein: Durch Abschattungen oder frequenzabhängige Beugung von Signalen durch den Kopf und Oberkörper kommt es zu akustischen Veränderungen.¹⁴ In geschlossenen Räumen kann auch durch Reflexionen und

⁸ (Vgl. Begault, 2000, S. 29)

⁹ (Vgl. Blauert, 1974, S. 37)

¹⁰ Ebd.

¹¹ (Vgl. Blauert, 1974, S. 96)

¹² (Vgl. Dickreiter et al., 2014, S. 132)

¹³ (Vgl. Blauert, 1974, S. 96)

¹⁴ (Vgl. Dickreiter u.a., 2014, S. 133)

Nachhall auf die Entfernung zurückgeschlossen werden. Allerdings sind auch hier Vorkenntnisse des Raumklangs durch den Hörer notwendig.

2.2.2. Im-Kopf-Lokalisation

Eine Besonderheit im Entfernungshören stellt die „Im-Kopf-Lokalisation“ (IKL)¹⁵ dar. Natürliche Schallquellen werden gemäß ihrer Position außerhalb des Kopfes lokalisiert. Bei der Wiedergabe von Schallquellen durch Kopfhörer oder Stereolautsprecher kann es allerdings zur Im-Kopf-Lokalisation kommen. Das Signal wird dann nicht mehr außerhalb des Kopfes verortet, sondern scheint aus dem Inneren des Kopfes zu kommen.¹⁶ Die Ursache dafür liegt in den Hörgewohnheiten. Wie in Kapitel 2.2.1. erwähnt, werden Signale, die vom Hörer auszuwerten sind, mit bereits bekannten verglichen. Bei unbekannten Reizmustern oder auch bei Signalen, deren Ursache visuell nicht ausgemacht werden kann, kann das Gehirn das Signal nicht zuordnen und lokalisiert es innerhalb des Kopfes. Dieser Fall tritt bei diotischer Stereokopfhörerwiedergabe auf, einer Wiedergabe, bei der auf beide Ohren das identische Signal gegeben wird. Hier werden sowohl die spektralen Verfärbungen der Ohrmuschel als auch die interauralen Unterschiede ausgeschaltet.¹⁷ Weitere Erklärungsversuche für die IKL waren der fehlende Knochenschall bei einer Wiedergabe über Kopfhörer oder Lautsprecher, die fehlende Dynamik der Wiedergabe und das statische Klangbild.¹⁸ Zusammenfassend lässt sich sagen, dass die beiden Stereosignale sowohl sehr ähnlich sein müssen, sowie der Hörereignisabstand kleiner sein muss als der Kopfradius des Hörers damit die IKL auftritt.¹⁹ Dass dieser Effekt bei diotischer Kopfhörerwiedergabe auftritt, kann wie bereits genannt daran liegen, dass die kopfbezogene Übertragungsfunktion des Kopfes ausgeschaltet wird, durch die ein Großteil unserer Lokalisation stattfindet.

¹⁵ (Dickreiter u.a., 2014, S. 133)

¹⁶ (Vgl. Dickreiter u.a., 2014, S. 133f.)

¹⁷ (Vgl. Nicol, 2010, S. 30)

¹⁸ (Vgl. Nicol, 2010, S. 30)

¹⁹ (Vgl. Blauert, 1974, S. 95 u. 108)

2.3. Lokalisation durch die Head-Related-Transfer-Function (HRTF)

„Der Mensch ist vorwiegend ein Augenwesen.“²⁰ Dementsprechend ist die Lokalisation von Schallereignissen schlechter als die visuelle Lokalisation.²¹ Blauert definiert die Lokalisation als Zuordnungsgesetz zwischen dem Ort eines Hörereignisses und dem Merkmal eines Schallereignisses. Die Lokalisationsunschärfe bezeichnet die Veränderung eines Merkmals des Schallereignisses, das gerade zu einer Veränderung des Hörereignisses führt.²² Die Einordnung von Schallquellen durch das Gehör ist schlechter als der Schallraum selbst, man spricht von verschmierter Lokalisation.²³ Da der Mensch in verschiedene Richtungen unterschiedlich gut lokalisiert, verändert sich auch die Lokalisationsunschärfe. Die maximale Lokalisation und somit minimale Lokalisationsunschärfe von 1° erzeugt das auditive System bei Schallereignissen aus der Vorwärtsrichtung.²⁴

Die Ortung von Schallereignissen durch das Gehör findet durch monaurale sowie binaurale Cues statt. Zu diesen zählen unter anderem Klangfärbung, interaurale Laufzeitdifferenzen sowie interaurale Pegeldifferenzen.²⁵ In der Summe bilden sie die Übertragungsfunktion des Kopfes und der Ohrmuschel, die so genannte Head-Related-Transfer-Function (HRTF). Die HRTF bildet die Basis der kopfbezogenen Wiedergabetechnik. Sie beschreibt die individuellen Verzerrungen durch Kopfform, Torso und der Ohrmuschel (Pinna) sowie interaurale Differenzen zu jeder Schallquelle, die im dreidimensionalen Raum platziert wird.²⁶ Das Gehör kann diese Transferfunktion dekodieren, um Rückschlüsse auf die Position des Schallereignisses zu bekommen.²⁷ Um die richtige Lokalisation zu gewährleisten, wird je nach Ursprung des Schallereignisses auf andere Cues zurückgegriffen. Dies hängt davon ab, auf welcher Ebene des Hörens sich das Schallereignis abspielt. Im Folgenden sollen nun erst das kopfbezogene Koordinatensystem und anschließend die Wirkungsweise der unterschiedlichen Lokalisationsparameter vorgestellt werden.

²⁰ (Blauert, 1974, S. 1)

²¹ (Vgl. Blauert, 1974, S. 31)

²² Ebd.

²³ (Vgl. Blauert, 1974, S. 30)

²⁴ (Vgl. Blauert, 1974, S. 31)

²⁵ (Vgl. Møller, 1992, S. 176)

²⁶ (Vgl. Nicol, 2010, S. 16)

²⁷ (Vgl. Nicol, 2010, S. 11)

2.3.1. Die Ebenen des Hörens

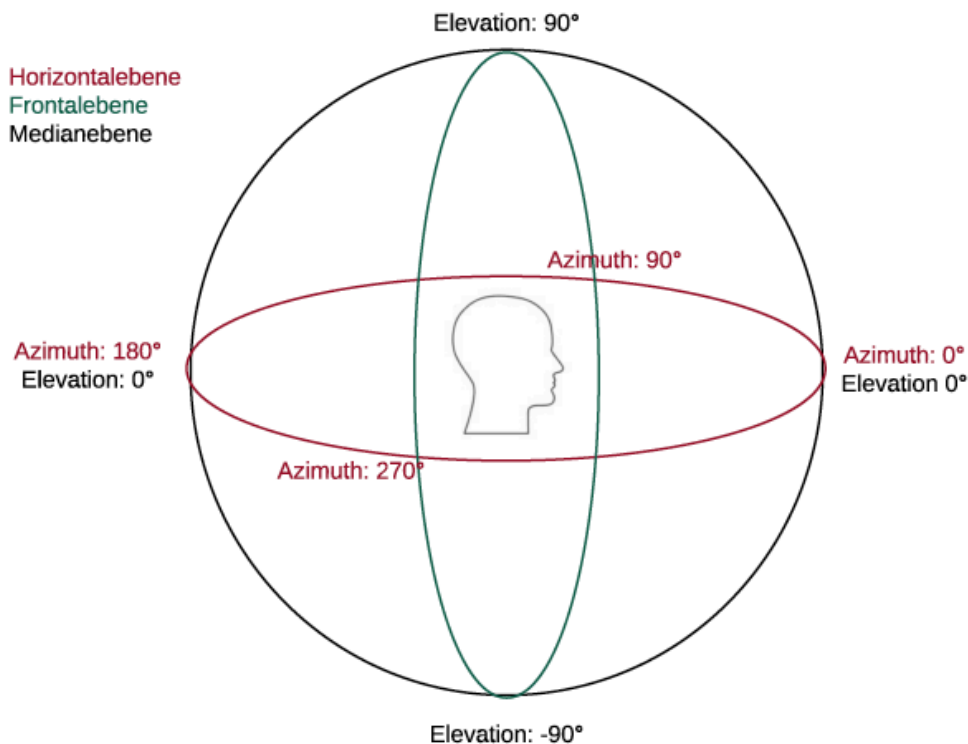


Abbildung 2: Ebenen des Hörens

Um Angaben über die Herkunft von Schallquellen zu machen, bedient man sich des kopfbezogenen Koordinatensystems und den Parametern δ für den Erhebungswinkel, φ für den Seitenwinkel oder Azimuth und r für die Entfernung zur Schallquelle. Dieses System wird aus drei Achsen bzw. Hörebenen gebildet: der Frontalebene zur Unterscheidung von vorne und hinten, der Horizontalebene zur Unterscheidung von rechts und links und der Medianebene, zur Einschätzung des Erhebungswinkels. Der Ursprung des Koordinatensystems liegt auf der Mitte der beiden Oberkanten der Gehörgangseingänge. Diese Verbindung spannt mit den Unterkanten der Augenhöhlen die Horizontalebene auf. Die Frontalebene steht senkrecht auf der Horizontalebene ebenfalls auf der Verbindung der Gehörgangseingänge. Mit diesen beiden Ebenen bildet die Medianebene einen rechten Winkel. Sie stellt bei einem symmetrischen Kopf die Symmetrieachse dar.²⁸

²⁸ (Vgl Blauert, 1974, S. 11)

Die beste Lokalisation erfolgt auf der Horizontalebene, die schlechteste mit einer Lokalisationsunschärfe von bis zu 17° auf der Medianebene.²⁹ Die Gründe dafür werden in Kapitel 2.3.3. erläutert. Auch in diesem Bereich gilt, dass bekannte Signale gegenüber unbekannten besser lokalisiert werden können.³⁰

2.3.2. Monaurale Cues

Einen Teil der HRTF stellen die monauralen Cues dar, die Kenngrößen, die auch mit nur einem Ohr empfangen und ausgelesen werden können. Diese entstehen aufgrund der Reflexion und Beugung von Schallwellen durch den Kopf, den Oberkörper sowie die Ohrmuschel. Frequenzen unterhalb von 3 kHz stehen in Abhängigkeit zu der Beugung durch Kopf und Oberkörper, Frequenzen darüber zu der Verzerrung durch die Ohrmuschel³¹. Durch ihre besondere Form bildet das Außenohr ein komplexes Resonanzsystem und ist in der Lage, das empfangene Schallsignal durch spektrale Veränderung stark zu prägen.³² Dies geschieht durch Reflexion, Beugung, Abschattung, Interferenzen, Streuung und Resonanzen.³³ Das Außenohr kann dabei durch die auftretenden Kammfiltereffekte je nach Einfallrichtung des Schalls Pegeldifferenzen von bis zu 20 dB herbeiführen.³⁴

Die Verzerrungen des Schallsignals können durch die Erfahrungswerte des Gehörs zur Richtungswahrnehmung genutzt werden. Monaurale Cues sind dann wirkungsvoll, wenn das Signal bekannt ist. Zusätzlich können länger andauernde Signale und auch breitbandigere besser lokalisiert werden als beispielsweise nur ein kurzer Pistolenschuss oder ein reiner Sinuston.³⁵

Die spektralen Cues sind vor allem für die vertikale Lokalisation von Bedeutung. Da hier Merkmale zwischen den Ohren nicht ausgewertet werden können, entspricht die Lokalisation über die spektralen Cues fast dem Hören auf der Medianebene. Neben ihrer geringen Bedeutung für die Lateralisation, der Lokalisation auf der

²⁹ (Vgl. Blauert, 1974, S. 35)

³⁰ Ebd.

³¹ (Vgl. Guillon, 2009, S. 65)

³² (Vgl. Nicol, 2010, S. 20)

³³ (Vgl. Blauert, 1974, S. 51)

³⁴ (Vgl. Görne, 2008, S. 103)

³⁵ (Vgl. Blauert, 1974, S.80)

Horizontalebene, kann durch die spektralen Kenngrößen der Erhebungswinkel abgeschätzt werden.³⁶ Je nach vertikaler Auslenkung werden verschiedene Frequenzen durch die Kopf- und Pinnaform unterschiedlich angehoben oder herausgefiltert. Das Gehör nimmt diese zwar nicht aktiv wahr, kann aber die Richtungsinformation verarbeiten und das Schallereignis entsprechend lokalisieren.³⁷

Jede Pinna trägt individualspezifische Merkmale und ist dabei so einzigartig wie ein Fingerabdruck. Diese HRTF und auch deren Unterschiede von Mensch zu Mensch können visuell dargestellt werden. Wie unterschiedlich diese sein können, zeigt Abbildung 3. Dort werden die gesammelten HRTFs von zehn verschiedenen Versuchspersonen bei einem frontal auftreffenden Schallsignal dargestellt.

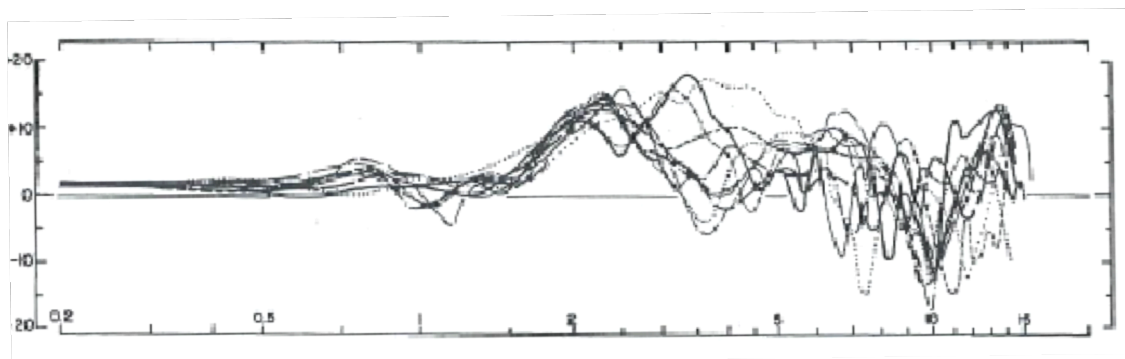


Abbildung 3: HRTF-Kurven von 10 Personen bei frontalem Schalleinfall

Dennoch lassen sich verschiedene Gemeinsamkeiten erkennen, die aus dem generellen Aufbau der Pinna resultieren und zur damit verbundenen Lokalisation führen. Das offensichtlichste ist die Helix, also die äußere Ohrmuschel. Wird diese umgedreht, indem man mit den Händen die Vorwärtsrichtung abschirmt, erscheint das Hörereignis im Rücken.³⁸

Auch bei den spektralen Verzerrungen können interindividuelle Gemeinsamkeiten beobachtet werden. Frequenzen unterhalb von 3,5 – 4 kHz sind kaum von der Verzerrung durch die Ohrmuschel betroffen.³⁹ Dazu ist die Wellenlänge dieser Frequenzen zu groß. Eine Beeinflussung durch die Pinna erfolgt erst, wenn $\frac{1}{4}$ der

³⁶ (Vgl. Blauert, 1974, S. 84)

³⁷ (Vgl. Dickreiter u.a., 2014, S. 133)

³⁸ (Vgl. Blauert, 1974, S. 80)

³⁹ (Vgl. Nicol, 2010, S. 22)

Wellenlänge den Abmessungen des Ohres entspricht.⁴⁰ Guillon beobachtete drei große Ausschläge in der Übertragungsfunktion des Ohres: einen zwischen 3,5 und 5,5 kHz, auf der vorderen ipsolateralen Seite. Ein weiterer Ausschlag besteht zwischen 8 und 10 kHz unterhalb der Horizontalebene und ein dritter zwischen 11 und 14 kHz in Vorwärtsrichtung. Ebenso wurden drei Einbrüche der Lautstärke beobachtet: der erste zwischen 5 und 10 kHz, ein weiterer zwischen 10 und 12 kHz und ein dritter, allerdings sehr individuell abhängiger, zwischen 16 und 17 kHz.⁴¹

Die Gemeinsamkeiten in der Frequenzkurve sind von der Ohrgröße und der Ausrichtung abhängig. Ausschläge und Einbrüche der HRTF sind lediglich frequenzverschoben. Middlebrooks entwickelt dazu ein Scaling-Modell, durch das die HRTFs von verschiedenen Menschen je nach Ohrmuschelgröße angepasst werden können.⁴²

Die monauralen spektralen Cues sind sehr individuell, unterliegen aber den genannten Grundmustern und weisen so auch interindividuell Gemeinsamkeiten auf. Sie dienen zur Ortung auf der Medianebene und basieren auf der Hörerfahrung. Zur vertikalen Lokalisation werden die monauralen Cues genutzt. Für die horizontale Ortung, auch Lateralisation genannt, bedient sich das Ohr der binauralen Unterschiede.

2.3.3. Binaurale Cues

Binaurale Cues bezeichnen die Unterschiede der Hörereignisse, die zwischen den Ohren entstehen. Diese Unterschiede treten auf, sobald sich das Schallereignis aus der Medianebene hinaus bewegt, der Einfallswinkel dementsprechend ungleich 0° ist.⁴³ Innerhalb der interauralen Differenzen, die für die Ortung wichtig sind, unterscheidet man zwischen drei verschiedenen Kenngrößen: den interauralen spektralen Cues, den interauralen Pegeldifferenzen und den interauralen Laufzeitdifferenzen. Diese sollen im Folgenden vorgestellt werden.

⁴⁰ (Vgl. Blauert, 1974, S. 51)

⁴¹ (Vgl. Guillon, 2009, S. 53f.)

⁴² (Vgl. Nicol, 2010, S. 56)

⁴³ (Vgl. Dickreiter u.a., 2014, S. 128)

2.3.3.1. Interaurale spektrale Cues

Die interauralen spektralen Cues treten ebenso wie die monauralen spektralen Cues aufgrund der Diffraktion und Reflexion durch Kopf, Torso und Ohrmuschel auf und können zur vertikalen sowie zur horizontalen Lokalisation dienen. Sie sind abhängig von Erhebungs- und Einfallswinkel, da die anatomischen Gegebenheiten die Schallsignale entsprechend verändern. Die jeweiligen Schallsignale werden je nach Einfallswinkel an beiden Ohren unterschiedlich verzerrt und gefiltert. Diese Differenzen zwischen den Ohrsignalen werden dann vom Gehör ausgewertet. Das auditive System kann so die Differenzen der Filterung analysieren und die Herkunft des Schallsignals ermitteln.⁴⁴ Dennoch wird angenommen, dass die spektralen Cues eher der Ermittlung des Erhebungswinkels dienen, als dass sie eine wichtige Rolle in der Lateralisation spielen.⁴⁵

2.3.3.2. Interaurale Pegeldifferenzen (ILD)

Zu einer der elementaren Kenngrößen der horizontalen Lokalisation gehören die interauralen Pegeldifferenzen. Erforscht wurden sie erstmals von John William Strutt, besser bekannt als Lord Rayleigh, der sie in der Duplex-Theorie verarbeitete.⁴⁶ Gemeinsam mit den Interauralen Laufzeitdifferenzen (ITD) sollten sie für das räumliche Hören des Menschen verantwortlich sein. Mittlerweile wurde auch die Rolle der Pinna gerade für die vertikale Lokalisation entdeckt, die Duplextheorie bildet allerdings nach wie vor die Basis für die Lateralisation.

Interaurale Pegeldifferenzen beschreiben den Unterschied des Schalldruckpegels zwischen den beiden Ohren. Dabei geht es nicht um den Pegelunterschied der sich aufgrund der längeren Strecke entwickelt, die von einem Ohr zum anderen vom Schall zurückgelegt werden muss. Die Differenz zwischen den Signalen entsteht durch die Abschattung der Schallquelle durch den Kopf. Der Einfluss der zurückgelegten Strecke ist im Vergleich zu den Abschattungseffekten, die Differenzen von bis zu 20 dB verursachen können, nicht relevant.⁴⁷ Da Signale je nach der Wellenlänge

⁴⁴ (Vgl. Schmicking, 2003, S. 126)

⁴⁵ (Vgl. Vassilakis, n.d.)

⁴⁶ (Vgl. Ansorge und Leder, 2011, S. 126)

⁴⁷ (Vgl. Plack, 2005, S. 178)

unterschiedlich stark gebeugt und reflektiert werden, ist die interaurale Pegeldifferenz frequenzabhängig. Tiefe Frequenzen bis ca. 300 Hz, deren Wellenlängen gegenüber dem Kopf groß sind, werden durch ihn oder den Torso lediglich gebeugt. Es kommen hier Pegelunterschiede von ca. 1 dB zu Stande, die vom Gehör bemerkt werden können. Ab dieser Grenze wächst die ILD mit zunehmender Frequenz und Auslenkung.⁴⁸ Höhere Frequenzen werden durch den Kopf reflektiert und es kommt zum Druckstau an dem der Schallquelle zugewandten Ohr. Parallel dazu entsteht am abgewandten Ohr ein Schallschatten durch den Kopf des Hörers wodurch die ILD ansteigt.⁴⁹ Die Intensitätsstereophonie bedient sich des gleichen Prinzips. Der Kopf ist also nicht nur ein Hindernis oder ein Dämpfer von Frequenzen, sondern hat hier auch eine verstärkende Funktion.

Die Lateralisation über ILDs unterliegt ebenfalls den Erfahrungen und Hörgewohnheiten. Zudem kann das Gehör „ermüden“. Findet die Beschallung des Gehörs durch eine Schallquelle konstant von einer Seite statt, gewöhnt sich das Gehör an das Signal und das Hörereignis rückt mehr in Richtung Medianebene.⁵⁰

2.3.3.3. Interaurale Laufzeitdifferenzen (ITD)

Interaurale Laufzeitdifferenzen sind der zweite elementare Teil der Duplextheorie. Diese entstehen durch die Strecke, die der Schall zurücklegen muss, um das der Schallquelle abgewandte Ohr zu erreichen. Der Schall trifft an diesem Ohr später auf, die dadurch entstandene Verzögerung kann das auditive System auswerten und die Information zur Lateralisation nutzen. Das Hörereignis rückt normalerweise auf die Seite, auf der das Schallereignis zuerst eintrifft. Das Hörereignis kann auf die entgegengesetzte Seite rücken, wenn die Wellenlänge dem Kopfdurchmesser entspricht.

Die dazu erforderlichen Abläufe finden in der „medium superior olive“ (MSO) im Gehirn statt. Dort werden die Ankunftszeiten des Signals durch ein Neuronenfeld miteinander verglichen und ausgewertet. Das Neuron, auf dem beide Ohrsignale gleichzeitig antreffen, wird stark angeregt. Findet dies in der linken Hemisphäre des

⁴⁸ (Vgl. Dickreiter u.a., 2014, S. 130)

⁴⁹ (Vgl. Görne, 2008, S. 119)

⁵⁰ (Vgl. Blauert, 1974, S. 131)

Gehirns statt, ordnet das Gehirn das Hörereignis auf die rechte Seite ein und umgekehrt.⁵¹ Auch dieser Mechanismus basiert auf Hörerfahrung, da die Neuronen mit der Zeit mit anderen Gehirnarealen verknüpft werden und so zum Beispiel lernen, Echosignale auszublenden.⁵²

Bei einem durchschnittlichen Kopfdurchmesser von 17 cm und der größtmöglichen seitlichen Auslenkung von 90° entsteht ein maximaler Weg von 21 cm, den der Schall auf dem Weg zum abgewandten Ohr zurücklegen muss und damit eine Laufzeitdifferenz von 0,61ms. Diese variiert je nach Kopfdurchmesser und anatomischen Gegebenheiten. Wie in 2.3.1. erwähnt, erreicht das Gehör die minimale Lokalisationsunschärfe in der Horizontalebene. Im Bezug auf die ITD ist das auditive System bereits in der Lage, eine Veränderung des Schallereignisses von 0.01ms aus der Vorwärtsrichtung zu registrieren und auszuwerten. Rayleigh betrachtete die interauralen Laufzeitdifferenzen unter dem Aspekt der Beschallung durch reine Töne. Das menschliche Gehör ist im Alltag allerdings selten reinen Tönen ausgesetzt, dementsprechend gelten Rayleighs Theorien im Alltag nur unter Einschränkungen. Das Spektrum des Signals spielt für die Ortung eine entscheidende Rolle.⁵³

Bei reinen Tönen kann es zu einer Mehrdeutigkeit in der Lokalisation kommen. Beispiel dafür ist ein reiner Ton, der aus einem Winkel von 90° und einer einzigen Frequenz auf die Ohren trifft, so dass an beiden ein phasengleiches Signal auftritt. Liegen keine Phasenunterschiede vor, kann das Signal durch den mangelnden Phasenunterschied frontal verortet werden.⁵⁴ Enthält das Signal hingegen mehrere Frequenzen, gestaltet sich die Ortung einfacher. Hierbei dienen insbesondere die tieferen Frequenzen zur eindeutigen Lateralisation. Die ITD besitzt eine frequenzspezifische Charakteristik. Laufzeitunterschiede bei höheren Frequenzen betragen lediglich 2/3 des tieffrequenten Werts.⁵⁵ Somit kann das Gehör aus Frequenzen bis zu 800 Hz, bzw. höchstens aber aus Frequenzen bis 1600 Hz, Rückschlüsse auf den Schallereignisort ziehen.⁵⁶ Grund dafür ist, dass die Wellenlänge in den tiefen Frequenzen größer als der

⁵¹ (Vgl. Plack, 2005, S. 180 f.)

⁵² (Vgl. Siveke, 2006, S. 30)

⁵³ (Vgl. Görne, 2008, S. 118 f.)

⁵⁴ Ebd.

⁵⁵ (Vgl. Nicol, 2010, S. 27)

⁵⁶ (Vgl. Dickreiter u.a., 2014, S. 129)

Kopfdurchmesser ist. Erreichen diese den Kopfdurchmesser, kann das Gehör keinen eindeutigen Phasenunterschied feststellen und so auch keine korrekte Lateralisation mehr vornehmen. Je nach Phasendifferenz kann es sogar zu einer Fehllocalisation auf die entgegengesetzte Seite kommen, wie Abbildung 4 verdeutlicht.⁵⁷

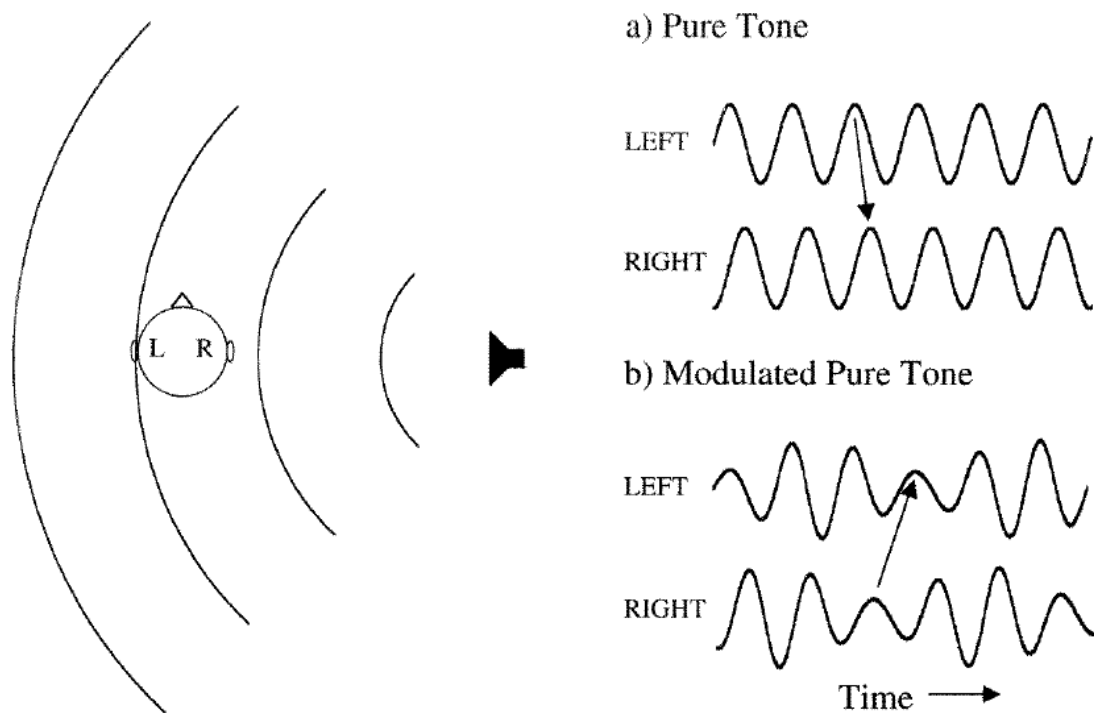


Abbildung 4: Fehllocalisation bei reinen Tönen

Dennoch gibt es Möglichkeiten für das Gehör, auch bei hohen Frequenzen durch die Phasenunterschiede eine korrekte Lateralisation vorzunehmen. Bei hohen reinen Tönen kann das auditive System die Hüllkurve der Frequenz analysieren und die Phasendifferenz auswerten.⁵⁸ Dagegen ist ein breitbandiges Signal gerade wegen seiner verschiedenen Frequenzen deutlich leichter auszuwerten als ein reiner Sinuston.

In der Duplextheorie wird das Zusammenwirken der beiden interauralen Kenngrößen definiert. Die ITD ist für die Lateralisation von tiefen Frequenzen, die ILD für die von höheren Frequenzen zuständig. Dies stimmt zwar für reine Töne, allerdings haben die meisten hochfrequenten Signale niederfrequente Anteile, die durch die ITD

⁵⁷ (Vgl. Plack, 2005, S. 177)

⁵⁸ (Vgl. Blauert, 1974, S. 122)

ausgewertet werden können. Daher dominiert bei hohen Frequenzen die Lateralisation durch die Interaurale Laufzeitdifferenz, die über das gesamte Spektrum hinweg Informationen herausfiltern kann.⁵⁹

Um die Möglichkeiten einer Wiedergabe darzustellen, die sich an den Erkenntnissen über das menschliche Hören orientiert, soll ein Überblick über den technischen Fortschritt zu einer realistischen Abhörumgebung gegeben werden.

⁵⁹ (Lindau u. a., 2010, S. 2)

3. Annäherung an ein realistisches Klangbild

Durch Fortschritte in der Technik hat sich die Tonwiedergabe über die Jahre geändert und verbessert. Auch der Verwendungszweck von Tonaufnahmen und –wiedergaben hat sich gewandelt: von einem Monosignal, das allein Informationszwecken diene, zu einer möglichst natürlich klingenden mehrkanaligen Wiedergabe, die neben der reinen Information auch kreative Elemente der Bearbeitung enthält.

Die Möglichkeiten der Distribution haben sich mit der Entwicklung von mehrkanalfähigen Medien vervielfältigt. Auch das Interesse an hochwertigen Soundanlagen für den Heimgebrauch ist mit ca. 20 Millionen HiFi-Anlagenbesitzern in Deutschland konstant hoch. Hochwertiger Ton und gute Wiedergabequalität sind inzwischen leichter zugänglich und nehmen einen hohen Stellenwert ein.⁶⁰

Im Folgenden soll nun ein Überblick über die technischen Möglichkeiten der Wiedergabetechnik gegeben werden.

3.1. Stereophonie

Der Begriff Stereophonie kommt aus dem griechischen und setzt sich aus den Wörtern „stereo“, plastisch, und „phonos“, der Ton, zusammen. Der Begriff beschreibt ein plastisches Hören. Damit wird ein Hörereignis suggeriert, das so natürlich und realistisch wie möglich klingen soll.⁶¹ Stereophonie bezeichnet damit theoretisch Zweikanal-, Vierkanal- oder generell Mehrkanalstereosysteme. In den 50er Jahren wurde der Begriff allerdings als Gegenpol zur einkanaligen Monophonie zum Synonym für zweikanalige Wiedergabetechnik und hat sich als solches bis heute im Sprachgebrauch gehalten.

Bei Stereophonie geht es um ein Übertragungssystem mit zwei Kanälen, also Informationsübermittlung durch einen linken und einen rechten Kanal. Dieses Mehrkanalsystem soll die Aufnahme räumlicher wirken lassen, die Möglichkeit zu einem breiteren Arrangement und damit zu mehr Durchsichtigkeit geben und ein rundes, volleres Hörereignis zustande bringen. Mittels Laufzeit- oder Intensitätsstereophonie können die dabei verwendeten Richtungsinformationen durch

⁶⁰ (IfD Allensbach, 2015)

⁶¹ (Vgl. Birkner, 2002, S. 13)

zwei Kanäle abgebildet werden. Die genannten Stereophonie-Methoden funktionieren ähnlich wie die in Kapitel 2.3.3. vorgestellten, interauralen Kenngrößen. Bei der Laufzeitstereophonie handelt es sich um das verzögerte, bei der Intensitätsstereophonie um das leisere Abspielen eines Signals auf einem Kanal.⁶² Durch die Aufstellung der beiden Lautsprecher, die mit dem Hörer ein gleichschenkliges Dreieck bilden sollten, kommt es zu Phantomschallquellen. Die Signale werden also nicht als zwei Schallereignisse, sondern ohne Phasen- oder Pegeldifferenzen als ein einziges Hörereignis in der Mitte der beiden Lautsprecher wahrgenommen. Dieses kann je nach Wunsch nach rechts oder links ausgelenkt werden und simuliert so ein räumliches Schallereignis.⁶³

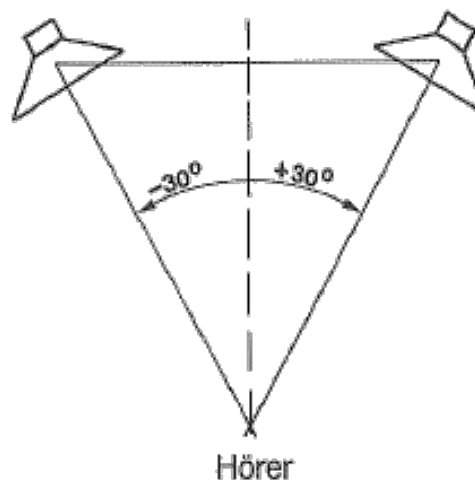


Abbildung 5: Stereo-Abhörumgebung

Stereophonie kann bei der Aufnahme oder bei der Nachbearbeitung erzeugt werden. Dies geschieht etwa für Laufzeitstereophonie mittels AB-Verfahren, für Intensitätsstereophonie mittels XY- oder MS-Verfahren. Hier gelten die gleichen Grundsätze wie bei den interauralen Kenngrößen. Im AB-Verfahren wird zur Darstellung der Stereobasis der Weg genutzt, den die Schallwelle von einem zum anderen Mikrophon zurücklegt. Für die Intensitätsstereophonie nutzt man die Pegel- und Phasendifferenzen zwischen den Druckempfängern.

Für Intensitätsstereophonie werden bei der XY-Stereophonie zwei Richtmikrophone gegeneinander angewinkelt übereinander montiert. So erreicht der Schall zwar beide

⁶² (Vgl. Görne, 2008, S. 297)

⁶³ (Vgl. Dickreiter u. a., 2014, S. 221)

Mikrophone gleichzeitig, unterscheidet sich aber im Pegel. Das MS-Verfahren ist ähnlich dem XY-Verfahren, wird aber zwingend mit einer Achtcharakteristik und einem beliebig wählbaren Richtmikrophon aufgezeichnet. Diese Arten der Stereomikrophonie lassen sich durch Matrizierung ineinander überführen. In Formaten wie ORTF oder NOS werden beide Methoden miteinander kombiniert. Diese werden als Äquivalenzstereophonie bezeichnet.⁶⁴

Weitere Möglichkeiten der Stereophonierung, auch in ihrer ursprünglichen Wortbedeutung, bietet die vierkanalige Quadrophonie. Diese konnte sich allerdings wegen Streitigkeiten über das Format und auch aus Mangel an Übertragungsmedien nicht durchsetzen.⁶⁵ Stattdessen setzte sich das Surround-System durch.

3.2. Surround

Die nächste Stufe der Mehrkanalstereophonie ist das Surround-System. Surround Sound bezeichnet zunächst den umgebenden Klang, also den Versuch, ein möglichst einhüllendes und dadurch noch realistischeres Klangbild zu schaffen. Die Entwicklung entstand im Kinobereich, gerade dort sollten die Zuschauer von dem Geschehen auf der Leinwand - unterstützt durch den realitätsgetreuen Ton - in eine fiktive Welt entführt werden. Nach Zwischenlösungen mit drei Frontkanälen und einem Effektkanal entstand das Todd-AO-70mm-Format mit fünf Frontkanälen sowie einem Surroundkanal. Daraus entwickelte sich das Dolby 5.1-Format, aus dem Tomlinson Holman wiederum für Lucasfilm und die „Star Wars“-Filme das THX-System kreierte. Mit dem digitalen Sechskanalverfahren von Dolby Digital und der Verbreitung von DVD-Datenträgern wurde das Surround-System auch für zu Hause zugänglich. Hier werden die sechs Kanäle wie folgt verteilt: drei Front- und zwei Surroundkanäle sowie ein Signal für einen Subwoofer.⁶⁶ Der Vorteil des 5.1-Systems war, dass sich ein Klangbild relativ realistisch durch zwei Surroundkanäle im Hintergrund abbilden ließ und die Hörzone (auch „Sweet Spot“ genannt) durch die Verteilung der Lausprecher vergrößert und stabilisiert wurde.⁶⁷ Weiterentwicklungen fanden durch das 7.1-

⁶⁴ (Vgl. Görne, 2008, S. 303 ff.)

⁶⁵ (Vgl. Birkner, 2002, S. 13)

⁶⁶ (Vgl. Birkner, 2002, S. 13f.)

⁶⁷ (Vgl. Theile und Wittek, 2011, S. 2 f.)

Verfahren durch Sony statt. Dabei existieren fünf Frontkanäle, zwei Surround-Kanäle sowie ein Subwoofer.

Die Nachteile des Surround Verfahrens bestehen in der Verkleinerung des Sweet Spots. Ein Abweichen von dieser idealen Hörposition kann eine Verschlechterung des Klangbilds von bis zu 80 % bedeuten, in der zweikanaligen Stereoabhöre beträgt diese Abnahme nur 50 %. Im Kinobereich konnte man dieser Entwicklung mittels Dekorrelation entgegenwirken. Dabei werden Surroundsignale auf mehrere Lautsprecher verteilt, so dass zwar die Lokalisationsunschärfe zunimmt, der Sweet Spot allerdings ebenfalls vergrößert werden kann. Zu Hause wird diese Aufgabe schwieriger: die Räume sind deutlich kleiner und können auch nicht speziell auf die akustischen Anforderungen ausgerichtet werden, da sie meist auch noch anderen Zwecken dienen. Die Wiedergabe erfolgt bestenfalls über fünf kreisförmig angeordnete Lautsprecher. So wird der Sweet Spot relativ klein und ist meist nur für ein oder zwei Personen möglich.⁶⁸ Eine Dekorrelation erfolgt im Heimgebrauch mit Delays, welche die Signale zueinander verschieben.⁶⁹

3.3. Auro-3D

Die 5.1-Technik deckt klanglich lediglich die Horizontalebene des Hörens ab. Um auch in der Medianebene für einen umhüllenden Klang zu sorgen, wurde durch Winfried van Baelen das Auro-3D-Format entwickelt. Dieses bezieht durch vier weitere Lautsprecher auch die Höhe mit ein. Baelen stellte ein 9.1-System vor, welches das klassische 5.1-Format durch vier Höhenlautsprecher ergänzt. Diese Basisversion des Auro-3D stellt eine kubusähnliche Anordnung dar, d.h. über allen Kanälen, außer dem Centerkanal, sind insgesamt vier Lautsprecher installiert. Mit diesem Aufbau ist es möglich, frühe Reflexionen aus dem oberen Hallraum sowie den räumlichen Nachhall wiederzugeben. Phantomschallquellen auf der Medianebene zwischen Höhen- und Horizontallautsprechern sowie Hörereignisse direkt über dem Hörer sind allerdings nicht möglich. Dennoch stellt die Auro-3D-Technik eine Möglichkeit dar, ein umhüllenderes, räumlicheres und auch räumlich tieferes Klangbild zu schaffen als es

⁶⁸ (Vgl. Birkner, 2002, S. 45)

⁶⁹ (Vgl. Birkner, 2002, S. 104f.)

die klassische 5.1-Technik vermag.⁷⁰ Aus den bereits genannten Gründen des Raumbedarfs mangelt es dieser Technik allerdings an Praktikabilität für den Heimgebrauch.

3.4. Binauraltechnik

Binaurale Technik bedeutet so viel wie „zweiohrige“ Technik. Hierbei wird versucht, das menschliche Hören mit zwei Ohren, also mit zwei getrennten Signalen, möglichst realistisch nachzuahmen. Die Grundidee liegt in der Trennkörper-Stereophonie. Diese Aufnahmetechnik bildet mit einem Mikrophonabstand von 17,5 cm den Kopfdurchmesser und damit die Entfernung der Ohren voneinander nach. Zwischen beiden Mikrofonen wird eine Trennplatte installiert, die den menschlichen Kopf und dessen Auswirkungen auf den Schall simulieren soll. Dabei gibt es unterschiedliche Varianten, beispielsweise die Kugelflächenmethode und die OSS-Scheibe (Optimales Stereo-Signal-Scheibe). Bei der Kugelflächenmethode werden zwei Mikrophone mit Kugelcharakteristik im 180°-Winkel in eine Kugel montiert. Die ITDs und ILDs, die dabei auftreten, sollen dem natürlichen Hören entsprechen. Bei einer OSS-Scheibe werden ebenfalls zwei Kugel-Mikrophone im geschätzten Kopfabstand von 17,5 cm montiert, allerdings jeweils nur 20° nach außen ausgerichtet. In der Mitte wird eine mit Schaumstoff bezogene Scheibe als Trennkörper installiert. So können der frontale Schall sowie der Diffusschall von hinten und die seitlichen Signale durch die Reflexion der Trennscheibe nahezu verzerrungsfrei aufgenommen werden. Skizzen beider Techniken werden in Abbildung 6 dargestellt. Die Klangfärbungen des natürlichen Hörens, die durch die Pinna entstehen, werden bei beiden Methoden vernachlässigt.⁷¹ Da diese wie oben bereits beschrieben für einen großen Teil unseres Lokalisationsvermögens verantwortlich sind, experimentierte die Firma Neumann in den 1970ern mit exakteren Kopfnachbildungen.⁷²

⁷⁰ (Vgl. Dickreiter u.a., 2014, S. 312ff.)

⁷¹ (Vgl. Birkner, 2002, S. 84f.)

⁷² (Vgl. Dickreiter u.a., 2014, S. 345)

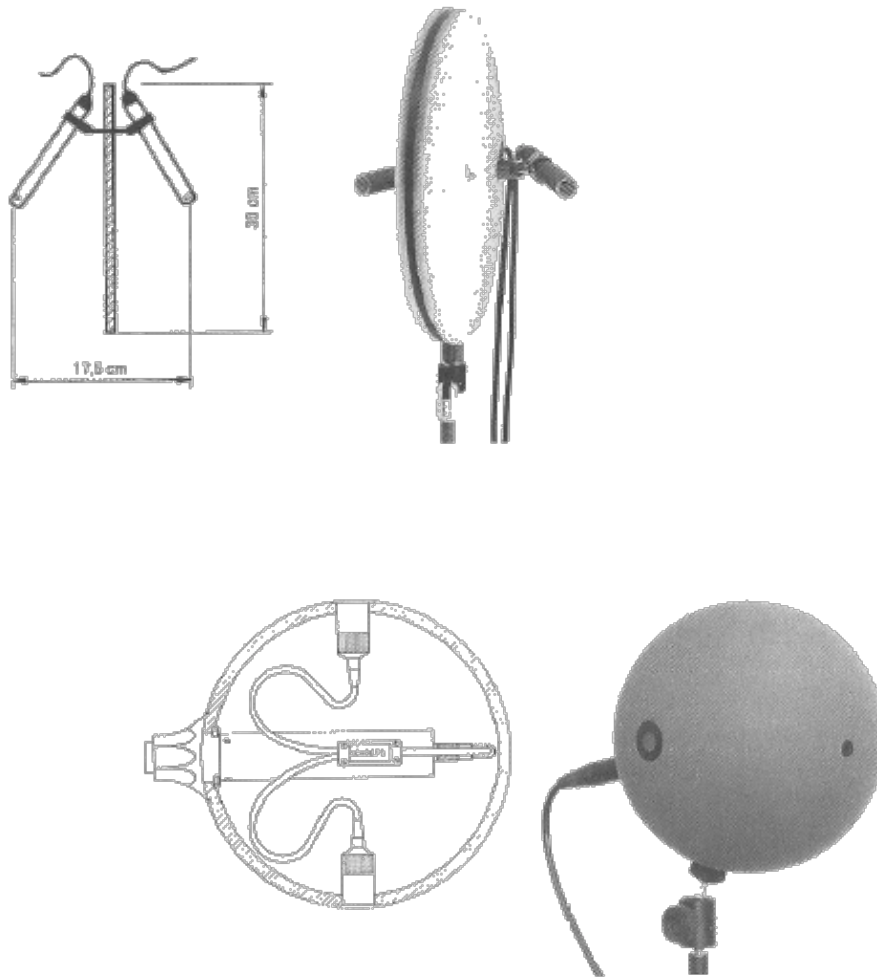


Abbildung 6: Trennkörper- und Kugelflächen-Stereophonie

3.4.1. Kunstkopfstereophonie

Kunstkopfstereophonie bezeichnet einen Teil der kopfbezogenen Binauraltechnik, also eine Form der Binauraltechnik, die möglichst viele Aspekte der menschlichen auditiven Wahrnehmung berücksichtigt. Dies soll gewährleisten, dass bei dem Verfahren ein natürliches Hörereignis entsteht, welches im besten Fall nicht mehr vom realen Hören zu unterscheiden ist.⁷³ Da in der Kunstkopftechnik die Einflüsse der Pinna in die Aufnahme mit eingebunden werden, müssen kopfbezogene Produktionen über Kopfhörer wiedergegeben werden. Dies bildet eine Einschränkung zur Trennkörperstereophonie, die monokompatibel ist.⁷⁴ 1970 experimentierte die Firma Neumann erstmals mit der Kunstkopftechnik. Der Kunstkopf ist eine möglichst exakte Nachbildung des menschlichen Kopfes, die alle akustischen Eigenschaften, d.h. Laufzeitdifferenzen, Pegeldifferenzen sowie spektrale Einflüsse aufweist. Durch letztere können auch Richtungen aus dem oberen Halbraum sowie Entfernungen gut nachgebildet werden. Die Signale werden durch Mikrophone aufgezeichnet, die anstelle eines Trommelfells montiert sind. Die Räumlichkeit beim Abhören einer Kunstkopfaufnahme über Kopfhörer sollte den Hörer im Idealfall einhüllen und eine akustische Realität simulieren.⁷⁵ Dass der Kunstkopftechnik damals nicht der Durchbruch gelang, lag vor allem an der fehlenden Wiedergabemöglichkeit über Lautsprecher. Kunstkopfaufnahmen sind kaum Lautsprecher-kompatibel, da das einzelne Signal nur für ein Ohr bestimmt ist und vom abgewandten Ohr nicht wahrgenommen werden soll. Es gibt die Möglichkeit durch Filter, die den binauralen Signalen vorgeschaltet werden, Übersprechen bei einer Wiedergabe über Lautsprecher an den Ohren zu verhindern. Dazu muss sich der Hörer je nach Lautsprecheranordnung in einer bestimmten Position befinden. Ein Drehen des Kopfes oder ein Verändern der Kopfposition erfordert auch eine Anpassung der vorgeschalteten Filter, daher ist ein authentisches Klangbild nicht ohne Headtracking



Abbildung 7: Kunstkopf KU 100 der Firma Neumann

⁷³ (Vgl. Dickreiter u.a., 2014, S. 345)

⁷⁴ (Vgl. Birkner, 2002, S. 84)

⁷⁵ (Vgl. Dickreiter u. a., 2014, S. 345)

möglich. Alternativ gibt es auch Möglichkeiten, ein mit dem Kunstkopf aufgenommenes Signal über Lautsprecher ohne Filter wiederzugeben. Dabei kommt es allerdings zu Klangverfärbungen, da die richtungsspezifischen Verzerrungen durch die Pinna so nicht wiedergegeben werden können. In den 70ern und 80ern war die Industrie aber noch nicht darauf vorbereitet, solche Aufnahmen durchzuführen. Erst als durch die Verbreitung des Walkmans und heute der mp3-Player Kopfhörer generell besser angenommen wurden, konnte die Technik erneut eingesetzt werden.⁷⁶

Die Kunstkopfaufnahme simuliert ein natürliches Hörerlebnis von den bisher vorgestellten Techniken am besten, da alle relevanten Informationen zur Lokalisation im 3D-Raum empfangen werden.⁷⁷ Die Güte des Hörbildes ist direkt von der Ähnlichkeit der Höreranatomie zu der des Kunstkopfes abhängig. Dies kann man durch die Aufnahme mit einem „echten Kunstkopf“ verbessern. Zu diesem Zweck kann ein echter Kopf verwendet werden, an dessen Gehörgangseingängen zwei Mikrophone platziert werden. Dementsprechend wird die Aufnahme dann die HRTF des verwendeten Kopfes erhalten. Dazu kann jeder Punkt im Gehörgang verwendet werden, da an allen Stellen bereits die ohrmuschelspezifischen Verzerrungen Informationen über die Richtung geben.⁷⁸

Diese ohrmuschelspezifischen Verzerrungen ersetzen die Funktionen des Außenohrs. Die richtungsspezifische Klangfärbung der Pinna kann bei einer Kopfhörerwiedergabe nicht mehr zur Lokalisation verwendet werden. Dabei ist es daher wichtig, dass die Kopfhörer-Übertragungsfunktion mit der Außenohrübertragungsfunktion übereinstimmt, die in einem richtungsneutralen Schallfeld gemessen wird. Richtungsspezifische Kopfhörerentzerrung kann zu spektralen Verfärbungen führen, die zu den binauralen Signalen in Kontrast stehen und somit nicht zu einem realistischen Hörbild führen.⁷⁹ Aus diesem Grund hat sich eine frequenzunabhängige Diffusfeldentzerrung für Kopfhörer durchgesetzt, die das Problem bei der Kopfhörerwiedergabe löst.

⁷⁶ (Vgl. Møller, 1992, S. 172 f.)

⁷⁷ (Vgl. Nicol, 2010, S. 1)

⁷⁸ (Vgl. Møller, 1992, S. 211)

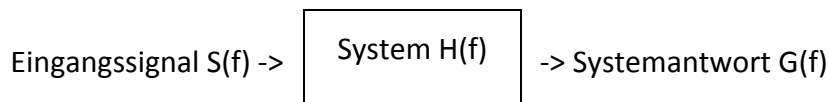
⁷⁹ (Vgl. Dickreiter u.a., 2014, S. 350)

3.4.2. Binauralsynthese

Die Binauralsynthese simuliert die Aufnahme eines Schallereignisses durch zwei Mikrophone in den Ohren der entsprechenden Person. Am Ende soll ein Hörereignis entstehen, das, ohne durch Kunstkopf- oder In-Ear-Mikrophone aufgezeichnet worden zu sein, einen räumlichen Höreindruck vermittelt. Dazu wird die individualspezifische Übertragungsfunktion der Ohren einer Person mit einem Schallereignis kombiniert. Die Übertragungsfunktion des linken sowie des rechten Ohrs wird zunächst in Abhängigkeit des Schallquellenorts ausgemessen. Alternativ können auch bereits vorhandene HRTFs benutzt werden, wenn die Abweichungen zur abhörenden Person in Kauf genommen werden können.⁸⁰

3.4.2.1. Das LTI-System

Die Binauralsynthese wird im LTI-System dargestellt: Ganz generell verändert oder überträgt ein System Signale. Jedes System erhält ein Eingangssignal, verarbeitet es und gibt eine Systemantwort, das Ausgangssignal, zurück.



Ein Linear-Time-Invariant-System (LTI-System) bezeichnet ein System, das lediglich eine lineare Verzerrung vornimmt, also Veränderungen in Amplitude und Phasenlage. Wenn neue Teiltöne durch das System entstehen, handelt es sich nicht mehr um ein LTI-System. Es gelten zwei Anforderungen:

1. Das Superpositionsprinzip: Die Summe zweier Signale ergibt als Systemantwort die Summe der Systemantworten auf das jeweils einzelne Signal. So ist die Linearität gegeben.
2. Eine zeitliche Verschiebung des Eingangssignals resultiert in einer ebenso zeitlich verschobenen Systemantwort. Damit ist das System zeitinvariant.⁸¹

⁸⁰ (Vgl. Nicol, 2010, S. 30)

⁸¹ (Vgl. Görne, 2008, S. 128)

Kein System in der Realität kann diese Anforderungen exakt erfüllen, kleinere Abweichungen von den LTI-Anforderungen können allerdings meistens vernachlässigt werden. Das LTI-System dient als Modell für Berechnungen in der Tontechnik.⁸²

3.4.2.1. Binaural Room Impulse Response (BRIR)

Auf der Suche nach einer Möglichkeit, ein Raum- und auch „Sweet-Spot“-unabhängiges System zu entwickeln, entstand die Überlegung, eine Raumsimulation über Kopfhörer zu generieren. Da jeder Raum in Bezug auf Hall oder Reflexionen eine eigene Charakteristik aufweist, soll diese durch binaurale Raum-Impulsantworten eingefangen und mit jedem beliebigen Signal verrechnet werden können. Eine Impulsantwort beschreibt das zeitliche Verhalten eines Signals im System. Betrachtet man den Raum als System, ist die Impulsantwort die Systemantwort, die entsteht, wenn man einen Diracstoß („ein schmaler Impuls mit unendlich großer Amplitude“⁸³) oder einen Sweep (Sinuston durch alle Frequenzen) zuführt. Die Charakteristika des Raumes verändern das Eingangssignal und ergeben ein Ausgangssignal, das alle notwendigen Informationen des Raums enthält. In der Praxis werden diese wie folgt aufgezeichnet:

Ein Kunstkopf wird in einem Raum platziert, dem aus den interessierenden Richtungen (meistens 5.1 oder eine andere Abhörsituation) einzeln Sweeps oder Diracstöße zugespielt werden. Die aufgenommenen binauralen Raumimpulsantworten enthalten zwei ohrspezifische Signale, die mit dem eigentlichen Signal gefaltet werden. Wird das bearbeitete Signal über Kopfhörer abgehört, wird durch die Berücksichtigung der Ohrmuschelfunktion nicht nur die IKL aufgehoben, es ermöglicht auch den Klangeindruck einer möglicherweise perfekten Abhörumgebung in einem Raum, der die entsprechenden Kriterien nicht erfüllen kann. Diese Raumsimulation bezeichnet man als „Binaural Room Scanning“ (BRS). Zwingend notwendig für diese Arbeitsweise ist allerdings ein Head-Tracking-System, dass vor den BRS-Prozessor geschaltet wird. Da der Mensch auch ständig durch kleinere Kopfbewegungen Orten vornimmt,

⁸² (Vgl. Görne, 2008, S. 128 f.)

⁸³ (Görne, 2008, S. 128)

insbesondere bei der Vorne-Hinten-Lokalisation, genügen die Informationen der spektralen Verzerrungen nicht aus.⁸⁴

Das BRS-Verfahren kann es ermöglichen, dass Räume für Abmischungen genutzt werden können, die raumakustisch eher ungeeignet sind. Durch eine virtuelle Abhörumgebung können dann an beliebiger Stelle die gewünschten, zuvor eingemessenen Raumbedingungen simuliert werden. Zum anderen können dadurch die Defizite einer Kopfhörerwiedergabe, wie die IKL oder eine mangelnde authentische Hallwiedergabe, reduziert werden.⁸⁵ Je nach anatomischer Abweichung zum Kunstkopf kann es allerdings zu Ortungsfehlern kommen.

3.4.2.2. *Personal Room Impulse Response (PRIR)*

Die Aufzeichnung der Impulsantworten für eine BRIR wird normalerweise an einem Kunstkopf durchgeführt. Es gibt die Möglichkeit, diese durch die Verwendung des eigenen Kopfes als Kunstkopfersatz zu personalisieren. Bei diesem Verfahren wird die Personal Room Impulse Response (PRIR) gemessen. Dabei werden kleine Mikrophone in den Gehörgang gesteckt, die als Impulsantworten die individuellen Verzerrungen der Pinna sowie die räumlichen Charakteristika aufnehmen. Durch eine Faltung von Signal und IR ist es möglich, eine für eine Person sehr realistische akustische Umgebung zu kreieren.

3.3.2.3. *Faltung*

In unserem System ist das Signal der Schallquelle das Eingangssignal, die Impulsantwort beschreibt das zeitliche Verhalten dieses Signals im System. Um die Systemantwort, also das Signal das auf den Kopfhörern wiedergegeben werden soll, zu erhalten, bedient man sich der Faltung. Mathematisch ist die Faltung zweier Signale im Zeitbereich folgendermaßen definiert:

$$g(t) = \int_{-\infty}^{\infty} H f(\tau) h(t - \tau) d\tau$$

⁸⁴ (Theile, 2015, S. 15)

⁸⁵ (Vgl. Theile, 2015, S. 15)

Das Ausgangssignal eines Systems ist das Eingangssignal gefaltet mit der Impulsantwort. Das verdeutlicht die vereinfachtere Schreibweise der Faltung, wobei der Operator " * " „gefaltet mit“ bedeutet:

$$g(t) = s(t) * h(t)$$

Die Impulsantwort beinhaltet alle relevanten Informationen des Raumes. Ist die IR bekannt, kann das Übertragungsverhalten des Systems auf jedes Signal angewendet werden.⁸⁶ Die Faltung lässt sich nicht nur in der Zeitebene darstellen, es gibt auch die Möglichkeit, diese mittels Fourier-Transformation in der Frequenzebene zu formulieren. Die Rolle der Impulsantwort und somit die Funktion des Systems, nimmt dann der komplexe Übertragungsfaktor $H(f)$ ein. Im Frequenzbereich wird die Faltung als Multiplikation des Eingangssignals mit dem Übertragungsfaktor H dargestellt:

$$G(f) = S(f) \cdot H(f)$$

Diese Darstellung der Multiplikation des Eingangsspektrums mit der Impulsantwort als Übertragungsfaktor entspricht der Faltung im Zeitbereich. Durch eine inverse Fouriertransformation lässt sie sich vom Frequenzbereich wieder in den Zeitbereich umwandeln.

Um eine Faltung vorzunehmen, gibt es zwei Methoden. Zum einen die Finite Impulse Response (FIR). Bei diesem Aufbau wird das Eingangssignal Sample für Sample mit entsprechenden Gewichtungsfaktoren multipliziert. In der Summe stellen diese Faktoren oder auch Filterkoeffizienten die Impulsantwort dar. Bei kurzen Impulsantworten ist das die einfachste Herangehensweise. Je länger die Übertragungsfunktion jedoch wird, desto länger dauert der Rechenprozess. So ist es aufgrund des Rechenaufwands teilweise unmöglich, auf diese Art und Weise eine Systemantwort zu erhalten. Für diese Fälle kann auf die schnelle Faltung zurückgegriffen werden. Dafür bedient man sich mittels Fourier-Transformation einer Auslagerung der Faltung in den Frequenzbereich. Dort werden dann die beiden

⁸⁶ (Vgl. Tarnow, 2009, S. 10)

Komponenten miteinander multipliziert und das Ausgangssignal durch die inverse Fourier-Transformation wieder in den Zeitbereich zurückgewandelt.

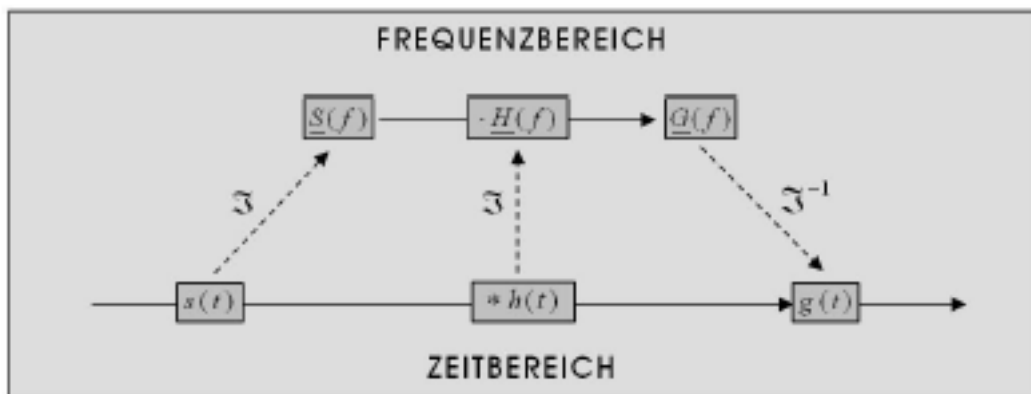


Abbildung 8: Darstellung der schnellen Faltung

Auch wenn diese Auslagerung in den Frequenzbereich einen Umweg darstellt, ist sie deutlich schneller als die FIR. Dies kann sich insbesondere in Anwendungen zu Nutze gemacht werden, in denen eine Echtzeitfaltung benötigt wird, das heißt in Situationen, in denen das Eingangssignal wechselnden Übertragungsfunktionen ausgesetzt wird und das Ausgangssignal möglichst schnell berechnet werden muss.⁸⁷

3.3.3. Probleme der Binauraltechnik

Wie in Kapitel 3.4.1. bereits erwähnt, ist die Qualität der Binauraltechnik stark davon abhängig, wie gut die anatomischen Eigenschaften mit dem in der Übertragungsfunktion festgehaltenen Modell übereinstimmen. Gerade die spektralen Verzerrungen durch die Pinna haben einen sehr großen Einfluss. Zu diesen zählen die Größe, die Form sowie die Auslenkung des Ohres, also der Winkel den die Ohrmuschel mit dem Kopf bildet. Es kommt nicht nur auf die Beschaffenheit der Ohren an, sondern auch auf den Durchmesser des Kopfes, da dieser maßgeblich die ITD beeinflusst. Die interaurale Laufzeitdifferenz stellt den zweiten großen Teil der Richtungsinformation dar. Ist der Durchmesser des Kopfes zu groß und das Headtracking ist aktiviert, verändern sich die Hörereignisse asynchron zur Kopfbewegung. Sind die Maße kleiner gewählt als der tatsächliche Wert, folgen bei einer Kopfdrehung die Hörereignisse der Bewegung. Ist die ITD zu groß, wird eine Bewegung der Hörereignisse in die

⁸⁷ (Vgl. Tarnow, 2009, S. 10 f.)

entgegengesetzte Richtung bemerkbar.⁸⁸ Dementsprechend kann eine Faltung mit einer standardisierten Übertragungsfunktion nie die gleiche Lokalisationsschärfe bilden, wie die mit der eigenen IR oder einer Aufnahme mit dem eigenen Kopf.

Eine Individualisierung durch die PRIR ist der Qualität wegen also vorzuziehen. Um eine solche Messung durchzuführen, sind besondere Bedingungen erforderlich: ein nachhallfreier Aufnahmeraum, eine möglichst hochwertige Lautsprecherausstattung, und eine Testperson, die je nach Messungsumfang eine halbe⁸⁹ bis hin zu zwei Stunden absolut still stehen kann.⁹⁰ Für Versuchsstände wird aus bis zu 1250 Richtungen gemessen⁹¹, in weniger exakten Varianten nur aus der klassischen 5.1-Abhöre. Diese Methoden erfordern zum einen viel Zeit, zum anderen sind sie nicht für alle Menschen zugänglich. Ein Benutzer müsste, um seine HRTF ausmessen zu lassen, extra ein Tonstudio besuchen und dort Messungen durchführen lassen, um die Dateien für die eigene Nutzung mitnehmen zu können. Dies würde bedeuten, dass zum einen viel Zeit, zum anderen relativ hohe Kosten für die Studiozeit und die Bezahlung eines Tontechnikers anfielen. Die Binauraltechnik muss für ein breiteres Publikum also benutzerfreundlicher gestaltet werden. Ein Ansatz dazu wird im Folgenden ermittelt.

⁸⁸ (Vgl. Lindau et al., 2010, S. 2)

⁸⁹ (Vgl. Litfin, 2014, S. 53)

⁹⁰ (Vgl. Nicol, 2010, S. 14)

⁹¹ (Vgl. Nicol, 2010, S.15)

4. Userfreundliche Binauraltechnik

Binauraltechnik ist durch den Gebrauch von Virtual Reality ein Format, das in Zukunft auf mobilen Endgeräten funktionieren muss. Im folgenden Kapitel soll nun eine Möglichkeit eruiert werden, wie eine nutzerfreundlichere Individualisierung der HRTF zu diesem Zweck zu erreichen ist.

4.1. Stand der Technik

Immersive Formate beinhalten sowohl visuelle als auch akustische Simulationen und können darüber hinaus auch die taktile Wahrnehmung ansprechen. Im Anschluss soll ein Überblick über momentane Möglichkeiten der virtuellen Technik gegeben werden.

4.1.1. Virtual Reality (VR)

Das Ziel von Virtual Reality ist eine Simulation der Welt und dem Erfahren dieser, die es dem Menschen nicht mehr ermöglicht, sie von der Wirklichkeit zu unterscheiden. Ein weiteres Merkmal der Virtuellen Realität ist die Immersion, das heißt ein möglichst umfassendes Ansprechen des VR-Nutzers durch die Ausgabegeräte. Für eine gelungene Immersion formulierten Slater und Wilbur 1997 folgende Bedingungen:⁹²

- 1) ausschließliche Computergenerierung der Sinneseindrücke
- 2) Ansprechen möglichst vieler Sinne
- 3) Vollständige Umgebung des Nutzers durch die Ausgabegeräte
- 4) möglichst lebendige Darstellung durch die Ausgabegeräte

Im Gegensatz zu herkömmlichen Anwendungen am Computer, also Systemen mit einem Graphic User Interface (GUI), bei dessen Benutzung Sinnesreize der realen Welt nach wie vor empfangen werden können, soll bei VR dieser Einfluss vollständig ausgeblendet werden. Virtual Reality Headsets umschließen den Hör- und den Sehsinn, „Data Gloves“ können die Bewegungsdaten der Hand erfassen, mit denen dann in der virtuellen Welt navigiert und agiert werden kann. Durch Headtracking, das auf dem Headset befestigt ist, kann der VR-Nutzer durch Kopfdrehungen das gesamte Panorama erfassen und die simulierte Realität, die ihn umgibt, erkunden. Dies ist zum einen durch Aufnahmen mit 360°-Kameras möglich, zum anderen werden

⁹² (Vgl. Dörner et al., 2013, S. 14)

verschiedene Anwendungen computeranimiert. Damit der VR-Anwender sich interaktiv in der VR bewegen kann, muss ermöglicht werden, dass die Umgebung sich in Echtzeit anpassen kann. Diese Wechselwirkung von Software und Nutzer bezeichnet man als Human-Computer-Interaction (HCI).⁹³ Einen Teil der HCI oder MCI (Mensch-Computer-Interaktion) stellt die Usability dar. Dies bezeichnet „das Ausmaß, in dem ein Produkt durch bestimmte Nutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen“⁹⁴. Dazu zählt die Praktikabilität der Benutzung, der Aufwand der betrieben werden muss, um das Ziel zu erreichen und auch die Zufriedenheit des Nutzers. Virtual Reality erlebt in den letzten Jahren mehr und mehr Zuspruch. Mit dem Kauf von Oculus Rift durch Facebook im Frühjahr 2014 sollte, laut Oculus-Rift-Mitbegründer Luckey Palmer, VR verfügbar und vor allen Dingen bezahlbar werden, so dass eine breitere Masse angesprochen werden kann.⁹⁵ Die Nachfrage nach solchen Systemen steigt tatsächlich, wobei dieser Trend laut Prognosen auch noch deutlich stärker wird, wie im Diagramm gezeigt wird.

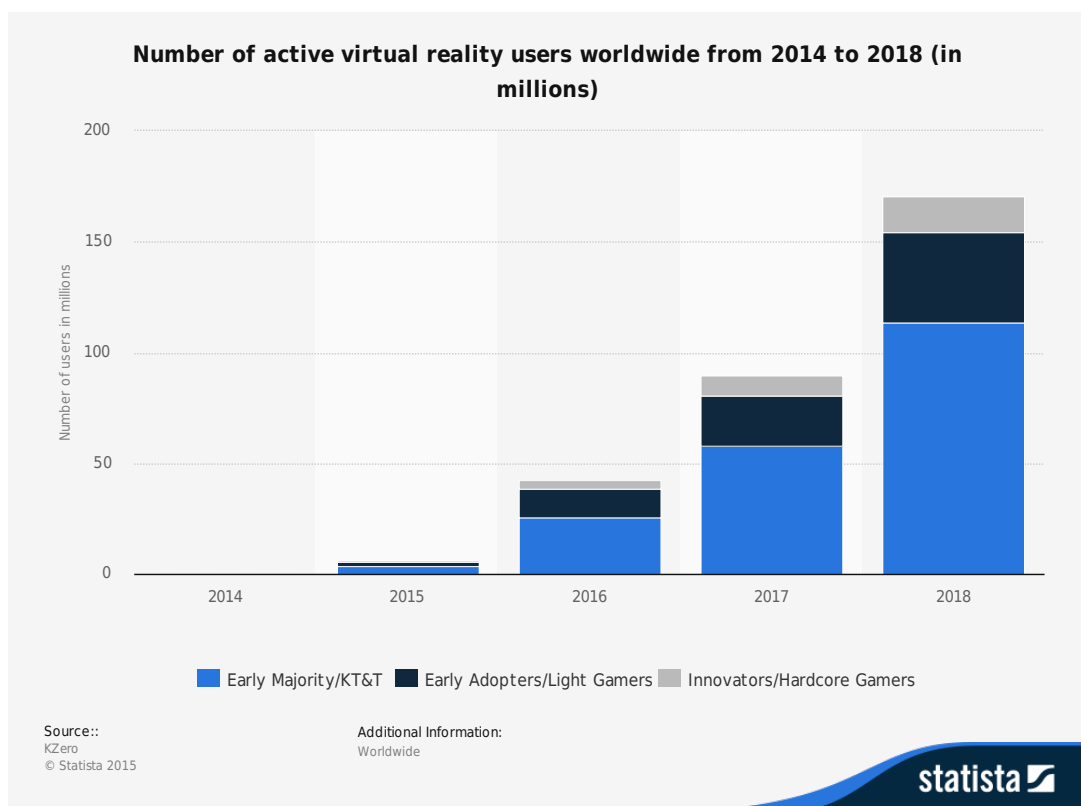


Abbildung 9: Prognostizierter Anstieg von VR-Nutzern bis 2018

⁹³ (Vgl. Dörner et al., 2013, S. 158)

⁹⁴ Ebd.

⁹⁵ (Vgl. Palmer, 2014)

Facebook kündigte am 21. Februar 2016 die Zusammenarbeit mit Samsung an, deren Smartphone Samsung Galaxy S6 in Zukunft in der Lage sein soll, VR selbst zu produzieren. Teil des Systems ist eine 360°-Kamera, deren Daten vom Smartphone zu auf der Brille abbildbaren VR-Erlebnissen umgerechnet werden sollen.⁹⁶

Von Augmented Reality (AR) spricht man, wenn die Wirklichkeit nicht völlig ausgeblendet wird, sondern durch die AR lediglich aufgewertet wird. Beispiele dafür existieren momentan auf Smartphones, die durch das Scannen von verschiedenen Oberflächen diese auf dem Display mit zusätzlichen Informationen ergänzen. Hier wird das Blickfeld der Kamera in die Realität zugelassen und durch die AR erweitert. Headsets in diesem Feld existieren noch nicht. Ein erster Versuch, diese Dinge zu kombinieren, war die Google Glass, die sich aufgrund mangelnder Praktikabilität und Einwänden durch Datenschützer nicht durchsetzen konnte.

4.1.2. Virtualisierung von Räumen

Die Problematik der nicht einblendbaren Realität ergibt sich bei Audioproduktionen nicht. Hier kann mit je nach Anwendung geschlossenen oder offenen Kopfhörern gearbeitet werden und somit Information von außen zugelassen werden oder nicht.

Die aktuelle Audiotechnik umfasst allerdings noch keine nutzerfreundliche Individualisierung der HRTF. Dennoch soll hier gezeigt werden, welche bisherigen Ansätze in der Binauraltechnik unternommen wurden, um ein realistisches Klangbild zu erhalten.

Der Spatial Audio Designer (SAD) von New Audio Technologies aus Hamburg stellt eine virtuelle Arbeitsumgebung dar, in der es Toningenieuren möglich ist, Surround oder 3D-Sound zu produzieren. Dabei kann auf unterschiedlichen Surroundformaten sowie 3D-Formaten wie 13.1 gemischt werden. Für eine mangelnde Abhörumgebung wurde dort die binaurale Abhörmöglichkeit über Headphone Surround 3D™ entwickelt.⁹⁷ Der SAD nutzt eine standardisierte HRTF, die mit den Signalen entsprechend ihres Ursprungs gefaltet werden. Diese entspricht bei manchen mehr, bei anderen weniger gut dem persönlichen Hörbild. Mit dieser HRTF wurden unterschiedliche Räume ausgemessen, die dann im SAD abgerufen werden können und zur Mischung

⁹⁶ (Vgl. Jung, 2016)

⁹⁷ (Ammermann, Website von NewAudioTechnologies)

verwendet werden. Zusätzlich können noch weitere Einstellungen getroffen werden, mit denen der Raum durch eine Veränderung des Reverbs individuell angepasst werden kann. Das Endprodukt kann von jedem Kopfhörer wiedergegeben werden.⁹⁸

Dementsprechend ist es zwar für den Endkunden wenig Aufwand, 3D-Sound zu Hause abzurufen, die Anpassung an den jeweiligen Kopf wird aber nicht berücksichtigt.

Eine Möglichkeit, die PRIR zu generieren und damit zu arbeiten, stammt von dem Unternehmen Smyth. Der Smyth Realiser A16 ist ein Interface zur Generierung der PRIR sowie einer Abhörumgebung von 32 Lautsprecherkanälen. Die Messung der persönlichen Impulsantwort funktioniert durch kleine Mikrophone, die mithilfe einer kleinen Einführschiene im Ohr platziert werden. Diese werden dann von einem 5.1- oder 7.1-Abhörsystem mit Sweeps beschallt. Zusätzlich zu der statischen Kopfposition werden auch noch zwei weitere Winkel der Kopfdrehung eingemessen, damit später eine Interpolation der gemessenen Werte möglich ist, sobald der Kopf bewegt wird. Durch einen Headtracker, der an einem Bügel auf dem Kopf befestigt wird, kann die entsprechende Kopfposition ermittelt und auch später bei der Wiedergabe berücksichtigt werden. Die interpolierten Impulsantworten werden dabei mit dem tatsächlichen Signal gefaltet, so dass ein räumliches Klangbild entsteht, das sich den Kopfbewegungen des Hörers anpasst.⁹⁹ Die Technik für dieses Vorgehen ist aufwändig, eine Ausmessung kann wie in 3.3.3. erwähnt nur unter bestimmten Bedingungen durchgeführt werden. Eine ausführliche Beschreibung des Smyth Realisers und ein Vergleich mit der reellen Arbeitsumgebung kann in der Arbeit von Philipp Reineboth nachgelesen werden.¹⁰⁰

Im Zuge der Entwicklung von Virtueller Realität hat auch Android die binaurale Technik in das Software Development Kit (SDK) übernommen.¹⁰¹ Da es mit Google Cardboard bereits möglich ist, VR-Apps auf seinem Smartphone zu erfahren, ist die Weiterführung des immersiven Formats durch binaurales Audio ein logischer Schritt. Google Cardboard ermöglicht dem Nutzer einen 360°-Blick in der VR, der zwar nicht an die Brillanz der eigens dafür entwickelten Headsets von Oculus Rift oder Sony Playstation VR herankommt, dennoch für interessierte VR-Nutzer eine preiswerte Alternative

⁹⁸ (Vgl. Wisse und Schuster, 2015, S. 44 f.)

⁹⁹ (Vgl. Smyth, 2015, S. 41 ff.)

¹⁰⁰ (Reineboth, 2012)

¹⁰¹ (Martz, 2016)

bietet. Während das Oculus-Rift-Headset bereits mit binauralem Ton arbeitet, war das für Google Cardboard noch nicht möglich. Seit Januar 2016 wird die binaurale Technik auch im Android SDK zur App-Programmierung angeboten. Entwickler können dementsprechend auch bei Google-Cardboard-Anwendungen mit 3D-Sound arbeiten. Hier kann der Umgebungsklang gewählt werden, je nach Größe und Material des Raumes. Ebenso wird die Anatomie eines standardisierten menschlichen Ohrs berücksichtigt. Das SDK ist für mobile multi-core CPUs optimiert. Da die Echtzeitfaltung in einen separaten Thread ausgelagert ist, wird die primäre CPU nicht belastet. Die binaurale Technik ist also in Smartphone-Anwendungen angekommen, eine Optimierung dieser durch Individualisierung der HRTF liegt allerdings noch nicht vor.

4.2. Binauraltechnik für mobile VR

Angeichts des prognostizierten steigenden Interesses an Virtual Reality soll überlegt werden, wie binaurale Technik für den VR-Anwender optimiert werden kann. Da es sich bei Google Cardboard um ein Medium handelt, das leicht und günstig erworben werden kann und das in den letzten Monaten auf großes Interesse gestoßen ist, wird eine Optimierung für die mobile Nutzung auf Smartphones thematisiert. Neben 5 Millionen Google-Cardboard-Käufen wurden innerhalb von zwei Monaten 10 Millionen mehr immersive Apps heruntergeladen als in den Monaten zuvor.¹⁰²

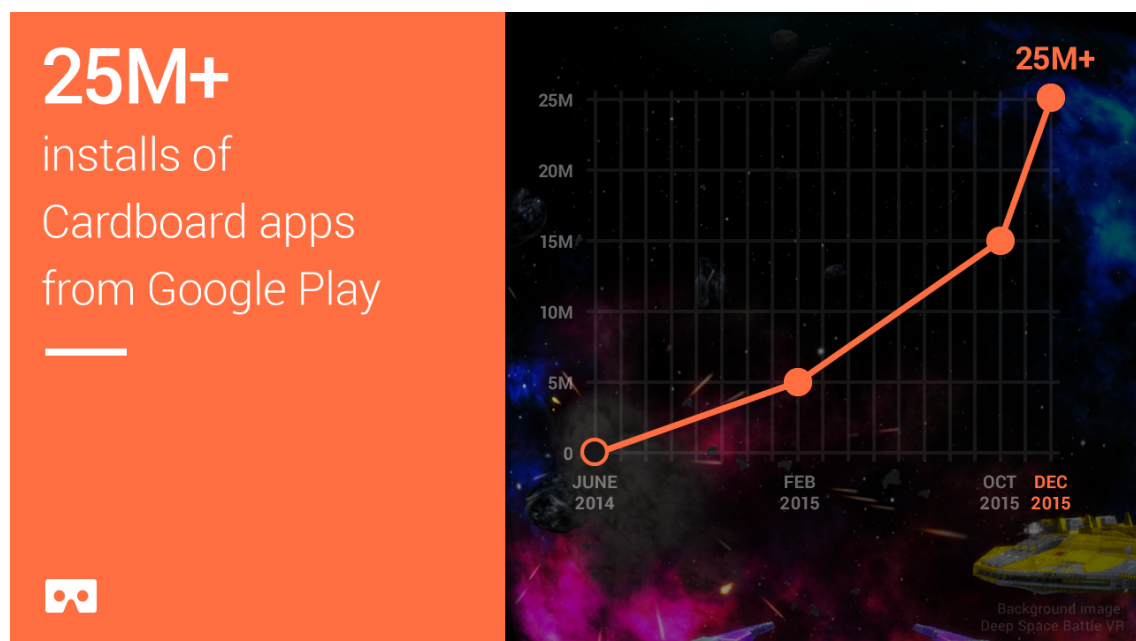


Abbildung 10: Anzahl der heruntergeladenen VR-Applikationen 2014/15

4.2.1. Individualisierung

Die Immersion von Virtual Reality hängt von der Authentizität der Darbietung ab. Dies bezieht sich nicht nur auf die optische Darstellung, sondern auch auf die akustische Simulation. Je realistischer der Raum akustisch wiedergegeben wird, desto eher sind wir dazu bereit, uns ganz in die virtuelle Welt zu vertiefen. Umgekehrt: Erscheint uns ein Geräusch unnatürlich, holt es uns aus der simulierten Kulisse und wir beginnen über die Darstellung nachzudenken, was uns daran hindert, uns auf die VR einzulassen.

¹⁰² (Vgl. Bavor, 2016)

Ein großer Störfaktor hierfür kann die IKL sein, die insbesondere bei der Kopfhörerwiedergabe von diotischen Signalen hervorgerufen werden kann. Sie tritt aber auch auf, wenn die gehörte HRTF nicht mit der eigenen übereinstimmt. Je nach Abweichung nimmt die „Nach-Außen-Verlagerung“ der Signale ab, bzw. kommt es zu einer Im-Kopf-Lokalisierung.¹⁰³

Auch der Elevationswinkel kann bei abweichender Pinna nicht mehr korrekt eingeschätzt werden, ebenfalls nimmt die Lokalisationsunschärfe zu. Bei einer optimal eingestellten und individualisierten HRTF hingegen nimmt die Lokalisationsunschärfe lediglich um 1° bis 2° zu. Die Verwendung eines echten Kopfes als Kunstkopf im Gegensatz zu einer Nachbildung führt zu besseren Ergebnissen.¹⁰⁴ Wie stark die Lokalisation unter Verwendung von individualisierten von nicht-individualisierten HRTFs schwankt, ist in Abbildung 11 zu sehen. Es sind die Auswirkungen von nicht individualisierten HRTFs auf das räumliche Empfindungsvermögen der Hörrinde eines Frettchens bei einer horizontalen Abdeckung von 360° und einem Elevationsspektrum von -60° bis 90° dargestellt. Dabei sind die hell markierten Punkte jeweils der Schwerpunkt der neuronalen Anregung. Die Balken stellen die durchschnittliche Anregung der Neuronen pro Impuls dar. Es wird ersichtlich, dass es bei nicht-individualisierten HRTFs zu großen Abweichungen kommen kann. Mitunter wird das Hörereignis aufgesplittet. Die stärksten Schwankungen sind in den Abbildungen C16, C49 und C92 zu sehen.

¹⁰³ (Vgl. Nicol, 2010, S. 45)

¹⁰⁴ (Vgl. Nicol, 2010, S. 46)

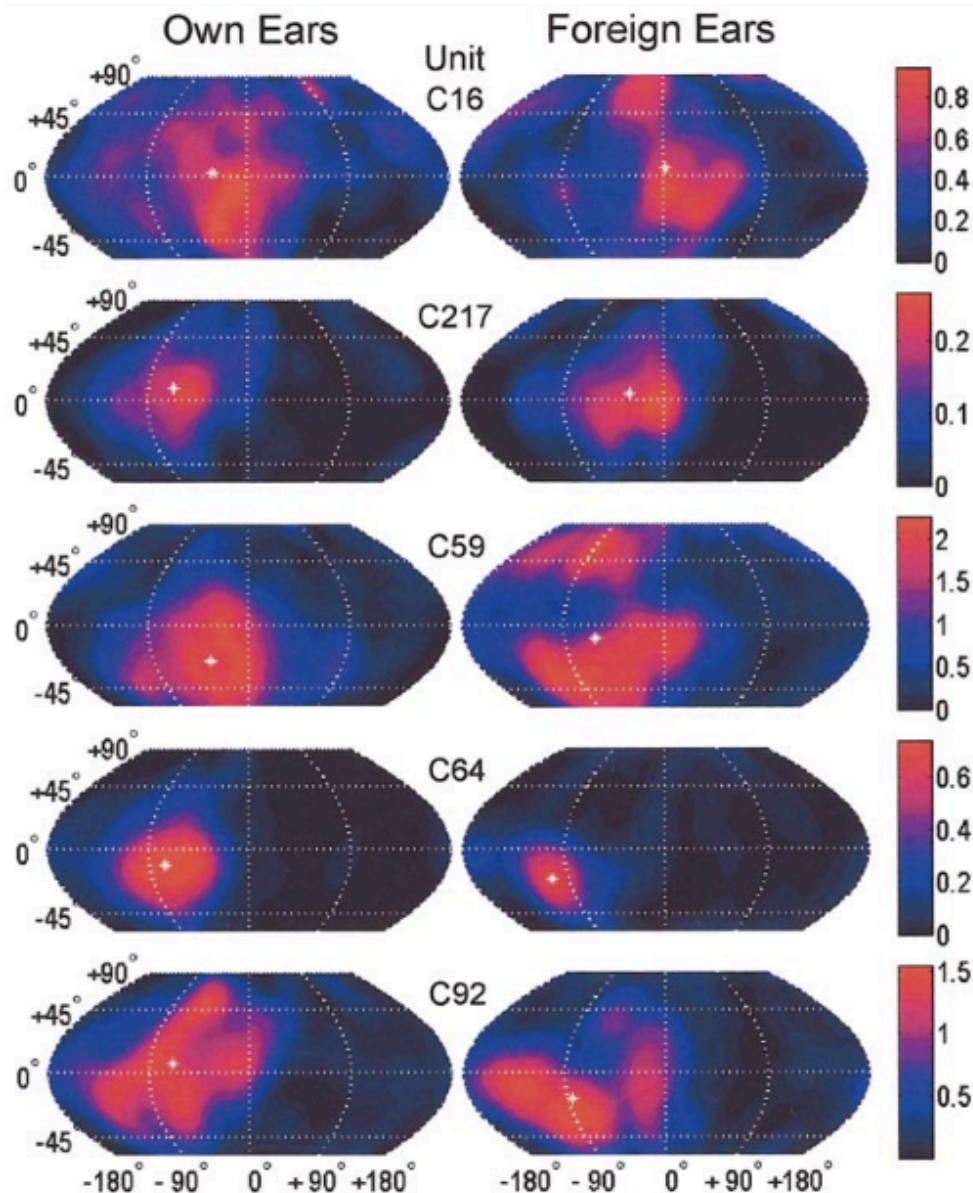


Abbildung 11: Lokalisationsunschärfe bei nicht individualisierter HRTF

4.2.2. Wahl des Individualisierungsfaktors

Um sich einer nutzerfreundlichen Individualisierung der HRTF zu nähern, die ohne zeitlichen und technischen Aufwand auskommt, muss betrachtet werden, welche der Kenngrößen der Übertragungsfunktion sich dafür eignet. Die Pinna ist durch ihre einzigartige Form nur sehr aufwändig individuell darzustellen. Durch die Vielzahl an Parametern, die den Schallwellengang beeinflussen können, ist es kaum möglich, eine Formel zu beschreiben, in der diese berücksichtigt werden. Wie in 3.3.3. bereits

beschrieben ist eine langwierige und aufwändige Ausmessung der HRTF für den Nutzer nicht leicht zugänglich. Eine Lösung für dieses Problem kann eine Auswahl einer HRTF aus einer Datenbank darstellen. Dabei kann eine Sammlung von möglichen HRTFs zur Verfügung gestellt werden, aus der Nutzer dann ein Modell für sich auswählen kann. Nicol schlägt dazu zwei verschiedene Wege vor:¹⁰⁵

- 1) Eine Auswahl aus nicht-individuellen HRTFs aus der durch einen Hörtest oder auch nur durch die optische Erscheinung der ausgemessenen Ohren die beste HRTF ausgewählt wird.
- 2) Eine Anpassung von nicht-individuellen HRTFs an die des Nutzers. Der Nutzer wählt wie in Modell 1) eine HRTF aus, die seinem Empfinden nach der seinen am nächsten kommt. Diese wird dann individualisiert. Da die HRTFs, wie in 2.3.3.1. erläutert, gewisse Gemeinsamkeiten aufweisen, kann die ausgewählte Übertragungsfunktion durch eine Skalierung nach Middlebrooks¹⁰⁶ je nach Größe und Ausrichtung angepasst werden. Möglich ist auch eine nutzerkontrollierte Anpassung mit einem Bandfilter.¹⁰⁷

Beide Individualisierungsansätze sind weniger genau als eine exakte Ausmessung in einem Studio, stellen allerdings einen Fortschritt zu einem Kunstkopf oder einer vorgegebenen HRTF dar. Da die spektralen Kenngrößen essentielle Information über die vertikale Lokalisation beinhalten, ließe sich die Problematik der Vorne-Hinten-Lokalisation durch die Individualisierung verringern. Je ungenauer aber die Einschätzung und Auswahl des Nutzers ist, desto stärker schlägt sich das auf die Rate der Verwechslungen und Lokalisationsfehler nieder. Die spektralen Filter der HRTF können also nur durch Annäherung individualisiert werden, eine exakte Einflussnahme ist – außer durch ausgemessene HRTFs – nicht realisierbar.

¹⁰⁵ (Vgl. Nicol, 2010, S. 56)

¹⁰⁶ Ebd.

¹⁰⁷ (Vgl. Nicol, 2010, S. 57)

Wie in 2.3.1. bereits erwähnt, erreicht der Mensch in der Horizontalebene die kleinste Lokalisationsunschärfe. Die möglichst realistische Abbildung dieser ist demzufolge besonders wichtig und soll im Folgenden behandelt werden. Die interauralen Laufzeit- und Pegeldifferenzen sollen auf Individualisierung und deren Sinnhaftigkeit geprüft werden. Macpherson und Middlebrooks versuchten 2002 in verschiedenen Versuchen, die Gewichtung der interauralen Kenngrößen auf die horizontale sowie die vertikale Lokalisation zu eruieren. Dabei modulierten sie unabhängig voneinander die ITD und die ILD um 600 μ s oder 60dB. Es ergab sich folgendes Ergebnis:¹⁰⁸

- 1) Interaurale Laufzeitdifferenzen sind bei tiefen Frequenzen besonders wichtig für die Lokalisation. Ihr Einflusspektrum kann je nach Breitbandigkeit des Signals auch in den höheren Frequenzen zunehmen.
- 2) Die ILD hat in den höheren Frequenzen Einfluss auf die Ortung. Je nach Signalart wird sie generell als weniger wichtig oder gleich wichtig wie die ITD eingeschätzt.

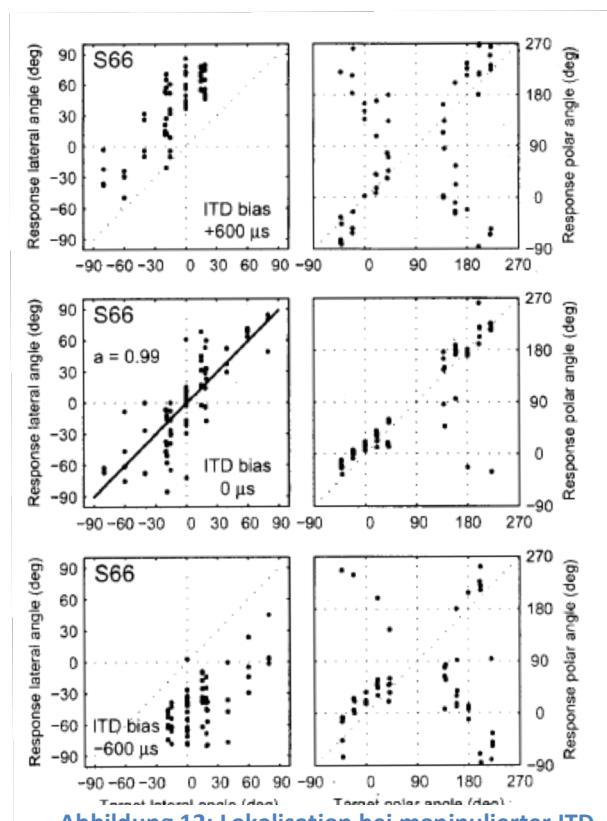


Abbildung 12: Lokalisation bei manipulierter ITD

- 3) Die ITD nimmt auch auf die vertikale Lokalisation Einfluss. In dem Versuch, bei dem die interaurale Laufzeitdifferenz manipuliert wurde, nahm die Vorne-Hinten-Lokalisationsunschärfe maßgeblich zu.

Zusammenfassend lässt sich sagen, dass eine eindeutige Gewichtung der interauralen Kenngrößen nicht vorgenommen werden kann, die interaurale Laufzeitdifferenz allerdings eine spezielle Rolle einnimmt. Da auch hohe Frequenzen über ihre Hüllkurve

¹⁰⁸ (Vgl. Macpherson und Middlebrooks, 2002, S. 2219)

zur Differenzierung der Phasenverschiebung geeignet sind, deckt die ITD ein größeres Spektrum ab als die ILD. Zudem besitzt sie Auswirkungen auf die vertikale Lokalisation, das Problem der Vorne-Hinten-Vertauschung kann so etwas eingegrenzt werden. Daher soll sie als Faktor für eine nutzerfreundliche Individualisierung verwendet werden.

Eine Kombination aus der zuvor behandelten Individualisierung der spektralen Filter und der Individualisierung der ITD kann ein noch besseres Ergebnis für die Vorne-Hinten-Lokalisation versprechen.

4.2.3. Modell der Anwendung

Die Individualisierung soll für Nutzer zur Verfügung stehen, die eine binaurale Applikation auf dem Smartphone nutzen. Möglich sind dabei Virtual-Reality-Anwendungen über Google Cardboard oder andere mit umgebendem Klang, die am Handy aufgerufen und dort auch mobil gespielt und genutzt werden können.

Google Android setzt mit Android Studio auf Appentwickler, die ohne großen Aufwand ihre Programme im Google Play Store hochladen können. Ein ähnliches Modell wird durch den Appstore von Apple dargestellt, dort wird die Veröffentlichung der App durch eine Prüfinstanz beurteilt.¹⁰⁹

Das vorgestellte Modell wird mit BPMN2.0 und Pseudocode beschrieben, um eine Umsetzung für verschiedene Betriebssysteme leicht zu ermöglichen. Je nach Betriebssystem oder Anwendung kann dieser Baustein in die App eingebunden sein, so dass die Individualisierung der ITD im Look der App und an geeigneter Stelle durch den Nutzer durchgeführt werden kann.

Alternativ wäre eine App-to-App Implementierung ebenfalls möglich. In diesem Fall würde die Berechnung der ITD in einer gesonderten App stattfinden, die eigentliche binaurale Anwendung würde dann auf die Ergebnisse des ITD-Berechners zugreifen. Der Nachteil hierbei ist, dass der Nutzer eine extra App herunterladen müsste, über deren Nutzen und Existenz er sich zunächst nicht im Klaren ist. Dies wäre ein umständlicherer Vorgang, der die Usability schmälert. Ein weiterer Nachteil wäre der Rechenvorgang im Hintergrund und die Tatsache, dass zwei Apps geöffnet werden

¹⁰⁹ (Vgl. Post, 2014, S. 25)

müssen, was die CPU-Auslastung des Handy unnötig erhöht. Dies kann zu nicht mehr vertretbaren Latenzen führen.

4.2.4. Anwendungsmöglichkeiten

Die Möglichkeiten für die Anwendung sind weit gestreut. Zunächst soll die in Kapitel 4.1.1. vorgestellte VR durch angepasstes Audio unterstützt werden. In jedem Szenario, in der Google Cardboard Brillen genutzt werden, seien es Virtual-Reality-Spiele, Filme oder interaktive Tutorials, die über die Brille erlebt werden, kann die Bibliothek eingesetzt werden.

Aber auch für Anwendungen ohne Virtual Reality kann das Konzept greifen. Da der Ton in der Lage ist, auf Dinge aufmerksam zu machen, die das Auge durch seinen eingeschränkten Blickwinkel nicht erfassen kann, sind auch rein akustische Anwendungen denkbar. Dabei kann die Anwendung bei Stadtführern zum Einsatz kommen, wenn der Nutzer auf Sehenswürdigkeiten in seinem Rücken aufmerksam gemacht werden soll. Eine entsprechende Anwendung mit GPS-Tracking hat Markus Hädrich vorgestellt.¹¹⁰

Ebenso denkbar sind Anwendungen im Museum, bei denen inzwischen auch immer häufiger auf ein klassisches Audioguide verzichtet wird, sondern mit Smartphone-Apps gearbeitet wird. Da diese Apps bereits in Richtung AR tendieren, in denen die eigentliche Hörumgebung mit akustischen Informationen aufgewertet wird, dient das tatsächliche akustische Milieu also als Referenz. Umso wichtiger ist eine möglichst reale Darstellung der HRTF, unterschiedliche HRTFs können sonst zu Verwirrung des Nutzers führen und zu unerwünschten Lokalisationsergebnissen führen.

Auch in der Raumfahrt nutzt die NASA bereits binaurale Technik für Piloten. Ziel hierbei ist es, die akustischen Signale, wie Funk, Warnsignale oder Bordsignale besser differenzieren zu können. Eine Verschiebung von Hörereignissen in der Medianebene ist dementsprechend wünschenswert. Auch hier kann eine Individualisierung bessere Ergebnisse schaffen.¹¹¹

Tonmeister können mit einer optimierten HRTF-Übertragungsfunktion mobiler arbeiten und sind nicht mehr an Studioräumlichkeiten gebunden. In Verbindung mit

¹¹⁰ (Hädrich, 2015)

¹¹¹ (Vgl. Begault, 2000)

den Impulsantworten von ausgemessenen Regieräumen oder Räumen mit geeigneter Akustik ist es möglich, unterwegs oder unter schlechteren Bedingungen zu produzieren.¹¹² In diesem Fall müsste die Anwendung in entsprechende Programme wie den SAD eingebaut werden. Dieser arbeitet wie in 4.1.2. besprochen mit einer vorgegebenen HRTF. Eine Anpassung könnte auch hier greifen und die Arbeit erleichtern. Vorrangige Einsatzmöglichkeit soll aber das Smartphone sein, dementsprechend muss geprüft werden, welche Anforderungen an das Gerät notwendig sind.

4.3. Technische Umsetzung

Die Umsetzung einer dynamischen Binauralsynthese auf mobilen Endgeräten mit einer individualisierten ITD beinhaltet die folgenden zu untersuchenden Aspekte: die technischen Anforderungen an das Gerät, die Methode der Individualisierung, sowie die Extraktion der ITD.

4.3.1. Anforderung an die Technik

Eine dynamische Binauralsynthese ist ein aufwändiger Prozess: die Schallquellenposition sowie die Menge der Schallquellen verändert sich, durch das Headtracking muss immer wieder der Winkel des auftreffenden Schalls berechnet werden und auch die Umgebung, in der sich der Hörer befindet muss berücksichtigt werden. Je nach Ort wird die Faltung mit der Übertragungsfunktion neu berechnet und ausgegeben. Dass dies auf einem Smartphone möglich ist, haben zum einen Sander u.a.¹¹³ bewiesen zum anderen hat Litfin¹¹⁴ ein Konzept zur Faltung auf Mobilgeräten vorgestellt. Die technischen Anforderungen an eine mobile dynamische Binauralsynthese ist eine geringe Latenz bei der Faltung des Audiosignals, um die Immersion der Anwendung nicht zu beeinträchtigen. In der Literatur genannte Grenzwerte liegen zwischen 85 und 96 ms.¹¹⁵ Zusätzliche Immersion soll durch Headtracking generiert werden, demzufolge ist diese Komponente in der

¹¹² (Vgl. Wisse and Schuster, 2015, S. 45)

¹¹³ (Sanders u.a., 2012)

¹¹⁴ (Litfin, 2014)

¹¹⁵ (Vgl. Sandvad und Mackensen in Hädrich, 2015, S. 86)

Latenzberechnung zu berücksichtigen. Ein separates Head-Tracking System ist dabei aber nicht miteinzubeziehen, da Smartphones durch ihren Gyroskop-Sensor dazu in der Lage sind, horizontale Bewegungen sowie Veränderungen in der Elevation durch das Magnetfeld der Erde zu erkennen. Diese Technik wird bereits für die VR-Anwendungen benutzt, um visuell und teilweise auch akustisch auf die Kopfbewegungen des Nutzers zu reagieren.

Ein weiterer Aspekt ist die Rechenkapazität der mobilen Endgeräte. Der momentane Stand bei Samsung ist ein Octa-core-Prozessor, ein Arbeitsspeicher von 3GB sowie eine Prozessorgeschwindigkeit von bis zu 2,1 GHz.¹¹⁶ Dies ist ein deutlicher Fortschritt zu der Technik, unter der Sanders u. a. 2012 ihre mobile Binauralsynthese implementierten. Das dabei verwendete iPhone 4s verfügte über einen ARM Cortex-A9-Prozessor bei einer Prozessorgeschwindigkeit von 800 MHz und einem Arbeitsspeicher von 512 MB.¹¹⁷ Es kann also davon ausgegangen werden, dass sämtliche Ergebnisse auch auf die aktuellen Smartphones übertragen werden können, bzw. mit einer deutlich besseren Leistung gerechnet werden kann. Definierte Qualitätsparameter, die Sanders u.a. 2012 in ihre mobile Binauralsynthese miteinbezogen, sind die Folgenden:¹¹⁸

- Kopfhörer Equalization/ Headphone Equalization, um spektrale Verzerrungen zu vermeiden, die an den Ohrmuscheln entstehen können. (HEQ)
- Frühe Reflexionen/ Early Reflections, die für den Raumklang von BRIRs essentiell sind (ER)
- Individuelle HRTF
- Filter-Interpolation im Frequenzbereich, um die in geringerer Auflösung ausgemessenen HRTFs zu verbessern (FIF)
- Filter-Interpolation im Zeitbereich (FIT)
- Head-Tracking
- Ansprechschwelle zur Ermittlung von interaktiver Latenz

¹¹⁶ ("Samsung Galaxy S6," 2016)

¹¹⁷ (Vgl. Sanders u. a., 2012, S. 5)

¹¹⁸ (Sanders u. a., 2012, S. 3)

Bei ihrer implementierten Anwendung zur mobilen Binauralsynthese konnten bei der Nutzung auf dem iPhone 4s folgende Parameter in Abstufung erreicht werden.

Quality level	Activated quality parameters	Render time (avg)	Memory allocation	Max channels
1	static; HEQ	6001 μs	2,26 MB	28
2	static; HEQ; ER	6021 μs	2,28 MB	27
3	DS; FUR = 2; LA = 180°; LE = 0°	7001 μs	2,04 MB	26
4	DS; FUR = 2; LA = 360°; LE = 0°; SM; ER	7186 μs	2,08 MB	25
5	DS; FUR = 2; LA = 180°; LE = 90°; SM; ER; HEQ	7453 μs	2,92 MB	24
6	DS; FUR = 2; LA = 180°; LE = 180°; FC; ER; HEQ	8428 μs	4,87 MB	21
7	DS; FUR = 1; full-sphere; SM; ER; HEQ	8661 μs	4,48 MB	20
8	DS; FUR = 1; full-sphere; SM; ER; HEQ; FIT	9149 μs	5,09 MB	19
9	DS; FUR = 1; full-sphere; FC; ER; HEQ	10617 μs	6,11 MB	17
10	DS; FUR = 1; full-sphere; FC; ER; HEQ; FIT	11093 μs	6,43 MB	16
11	DS; FUR = 1; full-sphere; SM; ER; HEQ; FIF	33089* μs	5,87 MB	7
12	DS; FUR = 1; full-sphere; FC; ER; HEQ; FIF	35449* μs	8,38 MB	6

Abbildung 13: Qualitätsniveaus der implementierten Anwendung durch Sanders u.a.

Dabei überschritten Stufe 11 und 12 die verfügbare Rechenzeit von 23220 μs , was auf die Interpolation im Frequenzbereich zurückzuführen ist. Außer dem Interpolationsverfahren ließen sich alle Qualitätsparameter durchführen. Dieses lässt sich aber durch das FIT ersetzen.

4.3.2. ITD-Errechnung

Um auf die interaurale Laufzeitdifferenz Einfluss nehmen zu können, muss diese aus der HRTF extrahiert werden und erst nach der Faltung mit den Audiosignalen wieder als Delay auf die Signale aufgerechnet werden. Im Folgenden sollen die einzelnen Aspekte der Methode erläutert werden.

4.3.2.1. Extraktion der Interauralen Laufzeitdifferenz

Die HRTF lässt sich durch die Hilbert-Transformation in drei Bestandteile zerlegen: den Betrag, den minimalphasigen Anteil, sowie den exzessphasigen Anteil. Dieser wiederum besteht aus einem frequenzabhängigen Allpassanteil sowie einem reinen Delay- Term.¹¹⁹

$$H(j\omega) = |H(\omega)| \cdot e^{j\varphi_{min}(\omega)} \cdot e^{j\varphi_{all}(\omega)} \cdot e^{j\varphi_{lin}(\omega)}$$

¹¹⁹ (Vgl. Lindau u. a., 2010, S. 3)

Durch diverse Versuche¹²⁰ konnte festgestellt werden, dass ein Verzicht auf den Allpass-Filter kaum hörbare Auswirkungen auf die räumliche Wahrnehmung hat. Demzufolge kann dieser, wie bereits in weiten Teilen so angewendet, weggelassen werden.¹²¹ Die in minimalphasige und Delay-Anteile zerlegte und reduzierte Formel der HRTF ist dann die folgende:

$$H_{ml}(j\omega) = |H(\omega)| \cdot e^{j\varphi_{min}(\omega)} \cdot e^{j\varphi_{lin}(\omega)}$$

Der linear-phasige Anteil kann nun durch ein individuelles Delay ersetzt werden, solange dieses der ITD in adäquatem Maße entspricht.¹²²

Die HRTF stellt eine komplexe und mathematisch kaum darstellbare Funktion dar. Mathematische Separationen der ITD werden ausgeschlossen. Eine Extraktion kann aber durch möglichst genaue Messungen vorgenommen werden.

Um die interaurale Laufzeitdifferenz zu extrahieren, wurden von Estrella bereits verschiedene Verfahren getestet¹²³, bewährt hat sich dabei die Onset Detection Methode. Dabei werden an beiden Ohren die Samplewerte gemessen, die ein Schallsignal zu dem jeweils linken und rechten Ohr benötigt. Aus der Differenz der beiden wird dann die interaurale Laufzeitdifferenz ermittelt. Für eine bessere und genauere Bestimmung der Werte wird empfohlen, die Messung im überabgetasteten Bereich vorzunehmen. Genutzt wird hier eine zehnfache Überabtastung.¹²⁴ So kann die eine bei Lindau maximal aufgetretene Abweichung von drei Samples¹²⁵ im überabgetasteten Bereich in der wieder heruntergerechneten Version vernachlässigt werden.

¹²⁰ (Kulkarni u. a., 1999)

¹²¹ (Vgl. Lindau u. a., 2010, S. 3)

¹²² (Vgl. Jorgos Estrella, 2011, S. 20)

¹²³ (Vgl. Estrella, 2010)

¹²⁴ Ebd.

¹²⁵ (Vgl. Lindau u. a., 2010, S. 3)

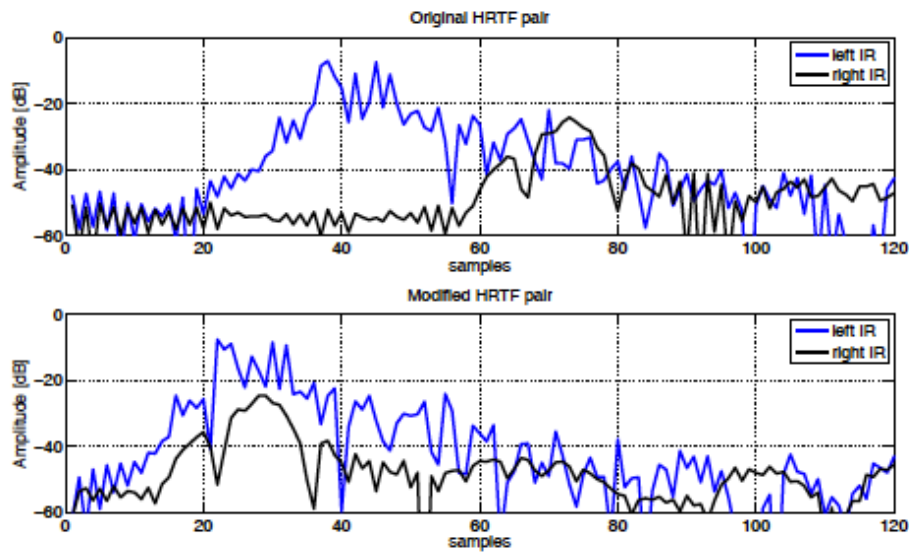


Abbildung 14: Extraktion der ITD mittels Onset Detektion

Durch die Extraktion der ITD kann nun die manipulierte ITD aufgerechnet werden und die Kanäle mit dem jeweiligen Delay versehen werden. Dazu wird die individuelle Laufzeitdifferenz ermittelt.

4.3.2.2. Individualisierung der Interauralen Laufzeitdifferenz

Eine Möglichkeit, um die interaurale Laufzeitdifferenz zu ermitteln, ist die einfache, aber wirkungsvolle Woodworth-Schlosberg-Formel.

$$ITD(\theta) = \frac{r_{Kopf}}{c_0} (\sin(\theta) + \theta)$$

r = Kopfradius

θ = Azimuth

c = Schallgeschwindigkeit

Durch sie ist eine individuelle Manipulation der ITD auszuführen. Zu dieser Formel sind verschiedene Kritikpunkte hervorgebracht worden: Zum einen richtet sich die Formel an einem runden Kopf aus, zum anderen sitzen die Ohren in der Annahme von Woodworth-Schlosberg an den Enden des Kopfdurchmessers und nicht, wie

anthropometrisch korrekt, leicht nach hinten versetzt.¹²⁶ Eine weitere Schwachstelle der Formel ist, dass die Elevation des Schallereignisses nicht in die Berechnung mit einbezogen wird. Diese ist aber zur charakteristischen Bildung der individuellen ITD essentiell.¹²⁷

Um die frontale Ebene sowie die Elevation mit einzubeziehen, wurde Woodworth-Schlosberg von Lorcher und Jot zu folgender Formel ausgeweitet:¹²⁸

$$ITD(\theta) = \frac{r_{Kopf}}{c_0} (\arcsin(\cos\theta \sin\varphi) + \cos\theta \sin\varphi)$$

Für diese Art von Formel ist es wichtig, dass eine exakte Position der Schallquelle bekannt ist. Ob diese bekannt ist, hängt von der Anwendung ab.

Da die interaurale Laufzeitdifferenz frequenzabhängig ist, wurden von Kuhn¹²⁹ eine weitere Formel für niedrige Frequenzen vorgestellt:

$$ITD_{lf} = \frac{3r}{c} \sin\theta$$

Wightman und Kistler fanden in Versuchen heraus, dass man auch mit der Formel für hohe Frequenzen zufriedenstellende Ergebnisse erhält, weswegen die Formel von Kuhn und die gesonderte Bearbeitung von niederfrequenten Signalen vernachlässigt werden kann.¹³⁰

Abweichungen von der ITD, die für das menschliche auditive System nicht mehr hörbar sind, misst die Just Noticeable Difference (JND) der ITD. Bei einem Signal aus der Medianebene beträgt diese nur 10 µs, je größer der Wert der ITD wird, desto mehr wächst auch die JND ITD. Bei einer interauralen Laufzeitdifferenz von 790 µs beträgt diese bereits 50 µs.¹³¹ In Versuchen von Estrella überschritt die Differenz einer mit Woodworth-Schlosberg manipulierten ITD zu einer Vergleichs-ITD, die aus der HRTF

¹²⁶ (Vgl. Lindau u. a., 2010, S. 3)

¹²⁷ (Vgl. Estrella, 2010, S. xii f.)

¹²⁸ Ebd.

¹²⁹ (Vgl. Kuhn, 1977, S. 157 ff.)

¹³⁰ (Vgl. Nicol, 2010, S. 36)

¹³¹ (Vgl. Nicol, 2010, S.39)

extrahiert wurde, lediglich 35 μ s bei einer seitlichen Auslenkung von 30°. ¹³² Daraus wird geschlossen, dass die Formel nach wie vor ihre Berechtigung hat und zur Individualisierung genutzt werden kann. Die Realitätsstreue der ITD hängt von dem genutzten Kopfradius ab. Die Möglichkeiten, diesen realistisch zu bestimmen, werden im Folgenden erörtert.

4.3.2.3. Der optimale Kopfradius

Um den Kopfradius zu erhalten, muss der Durchmesser des Kopfes halbiert werden. Dies stellt mit Sicherheit die nutzerfreundlichste Lösung dar, da lediglich ein Parameter gemessen werden muss. Den optimalen Kopfradius, der am ehesten der Anthropometrie entspricht, formuliert Algazi unter Berücksichtigung von drei ausgemessenen anatomischen Werten.

$$r_{opt} = w_1 x_1 + w_2 x_2 + w_3 x_3 + b \text{ [cm]}$$

x_1 = Kopfbreite/2

x_2 = Kopfhöhe/2

x_3 = Kopftiefe/2

Durch Regression zwischen verschiedenen Nationalitäten ermittelte Messungen ergaben die folgende empirische Gewichtung der Parameter: ¹³³

$$r_{opt} = 0.51 x_1 + 0.019 x_2 + 0.18 x_3 + 3.2 \text{ [cm]}$$

Um die Anwendung nutzerfreundlich zu gestalten, ist allerdings eine Ausmessung von drei Parametern durch den VR-User nicht sinnvoll. Es soll daher der erstgenannte Ansatz verwendet werden. Wichtig bei der Messung ist der Abstand zwischen den Ohren und somit die Strecke, die auf der Frontalebene liegt und den Kopf in Vorne und Hinten teilt.

¹³² (Vgl. Estrella, 2010, S. xiii)

¹³³ (Jorgos Estrella, 2011, S. 17)

4.3.3. Aufbau der Anwendung

Um ein möglichst betriebssystemunabhängiges Modell zu erstellen, wird der folgende Algorithmus zur Berechnung der ITD im Pseudocode verfasst.

Die Einbettung der Berechnung wird im Folgenden dargestellt. Um einen Wert zur ITD-Berechnung zu erhalten, wird der Nutzer zur Eingabe seines Kopfdurchmessers aufgefordert. Alternativ kann auch über eine Auswahl einer BRIR mit extrahierter ITD nachgedacht werden, so dass die Eingaben von zwei Parametern vom Nutzer getroffen werden. Die Daten des Headtrackings durch die Gyroskop-Sensoren müssen miteinbezogen werden. Die ausgewählte BRIR oder auch eine generalisierte BRIR sowie die durch die App und Headtracking ermittelten Positionsdaten werden dem Faltungsalgorithmus übergeben. Dieser bedient sich der in 3.3.2.3. beschriebenen schnellen Faltung. Die Faltung wird hier in zwei Teile aufgesplittet, erstens in die Faltung mit dem Betrag der HRTF und zweitens in die Aufrechnung der Phasendifferenz auf die Signale. Die beiden Signale für den linken und den rechten Kanal, die bei der Faltung entstehen, werden anschließend mittels der errechneten ITD mit einem entsprechenden Delay versehen.

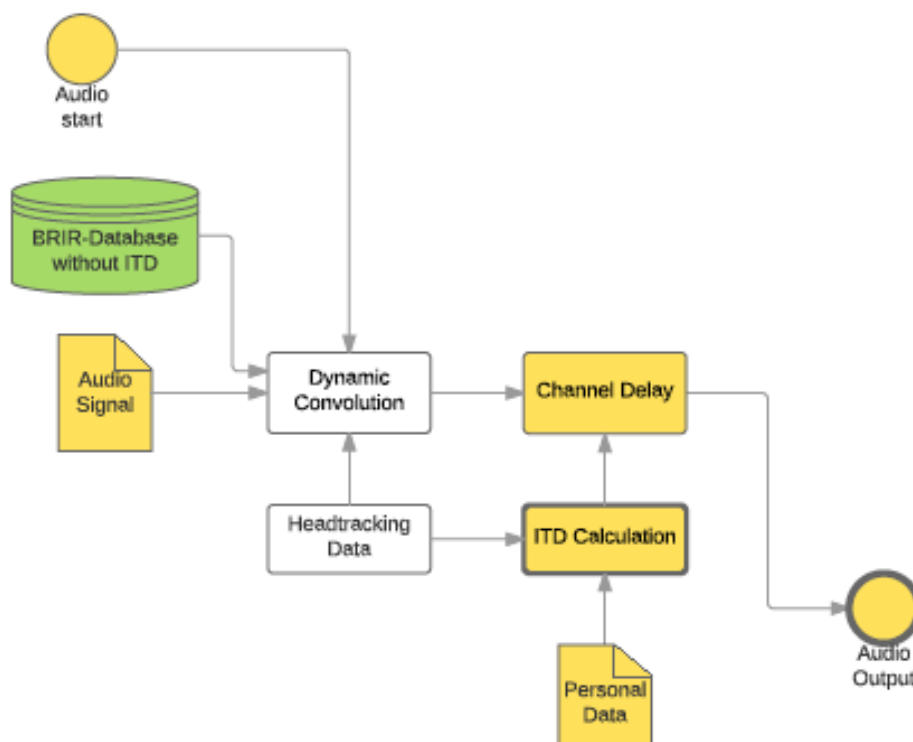


Abbildung 15: Grafische Darstellung der Anwendung

Die für die Errechnung der Interauralen Laufzeitdifferenz zu nutzenden Parameter erhält man durch die Informationen des Headtrackings sowie der Eingabe durch den Nutzer. Ziel ist es, dass der Nutzer nur einmal seinen Kopfdurchmesser eingeben muss, dieser wird dann pro Anwendung in einem Profil gespeichert. Die Errechnung der ITD erfolgt über die Woodworth-Schlosberg-Formel, das entsprechende Delay wird auf den linken oder rechten Kanal gegeben. Um unnötigen Rechenaufwand zu vermeiden, soll mit einer Abtastrate von 25 Hz abgefragt werden, ob sich die Position des Nutzers verändert hat. Ist dies der Fall, wird eine neue ITD berechnet. Die Klasse zur Errechnung der interauralen Laufzeitdifferenz könnte im Pseudocode so aussehen:

Program ITD-Calculation (*angle*, *diameter*)

$c = 343 \text{ m/s}$

$radius = diameter / 2$

If $0^\circ < angle < 90^\circ$

$ITD = \frac{radius}{c} (\sin(angle) + angle)$

return („Delay right“, ITD)

If $180^\circ < angle < 360^\circ$

$calculated\ angle = 360^\circ - angle$

$ITD = \frac{radius}{c} (\sin(calculated\ angle) + calculated\ angle)$

return („Delay left“, ITD)

In der Methode werden die nötigen Parameter durch die App bzw. das Headtracking übergeben. Durch eine if-Schleife werden die entsprechenden Winkel ihrem späteren Delay zugeordnet. Dadurch kann das Delay dann auf den richtigen Kanal aufgerechnet werden. Die Rückgabewerte der ITD sowie der Information für den Kanal werden später von dem Channel Delay verarbeitet.

5. Fazit und Ausblick

In der Arbeit sollte das Problem einer Individualisierung der Binauralsynthese auf Smartphones behandelt werden. Bisherige Modelle der räumlichen Darstellung lassen weder Mobilität noch platzsparende Lösungen zu. Die Binauraltechnik bietet die beste Räumlichkeit sowie durch die Kopfhörerwiedergabe eine mobile anwendbare Lösung. Die aufwändige Messmethode der eigenen HRTF stellt für die Individualisierung nach wie vor ein Problem für den Nutzer dar. Es wurde eine Methode zur nutzerfreundlichen Individualisierung vorgestellt, um eine realistische Abbildung des natürlichen Hörbildes zu schaffen. Da durch die zunehmenden Veröffentlichungen von VR-Anwendungen dem Konsumenten ein immer größeres Angebot zur Verfügung gestellt wird, soll dieser die Möglichkeit haben, die akustische Simulation auf sein Hörbild anzupassen. Dabei wurde als Individualisierungslösung eine Datenbank von verschiedenen HRTFs vorgeschlagen, deren ITD extrahiert wird. Der Individualisierungsfaktor wurde je nach Möglichkeit der Einflussnahme sowie Gewichtung für die Lokalisation ermittelt. Die nutzerspezifische Manipulation findet über die Berechnung der interauralen Laufzeitdifferenz statt, da diese eine der wichtigsten Kenngrößen zur Lokalisation darstellt und durch Woodworth-Schlosberg einfach verändert werden kann. Durch die Eingabe des Kopfdurchmessers des Nutzers kann die ITD errechnet und mit einer HRTF mit extrahierter ITD gefaltet werden. Eine geeignete Methode zur Extraktion der interauralen Laufzeitdifferenz mittels Onset Detektion wurde vorgestellt. Dazu wurde der Aufbau einer App, in welche die Anwendung eingebettet werden kann, ermittelt und in BPMN2.0 visualisiert, sowie die technischen Anforderungen geklärt. Es wurden diverse Möglichkeiten vorgestellt, in denen eine nutzerfreundliche Individualisierung Verwendung finden kann. Insbesondere für weitere Formen von Virtueller Realität, die aufgrund ihrer Beschaffenheit mobil genutzt werden sollen, wie Stadtrundgänge oder ähnliches, ist eine derartige Form der Anpassung an den Nutzer notwendig. Die Methode zur ITD-Errechnung wurde in Pseudocode dargestellt.

Die vorgestellte Lösung stellt einen Kompromiss zwischen möglichst genauer Abbildung und Nutzerfreundlichkeit dar. Viele Möglichkeiten, auf die HRTF Einfluss zu nehmen, werden aufgrund von mangelnder Usability vernachlässigt. Darunter fällt eine Skalierung der spektralen Filter durch den Nutzer oder die Eingabe von weiteren

Werten für die Kopfradius-Berechnung. Dennoch wird ein maßgeblicher Teil der HRTF bestimmt und individualisiert. Dies kann den ersten Schritt zu einer persönlich individualisierbaren HRTF bilden, die mobil und in kurzer Zeit durchgeführt werden kann.

Für weitere Individualisierungen, die auch genauere Ergebnisse liefern, muss die Form der Pinna miteinbezogen werden. Zukünftige Lösungen könnten über eine Erfassung der Ohrmuschelform per Smartphone-Fotos aus entsprechenden Winkeln funktionieren, welche dadurch die Aufgabe einer 3D-Lasererfassung übernehmen könnten und Form und Ausrichtung für die HRTF berücksichtigen. Methoden zur Lasererkennung der Ohrmuschel existieren bereits, allerdings gilt hier ebenso wie bei HRTF-Ausmessungen, dass diese nur unter größerem Aufwand und mit entsprechender Technik zu bewerkstelligen sind. Weitere Individualisierungen können auch die ILD betreffen, hier muss die Anatomie ebenfalls miteinbezogen werden.

Beim jetzigen Stand der Technik stellt die hier vorgestellte Methode allerdings die optimale Synthese aus einem möglichst individualisierten Hörerlebnis und einer nutzerfreundlichen Anwendung dar.

Literatur- und Quellenverzeichnis:

Bücher:

- Ansorge**, Ulrich, und Helmut Leder. *Wahrnehmung und Aufmerksamkeit*. 1. Aufl. Basiswissen Psychologie. Wiesbaden: VS-Verl, 2011.
- Birkner**, Christian. *Surround: Einführung in die Mehrkanalton-Technik ; das Audioformat der Zukunft ; Standards - Aufnahmepraxis - Weiterverarbeitung*. Ein Fachbuch von PPV-Medien. Bergkirchen: PPV, Presse-Project-Verl.-GmbH, 2002.
- Blauert**, Jens. *Räumliches Hören*. Monographien der Nachrichtentechnik. Stuttgart: Hirzel, 1974.
- Dickreiter**, Michael, Volker Dittel, Wolfgang Hoeg, und Martin Wöhr. *Handbuch der Tonstudientechnik*. 8., überarbeitete und erweiterte Auflage. De Gruyter Saur reference. Berlin: De Gruyter Saur, 2014.
- Dörner**, Ralf, Wolfgang Broll, Paul Grimm, und Bernhard Jung, Hrsg. *Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. eXamen.press. Berlin: Springer Vieweg, 2013.
- Görne**, Thomas. *Tontechnik: [Schwingungen und Wellen ; Hören ; Schallwandler ; Impulsantwort, Faltung ; Sigma-Delta-Wandler ; Stereo, Surround, WFS ; Regiegeräte, tontechnische Praxis] ; mit 207 Bildern und 33 Tabellen*. 2., aktualisierte Aufl. München: Hanser-Verl, 2008.
- Mednieks**, Zigurd, Hrsg. *Android-Programmierung: [Java-Programmierung für moderne mobile Endgeräte]*. 1. Aufl. Köln: O'Reilly, 2013.
- Macpherson**, Ewan A., und John C. Middlebrooks. „Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited“. *Journal of the Acoustical Society of America* 111 (5. Januar 2002): 2219–36.
- Møller**, Henrik. „Fundamentals of Binaural Technology“. *Applied Acoustics*, Nr. 32 (3. März 1992): 171–218.
- Nicol**, Rozenn. *Binaural Technology*. New York: Audio Engineering Society, 2010.
- Plack**, Christopher J. *The sense of hearing*. Mahwah, N.J: Lawrence Erlbaum Associates, 2005.

Post, Uwe. *Android-Apps entwickeln: eine Spiele-App von A bis Z ; [Schritt für Schritt zu eigenen Apps und Spielen ; inkl. Crashkurs zu Java und objektorientierter Programmierung ; Animationen, Sounds, Zeichnen, Bewegungssensoren, Kamera, Highscores u.v.m].* 4., aktualisierte Aufl. Galileo Computing Für Einsteiger. Bonn: Galileo Press, 2014.

Randall, R. B. *Application of B & K Equipment to Frequency Analysis.* S.l.: Brüel & Kjaer, 1977.

Schmicking, Daniel. *Hören und Klang: empirisch phänomenologische Untersuchungen.* Epistemata, Bd. 336. Würzburg: Königshausen & Neumann, 2003.

Zwicker, Eberhard. *Psychoakustik.* Hochschultext. Berlin: Springer, 1982.

Fachartikel und Paper:

Estrella, Jorgos. „Zur Extraktion von interauralen Laufzeitdifferenzen in binauralen Raumimpulsantworten“. Technische Universität Berlin, 2010.

Estrella, Jorgos. „Real Time Individualization of Interaural Time Differences for Dynamic Binaural Synthesis“. Diplomarbeit, Technische Universität Berlin, 2011.

Begault, Durand R. „3D Sound for Virtual Reality and Multimedia“, 2000.

Glyde, Helen, Jörg M. Buchholz, Harvey Dillon, Sharon Cameron, und Louise Hickson. „The Importance of Interaural Time Differences and Level Differences in Spatial Release from Masking“. *The Journal of the Acoustical Society of America* 134, Nr. 2 (2013): EL147. doi:10.1121/1.4812441.

Guillon, Pierre. „Individualisation des indices spectraux pour la synthèse binaurale (recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF)“. Université du Maine, 2009.

Hädrich, Markus. „Konzeption und Implementierung einer mobilen App für geolokalisierte dynamische Binauralsynthese“. TU Berlin, 2015.

Kuhn, George F. „Model for the Interaural Time Differences in the Azimuthal Plane“. *The Journal of the Acoustical Society of America* 62, Nr. 1 (1977): 157. doi:10.1121/1.381498.

- Kulkarni**, Abhijit, S. K. Isabelle, und H. S. Colburn. „Sensitivity of Human Subjects to Head-Related Transfer-Function Phase Spectra“. *The Journal of the Acoustical Society of America* 105, Nr. 5 (1999): 2821. doi:10.1121/1.426898.
- Lindau**, Alexander, Jorgos Estrella, und Stefan Weinzierl. „Individualization of dynamic binaural synthesis by real time manipulation of the ITD“. London, 2010.
- Litfin**, Dominik. „Binauralsynthese von Mehrspurproduktionen auf mobilen Endgeräten“, 28. Februar 2014.
- Mrsic-Flogel**, Thomas, Andrew King, Rick Jenison und Jan Schnupp. „Listening through different ears alters spatial responsefields in primary auditory cortex“. *Journal of Neurophysiology* 86, Nr.2 (01. Juli 2001): 1043-1046.
- Reineboth**, Philipp. „Virtuelle Professionalität - kann eine simulierte Abhörumgebung eine reale Tonregie ersetzen?“ Bachelorarbeit, Hochschule der Medien, 2012.
- Sanders**, Christian, Frank Wefers, und Dieter Lockschat. „Scalable binaural synthesis on mobile devices“. In *Convention Paper 8783*. San Francisco, USA, 2012.
- Siveke**, I. „Binaural Response Properties of Low-Frequency Neurons in the Gerbil Dorsal Nucleus of the Lateral Lemniscus“. *Journal of Neurophysiology* 96, Nr. 3 (11. Januar 2006): 1425–40. doi:10.1152/jn.00713.2005.
- Smyth**, Mike. „Headphone-Monitoring of the new Immersive Audio Formats“. *vdt Magazin*, Nr. Heft 2/2015 (2015): 41–43.
- Tarnow**, Christopher und Daniel Keinath. „Entwicklung und Erprobung einer Methode zur Integration binauraler Raumimpulsantworten bei der Mischung von Mehrspurproduktionen“. Hochschule für Musik Detmold, 2009.
- Theile**, Günther. „Der Studio-Kopfhörer“. *vdt Magazin*, Nr. Heft 2/2015 (2015): 9–16.
- Theile**, Günther, und Helmut Wittek. „Die dritte Dimension für Lautsprecher-Stereofonie“, Mai 2011. http://www.hauptmikrofon.de/doc/Auro3D_D_0311.pdf.
- Vassilakis**, Pantelis. „Acousticslab: Auditory Localisation Cues“. *Columbia College, Chicago*. Zugriffen 26. Januar 2016. <http://acousticslab.org/psychoacoustics/PMFiles/Module07b.htm#12c>.
- Wisse**, Elke, und Denis Schuster. „Kopfhörer-Virtualisierung“. *vdt Magazin*, Nr. Heft 2/2015 (2015): 43–45.

Homepages und Blogs:

Ammermann, Tom. „Firmenhomepage: New Audio Technology, SAD“. *Firmenhomepage: New Audio Technology, SAD*. (Stand: 15.01.2016)

<http://www.newaudiotechnology.com>

Bavor, Clay. „(Un)folding a virtual journey with Google Cardboard“. *Googleblog*, 27. Januar 2016. (Stand: 15.02.2016)

<https://googleblog.blogspot.com/2016/01/unfolding-virtual-journey-cardboard.html>.

IfD Allensbach. „Anzahl der Personen in Deutschland, die HiFi-Lautsprecherboxen im Haushalt besitzen, von 2012 bis 2015“, Juli 2015. (Stand: 20.02.2016)

Kuchera, Ben. „Google Cardboard Adds ‚Inexpensive‘ Spatial Audio to Its SDK“. *Gaming magazine. Polygon*, 13. Januar 2016. (Stand: 15.01.2016)

<http://www.polygon.com/2016/1/13/10759950/google-cardboard-sdk-spatial-audio>.

Martz, Nathan. „Spatial Audio Comes To The Cardboard SDK“. *Google Developers Blog*, 13. Januar 2016. (Stand: 15.01.2016)

<http://googledevelopers.blogspot.de/2016/01/spatial-audio-comes-to-cardboard-sdk.html>.

Moynihn, Tim. „Why Spatial Audio Is Such A Big Deal For Google Cardboard“. *Technologie-Magazin. Wired*, 17. Januar 2016. (Stand: 17.01.2016)

<http://www.wired.com/2016/01/why-spatial-audio-is-a-big-deal-for-google-cardboard/>.

Palmer, Luckey. „The Future of VR“. *www.reddit.com*, 25. März 2014. (Stand: 24.02.2016)

<https://redd.it/21cy9n>.

University of California, Davis, „Psychoacoustics of Spatial Hearing“.

Universitätshomepage, 25. Februar 2011. (Stand: 23.01.2016)

<http://interface.cipic.ucdavis.edu/sound/tutorial/psych.html>.

Schmid, Holger. „Anzahl der Smartphone-Nutzer in Deutschland in den Jahren 2009 bis 2015 (in Millionen)“. *Statista*, September 2015 (Stand: 18.01.2016).

<http://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenuutzer-in-deutschland-seit-2010/>.

„Samsung Galaxy S6“. *Samsung Homepage*, 2016. (Stand: 13.03.2016)

<http://www.samsung.com/de/consumer/mobile-devices/smartphones/galaxy-s/SM-G920FZKADBT>.

Zeitungsartikel:

Jung, Helmut Martin. „Samsung zeigt Galaxy S7 mit Virtual Reality: ‚Es ist, als wäre man da‘“. *Sueddeutsche Zeitung*. 21. Februar 2016. (Stand: 13.03.2016)

<http://www.sueddeutsche.de/digital/mobile-world-congress-samsung-praesentiert-galaxy-s-wasserdichte-virtual-reality-mit-facebook-1.2874250>.

Bildquellen

Abbildung 1: <http://www.dhi-online.de/DhiNeu/Harten/doku/8-2.htm> (24.03.2016)

Abbildung 2: Ebenen des Hörens

Abbildung 3: Görne, 2008, S. 104

Abbildung 4: Plack, 2005, S. 177

Abbildung 5: Birkner, 2002, S. 48

Abbildung 6: Birkner, 2002, S. 84f.

Abbildung 7: www.neumann.com (24.03.2016)

Abbildung 8: Tarnow und Keinath, 2009, S. 12

Abbildung 9: Bavor, 2016 (15.02.2016)

<https://googleblog.blogspot.com/2016/01/unfolding-virtual-journey-cardboard.html>.

Abbildung 10: Kzero.co.uk, 2014

<http://de.statista.com/statistik/daten/studie/426237/umfrage/prognose-zur-anzahl-der-aktiven-virtual-reality-nutzer-weltweit/> (25.02.2016)

Abbildung 11: Mrsic-Flogel u.a., 2001, S. 1044

Abbildung 12: Macpherson, 2002, S. 2221

Abbildung 13: Sanders, 2012, S. 6

Abbildung 14: Estrella, 2011, S. vii

Abbildung 15: Grafische Darstellung der Anwendung

Elektronisches Exemplar der Arbeit:



An dieser Stelle ist eine CD mit einem PDF der
vorgelegten Arbeit eingelegt.