Ein Vergleich von Virtual Reality Ton auf Basis von Kopfhörern und Lautsprechern.

Bachelorarbeit
im Studiengang
Audiovisuelle Medien

Vorgelegt von
Nils Beermann
Matrikelnummer: 27196
am 28. Februar 2017
an der Hochschule der Medien Stuttgart

Erst-Prüfer: Prof. Oliver Curdt
Zweit-Prüfer: Prof. Dr. Michael Felten

Eidesstattliche Versicherung:

Hiermit versichere ich, Nils Beermann, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Ein Vergleich von Virtual Reality Ton auf Basis von Kopfhörern und Lautsprechern" selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn 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 ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

_____

Unterschrift

Kurzfassung:

Diese Bachelorarbeit befasst sich mit der Tonwiedergabe von Virtual Reality Anwendungen über Kopfhörer sowie Lautsprecher. Das Ziel der Arbeit ist ein Vergleich zwischen der Kopfhörer- und Lautsprecherwiedergabe auf Basis von Literatur.
Zuerst werden die Funtionsweisen des Gehörs und des räumlichen Hörens betrachtet, sowie mehrere Technologien für die Produktion von immersivem Ton und dessen Wiedergabe untersucht. Weiter wird ein überblick über das Thema Virtual Reality und dessen Anwendungen gegeben. Zuletzt werden die Stärken und Schwächen der verschiedenen Verfahren, sowie ihre Eignung für verschiedene Virtual Reality Anwendungen untersucht. Es wird die Eignung der beiden Wiedergabemedien, in Kombination mit den gängigsten Technologien zur Darstellung von Virtual Reality, diskutiert.

Abstract:

This bachelor thesis deals with the reproduction of sound in virtual reality applications using headphones and loudspeakers. The objective is a comparison between the reproduction with headphones and the reproduction with loudspeakers based on literature.
Initially the workings of the human ear and spatial hearing are regarded as well as multiple technologies for the production and reproduction of immersive sound. Further an overview of the topic of virtual reality and its applications is given. Finally strengths and weaknesses of the different technologies as well as their suitability for various virtual reality applications are examined. The qualification of the two reproduction methods in combination with the most common technologies for the presentation of virtual reality is discussed.

Abbildungsverzeichnis:

Abbildung 1: Schematische Darstellung des menschlichen Hörapparates

Abbildung 2: Schematische Darstellung des kopfbezogenen Koordinatensystems

Abbildung 3: Die Richtungsbestimmenden Bänder nach Jens Blauert

Abbildung 4: Beispiel eines HRTF-Paares

Abbildung 5: Der Kunstkopf KU-100 der Firma Neumann

Abbildung 6: Grafische Darstellung der Kugelflächenfunktionen nullter bis dritter Ordnung (Urheber: I, Sarxos)

Abbildung 7:Links: Sennheiser Ambeo® VR-Mic mit vier Kapseln für FOA | Mitte: MH Acoustics Eigenmike® mit 32 Kapseln für HOA bis vierter Ordnung | Rechts: Visisonics Audio/Visual Camera mit 64 Kapseln für HOA bis zu siebter Ordnung

Abbildung 8: Außenansicht des CAVE an der Fachhochschule Seinäjoki

Abbildung 9: Google Cardboard VR-Brille

Abbildung 10: Samsung GEAR VR

Abbildung 11: Infografik zur VR Audio Engine

Abkürzungsverzeichnis:

CAVE	Cave Automatic Virtual Environment
dB	Dezibel
FOA	First Order Ambisonics
GPS	Global Positioning System
HMD	Head-Mounted-Display
HOA	Higher Order Ambisonics
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function
Hz	Herz
ILD	Interaural Level Difference
ITD	Interaural Time Difference
kbit/s	Kilobit pro Sekunde
kHz	Kiloherz
LCD	Liquid Crystal Display
mbit/s	Megabit pro Sekunde
SDK	Software Development Kit
VR	Virtual Reality

Inhaltsverzeichnis

Eidesstattliche Versicherung:.....3
Kurzfassung:.....4
Abstract:.....4
Abbildungsverzeichnis:.....5
Abkürzungsverzeichnis:.....5
1 Einleitung.....8
2 Hören.....9
2.1 Das menschliche Gehör.....9
2.1.1 Aufbau und Funktion des Hörapparates.....9
2.1.2 Wahrnehmung von Frequenz und Schalldruck.....11
2.2 Räumliches Hören.....12
2.2.1 Die Lage von Schallereignissen im Raum.....12
2.2.2 Horizontales Richtungshören.....13
2.2.2.1 Laufzeitunterschiede.....13
2.2.2.2 Pegelunterschiede.....14
2.2.3 Vertikales Richtungshören.....15
2.2.3.1 Klangfarbe.....15
3 Binauraler Ton.....17
3.1 Grundlagen.....17
3.2 HRTF / HRIR.....17
3.3 Anwendungen.....21
3.3.1 Aufnahme.....21
3.3.2 Synthese.....22
3.3.3 Wiedergabe.....22
4 Ambisonics.....24
4.1 Basics oder so.....24
4.1.1 Grundlagen und Kugelflächenfunktionen.....24
4.1.2 FOA und HOA.....26
4.2 Produktion.....27
4.2.1 Aufnahme.....27
4.2.2 Bearbeitung und Transfer.....28
4.2.3 Wiedergabe.....29
5 Objektbasierter Ton.....31
5.1 Prinzip.....31

5.2 Produktion....32
5.2.1 Aufnahme und Bearbeitung....32
5.2.2 Übertragung und Wiedergabe....33
6 Virtual Reality....35
6.1 Grundlagen....35
6.1.1 VR-Systeme....35
6.1.2 Geschichte der virtuelle Realität....36
6.2 Anwendungen....38
6.2.1 Lineare VR-Anwendungen....39
6.2.2 Interaktive VR-Anwendungen....40
6.2.3 Mobile und Stationäre VR....41
7 VR-Sound....44
7.1 Ambisonics....44
7.1.1 Stärken&Schwächen....44
7.1.2 Mögliche Anwendungen....46
7.2 Binaurale Aufnahme und Synthese....47
7.2.1 Stärken & Schwächen....47
7.2.2 Mögliche Anwendungen....48
7.3 Objektbasiert....49
7.3.1 Stärken & Schwächen....49
7.3.2 Mögliche Anwendungen....50
7.4 Andere und Mischformen....51
7.4.1 Stärken & Schwächen....51
7.4.2 Mögliche Anwendungen....52
7.5 Kopfhörer oder Lautsprecher....53
7.5.1 Mobile VR-Anwendungen....53
7.5.2 3D-Bildschirme und Projektionen....54
7.5.3 Cave Automatic Virtual Environment....54
7.5.4 HMDs....55
8 Fazit....56
Literaturverzeichnis....58
Artikel:....58
Bücher:....59
Paper:....59
Internetseiten:....61
Abbildungsquellen:....63
Elektronisches Exemplar:....64

1 Einleitung

Virtuelle Realität ist längst keine Zukunftsmusik mehr. Schon seit den 1980er Jahren gibt es Videobrillen, Datenhandschuhe und Bewegungsverfolgung. Mit den Markteinführungen mehrerer kommerzieller Virtual Reality (VR) Brillen 2016 hat diese Technologie einen weiteren großen Schritt nach vorne und Virtuelle Realität für den durchschnittlichen Konsumenten verfügbar gemacht.

Bis vor wenigen Jahren stand bei der Weiterentwicklung in diesem Bereich vor allem der visuelle Aspekt im Vordergrund. Doch der Ton konnte in den letzten Jahren aufholen und bietet inzwischen vielversprechende sowie bereits beeindruckende Möglichkeiten den Nutzer in eine virtuelle Welt zu transportieren. Dabei bedienen sich diese Techniken auf der einen Seite Lautsprechern, wie das objektbasierte System Dolby Atmos oder das fast in Vergessenheit geratene Ambisonics und auf der anderen Seite Kopfhörern, wie die Binauralsynthese.

Es soll verglichen werden, ob sich die Wiedergabe über Lautsprecher oder Kopfhörer besser bei Virtual Reality Anwendungen eignen. Dazu werden zunächst Aufbau und Funktion des menschlichen Gehörs im Hinblick auf räumliches Hörern betrachtet. Weiter werden binaurale Techniken sowie Ambisonics und objektbasierte Verfahren untersucht. Anschließend wird ein Überblick über die Entwicklungsgeschichte und Anwendungen für Virtual Reality gegeben. Zuletzt werden die Stärken und Schwächen der genannten Technologien für mögliche Anwendungen aus dem Bereich Virtual Reality betrachtet sowie die Frage nach Kopfhörern oder Lautsprechern geklärt.

2 Hören

2.1 Das menschliche Gehör

Bei vielen Virtual Reality Anwendungen soll eine möglichst vollständige Immersion des Nutzers stattfinden. Um diese zu erreichen müssen mehrere Sinne angesprochen werden. So wird ein höherer Grad an Immersion erreicht, wenn neben visuellen Reizen auch auditive hinzukommen.[1] Dabei ist wichtig, dass die verschiedenen Reize zeitlich und inhaltlich kongruent sind.
Nachfolgend soll das menschliche Gehör im Hinblick auf räumliches Hören untersucht werden, um festzustellen, welche Informationen geliefert werden müssen, damit der Nutzer überzeugend in eine virtuelle Welt versetzt werden kann.

2.1.1 Aufbau und Funktion des Hörapparates

Das menschliche Gehör dient zur Wahrnehmung von Schallereignissen in der Umgebung. Es kann in drei Bereiche aufgeteilt werden, die unterschiedliche Bestandteile sowie Funktionen haben. Das Außenohr besteht aus der Ohrmuschel sowie dem Gehörgang und dient der Bündelung von eintreffenden Schallwellen und leitet diese zum eigentlichen Hörapparat weiter. Außerdem findet durch das Außenohr eine Filterung von Schallereignissen statt, die bei der Ortung von Schallquellen eine wichtige Rolle spielt (siehe Kapitel 2.2).[2]
Der zweite Teil, das Mittelohr, beginnt mit dem Trommelfell und umfasst des weiteren die Paukenhöhle mit den drei Gehörknöchelchen und die Eustachische Röhre. Das Trommelfell bildet eine luftdichte Abtrennung zwischen Außen- und Mittelohr und wird durch Schallwellen in Schwingung versetzt. Diese Schwingungen werden über die Gehörknöchelchen an das Innenohr übertragen. Außerdem fungiert das Mittelohr als Impedanzverstärker wobei durch Hebelwirkungen zwischen den Gehörknöchelchen und den Größenunterschied zwischen Trommelfell und Steigbügelfußplatte die Amplitude der Schwingung

1 (Vgl. Larsson et al., 2002, S.6)
2 (Vgl. Ulrich und Hoffmann, 2007, S.496f., S.586)

abnimmt und die übertragene Kraft verstärkt wird. Die Eustachische Röhre dient der Belüftung des Mittelohres, damit das Trommelfell möglichst frei schwingen kann.[3]

Das Innenohr wird aus der Schnecke und den Bogengängen des Gleichgewichtsorgans aufgebaut. Die Schnecke beherbergt mit dem Corti-Organ das eigentliche Hörorgan, welches mit Hilfe von Sinneszellen, den sogenannten Haarzellen, Schallwellen in Nervenimpulse umwandelt. Die Flüssigkeit mit der die Schnecke gefüllt ist, weist eine deutlich höhere Schallimpedanz auf als Luft. Ohne die Impedanzwandlung im Mittelohr würden über 98% des Schalls reflektiert werden, durch die Funktion des Mittelohrs können etwa 60% der Schallenergie auf das Innenohr übertragen werden.[4]

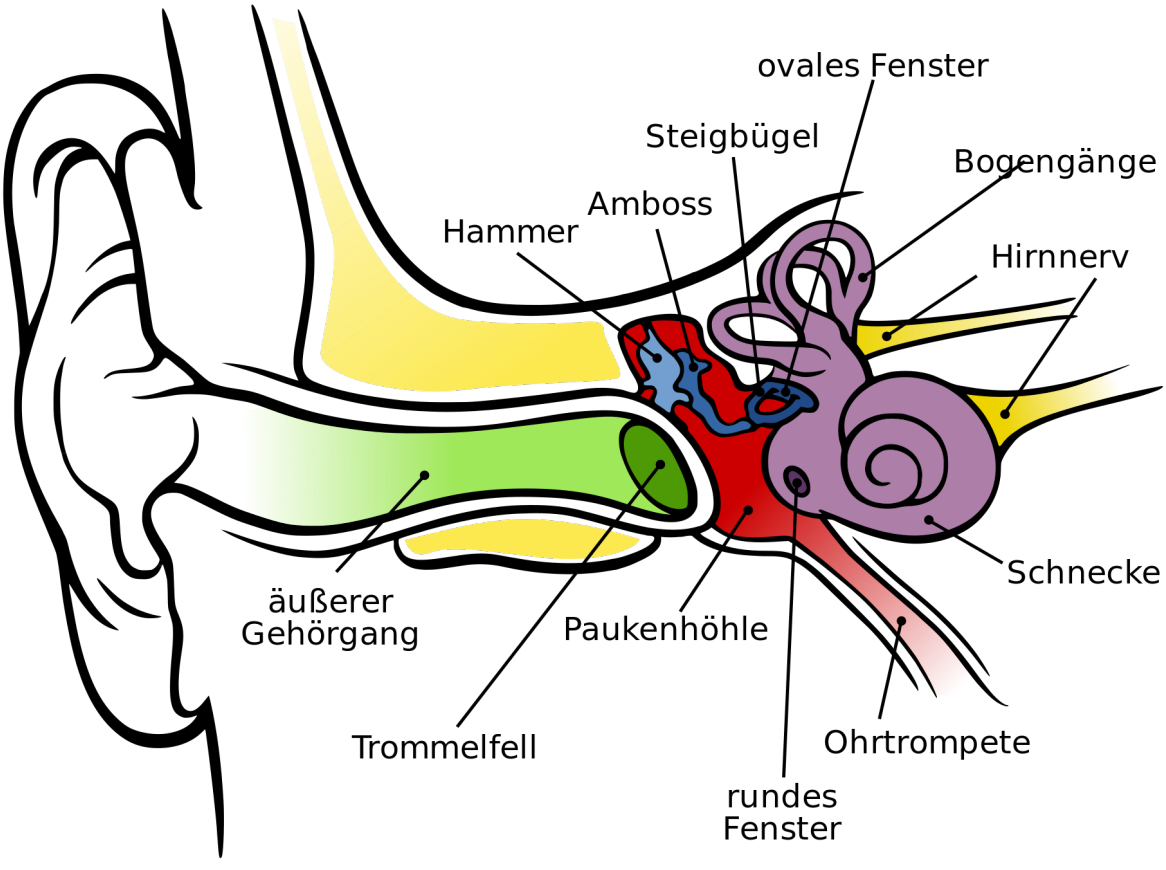

Abbildung 1: Schematische Darstellung des menschlichen Hörapparates

[3] (Vgl. Ulrich und Hoffmann, 2007, S. 502)
[4] (Vgl. Ulrich und Hoffmann, 2007, S. 509ff.)

2.1.2 Wahrnehmung von Frequenz und Schalldruck

Das Hörorgan des Menschen kann Frequenzen von etwa 20Herz (Hz) bis 18.000Hz erkennen. Bei einer Schallgeschwindigkeit von 1487m/s in 20°C warmen Wasser entspricht dieser Frequenzbereich Wellenlängen von ca. 8cm bis 74m. Das Corti-Organ und die Basilarmembran, auf der dieses sitzt, haben eine Länge von rund 32mm. Die Basilarmembran ist am Anfang der Schnecke aufgrund ihrer Abmessungen eher steif und wird wird zur Spitze der Schnecke hin fortlaufend weicher.[5] Aufgrund dieses Aufbaus hat jeder Punkt auf der Basilarmembran eine bestimmte Resonanzfrequenz. Schallwellen, die in das Innenohr gelangen, werden durch die Flüssigkeitsbewegung auf die Basilarmembran übertragen und wandern diese entlang. Wenn eine solche Welle den Punkt erreicht, dessen Resonanzfrequenz ihrer eigenen Frequenz entspricht verlangsamt sich ihre Ausbreitungsgeschwindigkeit und es tritt ein Maximum der Amplitude auf. Danach fällt die Amplitude der Welle stark ab. Somit wird jede Frequenz einem Amplitudenmaximum an einer bestimmten Stelle entlang des Corti-Organs zugeordnet. Das menschliche Gehör kann dabei eine Verschiebung des Amplitudenmaximums um 50µm feststellen und so etwa 640 Frequenzschritte wahrnehmen.[6]

Die Haarzellen mit deren Hilfe das Corti-Organ die Frequenzen von Schallwellen ermittelt spielen auch eine wichtige Rolle bei der Bestimmung der Lautstärke von Schallereignissen. Diese sind über Synapsen mit je einer Nervenfaser verbunden. Die Nervenfasern des Hörnervs senden, wenn die Haarzellen nicht gereizt werden Aktionspotenziale mit einem bestimmten Ruhetakt aus. Wenn Haarzellen gereizt werden, werden mehr Aktionspotenziale ausgesendet, wenn sie gehemmt werden weniger. Dadurch kann die Rate der Aktionspotenziale zur Bestimmung der Lautstärke verwendet werden. Bei höheren Pegeln ist die Rate insgesamt höher. Allerdings ist die maximale Rate begrenzt. Bei sehr großen Schalldrücken senden auch die Nervenfasern angrenzender Haarzellen Aktionspotenziale mit höherer

[5] (Vgl. Ulrich und Hoffmann, 2007, S. 514f.)
[6] (Vgl. Ulrich und Hoffmann, 2007, S. 516)

Rate. Dies hat zur Folge, dass das Gehör weniger Frequenzselektivität aufweist und eine Unterscheidung zwischen Nutz- und Störschall erschwert wird.[7]

2.2 Räumliches Hören

2.2.1 Die Lage von Schallereignissen im Raum

Schallereignisse können an beliebigen Stellen im Raum auftreten, daher wird zur Beschreibung ihrer Lage im Raum meist ein kopfbezogenes Kugelkoordinatensystem verwendet. Da Menschen ihre Ohren nur minimal relativ zum Kopf bewegen können ist dieses Koordinatensystem zugleich auch ohrenbezogen. Der Ursprung dieses Koordinatensystems liegt in der Mitte einer Geraden, die die beiden Öffnungen der Gehörgänge miteinander verbindet. In diesem Koordinatensystem wird der Raum durch drei Ebenen gegliedert. Die Horizontalebene wird horizontal um den Ursprung aufgespannt. Auf ihr liegt die Gerade zur Bestimmung des Ursprungs und sie unterteilt das Koordinatensystem in oben und unten. Die Gerade zwischen den Gehörgängen liegt auch auf der Frontalebene, welche orthogonal auf der Horizontalebene steht. Die Frontalebene gliedert das Koordinatensystem in vorne und hinten. Orthogonal auf beiden anderen Ebenen steht die Median- oder Vertikalebene und teilt das Koordinatensystem in links und rechts auf. Der Ursprungspunkt eines Schallereignisses wird durch die Koordinaten r, δ und φ beschrieben. Die Distanz zum Ursprung wird durch den Radius r, der Horizontalwinkel , auch als Azimut bezeichnet, durch den Winkel φ und der Vertikalwinkel, auch als Elevation bezeichnet, durch den Winkel δ beschrieben.[8] [9]

[7] (Vgl. Ulrich und Hoffmann, 2007, S. 521)
[8] (Vgl. Ulrich und Hoffmann, 2007, S. 581f.)
[9] (Vgl. Blauert, 1974, S. 11f.)

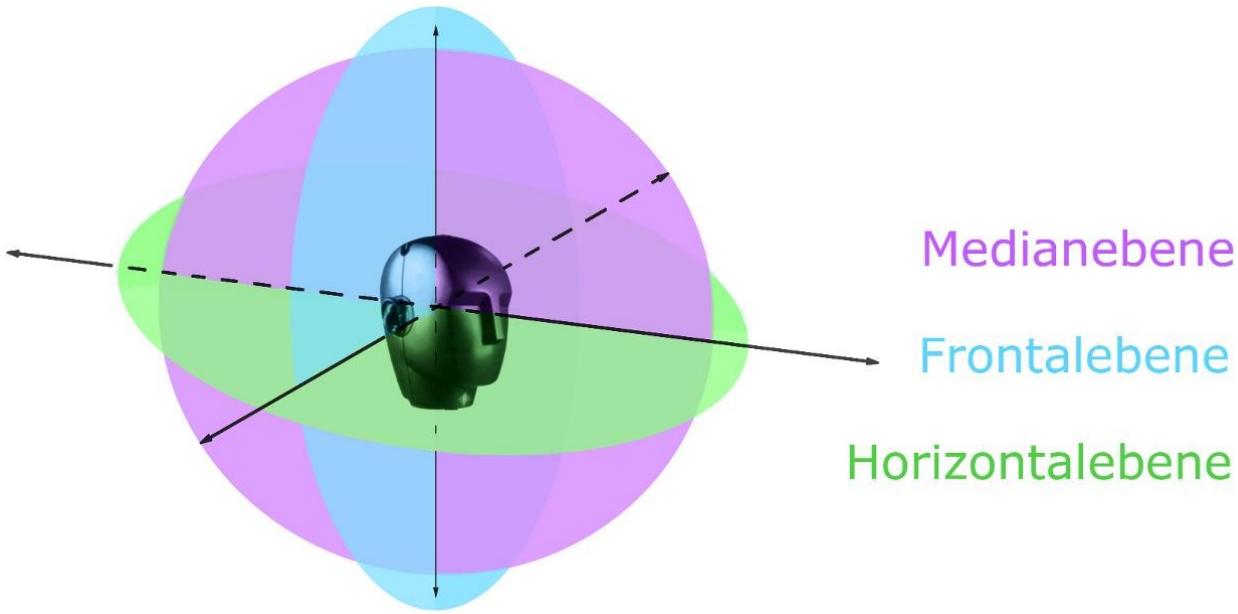

Abbildung 2: Schematische Darstellung des kopfbezogenen Koordinatensystems

2.2.2 Horizontales Richtungshören

2.2.2.1 Laufzeitunterschiede

Eine Möglichkeit des menschlichen Gehörs den horizontalen Winkel eines eintreffenden Schallereignisses zu bestimmen ist die Analyse von interauralen Zeitunterschieden (ITD). Bei einem Schallereignis, dessen Ursprung nicht auf der Medianebene liegt, erreichen die Schallwellen zuerst das näher gelegene Ohr und kurz darauf das weiter Entfernte.[10] Im Fall der Lateralisation, also der Wahrnehmung eines Schallereignisses im Kopf, die beispielsweise bei der Verwendung von Kopfhörern erfolgt, wird bei einer frequenzunabhängigen Verzögerung um etwa 630µs das Schallereignis im Eingangsbereich des Gehörganges lateralisiert dessen Ohr der Schall zuerst erreicht hat.[11] Eine solche volle Auslenkung kann bei Tönen nur dann stattfinden, wenn die halbe Periodendauer (T/2) 630µs nicht unterschreitet. Dadurch verringert sich bei Tonhöhen über etwa 800Hz die maximale Auslenkung und tritt bei T/2 auf.[12] Bezüglich der ITDs ist das Gehör in der Lage einzelne spektrale Anteile

[10] (Vgl. Ulrich und Hoffmann, 2007, S. 583f.)
[11] (Vgl. Blauert, 1974, S. 115)
[12] (Vgl. Blauert, 1974, S. 119)

einfallenden Schalls getrennt auszuwerten[13] und die Lateralisationsunschärfe, die kleinste Änderung der Verzögerung, die zu einer Änderung der wahrgenommenen Auslenkung führt, sinkt in der Regel mit steigendem Pegel und wachsender Signaldauer.[14] Des weiteren werden oberhalb von etwa 1,6 Kiloherz (kHz) in der Regel keine seitlichen Auslenkungen durch frequenzunabhängig verzögerte Töne erkannt. Bei Oktavrauschen, sowie amplitudenmodulierten Sinustönen hingegen, kann das Gehör Verschiebungen der Hüllkurven detektieren. Die Analyse der Hüllkurven beginnt ab einer Frequenz der Trägerschwingung von etwa 500Hz und wird mit steigender Trägerfrequenz genauer. Bei derartigen Signalen ohne spektrale Anteile unter etwa 1,6kHz werden vom Gehör ausschließlich Verschiebungen der Hüllkurven ausgewertet.[15]

2.2.2.2 Pegelunterschiede

Eine weitere Möglichkeit die seitliche Auslenkung eines Schallereignisses festzustellen bietet die Analyse von interauralen Pegelunterschieden (ILD). Diese entstehen vor allem dadurch, dass der Kopf, je nach Größe, für Schallwellen ab etwa 2 bis 3kHz ein Hindernis darstellt.[16] Auch bei der Bewertung von ILDs verarbeitet das Gehör spektrale Anteile getrennt, so kann es zu mehreren Hörereignissen kommen, wenn spektrale Anteile eines Signals mit unterschiedlichen Pegeldifferenzen präsentiert werden.[17] Die Lateralisationsunschärfe bei ILDs nimmt, vor allem bei tiefen Tönen, bei weiter Horizontaler Auslenkung zu. Diese Zunahme beginnt bei etwa 8-10Dezibel (dB) Pegelunterschied. Außerdem ist die Lateralisationsunschärfe pegelabhängig. Sie ist bei niedrigen und hohen Pegeln größer als bei mittleren.[18] Durch Anpassung und Ermüdung sinkt bei längerer Belastung die Empfindlichkeit des Gehörs mit der Zeit. Dadurch wandern seitlich Lateralisierte Signale mit der Zeit in Richtung der Kopfmitte, da das Ohr, das dem lauteren Signal ausgesetzt ist sich stärker

13 (Vgl. Blauert, 1974, S. 118)
14 (Vgl. Blauert, 1974, S. 125)
15 (Vgl. Blauert, 1974, S. 120ff.)
16 (Vgl. Ulrich und Hoffmann, 2007, S. 585)
17 (Vgl. Blauert, 1974, S. 128)
18 (Vgl. Blauert, 1974, S. 127ff.)

anpasst. Bei einer künstlichen Veränderung der Empfindlichkeit eines Ohres ist die Lateralisation zunächst zum empfindlicheren Ohr hin verschoben. Es kommt zu einer Anpassung über Zeit, die wieder zu einem normalen Höreindruck führt und durch Training beeinflussbar ist.[19]

Im Falle der Lateralisation hat die Auswertung von Verschiebungen der Trägerschwingungen einen wesentlichen Einfluss, wenn es keine Signalanteile über etwa 1,6kHz gibt und die Auswertung von ILDs und Hüllkurvenverschiebungen dominiert sobald ein Signal Anteile oberhalb von etwa 1,6kHz enthält. Damit sind im freien Schallfeld die Bewertungen von ILDs und Hüllkurvenverschiebungen dominant, da hier fast ausschließlich Signale mit Anteilen über 1,6kHz vorkommen. Hinzu kommt, dass die relative Rolle der Mechanismen interindividuell schwanken kann.[20]

2.2.3 Vertikales Richtungshören

2.2.3.1 Klangfarbe

Wird vereinfachend davon ausgegangen, dass der Kopf eines Menschen symmetrisch ist, kommt es bei Schallereignissen, die ihren Ursprung auf der Medianebene haben zu identischen Signalen an beiden Ohren.[21] In diesem Fall entstehen keine ITDs oder ILDs, die das Gehör auswerten könnte. Die Ohrmuschel ist durch ihre Form akustisch ein lineares Filter, dessen Übertragungsfunktion von Richtung und Entfernung einer Schallquelle abhängt.[22] Weiter bildet die Ohrmuschel zusammen mit dem Gehörgang ein akustisches Resonatorsystem. Auch die Anregung dieses Systems ist abhängig von Richtung und Entfernung der Schallquelle.[23] Diese beiden Mechanismen sorgen dafür, dass abhängig von der Einfallsrichtung des Schalls bestimmte Frequenzen verstärkt und andere gedämpft werden. Die dadurch entstehenden Minima und Maxima kann das Gehör bestimmten Richtungen zuordnen. Blauert hat durch Studien zu diesem

[19] (Vgl. Blauert, 1974, S. 130f.)
[20] (Vgl. Blauert, 1974, S. 139)
[21] (Vgl. Blauert, 1974, S. 78)
[22] (Vgl. Blauert, 1974, S. 50)
[23] (Vgl. Blauert, 1974, S. 55)

Mechanismus mehrere sogenannte „Richtungsbestimmende Bänder" entdeckt (siehe Abbildung 3). Diese Bänder stellen Frequenzbereiche dar, in denen eine Pegelerhöhung einen bestimmten Richtungseindruck bewirkt; so haben Pegelmaxima in den Bereichen um 1kHz und 11kHz zur Folge, dass eine Schallquelle hinten wahrgenommen wird auch wenn sie sich nicht hinter dem Hörer befindet.[24] [25] Da die Richtungswahrnehmung in der Medianebene Pegel in bestimmten Frequenzbereichen vergleicht funktioniert sie mit breitbandigen Signalen deutlich genauer.[26] Auch die Dauer der Signale spielt eine Rolle, so kommt es bei sehr kurzen Signalen vermehrt zu Richtungsinversionen.[27] Weiter hat sich in Untersuchungen ein Lerneffekt gezeigt durch den Testpersonen in der Lage waren bekannte Signale genauer zu Lokalisieren.[28]

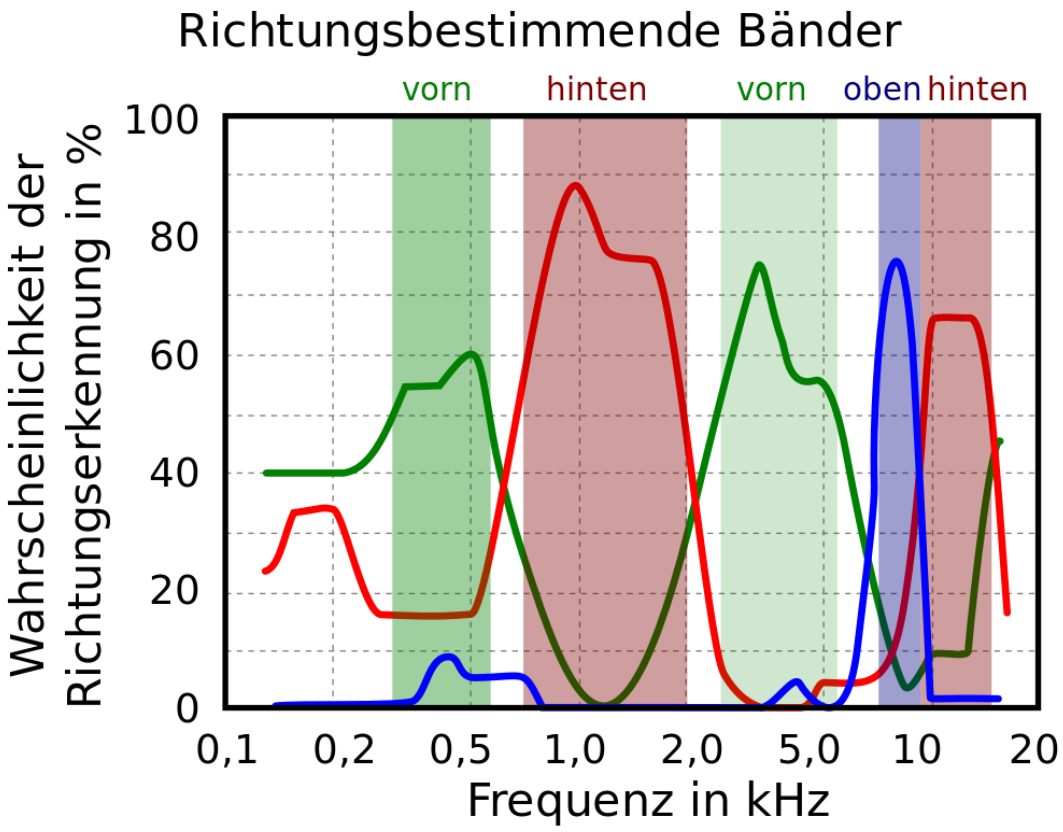

Abbildung 3: Die Richtungsbestimmenden Bänder nach Jens Blauert

[24] (Vgl. Blauert, 1974, S. 88ff.)
[25] (Vgl. Ulrich und Hoffmann, 2007, S. 590)
[26] (Vgl. Blauert, 1974, S. 83)
[27] (Vgl. Blauert, 1974, S. 84)
[28] (Vgl. Blauert, 1974, S. 85)

3 Binauraler Ton

3.1 Grundlagen

Eine verbreitete Methode zur Produktion von 3D-Audio ist die kopfbezogene Stereophonie, auch Kunstkopf-Stereophonie genannt. Wie in Kapitel 2.2 beschrieben wirken die Ohrmuscheln als richtungsabhängige Filter, die zusammen mit ITDs sowie ILDs für die Richtungsbestimmung von einfallenden Schallereignissen benötigt werden. Tonaufnahmen können so gemacht werden, dass bei Kopfhörerwiedergabe jedes Ohr das Signal bekommt, das es am Aufnahmepunkt bekommen hätte und somit auch der Höreindruck entsteht, der am Aufnahmepunkt entstanden wäre. Wenn die Filterfunktionen der Ohrmuscheln bekannt sind, können Monosignale mit entsprechenden Funktionen gefaltet werden. Bei der Wiedergabe werden diese Signale dann aus der entsprechenden Richtung wahrgenommen. Dafür ist notwendig, dass jedes Ohr ein bestimmtes Signal bekommt.[29] [30]

3.2 HRTF / HRIR

Die kopfbezogene Impulsantwort (head related impulse response / HRIR) beschreibt, wie ein Schallereignis in einer bestimmten Position durch Ohrmuschel und Gehörgang verändert wird.[31]

HRIRs werden in einem sehr aufwändigen Verfahren in einem reflexionsarmen Raum mit einem Lautsprecher als Punktschallquelle und Mikrofonen in den Gehörgängen einer Person oder eines Kunstkopfes gemessen. Dadurch entsteht eine Datenbank aus Paaren von HRIRs für diskrete Positionen der Schallquelle.[32]

Falls keine HRIRs für eine bestimmte Position einer Schallquelle vorhanden sind, werden diese durch interpolieren der nächstgelegenen vorhandenen HRIRs gewonnen. Auch für unterschiedliche Distanzen können HRIRs gemessen und

[29] (Vgl. Møller, 1992, S. 171f.)
[30] (Vgl. Hammershøi und Møller, 2005, S. 224)
[31] (Vgl. Mendonça et al., 2010, S. 2)
[32] (Vgl. Villegas, 2015, S. 203)

berechnet werden. Für die Berechnung von HRIRs mit unterschiedlichen Distanzen, wenn keine Daten zum interpolieren vorhanden sind, ist es möglich bestehende HRIRs in Kugelflächenfunktionen zu zerlegen oder eine Distanz-Varianzfunktion zu berechnen.[33] Durch die Messungen der HRIRs mit einer Punktschallquelle können mit diesen auch nur Punktschallquellen synthetisiert werden.[34]

Aus den beiden HRIRs für eine Position lassen sich auch ITDs und ILDs bestimmen. Dazu müssen zum einen die Zeitpunkte, zu denen der Impuls an den Ohren eintrifft und zum anderen Pegeldifferenzen im Frequenzbereich oberhalb von etwa zwei kHz verglichen werden. Wenn HRIRs aus dem Zeitbereich mit Hilfe von Fourier-Transformationen in den Frequenzbereich transformiert werden, entstehen sogenannte kopfbezogene Transferfunktionen (head related transfer function / HRTF). Eine beispielhaftes HRTF-Paar, das für eine frontale Schallquelle mit 15° Erhebungswinkel gemessen wurde, ist in Abbildung 4 dargestellt. In der Abbildung stellt die blaue Linie die Transferfunktion des rechten Ohres dar und die grüne Linie die des Linken. Die genaue Form von Kopf, Ohrmuscheln und Gehörgang bestimmt die HRIR und ist von Mensch zu Mensch unterschiedlich, damit sind auch HRIR und HRTF individuell.[35] Die Verwendung von unpassenden HRTFs führt zu falscher Lokalisation und kann Klangverfärbungen zur Folge haben.[36]

Da HRTFs individuell und sehr aufwändig zu messen sind, werden häufig gemittelte HRTFs verwendet. Es gibt verschiedene Möglichkeiten diese zu erstellen. Eine Möglichkeit ist für mehrere HRIRs vor der Fourier-Transformation den Durchschnitt zu berechnen. Dabei werden die Spitzen und Kerben der einzelnen HRIRs, die wichtig für die Richtungsbestimmung in der Medianebene sind, stark abgeschwächt.[37] Für eine weitere Methode wird eine durchschnittliche Ohrmuschel einer Gruppe von Menschen berechnet und aus dieser die

[33] (Vgl. Villegas, 2015, S. 203)
[34] (Vgl. Travis, 1996, S. 2)
[35] (Vgl. Mendonça et al., 2010, S. 2)
[36] (Vgl. Kaneko et al., 2016, S. 1)
[37] (Vgl. Kaneko et al., 2016, S. 2)

durchschnittliche HRTF. Diese, auf der durchschnittlichen Ohrform basierende HRTF, hat in subjektiven Tests von Kaneko et al. nur bei von unten einfallendem Schall merkbare Unterschiede im Vergleich zu Individuellen HRTFs produziert.[38]

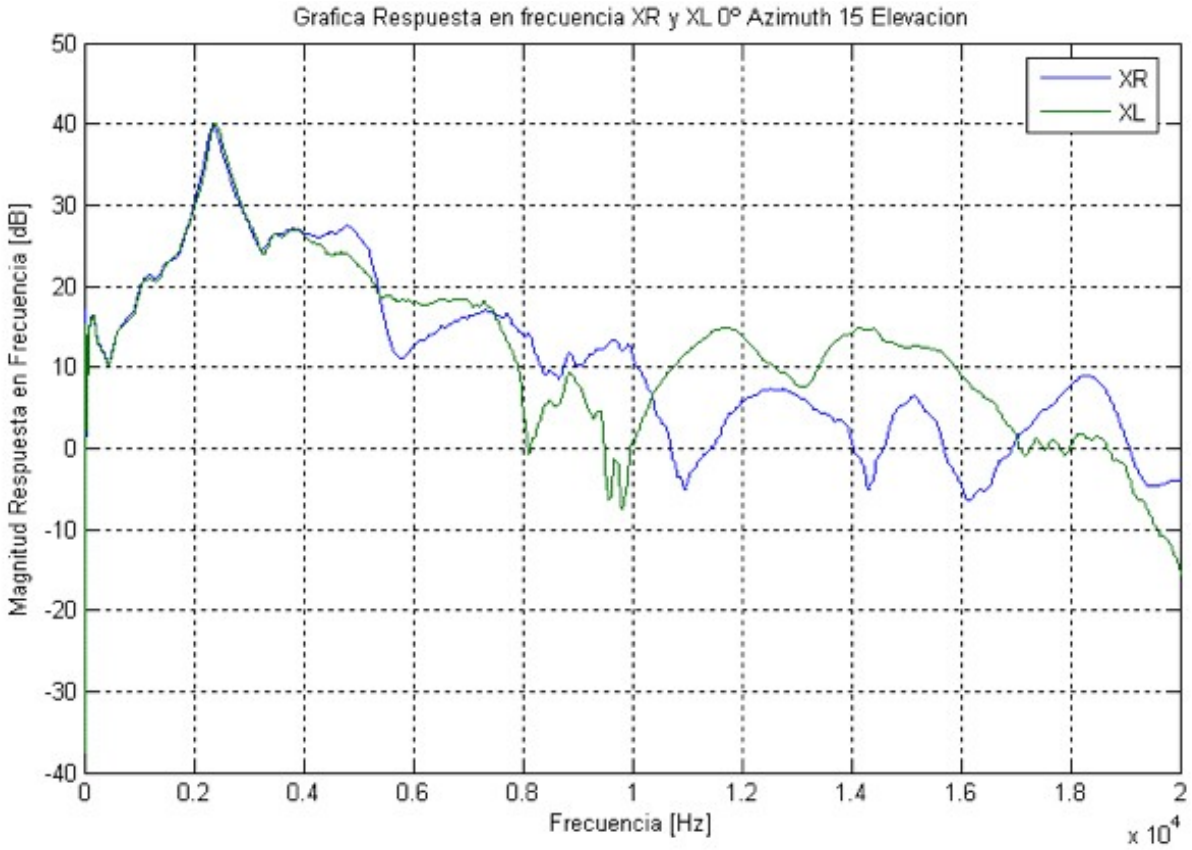

Abbildung 4: Beispiel eines HRTF-Paares

Wenn solche nicht individuellen HRTFs genutzt werden, treten häufig Probleme auf, die als „hole in the middle" und „front-back reversals" bekannt sind.[39] Letzteres beschreibt die Wahrnehmung von Schallquellen aus einer Richtung, die der eigentlichen Ursprungsrichtung gespiegelt an der Frontalebene entspricht. Diese Lokalisationsfehler treten vor allem im Bereich zwischen 30° und 150° Azimut auf, da die spektralen Unterschiede in diesem Bereich bei nicht individuellen HRTFs kleiner sind, als die Unterschiede zwischen mehreren individuellen HRTFs aus der gleichen Richtung.[40] Beim „hole in the middle" Problem sind Aufnahmen im Bereich um 0° Azimut zu leise und wenn sie über Lautsprecher

[38] (Vgl. Kaneko et al., 2016, S. 6)
[39] (Vgl. Lee et al., 2003, S. 1)
[40] (Vgl. Lee et al., 2003, S. 2f.)

wiedergeben werden, zerfällt das Panorama sehr schnell, falls der Hörer sich aus dem sweet spot entfernt.[41]

Um diese Probleme zu lösen kann eine so genannte „notch frequency exaggeration" verwendet werden um die spektralen Unterschiede zwischen Schallquellen aus unterschiedlichen Richtungen zu verstärken. Dies führt zu Beeinträchtigungen der Klangqualität und Schwierigkeiten beim Filtern von Übersprechen.[42] Lee et al. haben eine Methode entwickelt, die das gleiche Ziel verfolgt aber die neu entstehenden Probleme umgehen soll. Durch richtungsabhängige Gewichtung haben sie die einfallende Schallenergie aus frontaler Richtung, also zwischen -45° und 45° Azimut, verstärkt und damit das Auftreten von front-back reversals sowie des „hole in the middle"-Phänomens, wie ihre Experimente zeigen, deutlich verringert. Gleichzeitig umgeht diese Methode die zuvor genannten Probleme der „notch frequency exaggeration".[43] [44]

Abseits der aktiven Korrektur von nicht individualisierten HRTFs können Menschen durch Gewöhnung und Training ihre Lokalisationsgenauigkeit steigern. Mendonça et al. haben in ihren Versuchen gezeigt, dass sich neben der längerfristigen Gewöhnung durch aktives Training auch in kurzer Zeit deutliche Steigerungen bei der Lokalisation von simulierten Schallquellen beobachten lassen. Dabei hat sich nicht nur die Ortung von Stimuli an explizit trainierten Orten verbessert, sondern die Ortung allgemein. Die Verbesserung der Lokalisationsgenauigkeit ließ sich sowohl für horizontale als auch vertikale Einfallswinkel beobachten.[45]

41 (Vgl. Griesinger, 1989, S. 22)
42 (Vgl. Lee et al., 2003, S. 2)
43 (Vgl. Lee et al., 2003, S.4f.)
44 (Vgl. Lee et al., 2003, S. 7f.)
45 (Vgl. Mendonça et al., 2010, S. 5ff.)

3.3 Anwendungen

3.3.1 Aufnahme

Bei binauralen Aufnahmen soll das Schallfeld korrekt in zwei Kanäle transformiert werden, einen für jedes Ohr. Dabei ist vor allem die räumliche Information wichtig, die mit Hilfe der Aufnahmen festgehalten werden soll. Diese räumliche Information des eintreffenden Schalls ist auch schon am Eingang des Gehörgangs vorhanden, so können Aufnahmen mit Mikrofonen in den Eingängen der Gehörgänge einer Person gemacht werden. Dies ist deutlich einfacher umzusetzen als eine Mikrofonierung direkt vor dem Trommelfell.[46] Bei den genannten räumlichen Informationen handelt es sich um ILDs, ITDs und die spektralen Einflüsse der Ohrmuscheln, die korrekt aufgenommen werden müssen, um eine funktionierende binaurale Wiedergabe zu ermöglichen.[47] Solche Aufnahmen können außer mit Personen auch mit Kunstköpfen oder so genannten Kopf-Torso Simulatoren gemacht werden[48], die einen durchschnittlichen Kopf mit durchschnittlichen Ohrmuscheln nachbilden. Abbildung 5 zeigt ein Beispiel für einen solchen Kunstkopf. In Experimenten hat sich gezeigt, dass Aufnahmen in den Gehörgängen von Menschen, für Lokalisation[49] und Sprachverständlichkeit[50], bessere Ergebnisse liefern als Aufnahmen mit einem Kunstkopf.

Abbildung 5: Der Kunstkopf KU-100 der Firma Neumann

[46] (Vgl. Hammershøi und Møller, 2005, S. 225)
[47] (Vgl. Nykänen und Johnsson, 2011, S. 2)
[48] (Vgl. Nykänen und Johnsson, 2011, S. 1)
[49] (Vgl. Nykänen und Johnsson, 2011, S. 2)
[50] (Vgl. Nykänen und Johnsson, 2011, S. 7)

3.3.2 Synthese

Die Binauralsynthese verwendet Monosignale und faltet diese je einmal mit den HRTFs für eine bestimmte Richtung und Distanz. So werden virtuelle Schallquellen geschaffen, die von der entsprechenden Position zu kommen scheinen.[51]
Für eine andere Herangehensweise werden HRTFs in der optimalen Abhörposition eines Mehrkanalsystems, wie 7.0 Surround, gemessen und bei der Synthese werden die Signale der Lautsprecher mit den entsprechenden HRTFs gefaltet um virtuelle Lautsprecher binaural wiederzugeben.[52]
Der große Vorteil der binauralen Synthese gegenüber der Aufnahme ist die Möglichkeit HRTFs in Echtzeit zu aktualisieren, um sich bewegende Schallquellen zu simulieren oder eine gesamte Szene bei interaktiven Anwendungen auf Nutzereingaben reagieren zu lassen.

3.3.3 Wiedergabe

Die Wiedergabe von binauralen Aufnahmen oder per Binauralsynthese erstellten Signalen kann sowohl mit Kopfhörern als auch mit Lautsprechern erfolgen. Dabei darf das Signal für das linke Ohr nur am linken Ohr anliegen und das Signal für das rechte Ohr nur am Rechten.[53] Im Fall der Wiedergabe mit Kopfhörern ist dies automatisch gegeben. Weiter ist die räumliche Wiedergabe an den Kopf des Hörers gebunden, das heißt bei Kopfbewegungen bewegt sich das Gehörte mit. Falls es sich bei dem wiedergegebenen Material um eine binaurale Aufnahme handelt, kann diese Tatsache nicht umgangen werden.[54] Bei Material, das mit Hilfe von Binauralsynthese wiedergabeseitig gerendert wird, kann diese Bindung aufgehoben werden. Dazu muss die Ausrichtung des Kopfes kontinuierlich verfolgt und die HRTFs für die Faltung in Echtzeit aktualisiert werden.[55]
Bei der Wiedergabe über Lautsprecher kann der Hörer von zwei Lautsprechern mit ihn komplett einhüllenden Ton beschallt werden. Ohne besondere Maßnahmen

[51] (Vgl. Lentz und Behler, 2004, S. 7)
[52] (Vgl. Gorzel et al., 2012, S. 3)
[53] (Vgl. Lentz und Behler, 2004, S. 3)
[54] (Vgl. Travis, 1996, S. 4)
[55] (Vgl. Villegas, 2015, S. 210)

liegen jedoch beide Lautsprechersignale an beiden Ohren an. Dieses Übersprechen muss herausgefiltert werden, damit die binaurale Wiedergabe funktioniert. Wenn statische Filter verwendet werden, muss sich der Kopf eines Hörers in einem bestimmten Bereich befinden.[56] Lentz und Behler haben gezeigt, dass es möglich ist durch Positionsbestimmung und dynamische Filter Bewegung im Raum zu ermöglichen. Dies kann nur von einem Hörer gleichzeitig genutzt werden, wodurch sich dieses System gut für CAVE-Systeme[57] eignet, da auch diese nur von einer Person genutzt werden können und die Position des Nutzers verfolgen müssen.[58] Die dynamische Filterung von Übersprechen funktioniert nur stabil für Kopfdrehungen im Bereich zwischen den Lautsprechern. Lentz und Behler haben weiter gezeigt, dass dies mit Lautsprecheranordnungen mit einem Winkel zwischen den Lautsprechern von 90° sowie 180° funktioniert. Damit können beide Anordnungen kombiniert werden um komplette Kopfdrehungen mit vier Lautsprechern zu ermöglichen.[59]

[56] (Vgl. Choueiri, 2010, S. 1f.)
[57] (Vgl. https://www.evl.uic.edu/cave)
[58] (Vgl. Lentz und Behler, 2004, S. 2)
[59] (Vgl. Lentz und Behler, 2004, S. 4)

4 Ambisonics

4.1 Basics oder so

4.1.1 Grundlagen und Kugelflächenfunktionen

Der klassische Ansatz, der heute als „First Order Ambisonics" (FOA) bekannt ist, wurde 1973 von Michael Gerzon erstmals angedeutet.[60] Dieser Ansatz sieht vor, dass mit vier Kanälen das komplette Schallfeld aufgenommen wird. Ein ungerichtetes Mikrofon zeichnet den gesamten Schalldruck auf und drei Mikrofone mit Achtercharakteristik, die entlang der Achsen eines räumlichen Koordinatensystems ausgerichtet sind, erfassen die Schallschnellekomponenten. Das Signal, das sich aus diesen vier Kanälen zusammensetzt, wird „B-Format" genannt. Aus diesem Signal kann ein Dekodierer beliebige virtuelle Mikrofone berechnen, die im einfachsten Fall der Wiedergabe auf die entsprechenden Lautsprecher ausgerichtet sind. Die Überlagerung der Lautsprechersignale bildet dann wieder das aufgenommene, oder synthetisierte Schallfeld.[61] Mit den zuvor beschriebenen vier Kanälen kann FOA ein dreidimensionales Schallfeld kodieren, dabei werden mindestens sechs Lautsprecher in oktaedrischer Anordnung benötigt, um sicherzustellen, dass Phantomschallquellen bei der Wiedergabe korrekt dargestellt werden. Allgemein werden für eine möglichst korrekte Wiedergabe von Ambisonics mehr Wiedergabekanäle als Aufnahmekanäle benötigt.[62]

Wird ein Schallfeld an einem Punkt aufgenommen, kann es in Kugelflächenfunktionen zerlegt werden, ähnlich der Zerlegung eines Periodischen Signals durch die Fourier-Transformation.[63] Damit wird mit einer Darstellung über Zeit und Richtung beschrieben wie sich der Schalldruck im Schallfeld verändert.[64] Durch diese Darstellung des Schallfeldes in einem Kugelkoordinatensystem können Transformationsmatrizen gebildet werden, mit denen das Schallfeld um die Achsen

60 (Vgl. Gerzon, 1973)
61 (Vgl. Blauert und Rabenstein, 2012, S. 9f.)
62 (Vgl. Gerzon, 1985, S. 862)
63 (Vgl. Villegas, 2015, S. 202)
64 (Vgl. Shivappa et al., 2016, S. 4)

eines räumlichen Koordinatensystems rotiert werden kann.[65] Ambisonics nutzt diese Zerlegung des Schallfeldes und speichert oder überträgt das zerlegte Schallfeld in Form von Koeffizienten der beteiligten Kugelflächenfunktionen aus denen es auf der Wiedergabeseite rekonstruiert werden kann. FOA nutzt Mikrofone mit Kugel- sowie Achtercharakteristik, die auch durch Kugelflächenfunktionen nullter und erster Ordnung beschrieben werden können. Allgemein beschreiben Kugelflächenfunktionen, unter anderem, die Schallabstrahlung von Kugelschallquellen. Abbildung 6 zeigt eine grafische Darstellungen der Kugelflächenfunktionen nullter bis dritter Ordnung. Durch lineare Überlagerung solcher Funktionen gleicher Ordnung können sie rotiert werden. Mit Hilfe dieser Berechnungen werden auch die räumlichen Charakteristiken, der virtuellen Mikrofone, in die Richtungen der Lautsprecher für die Wiedergabe bewegt.[66]

Eine Weiterentwicklung der klassischen FOA, die Gerzon schon vorgeschlagen hat, bezieht weitere Ordnungen von Kugelflächenfunktionen mit ein. Sie wird meist als „Higher Order Ambisonics" (HOA)[67] bezeichnet und wird im nächsten Abschnitt genauer betrachtet.

65 (Vgl. Gorzel et al., 2012, S. 3)
66 (Vgl. Blauert und Rabenstein, 2012, S. 10)
67 (Vgl. Blauert und Rabenstein, 2012, S. 10)

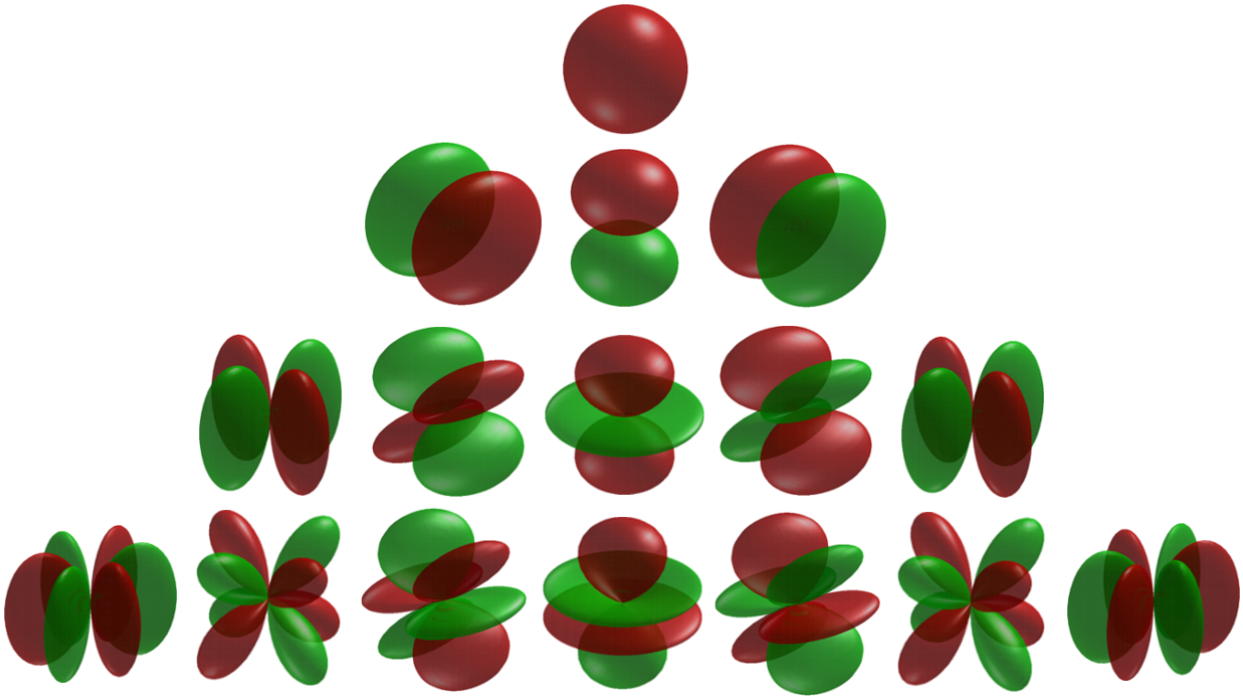

Abbildung 6: Grafische Darstellung der Kugelflächenfunktionen nullter bis dritter Ordnung

4.1.2 FOA und HOA

Neben den Kugelflächenfunktionen der nullten und ersten Ordnung können auch weitere Ordnungen hinzugenommen werden, dadurch erhöht sich die räumliche Auflösung. Für eine Aufnahme werden in diesem Fall mehr Mikrofone mit stärker gerichteter Charakteristik notwendig, um die hinzukommenden Koeffizienten der Kugelflächenfunktionen korrekt kodieren zu können. Es werden auch mehr Lautsprecher zur Wiedergabe benötigt. Da für Anwendungen im Bereich Virtual Reality vor allem dreidimensionale Wiedergabe interessant ist, wird hier nur von sphärischer Wiedergabe ausgegangen, für die gilt: Wenn M die höchste beteiligte Ordnung von Kugelflächenfunktionen ist, dann gilt für die minimal benötigte Anzahl von Lautsprechern $N=(M+1)^2$.[68]

Der Bereich in dem die Wiedergabe korrekt ist, genannt Sweetspot, wird mit steigender Ordnung größer und umfasst bei sehr hohen Ordnungen den gesamten Wiedergabebereich.[69]

[68] (Vgl. Blauert und Rabenstein, 2012, S. 10)
[69] (Vgl. Blauert und Rabenstein, 2012, S. 11)

FOA ist sehr effizient, allerdings ist die räumliche Auflösung nicht sehr groß. Dies lässt sich bei Aufnahmen vor allem in Räumen mit starkem Nachhall und wenn Schallquellen weiter entfernt sind beobachten.[70] Auch die Größe des Sweetspot ist bei FOA noch sehr klein, somit eignet sich FOA vor allem für die Aufnahme von Atmosphären.

4.2 Produktion

4.2.1 Aufnahme

HOA kann mit kleinen Mikrofonanordnungen oder Schallfeldmikrofonen aufgenommen werden. Dabei können auch bestehendes Material oder einzelne Mikrofone hinzugemischt werden. Weiter ist es möglich in der Postproduktion ein Schallfeld komplett synthetisch zu erstellen.[71]
Mehrkanalverfahren, wie HOA, erlauben auch die Produktion mit mehrkanaligem Quellmaterial und bessere Kontrolle der Größe von Schallquellen sowie deren Distanz.[72]
Derzeit gibt es Mikrofone auf dem Markt die Signale zur Codierung von FOA ohne vertikale Signale bis hin zu dreidimensionaler HOA siebter Ordnung liefern. Beispiele für solche Mikrofone sind in Abbildung 7 zu sehen. Eine andere Methode, für die Konvertierung von Mikrofonsignalen zu HOA, bedient sich gewichteter Addition einzelner Signale sowie Faltungen und hat den Vorteil, dass das HOA-Signal keine Informationen über Position und Eigenschaften der Mikrofone enthalten muss um die Ursprungsrichtung von Klängen zu kodieren. Es muss lediglich festgelegt werden in welcher Richtung die Quellen der Signale liegen sollen. Umgekehrt können auch Aufnahmen von mehreren einzelnen Mikrofonen, basierend auf ihren Positionen, zu einem HOA-Signal gemischt werden.[73]

[70] (Vgl. Bates und Boland, 2016, S. 5)
[71] (Vgl. Shivappa et al., 2016, S. 5)
[72] (Vgl. Travis, 1996, S. 5)
[73] (Vgl. Shivappa et al., 2016, S. 6)

Abbildung 7: Links: Sennheiser Ambeo® VR-Mic mit vier Kapseln für FOA. | Mitte: MH Acoustics Eigenmike® mit 32 Kapseln für HOA bis vierter Ordnung. | Rechts: Visisonics Audio/Visual Camera mit 64 Kapseln für HOA bis zu siebter Ordnung

4.2.2 Bearbeitung und Transfer

Zur Produktion und Bearbeitung von HOA gibt es Plug-ins, die eingangsseitig Signale von FOA- und HOA-Mikrofonen sowie mehreren einzelnen Mikrofonen zu Ambisonics-Signalen konvertieren und bestehende Ambisonics-Signale importieren können. Werden einzelne Mikrofonsignale konvertiert, können diese räumlich ausgerichtet und mit einem räumlichen Echoeffekt versehen werden. Zur Bearbeitung des kodierten Ambisonics-Signals kann das Schallfeld rotiert, verzerrt sowie an sämtlichen Achsen gespiegelt werden. Auch räumliche Filterung für eine vom Nutzer festgelegte Richtung ist möglich. Für die Ausgangsseite können verschiedene Renderer ausgewählt, Lautsprecher zugeordnet und das fertige Produkt als Datei gespeichert werden. Außerdem können Ambisonics-Signale direkt als Stream ausgegeben werden, dabei können durch mehrere Instanzen von

ausgangs Plug-ins und eine eigene Busstruktur auch mehr Lautsprecher angesteuert werden als die eigentliche Software Busse zur Verfügung stellt.[74]
HOA-Signale brauchen, je nach Ordnung, sehr viel Bandbreite für die unkomprimierte Übertragung. Beispielsweise würde die Übertragung eines HOA-Signals vierter Ordnung mit einer Abtastrate von 48 kHz und einer Samplingtiefe von 32bit etwa 40 Megabit pro Sekunde (mbit/s) benötigen. Allerdings ist es mit Hilfe von Kompression auf Basis des MPEG-H Standards möglich, die benötigte Bandbreite auf circa 400 Kilobit pro Sekunde (kbit/s) zu verringern; dies entspricht etwa der Bandbreite zur Übertragung eines 5.1 Surround Signals.[75]
HOA-Signale können auf statische Datenraten komprimiert werden, dabei ist die Datenrate abhängig von der Ordnung und nicht von der Komplexität des Schallfeldes.[76]
Neben der Kompression können HOA-Signale auch in mehreren Schichten übertragen werden, beispielsweise kann eine erste Schicht ein FOA-Signal und Informationen über weitere verfügbare Schichten enthalten und die weiteren Schichten transportieren die Signale für weitere Ordnungen. Damit kann ein Dekodierer die wiedergegebene Ordnung an die vorhandene Rechenleistung und Bandbreite anpassen.[77]

4.2.3 Wiedergabe

Das wichtigste Merkmal der Wiedergabe von Ambisonics Signalen ist, dass sie wiedergabeseitig gerendert werden. Dies bringt sehr hohe Flexibilität mit sich, da theoretisch Signale für beliebige Lautsprecheranordnungen gerechnet werden können, benötigt aber auch Leistungsstarke Hardware.[78] Diese Dekodierer teilen die Signale zunächst in die einzelnen Kanäle auf und filtern diese um frequenzabhängigen Eigenschaften des menschlichen Gehörs gerecht zu werden und einen ebenen Frequenzgang sicherzustellen. Weiter werden die Signale für die

[74] (Vgl. Shivappa et al., 2016, S. 6f)
[75] (Vgl. Shivappa et al., 2016, S. 7)
[76] (Vgl. Shivappa et al., 2016, S. 4)
[77] (Vgl. Shivappa et al., 2016, S. 7f.)
[78] (Vgl. Shivappa et al., 2016, S. 4)

nicht unendlich große Lautsprecheranordnung korrigiert und zuletzt durch eine Amplitudenmatrix für die exakte Anzahl und Positionen der Lautsprecher bearbeitet.[79]

Ambisonics kann nicht nur über Lautsprecher wiedergegeben werden, sondern mit Hilfe von so genannten virtuellen Lautsprechern auch effizient für die Kopfhörerwiedergabe binauralisiert werden. Dazu werden die Lautsprechersignale für eine Lautsprecheranlage berechnet und anschließend mit den HRTFs für die entsprechenden Lautsprecherpositionen gefaltet.[80] Dabei ist die benötigte Anzahl an Faltungen abhängig von der Ordnung und nicht von der Anzahl an virtuellen Lautsprechern. So kann HOA bei der Binauralisierung mit einer großen Anzahl an virtuellen Lautsprechern und Anordnungen, die mit realen Lautsprechern nicht praktikabel realisierbar sind, arbeiten.[81] Wie zuvor erwähnt erlaubt HOA einfache Rotationen des Schallfeldes. Dadurch kann bei der Binauralisierung mit statischen HRTF-Filtern gearbeitet werden, was die benötigte Rechenleistung senkt. Außerdem kann durch die hierarchische Struktur die benötigte Rechenleistung für den Dekodierer skaliert werden. So kann ein schwächerer Dekodierer auch ein Signal verarbeiten, in dem höhere Ordnungen kodiert sind und nur die Ordnungen dekodieren für die genug Rechenleistung zur Verfügung steht.[82]

79 (Vgl. Gerzon, 1985, S. 866)
80 (Vgl. Shivappa et al., 2016, S. 8)
81 (Vgl. Shivappa et al., 2016, S. 1)
82 (Vgl. Travis, 1996, S. 5f.)

5 Objektbasierter Ton

5.1 Prinzip

Die objektbasierte Tonproduktion stellt Schallereignisse in Form von Objekten dar. Diese Objekte bestehen aus einem Tonsignal und dazugehörigen Metadaten mit denen die Eigenschaften des Objektes festgelegt werden.[83] Die Metadaten enthalten, unter anderem, Informationen über Position, Lautstärke, Verzögerung und das Abstrahlverhalten der Schallquelle. Außerdem können die Metadaten dynamisch verändert werden, um beispielsweise sich bewegende Schallquellen darzustellen.[84] Bei der objektbasierten Tonproduktion sollen alle wichtigen Informationen einer Tonszenerie aufgenommen oder produziert und in Einzelteilen gespeichert oder übertragen werden, damit die Szene auf der Wiedergabeseite wieder vollständig zusammengesetzt werden kann. Mit Hilfe der Metadaten kann dem Nutzer auch die Möglichkeit gegeben werden, Änderungen an einzelnen Objekten vorzunehmen, wie die Lautstärke von Dialogen in Filmen anzupassen oder die gesamte Szene zu rotieren.[85] Letzteres ist eine wichtige Funktionalität für VR-Anwendungen, die Head-Mounted-Displays (HMD) verwenden oder mit Hilfe anderer Sensorik auf Kopfbewegungen des Nutzers reagieren, da in diesen Fällen, vorausgesetzt die Wiedergabe erfolgt über Kopfhörer, die Szene den Bewegungen entsprechend angepasst werden muss. Bei der Aufnahme mit Audioobjekten werden zwei Arten von Objekten unterschieden, zum einen explizite Objekte, die einzeln mikrofoniert werden können und deren Position entweder verfolgt wird oder statisch ist. Die andere Art sind implizite Objekte, die aus den Signalen mehrerer Mikrofone extrahiert werden müssen, dies kann beispielsweise mit Hilfe von Mustererkennung sowie Erkennung von Zeitunterschieden zwischen den Signalen erreicht werden.[86]

[83] (Vgl. Füg et al., 2014, S. 2)
[84] (Vgl. Gasull Ruiz et al., 2015, S. 1)
[85] (Vgl. Oldfield et al., 2014, S. 2)
[86] (Vgl. Oldfield et al., 2014, S. 5)

5.2 Produktion

5.2.1 Aufnahme und Bearbeitung

Zur Produktion von objektbasiertem Ton werden alle Schallquellen einzeln aufgenommen oder synthetisch hergestellt und mit ihren Metadaten versehen.[87] Allerdings gibt es auch andere Möglichkeiten objektbasierten Ton aufzunehmen. Eine davon besteht darin, eine Ambisonics Aufnahme in Objekte zu konvertieren, so kann zum Beispiel eine FOA Aufnahme in 22 Objekte konvertiert werden, die den Positionen der Lautsprecher eines 22.2-Systems entsprechen. Ein weiterer Weg zur Aufnahme einer kompletten Szene nutzt einzelne Mikrofone, die jeweils ein Objekt aufzeichnen. Bei dieser Methode kann der Raumhall entweder mit weiteren Mikrofonen aufgenommen werden oder es kann anhand der aufgenommenen Objekte künstlicher Hall, der nicht der Situation im Aufnahmeraum entsprechen muss, berechnet werden. Da alle Schallquellen direkt mikrofoniert werden müssen, um übersprechen zu verhindern, stellen sich bewegende Schallquellen eine besondere Schwierigkeit dar, da sich für eine korrekte Wiedergabe sowohl die Mikrofone, als auch die entsprechenden Objekte mit der Schallquelle mitbewegen müssen. Auch mit einer Mikrofonanordnung, bei der jedes Mikrofon einem Objekt entspricht, können Aufnahmen gemacht werden. Solche aufnahmen, wie auch konvertierte Ambisonics Signale, lassen sich mit der größten Genauigkeit wiedergeben, wenn die Positionen der Objekte möglichst genau mit den Positionen der Lautsprecher bei der Wiedergabe übereinstimmen. Denn Schallquellen, die von mehreren Objekten wiedergeben werden, bilden bei der Wiedergabe Phantomschallquellen[88], deren Ortung ungenauer ist und die sich bei Bewegungen aus dem Sweetspot verschieben.[89]

Bei der Bearbeitung von objektbasiertem Ton können die Eigenschaften sämtlicher Objekte, sowohl bei Liveübertragungen, als auch während der Produktion für eine spätere Wiedergabe, bestimmt oder angepasst werden. Dies wird meist durch Plug-ins mit grafischen Oberflächen bewerkstelligt, die neben der Bearbeitung

87 (Vgl. Altman et al., 2016, S. 3)
88 (Vgl. Messonnier et al., 2016, S. 5f.)
89 (Vgl. Messonnier et al., 2016, S. 3f.)

auch eine Automatisierung ermöglichen.[90] Neben den Eigenschaften können auch die Audiosignale der Objekte, wie bei der kanalbasierten Tonproduktion, bearbeitet werden, beispielsweise mit Filtern oder anderen Effekten. Das, vom Fraunhofer Institut entwickelte, Programm „SpatialSound Control", zum Produzieren und Bearbeiten von objektbasiertem Ton, erlaubt auch die gleichzeitige Bearbeitung durch verschiedene Benutzer an verschiedenen Geräten. So können zwei Nutzer zum Beispiel gleichzeitig die Audiosignale und die Eigenschaften von Objekten bearbeiten.[91]

5.2.2 Übertragung und Wiedergabe

Für die Übertragung von Objektbasiertem Ton gibt es bereits verschiedene Standards. Einer davon ist „Audio Definition Model", der weitere Metadaten zu einem „Broadcast Wave File" hinzufügt. Damit können Audiodateien als kanalbasiert, objektbasiert oder physikbasiert, wie Ambisonics, gekennzeichnet und die Metadaten für objektbasierte Dateien transportiert werden.[92] Eine weitere Möglichkeit bietet der 2014 veröffentlichte „MPEG-H 3D Audio" Standard.[93] Auch dieser Standard bietet die Übertragung von sowohl objektbasiertem Ton, als auch kanal- sowie physikbasiertem Ton an und ermöglicht die Kompression der zu transportierenden Daten.[94] Eine Schwierigkeit bei der Übertragung von objektbasiertem Ton ist die Tatsache, dass die benötigte Datenrate abhängig von der Anzahl an gleichzeitig aktiven Objekten ist, da alle Objekte sowie deren Metadaten nebeneinander übertragen werden müssen.[95] Dadurch bestimmt die Komplexität der Szene direkt die Datenmenge für Übertagung und Speicherung, wobei zum Beispiel bei Filmen große Menden an Objekten verwendet werden um die Produktion möglichst realistisch klingen zu lassen.

90 (Vgl. Gasull Ruiz et al., 2015, S. 3)
91 (Vgl. Gasull Ruiz et al., 2015, S. 4)
92 (Vgl. Messonnier et al., 2016, S. 1)
93 (Vgl. Füg et al., 2014. S. 3)
94 (Vgl. Oldfield et al., 2014 S. 7)
95 (Vgl. Messonnier et al., 2016, S. 5)

Auf der Wiedergabeseite wird für objektbasierten Ton ein Renderer benötigt, der die Positionen der Lautsprecheranlage kennt und dann anhand der Audio- sowie Metadaten die Lautsprechersignale berechnet.[96] Somit können solche Signale auf beliebigen Lautsprecheranlagen wiedergeben werden.[97]

Auch objektbasierter Ton kann für die Wiedergabe auf Kopfhörern binauralisiert werden. Dazu kann entweder jedes Objekt einzeln, anhand seiner Position, mit den entsprechenden HRTFs gefaltet werden, oder es werden die Signale für eine Lautsprecheranlage berechnet und diese anschließend mit den HRTFs für die Positionen der Lautsprecher gefaltet. Somit ist der Rechenaufwand für die Binauralisierung entweder von der Anzahl an Objekten oder an Lautsprechersignalen abhängig.[98]

[96] (Vgl. Gasull Ruiz et al., 2015, S. 3)
[97] (Vgl. Oldfield et al., 2014, S. 2)
[98] (Vgl. Shivappa et al., 2016, S. 4)

6 Virtual Reality

6.1 Grundlagen

6.1.1 VR-Systeme

Was ist VR überhaupt? Mit den Worten von Manfred Brill ausgedrückt, steht virtuelle Realität „[…] für eine neuartige Benutzeroberfläche, in der die Benutzer innerhalb einer simulierten Realität handeln und die Anwendung steuern und sich im Idealfall so wie in ihrer gewohnten realen Umgebung verhalten."[99] Aus dieser Ansicht leitet sich die Hoffnung ab, dass VR sich zu einer idealen Form der Mensch-Maschine-Kommunikation entwickeln wird, die jedem Nutzer eine einwandfreie Bedienung ohne Vorkenntnisse erlaubt. Während der Interaktion mit der virtuellen Welt soll der Nutzer ein Gefühl der tatsächlichen Anwesenheit in der virtuellen Welt, häufig „Immersion" genannt, haben. Dieses lässt sich deutlich verstärken, wenn die VR neben der sichtbaren Darstellung auch akustische sowie taktile Reize verwendet und die Wahrnehmung des Nutzers möglichst komplett eingehüllt wird.[100] Dabei ist vor allem die Synchronisation zwischen Bild und Ton wichtig. So wird die Qualität eines VR-Systems, das dem Nutzer synchrone Reize bietet und seine Bewegungen korrekt registriert, nicht unter Ungenauigkeiten beim Ton, wie nicht individualisierten HRTFs, leiden.[101] [102] Ein weiterer Faktor, der die Immersion unterstützt, ist ein hoher Grad an Interaktivität, der es dem Nutzer erlaubt viele Elemente der VR zu beeinflussen.[103] Außerdem ist die Steuerung mit Hilfe von Bewegungsverfolgung ein wichtiges Standbein, um Nutzern das Gefühl zu geben, tatsächlich in der virtuellen Welt zu sein.[104] Für die Bewegungsverfolgung stehen verschiedene Systeme zur Verfügung, die beispielsweise mit Hilfe von elektromagnetischer Sensoren und bekannten Magnetfeldern, Triangulation von Ultraschalltönen oder Kameras arbeiten. Dabei

99 (Vgl. Brill, 2009, S. 6)
100 (Vgl. Brill, 2009, S. 6)
101 (Vgl. Brill, 2009, S. 25)
102 (Vgl. Travis, 1996, S. 4)
103 (Vgl. Brill, 2009, S. 6f)
104 (Vgl. Brill, 2009, S. 7)

unterscheiden sich die verschiedenen Systeme deutlich in ihrer Genauigkeit, benötigten Datenraten und den Kosten.[105] So genannte Datenhandschuhe erlauben darüber hinaus eine sehr natürliche Interaktion mit der virtuellen Welt, bei der der Nutzer seine Hände wie in der realen Welt benutzen kann. Auch haptisches sowie „force feedback" sind möglich, die den Träger des Handschuhs spüren lassen, wenn in der VR etwas berührt wird.[106]

6.1.2 Geschichte der virtuelle Realität

Der Begriff „Virtual Reality" ist deutlich jünger als die eigentliche Idee dahinter, die Darstellung einer Welt, die im Idealfall so realistisch ist, dass sie von der Wirklichkeit nicht zu unterscheiden ist. Die erste Beschreibung einer solchen Darstellung stammt aus einer Arbeit von Ivan Sutherland aus dem Jahr 1965.[107] Der Begriff wurde erst in den 1980er Jahren von Jaron Lanier, einem Gründer der Firma Visual Programming Language, geprägt.[108]

Die technologische Entwicklung, die zum heutigen Stand der Technik in diesem Bereich geführt hat, begann schon deutlich früher. Allgemein sind die militärisch-technische Entwicklung sowie die Filmtechnik als Hauptantriebe für die VR-Technologie zu nennen. Die erste wichtige Errungenschaft auf dem Weg zur heutigen VR-Technologie ist das 1832 von Charles Wheatstone erfundene Stereoskop, dass räumliche Darstellungen ermöglichte. Ein weiterer wichtiger Schritt waren mechanische Flugsimulatoren, die erstmals während des ersten Weltkrieges eingesetzt wurden. Erst in den 1970er Jahren wurden sie durch Computergestützte Simulationen ersetzt.[109] Von den Fortschritten in der Informatik profitierte ab etwa Mitte der 1960er Jahre auch die Entwicklung von VR-Technologie, so wurde mit dem „Universal Digital Operational Flight Trainer" einer der ersten computerbasierten Flugsimulatoren geschaffen. Im Jahre 1970 wurde an der University of Utah das erste so genannte „Head-Mounted-Display" getestet,

[105] (Vgl. Brill, 2009, S. 30ff.)
[106] (Vgl. Brill, 2009, S. 34ff.)
[107] (Vgl. Brill, 2009, S. 2)
[108] (Vgl. Brill, 2009, S. 10)
[109] (Vgl. Brill, 2009, S. 8)

wobei die benötigte Rechenleistung für dessen Betrieb, die dem Forschungsteam zur Verfügung stehende Rechenleistung jedoch deutlich übertraf.[110] In den 1980er Jahren hatte vor allem das „NASA Ames Reserch Center" Einfluss auf die Weiterentwicklung der VR-Technologie. So wurde dort erstmals ein HMD mit LCDs gebaut und ein Datenhandschuh als Eingabegerät entwickelt. Auch die höhere Rechenleistung und bessere Verfügbarkeit von Computern beflügelte die Forschung.[111] 1992 wurde an der University of Illinois das „Cave Automatic Virtual Environment" (CAVE) konzipiert. Eine solche CAVE ist in Abbildung 8 zu sehen. Für dieses System werden die Wände eines Raumes mit 3D-Projektionen bespielt. Dies war ein grundsätzlicher Unterschied zu den bis dato vor allem verwendeten HMDs, die meist auch sehr schwer waren.[112] Aber es müssen für die 3D-Projektionen immer noch entsprechende Brillen getragen werden und die Position sowie die Blickrichtung des Nutzers müssen verfolgt werden, damit die Projektionen unabhängig von Bewegungen des Nutzers immer korrekt dargestellt werden. Seit den 2000er Jahren wird die virtuelle Realität in einigen Bereichen, wie zum Beispiel zur Visualisierung in der Automobilindustrie oder für medizinische Trainingssimulationen, schon fast selbstverständlich eingesetzt.[113] In weiteren Anwendungsbereichen sind VR-Anwendungen noch nicht weit verbreitet, bieten aber großes Potential, beispielsweise wird sich VR in der Unterhaltung durch die Markteinführungen diverser, erschwinglicher VR-Headsets weiter verbreiten.[114]

[110] (Vgl. Brill, 2009, S. 9)
[111] (Vgl. Brill, 2009, S. 10)
[112] (Vgl. Brill, 2009, S. 10f)
[113] (Vgl. Brill, 2009, S. 12)
[114] (Vgl. http://time.com/4277763/virtual-reality-buyers-guide/)

Abbildung 8: Außenansicht des CAVE an der Fachhochschule Seinäjoki

6.2 Anwendungen

Die Möglichkeiten für Virtual Reality Anwendungen sind sehr vielfältig. Schon heute werden für verschiedenste Zwecke virtuelle Realitäten eingesetzt und mehr befinden sich in Planung und Entwicklung. Dabei lassen sich diese Anwendungen in verschiedene Kategorien unterteilen. Im Folgenden werden lineare VR-Anwendungen, wie 360° Videoproduktionen oder -übertragungen sowie interaktive VR-Anwendungen, beispielsweise Spiele oder Trainingssimulationen, genauer betrachtet und hinsichtlich ihrer Anforderungen an die Auralisation untersucht. Außerdem werden die Unterschiede zwischen mobilen und stationären VR-Anwendungen betrachtet.

6.2.1 Lineare VR-Anwendungen

Im Bereich der linearen VR-Anwendungen finden sich alle Anwendungen bei denen der Nutzer nur Zuschauer ist und nicht mit der VR interagieren kann. Dabei kann sich der Nutzer umsehen und gegebenenfalls Bildausschnitte vergrößern, das Geschehen selbst ist nicht beeinflussbar. In diesem Bereich finden sich mit sphärischen Kameras produzierte 360° Videos und Filme. Bei der Tonproduktion für solche Inhalte müssen komplexe Szenen aus beim Dreh aufgenommenem, sowie bestehendem Material geschaffen werden. Auch Übertragungen von Events, wie Sportveranstaltungen oder Konzerten, sind möglich. Mit der entsprechenden Übertragungstechnologie können Events, live, einer unbegrenzten Anzahl von Zuschauern so dargestellt werden, als wären sie vor Ort. Daher rührt auch die Bezeichnung für diese Form der Übertragung: „Infinite Seat".[115] Für derartige Übertagungen muss der Ton so aufgenommen werden, dass er mit der Perspektive des Nutzers übereinstimmt und es ist, zum Beispiel bei Sportveranstaltungen, notwendig der Aufnahme noch einzelne Mikrofone wie die der Kommentatoren hinzuzufügen. Bei der Wiedergabe von solchen Übertragungen, wie auch 360° Videos, benötigt der Nutzer Möglichkeiten, die Blickrichtung zu steuern und die Wiedergabe von Bild und Ton muss entsprechend angepasst werden. Eine weitere Anwendungsmöglichkeit von linearer VR sind virtuelle Rundgänge. Durch diese können zum Beispiel Besucher eines Museums historische Bauten in ihrem ursprünglichen Zustand erleben.[116] Weiter können Architekten ihren Kunden ein geplantes Haus präsentieren, bevor es gebaut ist. Mit Hilfe von Nachhallsimulationen können auch die akustischen Eigenschaften von Gebäuden vor dem Bau untersucht werden.[117] Auch bei diesen Anwendungen muss der, zuvor produzierte, Ton der Blickrichtung des Nutzers entsprechen.

[115] (Vgl. Shivappa et al., 2016, S. 5)
[116] (Vgl. De Paolis, S. 19)
[117] (Vgl. Mazuryk und Gervautz, 1996, S. 6)

6.2.2 Interaktive VR-Anwendungen

Bei den interaktiven Anwendungen für VR hat der Nutzer die Freiheit selbst zu handeln und das Erlebnis zu beeinflussen. Dabei sind verschiedene Grade an Interaktivität möglich. Diese beginnen bei einfachen Entscheidungen, wie in interaktiven Filmen, in denen der Zuschauer beispielsweise die Möglichkeit hat, aus vorgegebenen Optionen Handlungsalternativen für den Protagonisten zu wählen. Da es sich bei solchen Filmen lediglich um 360° Filme mit mehreren möglichen Handlungsabläufen handelt, hat die Interaktivität in diesem Fall keine besonderen Auswirkungen auf den Ton und es gelten die selben Anforderungen wie für lineare 360° Videos und Filme. Weiter können mit Hilfe von VR Daten interaktiv visualisiert und bearbeitet werden; so können beispielsweise Ärzte Ergebnisse von Computertomographien als 3D-Modell untersuchen oder es können Produktionsanlagen in der Industrie bei der Planung sehr viel anschaulicher visualisiert und die Planungen mit sofortiger Sichtbarkeit der Auswirkungen modifiziert werden.[118] Bei der reinen Visualisierung von Daten ist normal kein immersiver Ton notwendig, daher kann der Ton in solchen Anwendungsfällen für einfache Informationszwecke genutzt oder komplett weggelassen werden.
Eine andere mögliche Anwendung findet sich beim Militär. Hier ist mit Flugsimulatoren zunächst eine der ältesten Anwendungen von VR zu nennen, aber auch andere Trainingssimulationen und die Fernsteuerung von Geräten durch so genannte „Telepresence", bei der der Steuernde sich fühlen soll, als wäre er vor Ort.[119] Im Fall von Trainingssimulationen muss der Simulator möglichst realistischen Ton generieren, um den Nutzer in die Simulation hineinzuversetzen. Ähnlich ist es bei der „Telepresence", hier werden Bild und Ton übertragen anstatt generiert. Auch außerhalb des Militärs lassen sich Trainingssimulationen nutzen. So kann VR auch sehr effektiv für Bildung und Ausbildung genutzt werden. Beispielsweise können angehende Ärzte riskante Operationen ohne Gefahr für tatsächliche Patienten durchführen, Schüler können komplizierte Versuche in einem virtuellen Labor durchführen und Sportler werden in Zukunft

[118] (Vgl. Brill, 2009, S. 2)
[119] (Vgl. De Paolis, S. 20

möglicherweise von virtuellen Trainern trainiert.[120] [121] Kooperative Nutzung von VR-Systemen ist eine Möglichkeit, die großes Potential hat. Dadurch können Nutzer, über weite räumliche Distanz an einem gemeinsamen Arbeitsplatz zusammenarbeiten und Training mit mehreren Nutzern wird somit ermöglicht.[122] Im Bereich der VR-Spiele gibt es unterschiedliche Ausprägungen der Interaktivität. Von Spielen mit fixer Position, in denen sich die Interaktion auf Handbewegungen beschränkt, lässt sich die Interaktion auf den ganzen Körper ausweiten, wobei die Navigation der virtuellen Welt, bei Consumer-Produkten, häufig mit Handcontrollern bewältigt wird und nur ein kleiner Bereich für tatsächliche Bewegung zur Verfügung steht. Die größtmögliche Interaktivität und natürlichste Interaktion bieten bisher so genannte „large-scale room scale video games", bei denen einzelne oder mehrere Spieler sich frei in der VR bewegen und auf verschiedenste Weise interagieren können.[123] Bei Spielen, wie auch Trainingssimulationen, können nur die Bausteine für die Szenen im Vorhinein produziert werden, da der Ton anhand von Interaktionen der Nutzer generiert wird. Dabei kommt es auf die Art von Simulation oder Spiel an, wie realistisch oder komplex die Szene sein muss. Beispielsweise benötigt die Simulation eines Operationsaals mit den Geräuschen medizinischer Geräte, wie Kardiografen, und den Stimmen der Ärzte und Helfer nur eine überschaubare Anzahl an Schallquellen, während in Spielen mit mehreren Spielern und einer Vielzahl an Gegnern die Zahl der Schallquellen erhebliche Ausmaße annehmen kann.

6.2.3 Mobile und Stationäre VR

Durch die rasante technologische Entwicklung der letzten Jahre sind Smartphones heute leistungsfähig genug, um Bilder und Ton für VR-Anwendungen zu rechnen und durch Beschleunigungsmesser sowie Magnetometer in der Lage, die Orientierung des Nutzers in Echtzeit zu bestimmen. Mit Hilfe von

[120] (Vgl Mazuryk und Gervautz, 1996, S. 9f.)
[121] (Vgl. De Paolis, S. 22)
[122] (Vgl. Mazuryk und Gervautz, 1996, S. 12)
[123] (Vgl. Kellaway, 2016, S. 2f.)

Positionsbestimmung wie GPS ist es möglich, auch die Position des Nutzers überall zu bestimmen und Anwendungen zu entwickeln, die diese Informationen nutzen.[124] Kombiniert mit Produkten wie zum Beispiel „Google Cardboard"[125] ermöglichen heutige Smartphones es, virtuelle Realitäten jederzeit und überall zu erleben. Abbildung 9 zeigt die zuvor genannte VR-Brille „Google Cardboard", die, bestehend aus Pappe und zwei Linsen, die Nutzung von VR-Anwendungen über ein Smartphone ermöglicht. Die VR-Brille „GEAR VR" des koreanischen Herstellers Samsung, zu sehen in Abbildung 10, bietet mit eignen Sensoren, Kopfband und Bedienelementen für spezielle Apps deutlich mehr Komfort, macht den Genuss von mobiler VR aber nur mit bestimmten Smartphones dieses Herstellers möglich. Dabei unterscheiden sich, gegenüber stationären VR-Anwendungen, besonders die Anforderungen an den Ton. Dieser wird bei mobilen VR-Anwendungen im einfachsten Fall als Monosignal vom Lautsprecher des Smartphones wiedergegeben. Die simpelste Lösung für 3D-Ton mit hoher Qualität nutzt in diesem Fall Kopfhörer, die die Mobilität nicht einschränken und helfen können störende Umgebungsgeräusche zu unterdrücken. Um über Kopfhörer 3D-Ton zu übertragen sind binaurale Signale notwendig, deren Berechnung mit Hilfe von Faltungen, rechenintensiv ist. Auch wenn heutige Smartphones beeindruckend leistungsstark sind, müssen noch immer Kompromisse bei der Berechnung von komplexen Szenen gemacht werden. Umsetzungen mit externen Lautsprechern sind für mobile VR-Anwendungen nicht praktikabel.

[124] (Vgl. Mazuryk und Gervautz, 1996, S. 53)
[125] (Vgl. https://vr.google.com/cardboard/)

Abbildung 9: Google Cardboard VR-Brille

Abbildung 10: Samsung GEAR VR

7 VR-Sound

7.1 Ambisonics

7.1.1 Stärken&Schwächen

Die klassische Form von Ambisonics, FOA, schafft es zwar mit nur vier Kanälen ein drei dimensionales Schallfeld zu kodieren, allerdings kommt es aufgrund hoher Kohärenz zwischen den Lautsprechern zu Richtungsartefakten.[126] Weiter verfügt FOA über eine sehr geringe Richtungsauflösung und der Sweetspot ist sehr klein, sodass es durch den Kopf des Hörers zu Abschattungen und den dadurch ausgelösten falschen Richtungswahrnehmungen und Klangverfärbungen kommen kann.[127] Die Richtungsauflösung leidet zusätzlich unter Aufnahmebedingungen mit starkem Nachhall sowie weit entfernten Schallquellen.[128]
Durch hinzunehmen weiterer Ordnungen von Kugelflächenfunktionen lassen sich sowohl der Sweetspot vergrößern als auch die Richtungsauflösung verbessern.[129] Dabei kann der Sweetspot bei sehr hohen Ordnungen den gesamten Abhörbereich einschließen.[130] Allerdings steigt mit ansteigender Ordnung die Anzahl an beteiligten Kugelflächenfunktionen und damit auch die benötigten Aufnahmekanäle und Mindestanzahl an Lautsprechern für eine fehlerfreie Wiedergabe stark an.[131] Somit werden für die Aufnahme von HOA-Signalen entweder viele einzelne Mikrofone benötigt, wodurch das Schallfeld nur schwer an möglichst einem Punkt aufgenommen werden kann oder es werden spezielle Mikrofone mit vielen Kapseln in kugelförmiger Anordnung verwendet. Solche Mikrofone sind noch nicht sehr einfach verfügbar und neigen aufgrund der vielen Mikrofon-kapseln zu niedrigen Rauschabständen.[132] Neben Aufnahmen können HOA-Signale auch aus

126 (Vgl. Hiekkanen et al., 2007, S. 2)
127 (Vgl. Blauert und Rabenstein, 2012, S. 10)
128 (Vgl. Bates und Boland, 2016, S. 5)
129 (Vgl. Shivappa et al, 2016, S. 4)
130 (Vgl. Blauert und Rabenstein, 2012, S. 11)
131 (Vgl. Altman et al., 2016, S. 3)
132 (Vgl. Tsingos et al., 2016, S. 1)

bestehendem Material produziert werden oder bestehendes Material kann zu Aufnahmen hinzugemischt werden, was die Produktion sehr flexibel macht.[133]
Bei der weiteren Bearbeitung erlaubt HOA mit bestehenden Plug-ins schon vielfältige Möglichkeiten, wie Verzerrungen des Schallfeldes oder richtungsabhängige Filter. So können Effekte verwirklicht werden, die mit kanalbasierten Systemen nicht möglich und für objektbasierte Systeme zu komplex sind.[134] Auch Rotationen des Schallfeldes sind, aufgrund der Dekomposition in Kugelflächenfunktionen, mit wenig Rechenaufwand und geringer Fehleranfälligkeit umsetzbar.[135] Weiter kann die wahrgenommene Distanz von Schallquellen manipuliert werden, um dies fehlerfrei wiederzugeben müssen die Abstände der Lautsprecher bei der Wiedergabe bekannt sein, damit schränken diese Manipulationen die Flexibilität bei der Wiedergabe ein.[136] Diese Flexibilität kommt daher, dass Ambisonics nicht für eine bestimmte Wiedergabeform produziert wird, wie es beispielsweise bei kanalbasierten Formaten der Fall ist, sondern ein in Kugelflächenfunktionen aufgeteiltes Schallfeld kodiert, das auf der Wiedergabeseite für eine beliebige Lautsprecheranordnung dekodiert wird und auch für Kopfhörer kann Ambisonics wiedergegeben werden. Außerdem kann ein Dekodierer Rotationen des Schallfeldes durchführen um Kopfbewegungen des Nutzers zu Kompensieren, dies ist eine wichtige Funktionalität für VR-Anwendungen.[137]
Auch bei der Übertragung weist Ambisonics vorteilhafte Eigenschaften auf, so können Signale effizient komprimiert werden, wobei die benötigte Datenrate zur Übertragung nur von der Anzahl beteiligter Kugelflächenfunktionen, also der Ordnung, abhängt und nicht der Komplexität des Schallfeldes.[138] Darüber hinaus können Signale in mehreren Schichten übertragen werden, um die Übertragung

[133] (Vgl. Shivappa et al., 2016, S. 5)
[134] (Vgl. Shivappa et al, 2016, S. 7)
[135] (Vgl. Shivappa et al, 2016, S. 2)
[136] (Vgl. Blauert und Rabenstein, 2012, S. 12)
[137] (Vgl. Shivappa et al, 2016, S. 8)
[138] (Vgl. Shivappa et al, 2016, S. 2)

auch für Nutzer mit geringeren Bandbreiten und schwächeren Dekodierern zu ermöglichen.[139]

Auch bei der Wiedergabe zeigen sich Stärken von HOA. Vor allem bei der Binauralisierung kann HOA eine große Anzahl von virtuellen Lautsprechern verwenden, wobei die Anzahl der Lautsprecher die benötigte Rechenleitung nicht beeinflusst und es können dreidimensionale Anordnungen von Lautsprechern verwendet werden, die in der Realität sehr unpraktisch sind.[140] Darüber hinaus erlaubt die einfache Rotation von Schallfeldern die Verwendung statischer HRTF-Filter bei der Binauralisierung.[141]

Allerdings sind Mikrofone zur Aufnahme, sowie Programme zur Bearbeitung noch nicht ausgereift und einfach verfügbar und HOA ist noch nicht sehr weit verbreitet. So unterstützt die Audio SDK des VR-Brillenherstellers Oculus beispielsweise nur FOA.[142]

7.1.2 Mögliche Anwendungen

Aufgrund seiner Flexibilität bei der Produktion und Wiedergabe eignet sich Ambisonics für eine Vielzahl an Anwendungen. So kann der Ton für 360°-Video und andere lineare VR-Anwendungen beliebig komplex gestaltet werden, effiziente Kompression ermöglicht Liveübertragung und Streaming und die hierarchische Struktur ermöglicht die Skalierung der benötigten Rechenleistung und Bandbreite.

Bei interaktiven VR-Anwendungen eignet sich Ambisonics vor allem für Umgebungsgeräusche und kann diese per Rotation an Bewegungen des Nutzers anpassen. Allerdings ist HOA bei Videospielen auch für den Kompletten Ton schon seit einigen Jahren, wenn auch nur vereinzelt, in Verwendung.[143]

---

139 (Vgl. Shivappa et al, 2016, S. 7f.)

140 (Vgl. Shivappa et al, 2016, S. 1)

141 (Vgl. Travis, 1996, S. 5f.)

142 (Vgl. https://developer3.oculus.com/documentation/audiosdk/latest/concepts/audiosdk-features/#audiosdk-features-supported)

143 (Vgl. https://etiennedeleflie.net/2007/08/30/interview-with-simon-goodwin-of-codemasters-on-the-ps3-game-dirt-and-ambisonics/)

Bei allen Anwendungen lässt sich die Wiedergabe sowohl mit Lautsprechern als auch mit Kopfhörern implementieren.

7.2 Binaurale Aufnahme und Synthese

7.2.1 Stärken & Schwächen

Bei der binauralen Aufnahme und Synthese werden die benötigten Informationen für das räumliche Hören in ein Signal mit nur zwei Kanälen, einer für jedes Ohr, verpackt. Damit können die benötigten Datenraten für die Übertragung sehr klein gehalten werden. Wie genannt, können diese Signale im Gehörgang einer Person oder mit Hilfe eines Kunstkopfes aufgenommen, aber auch komplett synthetisch produziert werden. Dabei ist es wichtig, dass sowohl die Aufnahme- als auch die Wiedergabeseite korrekt entzerrt sind, um alle Informationen für das Gehör richtig festhalten und wiedergeben zu können.[144]

Die binaurale Aufnahme eignet sich sehr gut zur Aufzeichnung von Nachhall, wobei eine deutliche Trennung von Direkt- und Diffusschall bleibt.[145] Allerdings bewegen sich, bei der Wiedergabe von Aufnahmen, die Schallquellen bei Kopfbewegungen mit dem Kopf mit.[146]

Bei der Synthese werden Monosignale mit HRTFs gefaltet, um Schallquellen in bestimmter Richtung und Entfernung darzustellen. Hierbei müssen die verwendeten HRTFs möglichst genau mit denen des Nutzers übereinstimmen, da es sonst zu Fehllokalisationen und Klangverfärbungen kommen kann. Die Messung von individuellen Transferfunktionen ist jedoch sehr aufwändig.[147] Daher werden meist nicht individualisierte HRTFs verwendet, die dem durchschnittlichen Nutzer entsprechen und für möglichst viele Personen funktionieren sollen, denn HRTFs sind von Mensch zu Mensch mitunter sehr unterschiedlich.[148] Der Mensch ist aber auch in der Lage, sich an nicht exakt übereinstimmende Transferfunktionen anzupassen, wobei dieser Prozess mit Hilfe von Training stark beschleunigt werden

144 (Vgl. Hammershøi und Møller, 2005, S. 224f.)
145 (Vgl. Griesinger, 1989, S. 27)
146 (Vgl. Travis, 1996, S. 4)
147 (Vgl. Kaneko et al., 2016, S. 1)
148 (Vgl. Mendonça et al., 2010, S. 2)

kann[149] und die Verbesserung von durchschnittlichen HRTFs ist Gegenstand der Forschung.[150] [151]

Bei der Synthese von binauralem Ton können, wenn die Ausrichtung des Nutzers bekannt ist, die HRTF-Filter in Echtzeit aktualisiert werden, um die Kopplung der Schallquellen an die Ausrichtung des Kopfes zu lösen. Dies erlaubt dem Nutzer Kopfbewegungen zur Ortung von Schallquellen zu nutzen und verbessert so die räumliche Darstellung.[152] Eine Schwierigkeit dabei ist, dass die gemessenen HRTFs nur für diskrete Positionen gültig sind und bestehende Datenbanken nur eine begrenzte Anzahl von Positionen und Distanzen enthalten. Die nicht enthaltenen Positionen und Distanzen müssen aus den nächstgelegenen bekannten HRTFs interpoliert werden.[153]

Die Wiedergabe von binauralen Signalen erfolgt im einfachsten Fall mit Kopfhörern, da diese eine komplette Kanaltrennung bieten, aber mit Hilfe von Filtern, die das Übersprechen zwischen zwei Lautsprechern diminuieren, lassen sich auch Lautsprecher für die Wiedergabe einsetzen. Dazu müsste der Nutzer jedoch still an einem Punkt bleiben, aber wenn seine Position und Ausrichtung verfolgt werden, kann der Filter für die Kanaltrennung entsprechend angepasst werden. Mit einem solchen System ist bei Wiedergabe mit nur vier Lautsprechern eine volle Umdrehung des Nutzers möglich.[154] Andererseits ist diese dynamische Kanaltrennung auf einen simultanen Nutzer begrenzt.[155]

7.2.2 Mögliche Anwendungen

Binaurale Aufnahmen eignen sich für Musikaufnahmen im Studio oder bei Konzerten, um vor allem die Raumakustik authentisch für räumliches Hören wiederzugeben. Dabei zeigen sich gerade bei der Wiedergabe Hürden für eine einfache Anwendung, da ein Nutzer für die Wiedergabe mit Lautsprechern eine

[149] (Vgl. Mendonça et al., 2010, S. 5f.)
[150] (Vgl. Lee et al., 2003)
[151] (Vgl. Kaneko et al, 2016)
[152] (Vgl. Villegas, 2015, S. 210)
[153] (Vgl. Villegas, 2015, S. 203)
[154] (Vgl. Lentz und Behler, 2004, S. 3f)
[155] (Vgl. Lentz und Behler, 2004, S. 2)

Filterung von Übersprechen vornehmen muss. Die Rotation von binauralen Aufnahmen ist nur mit viel Aufwand möglich, daher ist die Verwendung für lineare VR-Anwendungen, die mit Richtungsbestimmung arbeiten, nicht sehr geeignet.
Die Synthese von binauralen Signalen hingegen lässt sich sowohl für lineare als auch für interaktive VR-Anwendungen nutzen, bei denen die dazugehörige Darbietung Kopfhörer bevorzugt oder nur die Verwendung weniger Lautsprecher erlaubt. So kann die Synthese mit den Daten der Richtungsbestimmung die räumliche Darstellung vom Kopf des Nutzers entkoppeln, um Bewegungen zu erlauben.

7.3 Objektbasiert

7.3.1 Stärken & Schwächen

Wie auch bei Ambisonics müssen objektbasierte Signale auf der Wiedergabeseite dekodiert und die einzelnen Lautsprechersignale berechnet werden. Somit besteht hier auch die Flexibilität bezüglich der genutzten Lautsprecheranordnung und auf der anderen Seite die Notwendigkeit für leistungsstarke Dekodierer. Objekte können sowohl aufgenommen, als auch aus bestehendem Material produziert und mit ihren Metadaten versehen werden. Im Gesamtsignal verwendet jedes Objekt eine eigene Spur und braucht seine Metadaten, damit ist die Datenmenge und die Bandbreite, die für eine Übertragung des Signals notwendig ist, abhängig von der Komplexität der Ton-Szene.[156]
Objektbasierte Tonproduktion eignet sich gut für lineare Anwendungen, wie die Verbreitung des objektbasierten Systems Dolby Atmos bei Kinofilmen zeigt.[157] Aber auch interaktive Anwendungen, zum Beispiel Videospiele, verwenden Objekte, wobei gerade in diesem Fall die Anzahl an Objekten die benötigte Rechenleistung beeinflusst.[158] Allerdings erlauben es komplexe Metadaten, Objekte mit individuellem Verhalten zu schaffen und die räumliche Auflösung unabhängig von der Anzahl an übertragenen Kanälen zu gestalten. Außerdem kann eine komplette

[156] (Vgl. Shivappa et al., 2016, S. 4)
[157] (Vgl. https://www.dolby.com/us/en/cinema/theatrical-releases.html)
[158] (Vgl. Altman et al., 2016, S. 3)

Szene durch Veränderungen an den Metadaten sehr simpel rotiert werden.[159] Neben Rotationen können Objekte und Szenen auch ohne großen Rechenaufwand bewegt werden, was Bewegungen mit sechs Freiheitsgraden ermöglicht.[160]
Bei der Aufnahme von Live-Events ist das Erstellen von Metadaten mit Schwierigkeiten verbunden und auch dynamische Objekte aufzunehmen ist nur mit Hilfe von Bewegungsverfolgung der entsprechenden Schallquellen möglich.[161]
Für die Wiedergabe werden Lautsprechersignale berechnet oder es wird eine Binauralisierung durchgeführt. Dafür können entweder sämtliche Objekte einzeln mit ihren Positionen entsprechenden HRTFs gefaltet werden, oder es werden Signale für virtuelle Lautsprecher berechnet, die dann mit den HRTFs ihrer Positionen gefaltet werden. Damit ist der benötigte Rechenaufwand für dieses Verfahren abhängig von der Anzahl der Objekte oder der virtuellen Lautsprecher.[162]

7.3.2 Mögliche Anwendungen

Objektbasierte Systeme eignen sich für die Tonproduktion und Wiedergabe für lineare VR-Anwendungen, wie beispielsweise 360°-Video oder Übertragungen von Live-Events. Bei Übertragungen jeglicher Inhalte müssen jedoch Kompromisse zwischen der benötigten Datenrate und der Komplexität der Ton-Szene gemacht werden.
Auch bei interaktiven VR-Anwendungen können objektbasierte Systeme eingesetzt werden. Gerade bei diesen Anwendungen können die Möglichkeiten dynamischer Objekte voll ausgenutzt werden. Diese Systeme werden im Bereich von Videospielen bereits eingesetzt, daher ist ein Einsatz in VR-Spielen zu erwarten[163]. Allerdings besteht auch hier die Verbindung zwischen der Anzahl an Objekten und dem Rechenaufwand.
Die große Flexibilität bei der Wiedergabe erlaubt den Einsatz von Lautsprechern oder Kopfhörern.

[159] (Vgl. Altman et al., 2016, S. 4)
[160] (Vgl. Altman et al., 2016, S. 5)
[161] (Vgl. Altman et al., 2016, S. 6)
[162] (Vgl. Shivappa et al., 2016, S. 4)
[163] (Vgl. https://www.audiokinetic.com/about/news/audiokinetic-wwise-releases-dolby-atmos-support/)

7.4 Andere und Mischformen

7.4.1 Stärken & Schwächen

Eine weitere Möglichkeit Ton zu Produzieren und Wiederzugeben, sind kanalbasierte Verfahren. Diese sind weit verbreitet, sowohl bei linearen Anwendungen wie Film und Fernsehen, als auch bei interaktiven Anwendungen wie Videospielen. Bei diesen Verfahren werden die fertigen Lautsprechersignale übertragen, somit wird auf der Wiedergabeseite keine Rechenleistung benötigt. Allerdings können so nur Signale für bestimmte Lautsprecheranordnungen geliefert werden, was zu wenig Flexibilität führt und die unterstützten Anordnungen bieten meist eine größere räumliche Auflösung in vorderer Richtung, was bei Kopfdrehungen des Nutzers zu Problemen führt. Auch sind solche Systeme, die drei dimensionale Tonwiedergabe unterstützen nur wenig verbreitet.
Es ist auch möglich die Vorteile von verschiedenen Systemen gemeinsam zu nutzen, wie es beispielsweise die hybride „VR Audio Engine" von „3DSoundlabs" tut. Sie verbindet HOA mit Ton-Objekten. Damit können die vielfältigen Möglichkeiten mit der Effizienz von HOA verbunden werden.[164] Abbildung 11 zeigt eine Infografik zur Aufteilung in Objekte und HOA in der zuvor genannten „VR Audio Engine" und wie sich die benötigte Rechenleistung im Vergleich zu einer objektbasierten Engine verhält.
Auch eine Form Hybrid-System ist die Kopfhörerwiedergabe von Ambisonics und objektbasiertem Ton, denn in beiden Fällen werden die entsprechenden Signale binauralisiert.

[164] (Vgl. http://www.pro.3dsoundlabs.com/category/vr-audio-engine/)

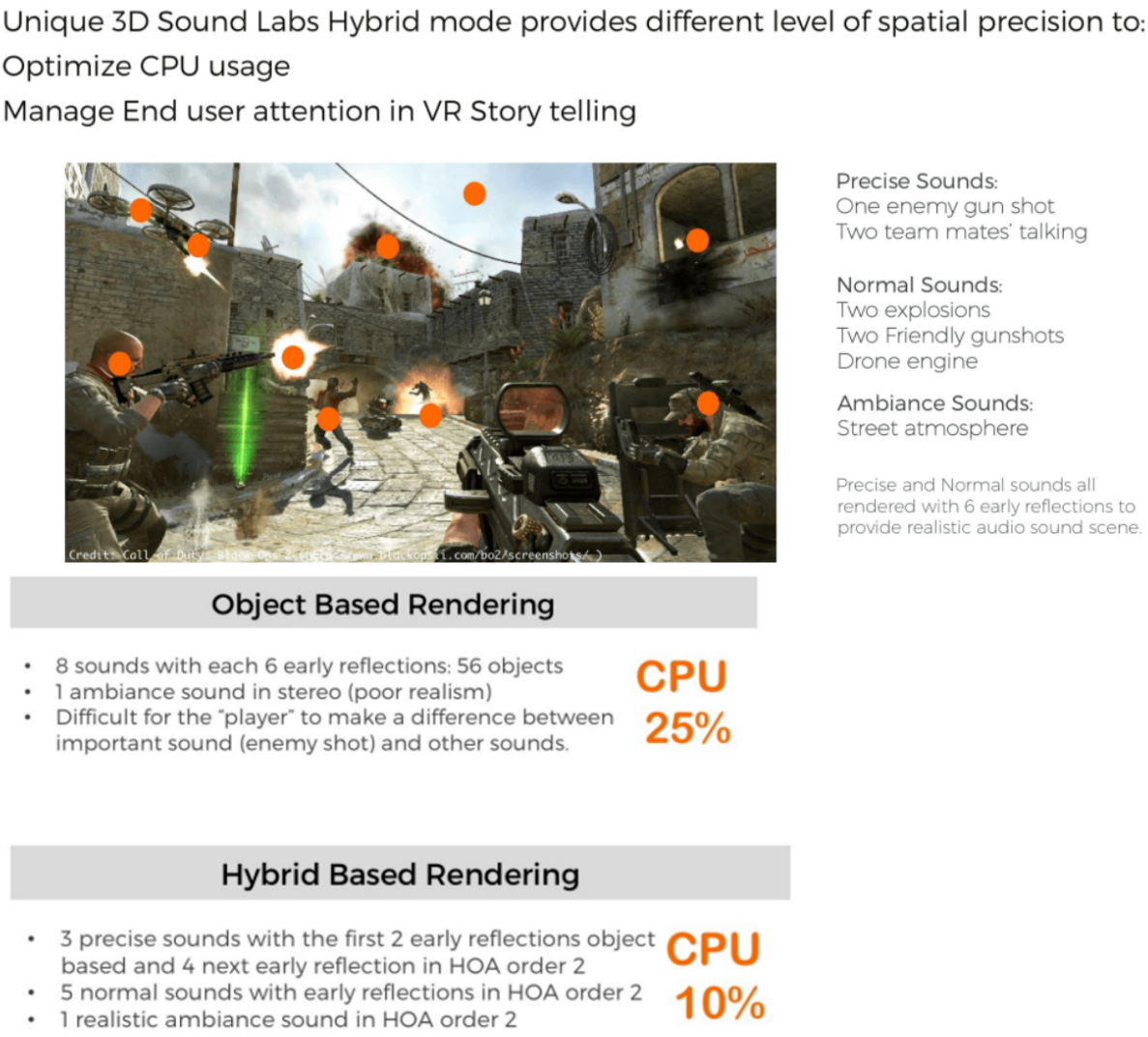

Abbildung 11: Infografik zur VR Audio Engine

7.4.2 Mögliche Anwendungen

Kanalbasierte Systeme eignen sich nur wenig für VR-Anwendungen. Wenn das System auf Bewegungen des Nutzers reagieren soll, müssen alle Lautsprecher virtuell verschoben werden, dazu ist eine Aktualisierung von HRTFs für jeden Lautsprecher, in Echtzeit, notwendig. Weiter sind kaum Systeme für die Wiedergabe von dreidimensionalem Ton geeignet.

Hybride Systeme, wie die „VR Audio Engine“, eignen sich vor allem für interaktive VR-Anwendungen, aber auch bei der Produktion von Material für lineare VR-Anwendungen können verschiedene Technologien gemeinsam verwendet werden.

7.5 Kopfhörer oder Lautsprecher

Alle zuvor betrachteten Systeme zur Produktion und Wiedergabe von dreidimensionalem Ton, lassen sich mit Hilfe der Binauralsynthese nicht nur auf Lautsprechern, sondern auch auf Kopfhörern nutzen. Binaurale Signale können mit Filtern, die das Übersprechen verringern, auf Lautsprechern genutzt werden. Damit ist die Frage nach Kopfhörern oder Lautsprechern eher von den Anforderungen der VR-Anwendungen anhängig, als von der verwendeten Tonwiedergabe.
Den größten Einfluss auf die Wahl des Tonwiedergabemediums hat das verwendete Medium für die Darstellung. Allerdings lassen sich in den meisten Fällen keine allgemeingültigen Aussagen machen, da mehrere Möglichkeiten bestehen und Vorlieben des Nutzers die Wahl beeinflussen. Im Folgenden werden die Darstellung mit HMDs, in CAVEs, mit Hilfe von 3D-Bildschirmen oder Projektionen und mobile VR-Anwendungen hinsichtlich ihrer Eignung für die Kopfhörer- oder Lautsprecherwiedergabe betrachtet und Empfehlungen gegeben.

7.5.1 Mobile VR-Anwendungen

Den einfachsten Fall stellen mobile VR-Anwendungen dar, die meist mit Smartphones ermöglicht werden. Dabei ist er Einsatz von Lautsprechern äußerst unpraktisch und die internen Lautsprecher der Smartphones bieten normal nur die Wiedergabe von Monosignalen.
Gleichzeitig können mit Kopfhörern binaurale oder binauralisierte Signale für eine dreidimensionale Wiedergabe genutzt werden. Daher sind für mobile VR-Anwendungen eindeutig Kopfhörer zu empfehlen.

7.5.2 3D-Bildschirme und Projektionen

Bei der Darstellung mit 3D-Bildschirmen und Projektionen wird hier nur von einem einzigen Nutzer ausgegangen, da zu jeder Zeit nur ein Nutzer die Blickrichtung in der virtuellen Realität, mit einem entsprechenden Eingabegerät, steuern kann. In diesem Fall sind Lautsprecher zur Wiedergabe geeignet, für dreidimensionale Wiedergabe werden entsprechende Lautsprecheranordnungen benötigt. Falls binaurale Signale wiedergeben werden, werden nur zwei Lautsprecher benötigt. Allerdings ist ohne Positionsbestimmung des Nutzers der Bereich der korrekten Wiedergabe sehr klein.
Auch Kopfhörer können verwendet werden, dabei empfiehlt sich eine Richtungsbestimmung des Nutzers, damit die Ursprungsorte von Schallquellen bei Kopfbewegungen stabil bleiben.
Ob Lautsprecher oder Kopfhörer verwendet werden sollten, ist bei Bildschirmen und Projektionen vor allem von den Vorlieben und Möglichkeiten der Nutzer abhängig.

7.5.3 Cave Automatic Virtual Environment

In einem CAVE müssen Position und Ausrichtung des Nutzers schon für die grafische Darstellung bestimmt werden. Die binaurale Wiedergabe über Lautsprecher kann diese Daten für eine dynamische Filterung des Übersprechens nutzen und so, mit nur wenigen Lautsprechern, dreidimensionalen Ton bieten. Andere Lautsprechersysteme benötigen für dreidimensionale Wiedergabe deutlich mehr Lautsprecher. Dies ist ein Problem, da Lautsprecher innerhalb des CAVE die Projektionsflächen verdecken und außerhalb auch nur die Bereiche, in denen die Lautsprecher den Projektionen nicht im Weg sind, genutzt werden können.
Auch die Wiedergabe über Kopfhörer kann die Positions- und Richtungsdaten des CAVE nutzen. In diesem Fall kann die Darstellungsfläche nicht von Lautsprechern verdeckt werden, allerdings können Kabel von kabelgebundenen Kopfhörern den Nutzer behindern.

Bei einem CAVE ist die Wahl zwischen Kopfhörern und Lautsprechern von den vorgesehenen Anwendungen sowie den Vorlieben der Nutzer abhängig. Lautsprechersysteme sind allerdings deutlich praktischer.

7.5.4 HMDs

Wenn für VR-Anwendungen HMDs verwendet werden, ist das Ziel meist eine möglichst vollständige Immersion des Nutzers. Mit Kopfhörern wird diese, durch die weitere Abschottung nach außen, unterstützt. Weiter können die Positions- und Richtungsdaten des HMD für eine stabile binaurale Wiedergabe genutzt werden. Bei einem einzelnen Nutzer ist dies auch mit Lautsprechern möglich, mit mehreren Nutzern muss für die Wiedergabe über Lautsprecher ein aufwändigeres System verwendet werden. Bei linearen Anwendungen, wie Liveübertragungen, erleichtert die Wiedergabe über Lautsprecher die Kommunikation zwischen den Nutzern. Bei interaktiven Anwendungen können von Nutzern generierte Geräusche andere Nutzer ablenken.
In den meisten Fällen lassen sich Kopfhörer für die Nutzung mit HMDs empfehlen, bei bestimmten Anwendungen überwiegen die Vorteile von Lautsprechern.

8 Fazit

Es sollte der Frage, ob Lautsprecher- oder Kopfhörersysteme besser für den Einsatz bei VR-Anwendungen geeignet sind nachgegangen werden. Die Technologien, die dreidimensionale Tonwiedergabe erlauben und damit für VR-Anwendungen nützlich sind, bedienen sich verschiedener Techniken. Es wurden Aufbau und Funktionsweise des Ohres im Hinblick auf räumliches Hören betrachtet, dabei hat sich gezeigt, dass ITDs, ILDs und spektrale Veränderungen durch das Außenohr die wichtigsten Informationen zur Ortung einer Schallquelle liefern. Dieser Informationen bedient sich die Binauraltechnik, um Höreindrücke an bestimmten Orten entstehen zu lassen. Diese Technik eignet sich besonders für die Wiedergabe mit Kopfhörern, da diese eine vollständige Kanaltrennung ermöglichen und wird auch verwendet, um andere Techniken, die mit Lautsprechern arbeiten, auf Kopfhörern wiederzugeben.
Nach der Binauraltechnik wurde auch Ambisonics genauer betrachtet. Ambisonics zerlegt ein Schallfeld in Kugelflächenfunktionen und kodiert deren Koeffizienten. Damit kann ein Dekodierer Lautsprechersignale berechnen, deren Überlagerung wieder das kodierte Schallfeld bildet. Dieser Ansatz bietet statische Datenmengen für Übertragung und Speicherung, die von der Anzahl an beteiligten Kugelflächenfunktionen bestimmt werden und unabhängig von der Komplexität der Szene sind.
Weiter wurde eine alternative zu diesem physikalisch basierten Ansatz untersucht. Bei objektbasiertem Ton werden auf der Aufnahme- und Produktionsseite Audiosignale mit Metadaten versehen, die die Eigenschaften der Quelle des Signals festlegen. Audiosignal und Metadaten bilden dann ein Objekt. Mit Hilfe solcher Audioobjekte können komplexe Szenen aufgenommen oder erstellt werden. Ein Dekodierer setzt auf der Wiedergabeseite aus den Objekten die Szene zusammen und berechnet, anhand der Lautsprecherpositionen, die Signale für die Lautsprecheranlage.
Weiter wurde ein Überblick über die Entwicklungsgeschichte und Anwendungen von Virtual Reality gegeben. Dabei hat sich gezeigt, dass sich die Anforderungen

an den Ton von Anwendung zu Anwendung deutlich unterscheiden und dass keine der verschiedenen Techniken eine Ideallösung für alle Anwendungen ist.
Zuletzt wurden die betrachteten Techniken auf ihre Eignung für den Einsatz bei bestimmten VR-Anwendungen, sowie die Frage nach dem Wiedergabemedium untersucht. Dabei stellte sich heraus, dass alle Techniken bei einer Vielzahl von Anwendungen gute Ergebnisse liefern, aber vor allem die Binauraltechnik wichtig ist. Diese liefert auch bei allen anderen Verfahren die Möglichkeit der Wiedergabe über Kopfhörer. Aber auch Hybride aus verschiedenen Technologien werden in nächster Zeit verfügbar sein. Somit ist es schwierig vorauszusagen, welche Techniken sich durchsetzten werden. Da aber in den meisten Anwendungsfällen sowohl Kopfhörer, als auch Lautsprecher eingesetzt werden können und es nur in wenigen Fällen deutliche Einschränkungen für das Wiedergabemedium gibt, werden Nutzer diese Entscheidung hauptsächlich auf Basis ihrer Vorlieben und finanziellen Möglichkeiten treffen.

Literaturverzeichnis

Artikel:

Blauert, J., & Rabenstein, R. (2012). Providing Surround Sound with Loudspeakers: A Synopsis of Current Methods. *Archives of Acoustics*, *37*(1), 5-18. doi:10.2478/v10168-012-0002-y

Dörner, R., Broll, W., Grimm, P., & Jung, B. (2016). Virtual Reality und Augmented Reality (VR/AR). *Informatik-Spektrum*, *39*(1), 30–37. doi:10.1007/s00287-014-0838-9

Gerzon, M. A. (1973). Periphony: With-Height Sound Reproduction. *J. Audio Eng. Soc*, *21*(1), 2–10.

Gerzon, M. A. (1985). Ambisonics in Multichannel Broadcasting and Video. *J. Audio Eng. Soc*, *33*(11), 859–871.

Griesinger, D. (1989). Equalization and Spatial Equalization of Dummy-Head Recordings for Loudspeaker Reproduction. *J. Audio Eng. Soc*, *37*(1/2), 20–29.

Hammershøi, D., & Møller, H. (2005). Binaural Technique: Basic Methods for Recording, Synthesis, and Reproduction. In J. Blauert (Hrsg.), *Communication Acoustics.* (S. 223-254). IEEE Computer Society Press. Aufgerufen von http://vbn.aau.dk/files/227876329/2005_Hammersh_i_and_M_ller.pdf

Lam, J., Kapralos, B., Kanev, K., Collins, K., Hogue, A., & Jenkin, M. (2015). Sound localization on a horizontal surface: Virtual and real sound source localization. *Virtual Reality*, *19*(3-4), 213–222. doi:10.1007/s10055-015-0268-2

Mazuryk, T., & Gervautz, M. (1996). *Virtual Reality History, Applications, Technology and Future* (No. TR-186-2-96-06). Favoritenstrasse 9-11/186, A-1040 Wien, Österreich. Abgerufen von Institute of Computer Graphics and Algorithms, Vienna University of Technology website: https://www.cg.tuwien.ac.at/research/publications/1996/mazuryk-1996-VRH/

Møller, H. (1992). Fundamentals of Binaural Technology. *Applied Acoustics*, 36(3/4), 171-218.

Turchet, L., Spagnol, S., Geronazzo, M., & Avanzini, F. (2016). Localization of self-generated synthetic footstep sounds on different walked-upon materials through headphones. *Virtual Reality*, *20*(1), 1–16. doi:10.1007/s10055-015-0272-6

Villegas, J. (2015). Locating virtual sound sources at arbitrary distances in real-time binaural reproduction. *Virtual Reality*, *19*(3-4), 201–212. doi:10.1007/s10055-015-0278-0

Bücher:

Blauert, J. (1974). *Räumliches Hören*. *Monographien der Nachrichtentechnik*. Stuttgart: Hirzel.

Brill, M. (2009). *Virtuelle Realität*. *Informatik im Fokus*. Berlin, Heidelberg: Springer-Verlag.

Ulrich, J., & Hoffmann, E. (2007). *Hörakustik: Theorie und Praxis* (1. Aufl.). Heidelberg: DOZ-Verl.

Research paper:[165]

Altman, M., Krauss, K., Susal, J., & Tsingos, N., 2016. *Immersive Audio for VR*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=18512

Bates, E., & Boland, F. 2016. *Spatial Music, Virtual Reality, and 360 Media*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=18496

Lentz, T., & Behler, G. 2004. *Dynamic Cross-Talk Cancellation for Binaural Synthesis in Virtual Reality Environments*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=12972

Füg, S., Hölzer, A., Borß, C., Ertel, C., Kratschmer, M., & Plogsties, J. 2014. *Design, Coding and Processing of Metadata for Object-Based Interactive Audio*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=17420

[165] Diese Quellen stammen aus der AES-Library. Sie befinden sich auch auf der CD mit dem elektronischen Exemplar dieser Arbeit.

Gasull Ruiz, A., Sladeczek, C., & Sporer, T. 2015. *A Description of an Object-Based Audio Workflow for Media Productions*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=17611

Gorzel, M., Corrigan, D., Squires, J., Boland, F., & Kearney, G. 2012. *Distance perception in real and virtual environments*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=18119

Hiekkanen, T., Lempiäinen, T., Mattila, M., Pulkki, V., & Veijanen, V. 2007. *Reproduction of Virtual Reality with Multichannel Microphone Techniques*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=14055

Kaneko, S., Suenaga, T., Fujiwara, M., Kumehara, K., Shirakihara, F., & Sekine, S. 2016. *Ear Shape Modeling for 3D Audio and Acoustic Virtual Reality: The Shape-Based Average HRTF*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=18091

Kellaway, S.-A. 2016. *Virtually Replacing Reality: Sound Design and Implementation for Large Scale Room Scale VR Experiences*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=18501

Larsson, P., Vastfjall, D., & Kleiner, M. 2002. *Better Presence and Performance in Virtual Environments by Improved Binaural Sound Rendering*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=11148

Lee, S.-l., Kim, L.-H., & Sung, K.-M. 2003. *Head Related Transfer Function Refinement Using Directional Weighting Function*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=12428

Mendonça, C., Santos, J., Campos, G., Dias, P., Vieira, J., & Ferreira, J. 2010. *On the Improvement of Auditory Accuracy with Non-Indivisualized HRTF-Based Sounds*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=15688

Messonnier, J.-C., Lyzwa, J.-M., Devallez, D., & Boisheraud, C. de 2016. *Object-Based Audio Recording Methods*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=18268

Nykänen, A., & Johnsson, R. 2011. *A Comparison of Speech Intelligibility for In-Ear and Artificial Head Recordings*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=15854

Oldfield, R., Shirley, B., & Spille, J. 2014. *An Object-Based Audio System for Interactive Broadcasting*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=17471

Shivappa, S., Morrell, M., Sen, D., Peters, N., & Akramus Salehin, S. M. 2016. *Efficient, Compelling, and Immersive VR Audio Experience Using Scene Based Audio/Higher Order Ambisonics*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=18493

Travis, C. 1996. *A Virtual Reality Perspective on Headphone Audio*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=7082

Tsingos, N., Govindaraju, P., Zhou, C., & Nadkarni, A. 2016. *XY-Stereo Capture and Up-Conversion for Virtual Reality*. Abgerufen von http://www.aes.org/e-lib/browse.cfm?elib=18497

Internetseiten:

Audiokinetic Inc. (2016). *Audiokinetic Wwise releases Dolby Atmos support, enabling immersive object-based audio for game development.* Abgerufen am 02. Februar 2017 von https://www.audiokinetic.com/about/news/audiokinetic-wwise-releases-dolby-atmos-support/

Choueiri, E. Y. (2010). *Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers.* Abgerufen am 10. Januar 2017 von https://www.princeton.edu/3D3A/Publications/BACCHPaperV4d.pdf

De Paolis, L. T. (2015). *Virtual and Aumented Reality Applications Introduction.* Abgerufen am 19. Januar 20017 von http://avrlab.it/wp-content/uploads/2015/03/lez-1-introduction.pdf

Dolby Laboratories, Inc. (2017). *Theatrical Releases in Dolby Vision and Dolby Atmos.* Abgerufen am 02. Februar 2017 von https://www.dolby.com/us/en/cinema/theatrical-releases.html

Eadicicco, L. (05. April 2016). Virtual Reality Buyer's Guide: Oculus Rift vs. HTC Vive vs. Samsung Gear VR. *TIME.* Abgerufen am 14. Februar 2017 von http://time.com/4277763/virtual-reality-buyers-guide/

electronic visualization laboratory. (2002). *The CAVE™ Virtual Reality Theater.* Abgerufen am 12. Januar 2017 von https://www.evl.uic.edu/cave

Etienne Deleflie. (30. August 2007). Interview with Simon Goodwin of Codemasters on the PS3 game DiRT and Ambisonics. [Web Log Eintrag] Abgerufen am 31. Januar 2017 von https://etiennedeleflie.net/2007/08/30/interview-with-simon-goodwin-of-codemasters-on-the-ps3-game-dirt-and-ambisonics/

Google Inc. (o.j.). Google Cardboard. Abgerufen am 16. Januar 2017 von https://vr.google.com/cardboard/

Oculus VR, LLC. (2016). *Oculus Audio SDK Guide. Abgerufen am 22. Januar 2017 von* https://developer3.oculus.com/documentation/audiosdk/latest/concepts/audiosdk-features/#audiosdk-features-supported

3D Sound Labs. (2016). *VR Audio Engine.* Abgerufen am 02. Februar 2017 von http://www.pro.3dsoundlabs.com/category/vr-audio-engine/

Abbildungsquellen:

Abbildung 1: https://commons.wikimedia.org/wiki/File:Anatomy_of_the_Human_Ear_en.svg

Abbildung 2: https://bildungsportal.sachsen.de/opal/auth/RepositoryEntry/1006567462/CourseNode/868818665770 51?1

Abbildung 3: https://commons.wikimedia.org/wiki/File:Akustik_-_Richtungsb%C3%A4nder.svg

Abbildung 4: https://en.wikipedia.org/wiki/Head-related_transfer_function#Technical_derivation

Abbildung 5: http://www.neumann.com/zoom.php?zoomimg=img/photosGraphics/Zooms/KU100_Z.jpg&zoomlabel=Kunstkopf%20KU%20100&w=415&h=600

Abbildung 6: https://de.wikipedia.org/wiki/Datei:Harmoniki.png (Urheber: I, Sarxos)

Abbildung 7: Zusammengesetzt aus: http://de-de.sennheiser.com/mikrofon-3d-audio-ambeo-vr-mic | https://mhacoustics.com/home | http://visisonics.com/products-2/#camera

Abbildung 8: http://www.seamk.fi/en/About-us/Faculties/School-of-Technology/Laboratories/Virtual-Reality-Laboratory

Abbildung 9: http://www.bloculus.de/google-cardboard-einsteiger-vr-oder-papiermuell/

Abbildung 10: http://www.samsung.com/de/wearables/gear-vr-r322/

Abbildung 11: http://www.pro.3dsoundlabs.com/category/intro-to-3d-audio/ (bearbeitet)

Elektronisches Exemplar: