

Relevante Faktoren zur Simulation der Schallausbreitung für eine realitätsnahe Immersion in virtuellen Räumen

Bachelorthesis im Studiengang Medienproduktion

Manik Möllers

Hermannstraße 2

38114 Braunschweig

Matrikel-Nr.: 15341002

Sommersemester 2020

Eine Bachelorarbeit im Fachbereich Medienproduktion

Technische Hochschule Ostwestfalen-Lippe

Liebigstraße 87

32657 Lemgo

Betreuer: Prof.Dr.rer.nat.Guido Falkemeier

Zweitprüfer: Jan Pieniak

Inhalt

1	Einleitung.....	3
2	Schallausbreitung in Räumen	4
3	Auditive Wahrnehmung	7
3.1	Raumwahrnehmung.....	7
3.1.1	Raumdimensionen.....	7
3.1.2	Raummaterial.....	11
3.2	Lokalisierung von Klangquellen	12
3.2.1	Richtungswahrnehmung	12
3.2.2	Distanzwahrnehmung	14
3.3	Zusammenfassung	17
4	Recording und Repräsentation.....	19
4.1	Kanalbasiertes Audio / Channel Based Audio	19
4.2	Objektbasiertes Audio / Object Based Audio.....	20
4.3	Szenenbasiertes Audio / Scene Based Audio	21
4.4	Besonderheiten bei der Wiedergabe mit Kopfhörern.....	25
4.5	Zusammenfassung	26
5	Hall-Effekte	27
5.1	Faltungshall / Convolution Reverb.....	27
5.2	Algorithmische Halleffekte	28
5.3	Gemodelte Räume.....	28
6	Fazit.....	30
7	Diskussion.....	32
8	Literaturverzeichnis	33
9	Abbildungsverzeichnis	37
10	Eidesstattliche Erklärung.....	38

1 Einleitung

Die Simulation realer Räume ist ein technisch und künstlerisch anspruchsvoller Prozess der Akustik und kommt zum Beispiel im Filmbereich häufig vor, wenn es darum geht, Sounds nachträglich an eine aufgenommene Szene anzupassen oder auch gänzlich neu zu erstellen und dem Bild anzupassen. Doch eine auditive Immersion spielt nicht nur im Filmbereich eine große Rolle, sondern kommt in den meisten audiovisuellen Medien zum Tragen. Seien es Videospiele, 360° Videos oder VR.

Um eine immersive Realität für den Konsumenten zu simulieren während eine Ortbarkeit der Klangquellen gewährleistet ist, spielen verschiedenste Faktoren zusammen. Sie setzen ein Verständnis der Wahrnehmung des Menschen und der verschiedenen Wiedergabesysteme und Simulationstechniken voraus. Ich möchte in dieser Arbeit die optimale Methode finden, mit der es möglich ist die größtmögliche Immersion für den Konsumenten zu schaffen. Durch die verschiedensten Medien, in denen dieses Thema wichtig ist, gibt es auch völlig verschiedene technische Anforderungen, sodass ich mich dazu entschieden habe, einzelne Software nicht genauer zu betrachten und stattdessen ein tieferes Grundverständnis zu schaffen.

Um dieses Thema im ganzen erfassen zu können, gehe ich zuerst auf den Prozess der Schallausbreitung in Räumen ein und analysiere danach, welche der im Schall enthaltenen Informationen Menschen nutzen, um die Eigenschaften eines Raums und der in ihm enthaltenen Klangquellen zu bestimmen und wo die die Grenzen der menschlichen Wahrnehmung liegen. Anschließend analysiere ich die Möglichkeiten, mit denen ein realistischer Höreindruck technisch simuliert und abgebildet werden kann, um damit eine klangliche Immersion zu schaffen und diese beim Konsumenten abzubilden.

2 Schallausbreitung in Räumen

Im Falle einer omnidirektionalen Schallquelle breiten sich die Wellen kugelförmig in alle Richtungen aus. Da er die kürzeste Entfernung zurücklegen muss, erreicht der Direktschall den Hörer als erstes. Dicht darauf folgen die ersten Reflektionen, die Teile der vom Raum reflektierten Wellen sind. Diese Reflektionen sind schwächer als der Direktschall, da die zurückgelegte Entfernung größer und die Dämpfung durch Oberflächenabsorption höher ist (Pop und Cabrera 2005, S. 116).

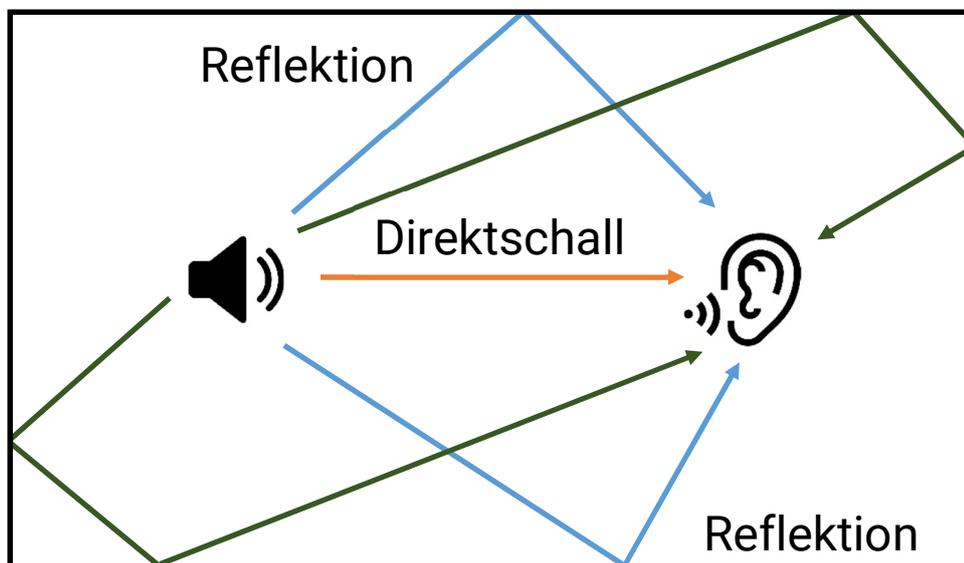


Abbildung 1 Schallausbreitung im Raum (Eigene Darstellung)

Die frühen Reflektionen werden zur Vereinfachung oft als die ersten 100ms des Halls bezeichnet. Konkret handelt es sich dabei um die Reflektionen bevor die Reflektionen eine asymptotisch statistischen Verhaltens aufweisen, in denen die einzelnen Echos nicht mehr unterschieden werden können (O. Smith 2010, EARLY REFLECTIONS). Diese späten Reflektionen bilden den Nachhall (Rungta et al. 2019, S. 1). Der Hall nimmt immer weiter in seiner Intensität ab, bis die gesamte Schallenergie absorbiert worden ist (Pop und Cabrera 2005, S. 116). Die Geschwindigkeit, mit der die Energie absorbiert wird, hängt von der Fläche aller Oberflächen und Objekte im Raum, sowie von

deren Absorptionskoeffizienten ab (Hameed et al. 2004, S. 2). Der Nachhall beeinflusst nahezu alle akustischen Eigenschaften der Signale, einschließlich der zeitlichen Struktur, des spektralen Inhalts und der Intensität (Shinn-Cunningham 2003, S. 2).

In der folgenden Grafik wird erkennbar, wie sich die Reflektionen nach dem Abspielen eines Impulses zeitlich verhalten. Nach dem Verstummen des Direktschalls sind die ersten Reflektionen zu erkennen (Pop und Cabrera 2005, S. 116). Sie bestehen aus einer relativ geringen Anzahl an Echos, während die folgenden späten Reflektionen eine so hohe Dichte an Reflektionen besitzen, dass man sie statistisch betrachten kann (O. Smith 2010, PERCEPTION OF ECHO DENSITY AND MODE DENSITY).

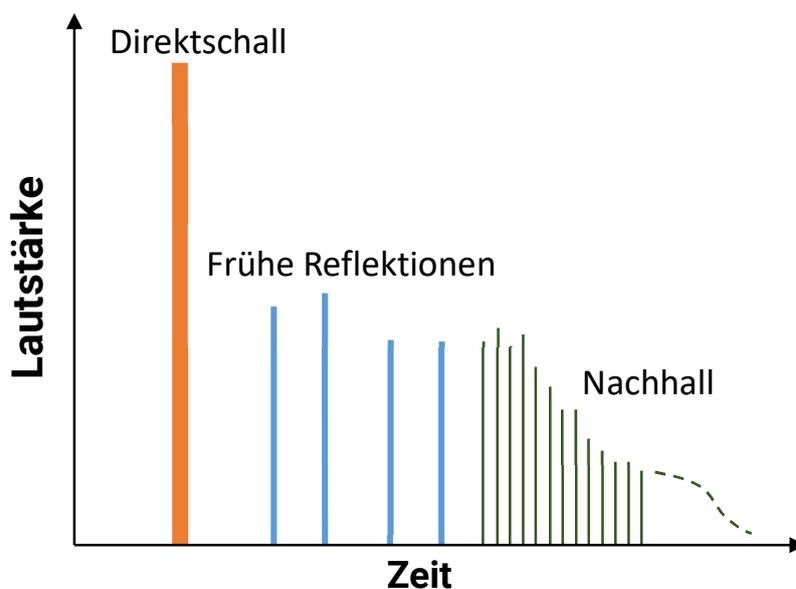


Abbildung 2 Reflektionen im zeitlichen Kontext (Eigene Darstellung)

So wie auch die Reflektionen im Zeitbereich, kann auch der Frequenzgang eines Raumes in zwei Segmente unterteilt werden. Das tieffrequente Intervall besteht aus einer relativ geringen Verteilung von Resonanzmoden. Bei höheren Frequenzen sind die Moden so dicht gepackt, dass sie statistisch betrachtet werden können (O. Smith 2010, PERCEPTION OF ECHO DENSITY AND MODE DENSITY).

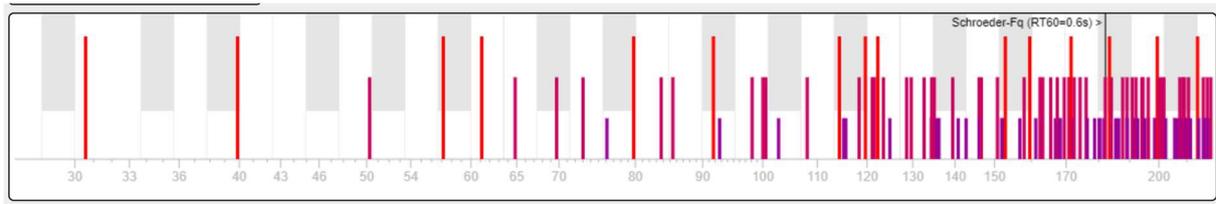


Abbildung 3 Berechnete Verteilung der Raummoden bei Raummaßen von 560x430x300m (amroc - THE Room Mode Calculator 2020)

Bevor der Direktschall und die Raumreflektionen auf das Trommelfell des Hörers treffen, werden sie durch den Kopf, die Ohrmuscheln und den Rumpf des Hörers beeinflusst (Iida 2019, S. 12). Das Signal, das die Ohren erreicht, ist mathematisch äquivalent zum unveränderten Quellsignal, das mit der Impulsantwort des Kopfes und des Raumes gefaltet ist (Shinn-Cunningham 2003, S. 5). Die auf dem jeweiligen Ohr eintreffenden Signale unterscheiden sich je nach Position der Klangquelle und des Zuhörers im Raum in ihren temporalen und spektralen Eigenschaften sowie ihrer Intensität (Shinn-Cunningham 2003, S. 2).

3 Auditive Wahrnehmung

In diesem Kapitel gehe ich auf die Wahrnehmungsfähigkeiten des Menschen ein, mit dem er die im Schall enthaltenen Informationen erfassen kann. Hierfür analysiere ich als erstes die Wahrnehmung des Raumes selbst und gehe dann auf die Lokalisierung von Klangquellen innerhalb eines Raums ein. Zu den Informationen, die durch Schall übertragen werden, gehören z.B. der Standort von Objekten, ihre Bewegung und ihr Zustand, sowie Eigenschaften der Umgebung (Pop und Cabrera 2005, S. 116).

3.1 Raumwahrnehmung

Da der Mensch in der Lage ist, die Größe des Raumes anhand dessen Klangs wahrzunehmen, ist es klar, dass einige Merkmale der Impulsantwort dafür verantwortlich sind und als Anhaltspunkte für die Einschätzung der Raumgröße dienen. Im Folgenden analysiere ich, welche dieser Informationen vom Menschen genutzt werden können, um einen Raum in Bezug auf seine Eigenschaften einzuordnen.

3.1.1 Raumdimensionen

Die Faktoren, die für die Wahrnehmung der Größe eines Raums verantwortlich sind, werden in der Literatur in vier Kategorien eingeteilt, die im Folgenden auf ihre Wirksamkeit untersucht werden.

Early Reflections

Die frühen Reflektionen gleichen dem Quellsignal spektral und temporal, enthalten aber zusätzlich Eigenschaften, die nicht mit dem Klangursprung in Verbindung stehen (Brown et al. 2015, S. 1). In einem kleinen Raum folgen die frühen Reflektionen sehr kurz nach dem Direktschall, in größeren Räumen treffen sie aufgrund der längeren Ausbreitungszeit der Schallwellen im Raum später beim Hörer ein (O. Smith 2010, EARLY REFLECTIONS). Mit der

Größe des Raums erhöht sich auch die zeitliche Dichte der einzelnen Echos aus der die frühen Reflektionen bestehen (Hameed et al. 2004, S. 2).

Für eine glaubwürdige Simulation der Raumgröße ist es trotzdem nicht nötig, die einzelnen Echos genau abzubilden. In einer Studie von Shinn-Cunningham (2003) zeigte sich, dass Menschen die feinen Details der frühen Reflektionen nicht unterscheiden können. Zuhörer können zwar die groben Charakteristiken der frühen Reflektionen zuordnen, das genaue Timing, sowie die Richtung der Echos allerdings nicht differenzieren (Shinn-Cunningham 2003, S. 7). Auf Grund dieser Studie schlossen Hamed et al. (2004) aus, dass frühe Reflektionen einen Einfluss auf die Wahrnehmung der Raumgröße haben (Hameed et al. 2004, S. 2).

Direct to Reverberant Ratio / D/R

Auch das Lautstärkenverhältnis der Reflektionen im Vergleich zum Direktschall (im Folgenden als Direct to Reverberant - D/R bezeichnet), ist in kleinen Räumen größer als in großen, da sich die Lautstärke gegenläufig proportional zur zurückgelegten Strecke verhält. Man kann sich einfach vorstellen, dass Schall in einem kleineren Raum, wie zum Beispiel einem Badezimmer, lauter und halliger klingt (Hameed et al. 2004, S. 5). In Experimenten stellte Sandvad (1999) fest, dass einige Zuhörer das Verhältnis zwischen Direktschall und Nachhallenergie, andere hingegen die Nachhallzeit zur Schätzung der Raumgröße verwendeten (Sandvad 1999). In einer nachfolgenden Studie kamen Hameed et al. (2004) zu dem Ergebnis, dass einige Testpersonen Räume bei einer niedrigen D/R kleiner einschätzten, andere diese aber sogar größer (Hameed et al. 2004, S. 5). Aus diesen Ergebnissen leiteten sie ab, dass D/R nur sehr vorsichtig oder sogar überhaupt nicht für die Simulation von Raumgrößen genutzt werden sollte (Hameed et al. 2004, S. 7).

Clarity Index

Der Clarity Index beschreibt das Lautstärkenverhältnis zwischen der Impulsantwort und der Schallenergie in den ersten 50 bis 80ms (Miśkiewicz et al. 2012, S. 41). Anders ausgedrückt: Ein Raum weist einen hohen Clarity Index auf, wenn er eine hohe Differenz zwischen früher und späterer Nachhallenergie aufweist. Der Name ergibt sich daraus, dass sich der Direktschall und die ersten Reflektionen spektrotemporal gleichen (Brown et al. 2015, S. 1). Ein hoher Clarity Index korreliert daher mit der Klarheit des Signals und erhöht z.B. die Sprachverständlichkeit in einem Raum. Doch auch im Bereich der Raumgrößenwahrnehmung spielt der Clarity Index eine Rolle, wie Cabreraa et al (2005) untersuchten. Sie führten Versuche in verschiedenen Testumgebungen durch, in denen sie die Nachhallzeit durch klangabsorbierendes Material kontrollierten (Cabreraa et al. 2005, S. 4). Es konnte festgestellt werden, dass in Räumen, bei denen die Nachhallzeit als Indikator zur Wahrnehmung der Raumgröße entfällt, der Clarity Index die nächstgrößere Rolle spielt (Cabreraa et al. 2005, S. 9, 2005, S. 5).

Quellen-Empfänger-Distanz

Ein weiterer Faktor, der in verschiedenen Studien auf seine Wahrnehmbarkeit in Bezug auf die Raumgröße überprüft wurde ist die Quellen-Empfänger-Distanz. In den hier betrachteten Versuchen wurde die Distanz zwischen Klangquelle und Empfänger physisch erhöht, was natürlich mit einer entsprechenden Raumgröße in Verbindung steht, damit der Versuch überhaupt möglich wird. Da ich im Kapitel [3.2.2 Distanzwahrnehmung](#) noch genauer auf die Faktoren eingehen werde, die für die Distanzwahrnehmung verantwortlich sind, werde ich hier nur die Versuchsergebnisse betrachten.

Zusammenfassend wurde ein positiver Effekt auf die wahrgenommene Raumgröße durch die Erhöhung der Distanz der Klangquelle bei gleicher realer

Raumgröße festgestellt (Cabreraa et al. 2005, S. 3). Die Raumgrößenwahrnehmung erhöht sich bei einer Zunahme der Distanz von 0,9m auf 2,7m um das Dreifache, während sie sich bei einer Distanzänderung von 2,7m auf 5,1m schon nur noch um das doppelte erhöht (Cabreraa et al. 2005, S. 5).

Nachhallzeit

Die Nachhallzeit ist die Zeit, die benötigt wird, bis die Lautstärke des Halls unter 60dB gefallen ist. Die Zeit bis zum Abklingen des Schalls hängt von den Absorptionskoeffizienten des Oberflächenmaterials und dem Volumen des Raumes ab (Abdullah et al. 2020). Sie bestimmt sich aus dem Verhältnis von Volumen zur absorbierenden Fläche (Pop und Cabrera 2005, S. 116). Wenn also ein großer Raum aus dem gleichen Material besteht wie ein kleinerer, hat der große Raum eine längere Nachhallzeit. Da die Nachhallzeit von vielen Faktoren wie zum Beispiel der Form des Raums und den Materialien der Wände und Möbel abhängt, korreliert sie nicht zwangsläufig mit der Größe des Raums (Hameed et al. 2004, S. 6). Es ist durchaus möglich in einem kleinen Raum eine größere Nachhallzeit zu erzeugen als in einem größeren, da die Nachhallzeit eines Raumes durch klangabsorbierende Materialien kontrolliert werden kann (Cabreraa et al. 2005, S. 4). In den meisten Fällen ist die Nachhallzeit im täglichen Leben in großen Räumen aber größer als in kleinen, da das Volumen normalerweise stärker ansteigt als die Absorptionsfläche (Pop und Cabrera 2005, S. 116; Hameed et al. 2004, S. 2, 2004, S. 6)

Im Vergleich zur Quellen-Empfänger-Distanz und D/R haben Cabrera et al. (2005) sowie Pop und Cabrera (2005) als auch Hameed et al. (2004) herausgefunden, dass die Nachhallzeit den größten Effekt auf die Wahrnehmung der Raumgröße hat (Cabreraa et al. 2005, S. 3; Hameed et al. 2004, S. 5; Pop und Cabrera 2005, S. 116). Nachhallzeit ist damit der

wichtigste Parameter für die Wahrnehmung der Raumgröße (Hameed et al. 2004, S. 1). Cabreraa et al. (2005) geben zudem an, dass sich ihre Studien über die Wichtigkeit von Nachhallzeit mit früheren Studien von Mershon et al. (1989), Hameed et al. (2004) und Sandvad (1999) decken (Cabreraa et al. 2005, S. 9).

3.1.2 Raummaterial

Zur Wahrnehmung von Material und Form eines Raumes liegen nach meinem Wissen kaum Forschungsergebnisse vor. In Studien zu der Wahrnehmung von Impact Sounds wurde aber herausgefunden, dass die relevantesten Merkmale, die eine Unterscheidung zwischen Materialkategorien ermöglichen, die Nachhallzeit und die Klangfarbe sind (Aramaki et al. 2011, S. 312). Aufgrund der bisher erfolgten Studien im Bereich der Raumwahrnehmung, gehe ich davon aus, dass Materialien auf Grund der begrenzten Klangwahrnehmung des Menschen schlecht unterschieden werden können. Räume, die aus stärker reflektierenden Materialien bestehen, haben einen insgesamt lautereren und längeren Nachhall wodurch sich D/R und Nachhallzeit erhöhen. Mit dem Material des Raums, verändert sich also auch die wahrgenommene Raumgröße. Es wäre zu überprüfen, wie sensibel der Mensch auf spektrale Veränderungen des Klangs reagiert.

3.2 Lokalisierung von Klangquellen

Im Folgenden analysiere ich die Wahrnehmungsfähigkeit des Menschen bei der Lokalisierung von Klangquellen in Bezug auf Distanz und Richtung. Durch Richtungs- oder Distanzänderungen einer Klangquelle ändert sich das Signal, das an den Ohren des Rezipienten ankommt. Menschen können Klänge ohne Zuhilfenahme visueller Eindrücke (Brown et al. 2015, S. 1) aus allen Richtungen hören und sind in der Lage diese auch zu lokalisieren (Hameed et al. 2004, S. 4).

3.2.1 Richtungswahrnehmung

Die menschlichen Ohren befinden sich auf beiden Seiten des Kopfes. Bei einer seitlichen Klangquelle ergeben sich daraus Unterschiede der Ankunftszeit und des Schalldrucks zwischen dem linken und dem rechten Ohr (Iida 2019, S. 29). Dieser Effekt erhöht sich mit der Lateralität der Schallquelle und wird als „interaural level difference“ ILD und „interaural time difference“ ITD bezeichnet. Sie spielen eine wesentliche Rolle bei der menschlichen Wahrnehmung der Schallrichtung (Iida 2019, v), liefern allerdings nicht den vertikalen Winkel der Schallquelle in der Sagittalebene.

In der folgenden Grafik ist zu erkennen, dass die interauralen Unterschiede an allen Punkten auf dem Kreis des gezeigten Kegels identisch sind. Die Information über die interauralen Differenzen gibt daher nur an, in welcher Sagittalebene sich eine Schallquelle befindet. Dies wird als “cone of confusion” bezeichnet (Iida 2019, S. 32). Auch für die Unterscheidung von hinten und vorne sowie oben und unten haben ILD und ITD keine Wirkung (Iida 2019, S. 33).

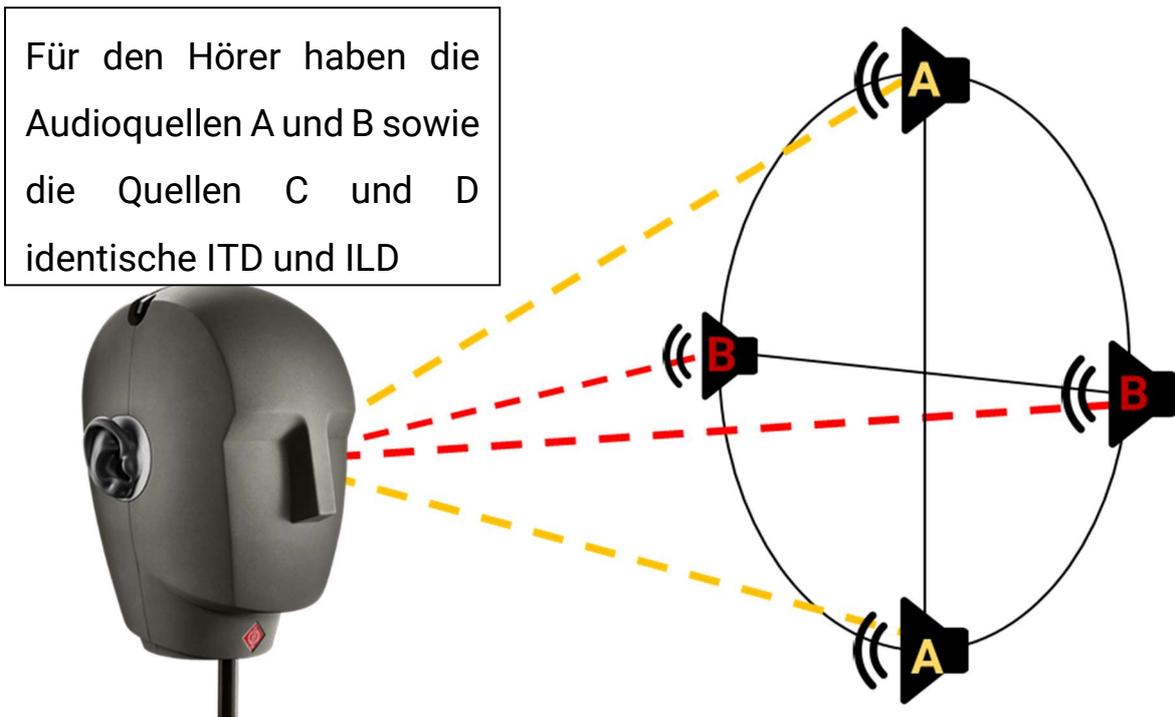


Abbildung 4 Cone of Confusion (Eigene Darstellung)

Sowohl der „Cone of Confusion“ als auch die Unterscheidung von vorne und hinten sowie oben und unten können durch Kopfbewegungen und Veränderungen im Frequenzspektrum durch die Interaktion des Schalls mit dem Kopf, der Ohrmuschel und des Rumpfes gelöst werden (Iida 2019, S. 51; R. Begault und M. Wenzel 2001, S. 913; Iida 2019, S. 13).

Doch auch Reflektionen haben eine Wirkung auf die Richtungswahrnehmung. Begault und Wenzel (2001) stellten einen positiven Effekt früher binauraler Reflektionen auf die Bestimmung des Azimut einer Klangquelle fest. Gleichzeitig wirken sie sich aber negativ auf die Höhenwahrnehmung aus (R. Begault und M. Wenzel 2001, S. 911).

3.2.2 Distanzwahrnehmung

Für die Distanzwahrnehmung spielen die durch den menschlichen Körper verursachten Änderungen nur eine geringe Rolle. Bei Distanzen von unterhalb von einem Meter ändern sich ILD und ITD zwar noch (Iida 2019, S. 136), sind aber nur dann signifikant, wenn die Quelle relativ weit seitlich verortet ist (Shinn-Cunningham 2001a, S. 2). Befindet sich die Klangquelle dagegen auf der Medianachse ist der ILD bei jeder Entfernung beinahe Null (Shinn-Cunningham 2001a, S. 2). Nach Shinn-Cunningham (2001) sind Interaurale Verschiebungen bei der Distanzwahrnehmung sogar komplett zu vernachlässigen (Shinn-Cunningham 2001a, S. 3–4).

Lautstärke

Da die Lautstärke eines Tons mit der Distanz abnimmt (Loomis, Klatzky, and Golledge, 1999), liegt die Vermutung nahe, dass sie einen Einfluss auf die Wahrnehmung der Distanz hat. Um Lautstärke zur Distanzbestimmung nutzen zu können, müssen Rezipienten allerdings das Verhältnis zwischen der Distanz der Quelle und der Lautstärke des Signals kennen (Iida 2019, S. 139, 2019, S. 137). Wenn der Hörer einen Ton vorher noch nie gehört hat und dementsprechend keine Erfahrungswerte für die Bestimmung der Lautstärke bei verschiedenen Distanzen besitzt, kann Lautstärke kein Hinweis auf die absolute Entfernung einer Klangquelle geben (Coleman 1962, S. 345–346). Für bekannte Umgebungsgeräusche, wie z.B. Schritte, einem Auto oder Flugzeug, sind die Lautstärken vertraut. Daher können sie eine grobe Abschätzung der Entfernung auf der Grundlage von Druckänderungen und erwarteten Druckpegeln liefern. Für Klangquellen, die zwar unbekannt sind, sich aber in ihrer Distanz ändern, können relative Distanzänderungen wahrgenommen werden (Philbeck und Mershon 2002, S. 1982).

Reflektionen

Nicht nur der Direktschall ändert sich mit der Position einer Klangquelle, auch der Hall in einem Raum ändert sich immer dann, wenn sich die Position der Klangquelle, des Rezipienten oder irgendeines Objekts im Raum ändert (O. Smith 2010, The Reverberation Problem). Infolgedessen hängen die zur Bestimmung der Quellenposition verwendeten akustischen Hinweise nicht nur von der Position der Quelle relativ zum Zuhörer ab, sondern auch von der akustischen Umgebung sowie von der Position und Ausrichtung von Quelle und Zuhörer im Raum (Shinn-Cunningham 2001b, S. 1).

Durch die Zunahme von Hall und der daraus resultierenden höheren Schallenergie, werden die Lautstärkenunterschiede zwischen dem rechten und dem linken Ohr verringert, was dazu führt, dass auch das ILD verringert wird (Shinn-Cunningham 2000, S. 8). Diese Verringerung ist in normalgroßen Räumen aber nur gering (Shinn-Cunningham 2003, S. 7). Es konnte sogar festgestellt werden, dass binaurale frühe Reflektionen einen positiven Effekt auf die Wahrnehmung des Azimuth haben (R. Begault und M. Wenzel 2001, S. 909), wenn die frühen Reflektionen aus Richtung der Klangquelle kommen (Hartmann 1983, S. 1388). Solange es sich um moderate Halllautstärken handelt, verbessern Raumreflektionen zusätzlich die Fähigkeit der Rezipienten, die Distanz einer Klangquelle zu bestimmen (Shinn-Cunningham 2003, S. 7). Gotoh et al. (1977) fand heraus, dass Menschen eine größere Distanz zwischen sich und einer Klangquelle wahrnehmen, wenn die Zeit zwischen dem Direktschall und den ersten Reflektionen größer ist. Die Distanzwahrnehmung nimmt weiter zu, wenn die folgenden Echos ebenfalls einen größeren zeitlichen Abstand haben (Gotoh et al. 1977, S. 670–671). Mendonça et al. (2013) fand heraus, dass mindestens drei Reflektionen nötig

sind, um einen positiven Einfluss auf die Distanzwahrnehmung zu erzeugen (Mendonça et al. 2013, S. 6).

Der Pegel des Direktschalls nimmt im Verhältnis zum Hall mit der Entfernung der Klangquelle gleichbleibend ab (Shinn-Cunningham 2001a, S. 2). Bei nahen Quellen unterhalb von einem Meter Entfernung vom Kopf variiert das Verhältnis von Hall zum Direktschall (D/R) nicht mehr linear (Shinn-Cunningham 2001a, S. 2). Es wurde lange angenommen, dass D/R hauptsächlich für die Distanzwahrnehmung verantwortlich ist (Mendonça et al. 2013, S. 2). Diese These wurde aber in späteren Experimenten in Frage gestellt (Shinn-Cunningham 2000, S. 7–8).

Auch Informationen über die eigene Position sind in den Raumreflektionen vorhanden. Es konnte gezeigt werden, dass Hörer unterscheiden können, ob sie sich dicht an einer Wand befinden. Diese Fähigkeit nimmt mit der Distanz zur Klangquelle zu. Dies lässt darauf schließen, dass das Vorhandensein sehr früher, starker Echos (wie sie auftreten, wenn ein Ohr der Wand zugewandt und die Quelle relativ weit entfernt ist) einen positiven Effekt haben (Shinn-Cunningham 2003, S. 5). Zusätzlich konnte festgestellt werden, dass sich bei einer mittigen Position des Hörers im Raum die Distanzwahrnehmungsleistung verbessert (Shinn-Cunningham 2003, S. 4).

Auch bei der Lokalisierung in Verbindung mit Reflektionen gewöhnen sich die Rezipienten an die Räume, in denen sie sich befinden. Die Wahrnehmungsleistung verbessert sich mit der Zeit (Shinn-Cunningham 2003, S. 4), selbst wenn die Probanden keine Rückmeldung über die Richtigkeit ihrer Vermutungen erhalten. Menschen passen ihre Wahrnehmung also automatisch an ihre Umgebung an (Shinn-Cunningham 2003, S. 7) und das sogar dann, wenn sich ihr eigener Standort ändert (Shinn-Cunningham 2003, S. 3).

Zusammen mit den besprochenen Lautstärkenunterschieden bietet Hall ausreichende Informationen um die Distanz einer Quelle zu bestimmen (Shinn-Cunningham 2001a, S. 4). Im Vergleich zwischen monauralem und binauralem Hall wurde eine im Wesentlichen gleichwertige Leistung in der Distanzwahrnehmung festgestellt. Diese Ergebnisse deuten darauf hin, dass ILD-Signale (die unter monauralen Hörbedingungen nicht verfügbar sind) nicht zur Wahrnehmung der Quellenentfernung beitragen, wenn Nachhall vorhanden ist (Shinn-Cunningham 2001a, S. 4).

3.3 Zusammenfassung

Alle Studien, die Nachhallzeit mit anderen Faktoren verglichen haben, kommen zu dem Ergebnis, dass Nachhallzeit den größten Einfluss auf die Wahrnehmung der Raumgröße darstellt. D/R hat zwar einen Einfluss, dieser führte allerdings bei verschiedenen Testpersonen teilweise sogar zu gegensätzlich Wahrnehmungen (Hameed et al. 2004, S. 5). D/R sollte daher für die Simulation der Raumgröße nicht verwendet werden. Clarity Index hat einen messbaren Einfluss auf die Wahrnehmung der Raumgröße, dieser ist jedoch vor allem dann wichtig, wenn Nachhallzeit durch stark absorbierende Oberflächen eines Raumes nicht genutzt werden kann (Cabreraa et al. 2005, S. 4). Im nahen Bereich bis etwa 5m ist es sinnvoll, die Klangquelle weiter entfernt zu platzieren, um die Raumgrößenwahrnehmung zu erhöhen (Cabreraa et al. 2005, S. 3). Doch auch hier hat die Nachhallzeit die größte Auswirkung (Hameed et al. 2004, S. 7). Cabreraa et al. (2005) stellt sogar fest, dass die Nachhallzeit einen größeren Effekt auf die Wahrnehmung der Raumgröße hat, als die reale Raumgröße (Cabreraa et al. 2005, S. 1). Die akustischen Charakteristiken eines Raumes machen es daher möglich, diesen größer oder kleiner klingen zu lassen, als er in Wirklichkeit ist (Cabreraa et al. 2005, S. 9). Die genauen Eigenschaften der frühen Reflektionen, wie Zeitintervalle und

Richtung zwischen den einzelnen Echos können zwar nicht differenziert wahrgenommen werden (Shinn-Cunningham 2003, S. 7), die grobe Richtung, aus der sie kommen, hat aber einen positiven Einfluss auf die Ortung von Klangquellen (Hartmann 1983, S. 1388). Wenn sie auch nicht die Wichtigkeit haben, wie lange angenommen, enthält D/R trotzdem viele Informationen über die Distanz von Klangquellen (Mendonça et al. 2013, S. 2). Dies kann wiederum passiv die wahrgenommene Raumgröße beeinflussen, da diese natürlich mit der Quelldistanz korreliert (Cabreraa et al. 2005, S. 3). Die Wahrnehmung der Distanz der Klangquelle ist im Wesentlichen von den Reflektionen und der Lautstärke des Direktschalls abhängig. Bei bekannten Geräuschen wie Sprache kann die absolute Distanz durch die Lautstärke wahrgenommen werden (Iida 2019, S. 139, 2019, S. 137). Bei Quellen, von denen kein vorheriges Lautstärkedistanzverhältnis bekannt ist, ist es bei sich bewegenden Quellen dennoch möglich durch die Lautstärkenveränderungen die Distanzänderungen einzuordnen (Philbeck und Mershon 2002, S. 1982). Zusammenfassend ist die Lautstärke zusammen mit den Raumreflektionen die wichtigste Quelle zur Einordnung von Distanz. Welche Eigenschaften des Halls aber genau für die Distanzwahrnehmung verantwortlich sind, ist nicht abschließend geklärt.

Die Richtungswahrnehmung erfolgt zu einem großen Teil durch interaurale Differenzen in der Zeit und der Lautstärke (Iida 2019, v). Diese kommen im „Cone of Confusion“ und bei der Unterscheidung von vorne-hinten und oben-unten an ihre Grenzen und können teilweise durch Kopfbewegungen des Hörers ausgeglichen werden (Iida 2019, S. 51; R. Begault und M. Wenzel 2001, S. 913; Iida 2019, S. 13).

4 Recording und Repräsentation

Für die auditive Abbildung von realen Räumen, ist die Wiedergabe ein zentraler Bestandteil. In diesem Kapitel gehe ich auf die verschiedenen Ausgabeformate ein und erläutere mit welchen Techniken diese Inhalte produziert werden können.

Die „International Telecommunication Union“ und die „European Broadcast Union“ (EBU) erkennen und unterstützen drei verschiedene immersive Audiokonzepte, die als kanalbasiertes, objektbasiertes und szenenbasiertes Audio bezeichnet werden. Jedes dieser Konzepte bietet unterschiedliche Merkmale für die Ersteller der auditiven Inhalte und auch für die Konsumenten. Sie unterscheiden sich grundlegend in der Form, wie die Inhalte übertragen und dann dekodiert und verarbeitet werden (Peters et al. 2016, S. 16).

4.1 Kanalbasiertes Audio / Channel Based Audio

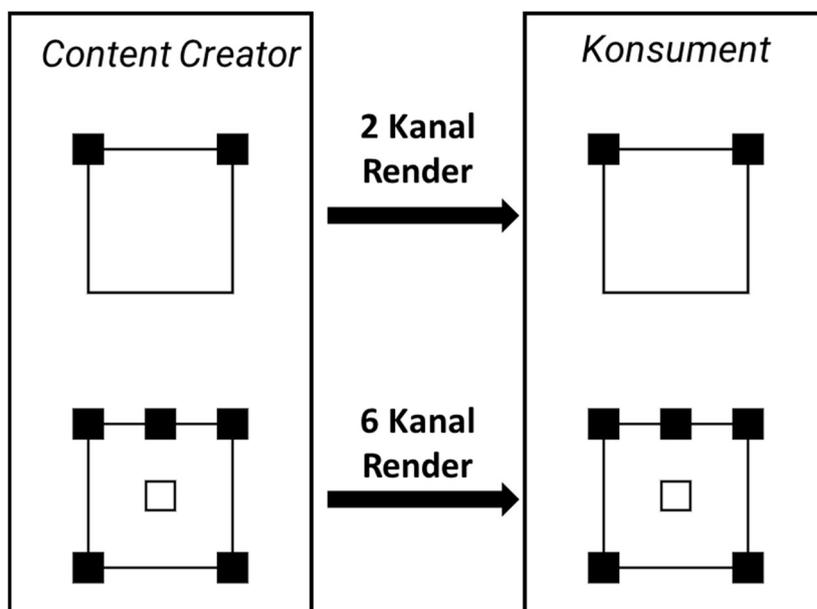


Abbildung 5 Kanalbasiertes Audio (Eigene Darstellung)

Die kanalbasierte Wiedergabe war mit Stereo 2.0 und Surround 5.1 sehr lange der Standard in der Broadcast Industrie (Peters et al. 2016, S. 16). Für die

kanalbasierte Wiedergabe wird für eine bestimmte Anzahl an Lautsprechern gemischt. Die Lautsprecher müssen auf eine vorgegebene Art aufgestellt werden, da die Inhalte bei der Ausgabe für jeden Kanal eingebunden werden. Der Vorteil dabei ist, dass am Decoder beim Konsumenten kein komplexes Rendering Equipment benötigt wird (Peters et al. 2016, S. 16). Nachteilig ist dabei, dass kanalbasierte Systeme darauf angewiesen sind, dass am Ort der Wiedergabe die Anzahl und Position der Lautsprecher dem Soundsystem gleichen, mit dem gemischt wurde, um das beabsichtigte Klangbild zu erhalten (Qualcomm Technologies, Inc. 2015, The story so far – and why it can get better; Peters et al. 2016, S. 16). Diese Voraussetzung ist aber häufig nicht gegeben und es kommt vor, dass Inhalte, die für ein 5.1 System produziert wurden, z.B. auf Kopfhörern gehört werden. (Olivieri et al. 2019, S. 4). Infolgedessen wird das Klangbild und damit auch der simulierte Raumeindruck verfälscht (Peters et al. 2016, S. 16).

4.2 Objektbasiertes Audio / Object Based Audio

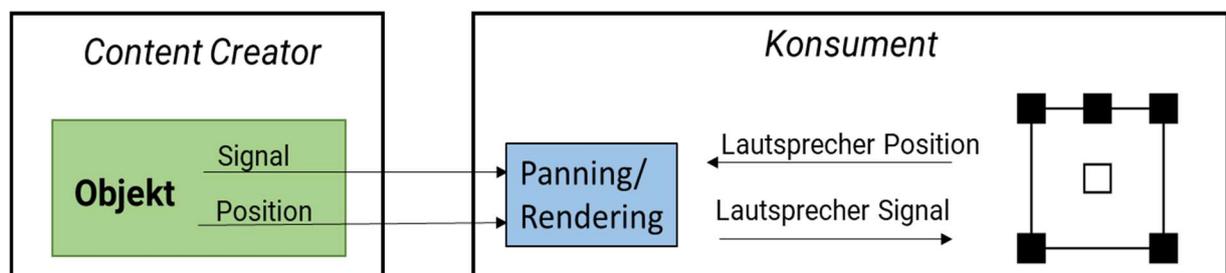


Abbildung 6 Objektbasiertes Audio (Eigene Darstellung)

Im Gegensatz zu kanalbasiertem Audio, werden bei objektbasierten Formaten einzelne Spuren in Form von Audioobjekten ausgegeben, die keinem bestimmten Kanal (Lautsprecher) zugeordnet sind. Diese Audioobjekte enthalten Metadaten mit Positionsinformationen, durch die sie dann in Relation zum Lautsprechersystem beim Konsumenten gerendert werden (Peters et al. 2016, S. 17). Der Renderer hat Informationen über die Anzahl und Positionen der Lautsprecher und kann ihnen die Audioobjekte mit Hilfe der Metadaten

zuweisen (Qualcomm Technologies, Inc. 2015, The story so far – and why it can get better). Mit objektbasiertem Audio ist es möglich jede Lautsprecherkonfiguration zu berücksichtigen. Hierzu zählen nicht nur standardisierte Lautsprecher Setups sondern auch alle früheren oder zukünftigen (Peters et al. 2016, S. 17).

Bei den Audioobjekten muss es sich nicht zwangsläufig um Punktquellen handeln. Sie können auch in ihrer Diffusität variieren und so einen größeren Bereich in der virtuellen Klangsphäre einnehmen (Atmos in studio | Dolby Developer 2020). So wäre es theoretisch möglich, eine Szene ausschließlich mit Audioobjekten zu rendern. Da aber mit jedem Objekt die benötigte Rechenleistung und Übertragungsbandbreite steigt (Peters et al. 2016, S. 17), ist es praktisch unmöglich komplexe Szenen (z.B.: Klatschen in einem Stadion oder Konzertsaal) rein objektbasiert wiederzugeben (Qualcomm Technologies, Inc. 2015, The story so far – and why it can get better). In diesen Fällen muss auch bei objektbasierten Formaten auf kanalbasiertes Audio zurückgegriffen werden.

4.3 Szenenbasiertes Audio / Scene Based Audio

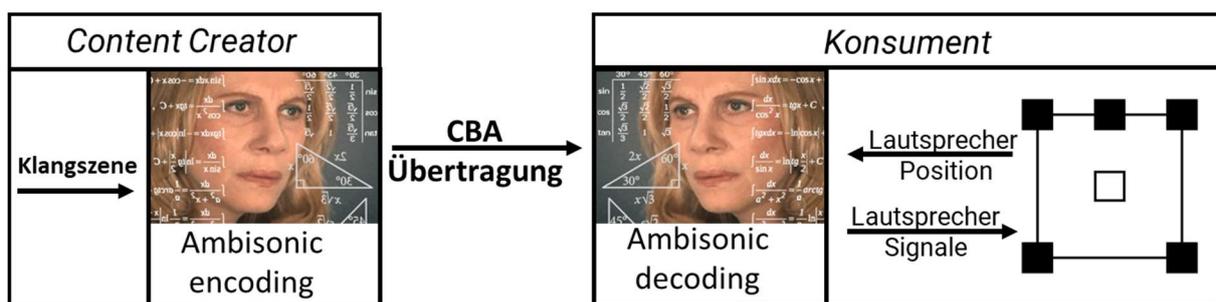


Abbildung 7 Szenenbasiertes Audio (eigene Darstellung)

Szenenbasiertes Audio gründet auf der Klangfeldtechnologie „Higher-Order-Ambisonics“ (Olivieri et al. 2019, S. 3). Mit dieser Technologie kann eine ganzheitliche Beschreibung einer live aufgenommenen oder künstlich erstellten Audioszene wie bei objektbasierten Formaten unabhängig vom

Lautsprecher Setup dargestellt werden (Peters et al. 2016, S. 16). SBA und HOA sind sehr eng miteinander verbunden. Während man sich SBA als Technologie für dreidimensionalen Klang vorstellen kann, ist HOA das zugrundeliegende Format, das die SBA Technologien möglich macht (Olivieri et al. 2019, S. 5).

HOA basiert auf Ambisonics und ist eine Methode, ein Klangfeld unter Berücksichtigung seiner direktionalen Eigenschaften zu kodieren (Arteaga 2018, S. 4). Es wird ein dreidimensionales Schallfeld auf der Oberfläche einer Kugel dargestellt (Olivieri et al. 2019, S. 5), das durch Berechnungen aus den verfügbaren Kanälen bestimmt wird. Jeder Kanal besitzt Informationen zum zeitlichen Verlauf des Schalldrucks der Klänge auf dem sphärischem Klangfeld (Arteaga 2018, S. 4). Die Anzahl der Kanäle bestimmt dabei die Genauigkeit der Abbildung dieser Eigenschaften im Klangfeld (Arteaga 2018, S. 22).

Für eine Aufnahme von Ambisonics erster Ordnung, benötigt man mindestens ein Array an vier Mikrofonen - ein omnidirektionales und drei bidirektionale. Mit dem omnidirektionalen Mikrofon erhält man Informationen über das gesamte Schalldruckfeld an der Aufnahmeposition. Die drei bidirektionalen Mikrofone müssen jede der Raumachsen abdecken (Arteaga 2018, S. 4). Durch ihre bidirektionale Richtcharakteristik nehmen diese Mikrofone nur Töne auf, die vor oder hinter ihnen positioniert sind. Wenn die drei Mikrofone richtig ausgerichtet sind, ist es möglich alle Quellen im dreidimensionalen Raum zu bestimmen (Garner 2018, S. 274–275).

Einer der größten Nachteile an Ambisonics erster Ordnung ist die ungenaue Richtwirkung (Arteaga 2018, S. 19). Bei höherer Ordnung, also durch das Hinzufügen weiterer bidirektionalen Mikrofonen in das Array, erhöht sich die

Auflösung und damit die Genauigkeit, mit der Klangquellen bestimmt werden können (Arteaga 2018, S. 22). Um szenenbasiertes Audio aufzunehmen ist es zusätzlich möglich, zu dem Ambisonics Mikrofonarray weitere individuelle Mikrofone zu nutzen, die an verschiedenen Stellen im Raum positioniert werden (Peters et al. 2016, S. 17). Das gewonnene Material wird dann in Form so genannter HOA Koeffizienten-Signale weiterverarbeitet, die alle Geräusche zusammen mit ihrer zeitlichen Änderung in der Klangszene enthalten (Peters et al. 2016, S. 17) und somit eine mathematische Repräsentation der Szene darstellen (Qualcomm Technologies, Inc. 2015, Break free from shackles).



Abbildung 8 HOA Mikrofon von Eigenmike (Products | mhacoustics.com 2020)

Genauso wie das Erstellen einer Szene durch reines Aufnehmen, ist es möglich, Audioszenen synthetisch aus bestehenden Audiokomponenten in einer DAW zu erstellen. Bei diesem Verfahren werden Audioelemente an der gewünschte Stelle platziert (Peters et al. 2016, S. 17). Anschließend wird die Szene in HOA-Koeffizienten umgewandelt, wie es auch bei einer Live-Aufnahme der Fall ist (Qualcomm Technologies, Inc. 2015, The big picture; Arteaga 2018, S. 6). Die berechneten HOA Koeffizienten können auch addiert werden. Hierdurch ist es möglich, kodierte Material miteinander zu mischen. (Qualcomm Technologies, Inc. 2015, The big picture). Beispielsweise ist die Kombination von Live-Aufnahmen mit synthetisch erstellten Szenen möglich.

Durch das Arbeiten mit Ambisonics, entstehen auch einige neue Möglichkeiten in der Bearbeitung. Da HOA-Koeffizienten auf Kugelflächenfunktionen basieren, ist es mathematisch einfach, die Klangszene zu drehen oder zu strecken, beziehungsweise zu komprimieren oder zu spiegeln (Frank et al. 2015, S. 4–5) (Zotter und Kronlacher 2014, S. 4). Es ist ebenfalls möglich, bestimmte Stellen auf der Klangsphäre zu fokussieren (Arteaga 2018, S. 12) und dadurch z.B. zu entfernen (Peters et al. 2016, S. 17). Auch ist mit Ambisonics ein weiches Panning möglich, ohne dass der Klang von einem Kanal zum nächsten „springt“ (Arteaga 2018, S. 24, 2018, S. 12). Dies ist ein großer Vorteil bei der Reproduktion in interaktive Medien oder bei der Kopfhörerwiedergabe mit Headtracking. Gerade Ambisonics erster Ordnung ist innerhalb der Onlinetechnologie viel vertreten. Es wird von Google/YouTube, Facebook 360° unterstützt und wird auch in Spielen häufig verwendet (Zotter und Frank 2019, S. 8).

Ambisonics ist ein hierarchisches Format. Dies ermöglicht die Entfernung von Koeffizienten einer höheren Ordnung, ohne dabei Audio Elemente zu verlieren, da sich mit höherer Ordnung nur die Auflösung erhöht (Peters et al. 2016, S. 18). Damit ist es auch bei szenenbasierte Formaten möglich, dieses auf jedem Lautsprecher Setup abzuspielen, ohne verschiedene Versionen rendern zu müssen (Peters et al. 2016, S. 17; Arteaga 2018, S. 13) . Bei der Wiedergabe von szenenbasiertem Audio auf Lautsprechersystemen spielt jeder Lautsprecher das Signal eines virtuellen Mikrofons ab, das in die Richtung des jeweiligen Lautsprechers zeigt (Arteaga 2018, S. 22).

Im Kern ist SBA kanalbasiert, weshalb die Komplexität des Renderings und der Übertragung nicht mit der Komplexität der Szene zunimmt, wie es bei objektbasierten Formaten der Fall ist (Peters et al. 2016, S. 17). Gerade in komplexen Kinomischungen kann das einen großen Unterschied machen

(Arteaga 2018, S. 5). Trotzdem haben die Kanäle bei Ambisonics nichts mit den Kanälen kanalbasierter Formate zu tun, da sie, auf den physikalischen Prinzipien akustischer Felder basieren (Arteaga 2018, S. 24).

4.4 Besonderheiten bei der Wiedergabe mit Kopfhörern

Die Wiedergabe auf Kopfhörern ist zwar kein Audiokonzept wie CBA, OBA oder SBA, hat aber einige Besonderheiten. Da Kopfhörer sich direkt am oder im Ohr befinden, gibt es keine durch den Körper des Hörers bedingten spektralen Klangveränderungen, keine binaurale Laufzeitverschiebung oder Lautstärkenunterschiede und auch Kopfbewegungen haben keinen Effekt. Um diese Effekte zu simulieren kann zum Beispiel direkt mit einem Dummy Head aufgenommen werden, in dem Mikrofone in den Ohren verbaut sind (Iida 2019, S. 79). Hierdurch werden ILD, ITD und die spektralen Veränderungen eingebacken und es gibt keine Möglichkeit mehr, die Richtung nachträglich zu ändern.



Abbildung 9 Dummy Head der Firma NEUMANN (NEUMANN 2020)

Zur nachträglichen Simulation der Spektralunterschiede sowie ITD und ILD können individuelle oder generische Head-Related-Transfer-Functions (HRTF) erstellt und nachträglich auf Audiomaterial angewendet werden. Diese können dann entweder beim Dekodieren auf der Konsumentenseite hinzugerechnet, oder direkt mit eingebacken werden, was natürlich die Nutzung individueller HRTFs ausschließt.

Ein solches HRTF kann aus einer Head-Related-Impulse-Response (HRIR) berechnet werden. Zum Erzeugen einer individuellen HRIR werden Mikrofone in den Ohren des Probanden genutzt, um mit diesen in einem schalltoten Raum einen breitbandigen Impuls aufzunehmen und daraus die Änderungen in

Laufzeit, Lautstärke und Frequenz zu erzeugen (Iida 2019, S. 21, 2019, S. 154). Diese Form der Erstellung ist sehr aufwändig und daher für die meisten Konsumenten ungeeignet. Inzwischen gibt es aber beispielsweise auch Möglichkeiten, eine Impulse Response aus Bildern des Kopfes und des Ohrs zu erzeugen (THX 2020). R. Begault und M. Wenzel (2001) fanden im Vergleich mit generischen HRTFs keine klaren Vorteile zur Nutzung individueller HRTFs. Obwohl man eine starke Korrelation zwischen dem Azimutfehler und der Größe der ITD-Differenz zwischen individuellen und generischen HRTFs für jedes Subjekt erwarten könnte, zeigte die Analyse der Daten keine Korrelation zwischen dem Azimutfehler und der Kopfgrößendifferenz. (R. Begault und M. Wenzel 2001, S. 912). Da aber einige Informationen über Kopfbewegungen wahrgenommen werden, können diese über Headtracking simuliert werden (R. Begault und M. Wenzel 2001, S. 915).

Ein weiteres Problem bei der Nutzung von Kopfhörern besteht darin, dass Klänge wirken, als würden sie sich innerhalb des Kopfes befinden. Um dies zu umgehen, können binaurale frühe Reflektionen genutzt werden, was nach R. Begault (2001) zu einer erhöhten Genauigkeit in der Wahrnehmung geführt hat (R. Begault und M. Wenzel 2001, S. 915).

4.5 Zusammenfassung

Obwohl alle Formate ihre Vor- und Nachteile haben, ist SBA das flexibelste und kann auch ohne viel Aufwand auf CBA und OBA übertragen werden. Alle genannten Formate können im MPEG-H Audio Standard kombiniert und übertragen werden (Olivieri et al. 2019, S. 4; Peters et al. 2016, S. 17).

5 Hall-Effekte

Die Simulation von Raumhall ist ein technisch komplexer Prozess, der so viel Rechenleistung benötigt, dass Halleffekte zur Nutzung in Echtzeit vereinfacht werden müssen (O. Smith 2010, COMPLEXITY OF EXACT REVERBERATION).

5.1 Faltungshall / Convolution Reverb

Die Faltung ist eine digitale Signalverarbeitungstechnik. Es handelt sich im Wesentlichen um eine mathematische Methode, zwei Signale zu einem dritten Signal zu kombinieren. Die beiden Signale, die gefaltet werden, sind das Eingangssignal und die Impulsantwort. Die Impulsantwort enthält Information darüber, wie ein Impuls in einem bestimmten System oder Raum abklingt. Durch Faltung mit einem Eingangssignal erhält man ein Ausgangssignal, das angibt, wie das Eingangssignal in diesem spezifischen System oder Raum abgeklungen wäre (Shriram 2011, S. 14). Diese Technik macht sich Faltungshall zu Nutze, um aus aufgenommenen oder simulierten Räumen eine Impulsantwort zu generieren, die anschließend mit dem Originalsignal gefaltet wird.

Um eine solche Impulsantwort aus einem realen Raum zu erstellen, muss ein breitbandiger Impuls aufgenommen werden. Aus den so gewonnenen Informationen lässt sich dann eine Impulsantwort ableiten und anschließend auf jedes Audiosignal anwenden (Iida 2019, S. 160). Eine Impulsantwort kann man sich als detailliertes Abbild der Raumreflektionen vorstellen. Es gibt dabei eine bestimmte Position der Klangquelle und eine bestimmte Position des Empfängers (Pop und Cabrera 2005, S. 115). Im Falle einer Aufnahme in einem realen Raum (Misic et al. 2016, S. 123), handelt es sich bei der Klangquelle um eine physikalische breitbandige Quelle, wie zum Beispiel einer Pistole oder auch einem Lautsprechers. Das Signal wird danach von

einem omnidirektionalen Mikrofonen aus der zu simulierenden Rezipientenposition aufgenommen (Hameed et al. 2004, S. 2). Aus diesem Grund ist es mit dieser Technik weder möglich ist, eine sich bewegende Quelle noch einen sich bewegenden Rezipienten zu simulieren. Um die Limitierung kanalbasierter Aufnahmen in Bezug auf das Ausgangsformat zu umgehen und dadurch eine möglichst vielseitig nutzbare Impulsantwort zu generieren, wäre es hier möglich, mit einem HOA Ambisonics Mikrofon-Array aufzunehmen.

5.2 Algorithmische Halleffekte

Reflektionen können aber auch mit einem rein synthetischen Delay Netzwerk simuliert werden. Die traditionelle Herangehensweise bei einem echtzeitfähigem synthetischen Hall basiert auf Feedback-Delay-Lines, dessen Ein- und Ausgänge sukzessiv miteinander verbunden sind (Jot und Chaigne 1991, S. 7; Jot 1997, S. 2). Auf diese Weise ist es möglich, die frühen Reflektionen und anschließend die späten zu rendern (Jot 1997, S. 1) und diese mit der Zeit abklingen zu lassen (Seetharaman und Pardo 2014, S. 739). Der Vorteil gegenüber aufgenommenen Impulsantworten besteht in der Möglichkeit, mit algorithmischen Halleffekten Quell- und Rezipientenposition anpassen zu können. Der Nachteil ist, dass es schwieriger ist, einen Raum akkurat zu simulieren. Denn es stehen anstatt physikalischer Größen wie Raumgröße oder Material lediglich abstrakte Werte wie zum Beispiel die Nachhallzeit zur Verfügung (Rungta et al. 2019, S. 1).

5.3 Gemodelte Räume

Um dieses Problem zu umgehen, kann auch mit gemodelten Räumen gearbeitet werden. Es wird also Raumgröße, Form und Material definiert und dadurch die Raumreflektionen berechnet. Viele dieser sogenannten „geometric sound propagation methods“ nutzen „Raytracing“. Hier wird die Annahme zu Grunde gelegt, dass sich Ton in geraden Linien ausbreitet. Mit diesen Rays, werden

dann Reflektionen und Streuungen berechnet. Durch Raytracing ist es inzwischen möglich, den kompletten Hall interaktiv zu berechnen (Schissler und Manocha 2017, S. 2; Rungta et al. 2019, S. 1). Da beim Raytracing aber von Strahlen anstatt von Wellen ausgegangen wird, bildet es die Wellen aus denen hörbare Wellenlängen bestehen, nicht genau ab. Was bei Objekten, die kleiner als die auftreffende Wellenlänge sind, zu physikalischen Ungenauigkeiten führt (Krokstad et al. 1968, S. 123, 1968, S. 122). Die genauesten Methoden zur Klangwiedergabe basieren daher auf der wellenbasierten Akustik. Ihr Rechen- und Speicheraufwand ist jedoch sehr hoch (Schissler und Manocha 2017, S. 1), weshalb in der Praxis auf Grund der Einfachheit in der Berechnung meist auf Raytracing zurückgegriffen wird (Rungta et al. 2019, S. 2).

6 Fazit

Möchte man Richtung und Distanz von Klangquellen in einem schalltoten Raum simulieren, würde es ausreichen diese zu pannen, um so Lautstärkenunterschiede zwischen dem linken und rechten Ohr zu simulieren. Zusätzlich muss die Gesamtlautstärke für die Distanzänderungen automatisiert werden. Kommt die Simulation eines Raums hinzu, wird eine Simulation wesentlich anspruchsvoller. Der physikalische Hintergrund von Raumklang ist sehr komplex, kann aber auf Grund der begrenzten Wahrnehmungsfähigkeit des Menschen so weit vereinfacht werden, dass eine Simulation in Echtzeit möglich ist. Die größtmögliche Flexibilität bei der Simulation von Raumreflektionen ist mit algorithmischen Halleffekten gegeben, da es so möglich ist, die Reflektionen bei Änderung der Rezipienten- und Quellposition anzupassen. Da der Nachhall aber keine Wirkung auf die Richtungswahrnehmung hat, kann dieser auch unabhängig von der Position generiert werden (Pop und Cabrera 2005, S. 116). Die Richtung der einzelnen Echos kann vom Menschen nicht unterschieden werden (Shinn-Cunningham 2003, S. 7), weshalb es ausreichend ist, eine vollständige Impulsantwort in frühe Reflektionen und Nachhall zu trennen und die frühen Reflektionen mit der Klangquelle zu pannen. Die Richtung der späten Reflektionen kann zwar für alle Klangquellen gleich sein, sollte aber, genauso wie die frühen Reflektionen, in ihrer Lautstärke angepasst werden können. Dies sollte sowohl unabhängig voneinander als auch gleichzeitig bei beiden möglich sein, um Clarity Index und D/R anpassen zu können.

Egal ob es sich um aufgenommene oder gemodelte Impulsantworten handelt, ist es immer am sinnvollsten in HOA zu arbeiten, da man so die größte Flexibilität in der Abbildung erreicht.

In der Postproduktion ist die klassische Methode eine Klangquelle in einem virtuellen Raum zu positionieren, die Anpassung aller Parameter - einzeln und unabhängig voneinander. Beispielsweise wird das Panning, die Diffusität, die Lautstärke und der Hall einer Quelle automatisiert. Eine völlig andere Herangehensweise ist es, den kompletten Workflow, der genutzt wird, um Klangquellen im Raum zu positionieren, nicht mehr von einzelnen Parametern abhängig zu machen, sondern diese durch Werte aus der echten Welt zu ersetzen. Anstatt in der Mischung mit Panning, D/R, Predelay, Highpass filtering zu arbeiten, würde man Raumdimensionen, Form und Material definieren und daraus die Änderungen im Schall berechnen, die sich durch Änderungen der Quell- und Rezipientenposition ergeben. Ein Beispiel für eine solche Software ist „Dspatial Reality“ oder „Spat Revolution“.

Im letzten Schritt kommt das gemischte Signal beim Rezipienten an. Hört dieser über ein Lautsprechersystem ab, ergeben sich die interauralen Zeitverschiebungen und alle spektralen Veränderungen, die durch den Körper hervorgerufen werden von selbst. Wird mit Kopfhörern abgehört, kann man die Interaktion des Schalls durch den Körper des Rezipienten zusätzlich mit einer HRTF simulieren und Kopfbewegungen mit einem Headtracker übertragen. Bestenfalls sollte hierbei in HOA gearbeitet werden, um das bestmögliche Ergebnis des durch Headtracking hervorgerufenen panning zu erzeugen (Arteaga 2018, S. 24). Das Ausgabeformat macht in jedem Fall einen großen Unterschied in der Raumsimulation. Vor allem beim Erstellen der Mischung sollte möglichst mit HOA gearbeitet werden. So kann für jedes denkbare Ausgabeformat eine passende Mischung generiert werden (Arteaga 2018, S. 24). Aber auch bei der Wiedergabe beim Konsumenten macht dieses Format Sinn, da hierdurch vermieden wird, dass bei einer falschen Aufstellung der Lautsprecher das gewünschte Klangbild verfälscht wird. Auch bei der Wiedergabe von 360° Videos oder auch für Spiele, die es erlauben, sich völlig

frei zu bewegen, muss das Panning so weich wie möglich funktionieren, weshalb auch hier Ambisonics die beste Wahl ist.

7 Diskussion

Ich bin immer auf der Suche nach neuen Methoden meinen Workflow in der Ton-Postproduktion zu vereinfachen und zu verbessern und teste daher oft neue Tools. Die Arbeit mit „Spat Revolution“ ist im Filmbereich zumindest bei komplexen Mischungen noch sehr kompliziert und aufwändig, da es keine gute Integration zwischen DAW und „Spat Revolution“ gibt. Außerdem sind die Anforderungen, die bei einer Mischung bestehen meist kein völliger Realismus und unterliegen auch einem großen Maß an künstlerischer Freiheit. Um eine Geschichte zu erzählen, ist es in vielen Fällen nötig die Grenzen der Physik zu missachten. „Spatial Reality“ konnte ich selbst nicht testen, es hat aber den Anspruch sich gerade im Filmbereich in die gewohnte Arbeitsumgebung zu integrieren. Auch werden hier gemodelte Räume genutzt, während bei „Spat Revolution“ mit nicht physikalischen Parametern gearbeitet wird.

Nach den Erkenntnissen, die ich über die Nutzung von Kopfhörern und die Möglichkeiten der Simulation von klassischen Mehrkanal Abhörsystemen gesammelt habe, werde ich in Zukunft testen in wie weit es möglich ist mit Hilfe von HRTFs und Headtrackern in der Mischung die gleichen Ergebnisse zu erzielen.

8 Literaturverzeichnis

Abdullah, R.; Ismail, S. I.; Nik Dzulkefli, N. N.S. (2020): Potential acoustic treatment analysis using sabine formula in unoccupied classroom. In: *J. Phys.: Conf. Ser.* 1529, S. 22031. DOI: 10.1088/1742-6596/1529/2/022031.

amroc - THE Room Mode Calculator (2020). Online verfügbar unter <https://amcoustics.com/tools/amroc?l=561&w=430&h=300&r60=0.6>, zuletzt aktualisiert am 24.08.2020, zuletzt geprüft am 24.08.2020.

Aramaki, Mitsuko; Besson, Mireille; Kronland-Martinet, Richard; Ystad, Sølvi (2011): Controlling the Perceived Material in an Impact Sound Synthesizer. In: *IEEE Trans. Audio Speech Lang. Process.* 19 (2), S. 301–314. DOI: 10.1109/TASL.2010.2047755.

Arteaga, Daniel (2018): Introduction to Ambisonics.

Atmos in studio | Dolby Developer (2020). Online verfügbar unter <https://developer.dolby.com/tools-media/studio-resources/Atmos-in-studio/>, zuletzt aktualisiert am 05.08.2020, zuletzt geprüft am 05.08.2020.

Brown, Andrew D.; Stecker, G. Christopher; Tollin, Daniel J. (2015): The precedence effect in sound localization. In: *Journal of the Association for Research in Otolaryngology : JARO* 16 (1), S. 1–28. DOI: 10.1007/s10162-014-0496-2.

Cabreraa, Densil; Jeongb, Daeup; Jeong Kwakc, Hyun; Kimd, Ji-Young (2005): Auditory Room Size Perception for Modeled and Measured Rooms.

Coleman, Paul D. (1962): Failure to Localize the Source Distance of an Unfamiliar Sound. In: *The Journal of the Acoustical Society of America* 34 (3), S. 345–346. DOI: 10.1121/1.1928121.

Frank, Matthias; Zotter, Franz; Sontacchi, Alois (2015): Producing 3D Audio in Ambisonics.

Garner, Tom A. (2018): *Echoes of Other Worlds: Sound in Virtual Reality*. Cham: Springer International Publishing.

Gotoh, Toshiyuki; Kimura, Yoichi; Kurahashi, Akira (1977): A Consideration of Distance Perception in Binaural Hearing. In: *The Journal of the Acoustical Society of Japan* (33).

Hameed, Sharaf; Pakarinen, Jyri; Valde, Kari; Pulkki, Ville (2004): Psychoacoustic Cues in Room Size Perception.

Iida, Kazuhiro (2019): Head-Related Transfer Function and Acoustic Virtual Reality. Singapore: Springer Singapore.

Jot, Jean-Marc (1997): Efficient models for reverberation and distance rendering in computer music and virtual audio reality.

Jot, Jean-Marc; Chaigne, Antoine (1991): Digital delay networks for designing artificial reverberators. In: *Journal of the Audio Engineering Society*.

Krokstad, A.; Strom, S.; Sørsdal, S. (1968): Calculating the acoustical room response by the use of a ray tracing technique. In: *Journal of Sound and Vibration* 8 (1), S. 118–125. DOI: 10.1016/0022-460X(68)90198-3.

Mendonça, Catarina; Lamas, João; Barker, Tom; Campos, Guilherme; Dias, Paulo; Pulkki, Ville et al. (2013): Reflection orders and auditory distance. In: *The Journal of the Acoustical Society of America* 133 (5), S. 3287. DOI: 10.1121/1.4805398.

Misic, Marko; Nikolov, Dusan; Tomasevic, Milo (2016): Analysis of CPU and GPU implementations of convolution reverb effect. In: *Telfor* 8 (2), S. 121–126. DOI: 10.5937/telfor1602121M.

Miśkiewicz, Andrzej; Rogala, Tomira; Rościszewska, Teresa; Rudzki, Tomasz; Fidecki, Tadeusz (2012): Concert Hall Sound Clarity: A Comparison of Auditory Judgments and Objective Measures. In: *Archives of Acoustics* 37 (1), S. 41–46. DOI: 10.2478/v10168-012-0006-7.

O. Smith, Julius (2010): Physical Audio Signal Processing for virtual musical instruments and audio effects. Artificial Reverberation. Online verfügbar unter http://ccrma.stanford.edu/~jos/pasp/Delay_Lines.html.

Olivieri, Ferdinando; Peter, Nils; Sen, Deep (2019): Scene-Based Audio and Higher Order Ambisonics: A technology overview and application to Next-Generation Audio, VR and 360° Video.

Peters, Nils; Sen, Deep; Kim, Moo-Young; Wuebbolt, Oliver; Weiss, S. Merrill (2016): Scene-based Audio Implemented with Higher Order Ambisonics. In: *SMPTE Mot. Imag. J* 125 (9), S. 16–24. DOI: 10.5594/JMI.2016.2623398.

Philbeck, John W.; Mershon, Donald H. (2002): Knowledge about typical source output influences perceived auditory distance. In: *The Journal of the Acoustical Society of America* 111 (5 Pt 1), S. 1980–1983. DOI: 10.1121/1.1471899.

Pop, Claudiu B.; Cabrera, Densil (2005): Auditory Room Size Perception for Real Rooms. In: *Proceedings of ACOUSTICS*.

Products | mhacoustics.com (2020). Online verfügbar unter <https://mhacoustics.com/products>, zuletzt aktualisiert am 23.08.2020, zuletzt geprüft am 23.08.2020.

Qualcomm Technologies, Inc. (2015): Scene Based Audio. A Novel Paradigm For Immersive And Interactive Audio User Experience.

R. Begault, Durand; M. Wenzel, Elizabeth (2001): Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. In: *Journal of the Audio Engineering Society*.

Rungta, Atul; Rewkowski, Nicholas; Klatzky, Roberta; Manocha, Dinesh (2019): P-Reverb Perceptual Characterization of Early and Late Reflections for auditory displays.

Sandvad, Jesper (1999): Auditory perception of reverberant surroundings. In: *The Journal of the Acoustical Society of America* 105 (2), S. 1193. DOI: 10.1121/1.425625.

Schissler, Carl; Manocha, Dinesh (2017): Interactive Sound Propagation and Rendering for Large Multi-Source Scenes. In: *ACM Trans. Graph.* 36 (1), S. 1–12. DOI: 10.1145/2943779.

Seetharaman, Prem; Pardo, Bryan (2014): Reverbalize. In: Kien A. Hua, Yong Rui, Ralf Steinmetz, Alan Hanjalic, Apostol Natsev und Wenwu Zhu (Hg.): Proceedings of the ACM International Conference on Multimedia - MM '14. the ACM International Conference. Orlando, Florida, USA, 03.11.2014 - 07.11.2014. New York, New York, USA: ACM Press, S. 739–740.

Shinn-Cunningham, Barbara (2000): Learning Reverberation: Considerations for Spatial Auditory Displays.

Shinn-Cunningham, Barbara (2001a): Distance Cues for Virtual Auditory Space.

Shinn-Cunningham, Barbara (2001b): Localizing sound in rooms. In: *Proceedings of the ACM SIGGRAPH and EUROGRAPHICS Campfire*.

Shinn-Cunningham, Barbara (2003): Acoustics and perception of sound in everyday environments.

Shriram, Alluri (2011): Perceived Differences between Natural and Convolution Reverberation in 5.0 Surround Sound.

THX (2020): THX Spatial Audio for Pros - THX. Online verfügbar unter <https://www.thx.com/thx-spatial-audio-for-pros/>, zuletzt aktualisiert am 03.06.2020, zuletzt geprüft am 17.08.2020.

Zotter, Franz; Frank, Matthias (2019): Ambisonics. A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement and Virtual Reality. Cham: Springer International Publishing (19).

Zotter, Franz; Kronlacher, Matthias (2014): Spatial transformations for the enhancement of Ambisonic recordings.

9 Abbildungsverzeichnis

Abbildung 1 Schallausbreitung im Raum (Eigene Darstellung).....	4
Abbildung 2 Reflektionen im zeitlichen Kontext (Eigene Darstellung).....	5
Abbildung 3 Berechnete Verteilung der Raummoden bei Raummaßen von 560x430x300m (amroc - THE Room Mode Calculator 2020).....	6
Abbildung 4 Cone of Confusion (Eigene Darstellung).....	13
Abbildung 5 Kanalbasiertes Audio (Eigene Darstellung).....	19
Abbildung 6 Objektbasiertes Audio (Eigene Darstellung)	20
Abbildung 7 Szenenbasiertes Audio (eigene Darstellung)	21
Abbildung 8 HOA Mikrofon von Eigenmike (Products mhacoustics.com 2020)	23
Abbildung 9 Dummy Head der Firma NEUMANN (NEUMANN 2020).	25

10 Eidesstattliche Erklärung

Hiermit Erkläre ich die vorliegende Bachelorarbeit mit dem Titel „Relevante Faktoren zur Simulation der Schallausbreitung für eine realitätsnahe Immersion in virtuellen Räumen“ selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen oder aus anderweitigen fremden Äußerungen entnommen wurden, sind als solche kenntlich gemacht.

Unterschrift und Datum des Verfassers