

# Masterarbeit

im Studiengang Audiovisuelle Medien

Chancen und Herausforderungen der  
binauralen Audiotechnik für auditive Medien.  
Eine technische und gestalterische Perspektive

Vorgelegt von

Bastian Kilper

37413

an der Hochschule der Medien Stuttgart

am 30.10.2020

zur Erlangung des akademischen Grades eines Master of Engineering

Erstprüfer: Prof. Dr. Frank Melchior

Zweitprüfer: Prof. Oliver Curdt

## **I. Eidesstaatliche Erklärung**

Hiermit versichere ich, Bastian Kilper, ehrenwörtlich, dass ich die vorliegende Masterarbeit mit dem Titel: „Chancen und Herausforderungen der binauralen Audiotechnik für auditive Medien. Eine technische und gestalterische Perspektive“ 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.

Stuttgart, den 30.10.2020

---

Bastian Kilper

## II. Kurzfassung

Die binaurale Audiotechnik ermöglicht die Wiedergabe von 3D-Audio über herkömmliche Kopfhörer. Die vorliegende Arbeit zeigt die Chancen und Herausforderungen bei der Anwendung von binauralem Audio für auditive Medien und berücksichtigt dabei unterschiedliche Aspekte der binauralen Audiotechnik aus technischer und gestalterischer Sicht. Als zentrale Forschungsfrage liegt der Vorgehensweise hierbei stets zugrunde, wie binaurales Audio aktuell angewendet wird und warum es bisher kein breites Angebot an binauralen Audioproduktionen gibt. Die vorliegende Arbeit zeigt auf, welchen Mehrwert und welche Herausforderungen eine Binauralmischung gegenüber einer Stereomischung bringt und für welche Inhalte die Anwendung von binauralem Audio geeignet ist.

Dies geschieht anhand einer literarischen Untersuchung von Sekundärquellen, die darstellen, wie Binauraltechnik funktioniert und wie sie in aktuellen binauralen Produktionssystemen umgesetzt wird. Diese Forschungsarbeit verdeutlicht, dass aus technischer Perspektive vor allem drei Faktoren die Qualität der Wiedergabe beeinflussen: die Individualisierung, die dynamische Nachführung der Mischung und das Verfahren, mit dem die binaurale Mischung erstellt wurde. Zur Stärkung und Ergänzung der These werden im zweiten Abschnitt der Arbeit die Gestaltungsmerkmale der auditiven Inhalte Musik, Radio und Hörspiel untersucht und Beispiele für binaurale Mischungen der jeweiligen Inhalte genannt. So wird exemplarisch aufgezeigt, welche neue Gestaltungsmöglichkeiten binaurales Audio liefert und verdeutlicht, dass sich für die meisten Audioproduktionen eine Kombination aus binauralisierten und nicht-binauralisierten Elementen in der Mischung am besten eignet, um so die klanglichen Einschränkungen von binauralem Audio auszugleichen.

### **III. Abstract**

The binaural audio technology enables the playback of 3D audio via conventional headphones. The present paper shows the chances and challenges of using binaural audio for auditory media and considers different aspects of binaural audio technology from a technical and creative point of view. As a central research question, the approach is always based on how binaural audio is currently used and why there is no wide range of binaural audio productions available. The present work shows the added value and challenges of binaural mixing compared to stereo mixing and for which content the use of binaural audio is suitable.

This is done by means of a literary examination of secondary sources that show how binaural technology works and how it is implemented in current binaural production systems. This research shows that from a technical perspective, three main factors influence the quality of the playback: the individualization, the dynamic tracking of the mix, and the method used to create the binaural mix. In order to strengthen and complement the thesis, the second part of the present paper examines the design characteristics of the auditory, contents music, radio and radio play and gives examples of binaural mixtures of the respective contents. This way, the new design possibilities of binaural audio are shown and it becomes clear that for most audio productions a combination of binaural and non-binaural elements in the mix is best suited to compensate for the sonic limitations of binaural audio.

## **IV. Danksagung**

Ich möchte mich herzlich bei meinen Freundinnen und Freunden, meinen Bandkolleginnen und -kollegen und meiner Familie bedanken, die mir den Rücken freigehalten haben, mich mental unterstützt haben und mir die nötige Zeit eingeräumt haben, mich voll und ganz der Realisierung dieser Arbeit zu widmen.

Mein Dank gilt außerdem meinem Erstprüfer Prof. Dr. Frank Melchior, der sein umfassendes Fachwissen mit mir teilte und zu jeder Zeit ein offenes Ohr für mich hatte, obwohl wir uns inmitten der Corona-Pandemie befanden.

Ein besonderes Dankeschön möchte ich meinem Zweitprüfer Prof. Oliver Curdt aussprechen, der meine Zeit an der Hochschule der Medien und den Verlauf meines Studiums durch lehrreiche Vorlesungen, Workshops und Gespräche entscheidend mitprägte.

# V. Inhaltsverzeichnis

<b>I. Eidesstaatliche Erklärung</b> .....	<b>II</b>
<b>II. Kurzfassung</b> .....	<b>III</b>
<b>III. Abstract</b> .....	<b>IV</b>
<b>IV. Danksagung</b> .....	<b>V</b>
<b>V. Inhaltsverzeichnis</b> .....	<b>VI</b>
<b>VI. Abkürzungsverzeichnis</b> .....	<b>IX</b>
<b>1 Einleitung</b> .....	<b>1</b>
<b>1.1 Problemstellung</b> .....	<b>1</b>
<b>1.2 Zielsetzung und Methodik</b> .....	<b>2</b>
<b>1.3 Vorgehensweise</b> .....	<b>2</b>
<b>2 Psychoakustik</b> .....	<b>4</b>
<b>2.1 Das menschliche Ohr</b> .....	<b>4</b>
<b>2.2 Außenohr-Übertragungsfunktion</b> .....	<b>5</b>
<b>2.3 Lokalisation einer Schallquelle</b> .....	<b>8</b>
2.3.1 Lokalisation in der Horizontalebene .....	10
2.3.1.1 Interaurale Laufzeitdifferenzen .....	10
2.3.1.2 Interaurale Pegeldifferenzen.....	11
2.3.1.3 Grenzen der horizontalen Lokalisation .....	11
2.3.2 Lokalisation in der Medianebene .....	13
2.3.3 Entfernungslokalisation .....	14
2.3.4 Lokalisation bei verzögerten Schallereignissen aus verschiedenen Richtungen .....	15
2.3.5 Auditorische Signalverarbeitung.....	15
<b>3 Binauraltechnik</b> .....	<b>18</b>
<b>3.1 Unterschiede zwischen Stereo- und Binauralmischungen</b> .....	<b>18</b>
3.1.1 Im-Kopf-Lokalisation .....	18
3.1.2 Ziele der Binauraltechnik.....	19
<b>3.2 Kunstkopfstereofonie</b> .....	<b>19</b>
3.2.1 Geschichte der Kunstkopfstereofonie .....	20
3.2.2 Eigenschaften von Kunstkopfmikrofonen .....	20
3.2.3 Entzerrung des Kunstkopfmikrofons .....	20
3.2.4 Nachträgliche Kunstkopfaufnahme einer raumbezogenen Aufnahme .....	21
3.2.5 Beispiele für Kunstkopfmikrofone.....	21
3.2.6 Originalkopfmikrofon .....	22
3.2.7 Einschränkungen bei der Kunstkopfstereofonie .....	23
<b>3.3 Kopfhörer</b> .....	<b>24</b>
3.3.1 Funktionsweise von Kopfhörern.....	24
3.3.2 Kopfhörerarten .....	25

3.3.3	Entzerrung der Kopfhörer .....	25
3.3.4	Headset .....	27
3.3.5	Aktive Geräuscherdrückung .....	27
<b>4</b>	<b>Binauralsynthese .....</b>	<b>29</b>
<b>4.1</b>	<b>Ideales System.....</b>	<b>29</b>
<b>4.2</b>	<b>Prinzip der Binauralsynthese .....</b>	<b>29</b>
4.2.1	Faltung .....	30
<b>4.3</b>	<b>Impulsantwort .....</b>	<b>31</b>
4.3.1	Binaurale Raumimpulsantwort .....	31
4.3.2	Individualisierung .....	31
4.3.2.1	Messbedingungen für eine HRTF-Messung.....	31
4.3.2.2	Auditive Messung.....	32
4.3.2.3	Visuelle Messung und Erstellung eines virtuellen Kopfmodells .....	33
4.3.2.4	Visuelle Messung und Typisierung .....	34
<b>4.4</b>	<b>Dynamische Binauralsynthese.....</b>	<b>35</b>
4.4.1	Virtual Reality .....	35
4.4.2	Mögliche Kopfbewegungen .....	36
4.4.3	Prinzip der dynamischen Binauralsynthese .....	37
4.4.4	Rasterauflösung des Datensatzes .....	38
4.4.5	Trennung der Nachführung von frühem Hallanteil und Diffusschall .....	38
4.4.6	Headtracking-Systeme .....	39
4.4.6.1	Elektromagnetische Sensoren.....	39
4.4.6.2	Optische Sensoren .....	40
4.4.6.3	Inertialsensoren.....	40
4.4.6.4	Beispiele für Headtracking-Systeme .....	41
<b>4.5</b>	<b>Verfahren zur Binauralsynthese.....</b>	<b>42</b>
4.5.1	Faltung mit der HRTF .....	42
4.5.2	Faltung mit der BRIR realer Räume .....	44
4.5.3	Faltung mit der BRIR virtueller Räume.....	45
<b>4.6</b>	<b>Reales System.....</b>	<b>47</b>
4.6.1	Vergleich realer Systeme .....	47
4.6.2	Diskussion.....	52
<b>5</b>	<b>Binaurale Inhalte.....</b>	<b>53</b>
<b>5.1</b>	<b>Binaurale Musik .....</b>	<b>53</b>
5.1.1	Gestaltungsmerkmale von Musik.....	53
5.1.2	Beispiele für binaurale Studioproduktionen .....	54
5.1.2.1	Studioproduktionen mit Kunstkopfstereofonie .....	54
5.1.2.2	Studioproduktionen mit Binauralsynthese .....	55
5.1.2.3	8D-Audio.....	56

5.1.3	Beispiele für binaurale Konzertmitschnitten.....	56
5.1.3.1	Becks „Sound and Vision“ .....	57
5.1.3.2	BBC Proms.....	58
5.1.3.3	VR-Konzerte.....	59
5.1.4	Mehrwert durch binaurale Audiotechnik .....	60
<b>5.2</b>	<b>Binaurales Radio und Podcasts.....</b>	<b>61</b>
5.2.1	Gestaltungsmerkmale von Radio und Podcasts .....	61
5.2.2	Sprachverständlichkeit .....	62
5.2.3	Mehrwert durch binaurale Audiotechnik .....	63
5.2.3.1	Dynamische Binauralsynthese von einem Sprechenden .....	63
5.2.3.2	Dynamische Binauralsynthese von mehreren Sprechenden.....	63
5.2.4	Individualisierung für binaurales Radio und Podcasts.....	64
<b>5.3</b>	<b>Binaurale Hörspiele.....</b>	<b>65</b>
5.3.1	Gestaltungsmerkmale von Hörspielen .....	65
5.3.2	BBC „The Turning Forest“ .....	66
5.3.3	Mehrwert durch binaurale Audiotechnik .....	67
<b>6</b>	<b>Konsum binauraler Inhalte.....</b>	<b>69</b>
<b>6.1</b>	<b>Welche Inhalte werden konsumiert? .....</b>	<b>70</b>
<b>6.2</b>	<b>Wie werden Inhalte konsumiert?.....</b>	<b>70</b>
<b>6.3</b>	<b>Wo werden Inhalte konsumiert?.....</b>	<b>71</b>
<b>6.4</b>	<b>Wo werden Inhalte angeboten?.....</b>	<b>72</b>
<b>6.5</b>	<b>Aktuelle Technologien.....</b>	<b>73</b>
6.5.1	Sony 360 Reality Audio.....	73
6.5.2	Dolby Atmos Music .....	73
6.5.3	Apple Spatial Audio.....	74
<b>7</b>	<b>Fazit.....</b>	<b>75</b>
<b>7.1</b>	<b>Ergebnisse .....</b>	<b>75</b>
<b>7.2</b>	<b>Diskussion und Ausblick.....</b>	<b>77</b>
<b>8</b>	<b>Abbildungsverzeichnis.....</b>	<b>80</b>
<b>9</b>	<b>Literaturverzeichnis.....</b>	<b>81</b>

## VI. Abkürzungsverzeichnis

ANC .....	Active Noise Cancelling
BRIR .....	Binaural Room Impulse Response
BRS .....	Binaural Room Synthesis
BRTF .....	Binaural Room Transfer Function
FFT .....	Fast Fourier Transformation
FIR .....	Finite Impulse Response
HOA .....	Higher Order Ambisonics
HRIR .....	Head Related Impulse Response
HRTF .....	Head Related Transfer Function
ILD .....	Interaural Level Differences
ITD .....	Interaural Time Differences
LTI-System .....	Linear Time Invariant System
SNR .....	Signal-to-noise Ratio
SOFA .....	Spatially Oriented Format for Acoustics
SRT .....	Speech Reception Threshold
VR .....	Virtual Reality

# 1 Einleitung

„3D ist nichts Besonders, nichts Außergewöhnliches, sondern unsere ganz natürliche, alltägliche, akustische Umgebung [...]. Wir nehmen unsere Umwelt in 3D wahr“ (Wisse, 2018).

## 1.1 Problemstellung

Der Mensch hört von Geburt an dreidimensional. Das Ziel, sich dieser „natürliche[n], alltägliche[n]“ (ebd.) Form des Hörens anzunähern verfolgte die Filmbranche seit der Einführung des Tonfilms. Mit dem Wunsch nach einem noch realistischeren Klang, wurde nach dem Umstieg von Mono auf Stereo in den 1950er Jahren im Jahr 1976 mit Jon Peters „A Star Is Born“ (siehe Pierson, 1976) der erste Film mit Raumklang veröffentlicht. Die Einführung des Formats Dolby Stereo war der erste Versuch, die frontale Beschallung durch eine Beschallung von hinten zu ergänzen. Als Quelle diente allerdings nach wie vor nur die zweikanalige Stereospur. Mit Tim Burtons „Batman Returns“ (siehe Burton, 1992) feierte 1992 der erste Film mit Surround-Sound seine Premiere. Der Filmtone konnte nun auf fünf voneinander unabhängigen Kanälen um den Zuschauerraum herum gemischt werden (vgl. El-Bawwab, 2015). 2012 erschien schließlich der erste Film mit 3D-Audio: Bei der Disney/Pixar-Produktion „Brave“ (siehe Anrews & Chapman, 2012) wurden erstmals Deckenlautsprecher und Effekte von oben eingesetzt. Seitdem steht 3D-Audio im Begriff, sich fest im Filmtone zu etablieren (vgl. El-Bawwab, 2015).

Abseits von Filmproduktionen fällt das Angebot von 3D-Audio deutlich geringer aus. Dabei schafft 3D-Audio auch für rein auditiven Medien wie Musik, Radio oder Hörspiele völlig neue gestalterische Möglichkeiten: „In 3D-Audio sehen wir die Chance, Musikern, Produzenten und nicht zuletzt der Musik selbst die Würde zurückzugeben“ (Kunz, 2019, S. 48f), so Stefan Zaradic, Musiker und Teilinhaber der Firma IAN solutions. Bei Konzerten und Klanginstallationen kommen immer häufiger 3D-Lautsprecher-Setups zum Einsatz. Der Heimkonsum von 3D-Audio wird bisher eher von Audiophilie praktiziert und hat sich noch nicht in der breiten Masse durchgesetzt. Begründet wird dies oft damit, dass nur wenige Konsumentinnen und Konsumenten den Platz und die finanziellen Mittel haben, ein 3D-fähiges Lautsprecher-Setup in ihrem Wohnzimmer aufzustellen (vgl. ebd., S. 50ff).

Die Binauraltechnik verspricht eine Lösung für dieses Problem: Zur Wiedergabe von 3D-Audio genügt prinzipiell ein herkömmlicher Stereo-Kopfhörer. Begünstigt wird die Technologie außerdem durch die steigende Nachfrage nach Kopfhörern in den letzten Jahren: „Die Kombination aus einer hohen Smartphone-Verbreitung und wachsenden Streaming-Angeboten hat einen Kopfhörer-Boom ausgelöst [...]. Waren Kopfhörer früher reine Gebrauchsgegenstände, sind sie inzwischen Teil der Popkultur“ (Krösmann & Klöß, 2019) so Dr. Sebastian Klöß, Referent des Digitalverbands Bitkom. Die Binauraltechnik scheint deshalb prädestiniert dafür zu sein, den Konsum von 3D-Audio für eine breite Masse an Hörerinnen und Hörern zu ermöglichen.

## **1.2 Zielsetzung und Methodik**

Die vorliegende Arbeit soll die Frage beantworten, was die Chancen und Herausforderungen der binauralen Audiotechnik sind. Dazu werden unterschiedliche Aspekte von binauralem Audio aus technischer und gestalterischer Perspektive betrachtet. Die zentrale Fragestellung hierbei ist, wie binaurales Audio aktuell angewendet wird und warum es bisher kein breites Angebot an binauralen Audioproduktionen gibt. Das Ziel der Arbeit ist es entsprechend aufzuzeigen, welchen Mehrwert und welche technischen und gestalterischen Herausforderungen eine Binauralmischung gegenüber einer Stereomischung bringt und für welche Inhalte die Anwendung von binauralem Audio geeignet ist. Grundsätzlich beschränkt sich die Arbeit dabei auf den Einsatz von binauralem Audio für rein auditive Medien. Mit dem Ziel, technische und gestalterische Ansätze für mögliche Anwendungen bei einer reinen Audioproduktionen zu adaptieren, werden ausgewählte Beispiele aus dem Bereich Film miteinbezogen.

Zur Beantwortung der zugrundeliegenden Fragestellung, wurde im deduktiven Ansatz eine Ableitung der Erkenntnisse durch Sekundärquellen in Form einer Literatarbeit verfolgt. Dabei wurde einschlägige Literatur unterschiedlicher Autorinnen und Autoren gewählt. Die Parameter zur Entscheidung waren hierbei insbesondere die Qualität und je nach Abschnitt besonders die Aktualität der Quelle, wobei die Qualität der Quelle an Kritiken, bewertender Literatur und dem eigenen Ermessen festgemacht wurde. Die zitierten wissenschaftlichen Publikationen sind peer-reviewed. Die Literatur wurde stets qualitativ untersucht, da das übergeordnete Ziel der Arbeit die Erkenntnisgewinnung, nicht aber die Inhaltsbewertung in den Fokus stellt. Zur Einordnung von teils kontroversen Meinungen und emotionalen Bewertungen von Klang, vor allem im Kapitel 5, wurde versucht, diverse gegensätzliche Quellen hinzuziehen. Im Kapitel 6 wurden mehrere Studien quantitativ ausgewertet und in Bezug zur Forschungsfrage gesetzt.

## **1.3 Vorgehensweise**

Kapitel 2 beschreibt zunächst die Psychoakustik des Menschen und untersucht die Eigenschaften des menschlichen Ohres, die für die Lokalisation einer Schallquelle verantwortlich sind. Dabei werden wesentliche fachliche und begriffliche Grundlagen beschrieben, die für das Verständnis der darauffolgenden Kapitel bedeutsam sind. Kapitel 3 legt die Unterschiede zwischen einer Stereo- und einer Binauralmischung dar, erklärt die Binauralaufnahme mittels Kunstkopfstereofonie und deren Einschränkungen und beschreibt die Eigenschaften von Kopfhörern, die als Wiedergabegerät für binaurales Audio dienen. Durch die Darlegung der historischen Entwicklung der Binauraltechnik werden grundsätzliche Eigenschaften und Technologien erklärt, die zentrale Komponenten der binauralen Audiotechnik sind. Von den Einschränkungen bei der Kunstkopfstereofonie ausgehend präsentiert Kapitel 4 zunächst ein ideales binaurales Produktionssystem und führt so anwendungsbezogen zum Prinzip der dynamischen Binauralsynthese, durch die binaurale Audioproduktionen aktuell umgesetzt werden. Dabei werden besonders die drei Faktoren Individualisierung, Headtracking und die

Qualität des Verfahrens zur Binauralsynthese hervorgehoben und eingeordnet, wie sie die Qualität der binauralen Wiedergabe beeinflussen. Die Untersuchung verschiedener existenter binauraler Produktionssysteme in einer Eigenschaftsmatrix liefert einen Überblick darüber, wie die zuvor beschriebenen Faktoren aktuell umgesetzt werden. Kapitel 5 untersucht binaurale Inhalte und betrachtet das Forschungsthema aus gestalterischer Sicht. Die Inhalte werden aufgrund ihrer unterschiedlichen Eigenschaften in die drei Kategorien Musik, Radio und Hörspiel eingeteilt. Um ein Verständnis für die unterschiedlichen Inhaltsformen zu schaffen, wird jede Form von Inhalt zunächst auf deren Gestaltungsmerkmale hin untersucht. Darauf folgende aktuelle Beispiele zeigen auf, ob und wie der jeweilige Inhalt von einer binauralen Mischung profitieren könnte und welchen Einfluss die technischen Faktoren auf die Inhaltsgestaltung haben. Kapitel 6 beschreibt die Audiokonsumtinnen und -konsumenten und deren Hörgewohnheiten unter Zuhilfenahme mehrerer Studien. Dadurch wird untersucht, ob und wie sich das bisherige Konsumverhalten von Stereomischungen für den Konsum von binauralen Medien ändern müsste und was es benötigt, um binaurales Audio als festen Bestandteil des Medienangebots zu etablieren. Zum Ende der Darlegungsarbeit werden drei aktuelle Technologien zur binauralen Audiowiedergabe vorgestellt und betrachtet, wie diese die in den vorangegangenen Kapiteln erarbeiteten technischen und gestalterischen Herausforderungen bewältigen. Eine Zusammenfassung der Ergebnisse und eine kritische Reflektion der Arbeit, sowie die Nennung möglicher zukünftiger und weitergehender Forschungsansätze schließen die vorliegende Arbeit ab.

## 2 Psychoakustik

„Die wissenschaftliche Disziplin der Psychoakustik untersucht, wie akustische Reize wahrgenommen werden“ (Weinzierl, 2008, S. 52). Die Psychoakustik kombiniert Erkenntnisse aus den Bereichen Psychologie, Physik und Biologie (vgl. Görne, 2015, S.110f). Im folgenden Kapitel soll die Funktionsweise des Gehörs und die auditive Wahrnehmung eines Schallereignisses erklärt werden. Dabei soll besonders auf die Lokalisation eines Schallereignisses und das Entfernungshören eingegangen werden, da diese Bereiche für die Anwendung von Binauraltechnik besonders wichtig ist.

### 2.1 Das menschliche Ohr

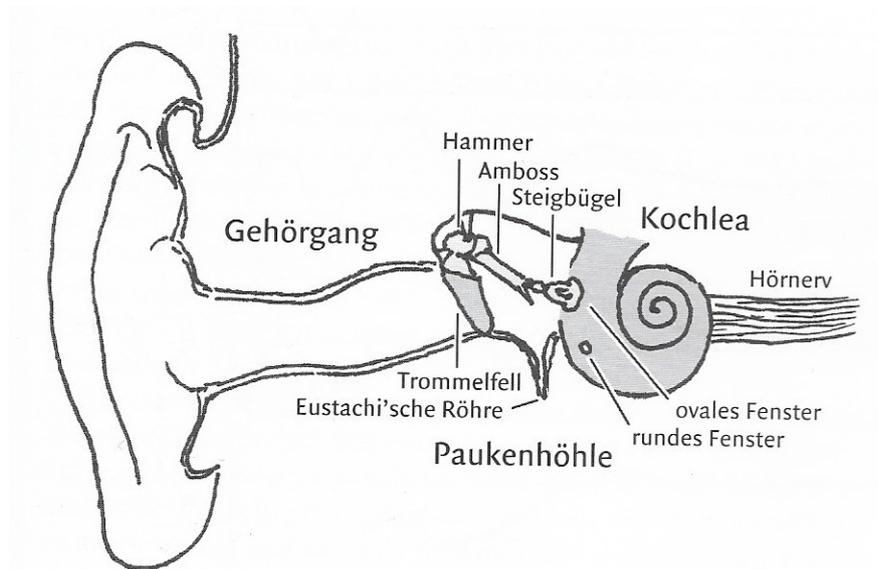


Abbildung 1: Aufbau des Ohrs (Görne, 2015, S. 111)

Abbildung 1 zeigt den Aufbau des menschlichen Ohrs. Es besteht aus Außen-, Mittel- und Innenohr. Zum Außenohr gehören die Ohrmuschel und der äußere Gehörgang. Die Ohrmuschel besteht größtenteils aus Knorpel und ist bei jedem Menschen sehr individuell geformt. Der äußere Gehörgang endet am Trommelfell, das Trommelfell grenzt das Außenohr vom Mittelohr ab. Die eustachische Röhre verbindet das Mittelohr mit dem Rachenraum und ermöglicht einen statischen Druckausgleich. Das Mittelohr selbst ist ein luftgefüllter Raum, der auch als Paukenhöhle bezeichnet wird. In der Paukenhöhle befinden sich die Gehörknöchelchen Hammer, Amboss und Steigbügel, die die Schwingung des Trommelfells an das Innenohr übertragen. Durch die Anordnung der Gehörknöchelchen entsteht eine Hebelwirkung, außerdem sind das Trommelfell und die Fußplatte des Steigbügels, das ovale Fenster, unterschiedlich groß. Dadurch wird die Schwingung des Trommelfells verstärkt ans Innenohr übertragen. Das Innenohr besteht aus Gleichgewichtsorgan und Hörorgan, das auch als Kochlea bzw. Schnecke bezeichnet wird. Die Kochlea ist ein zweieinhalbmal aufgewickelter Kanal.

Sie ist durch die Reissner'sche Membran und die Basilarmembran in drei Kanäle aufgeteilt. Auf der Basilarmembran befindet sich das cortische Organ mit den Sinneszellen, die aus äußeren und inneren Haarzellen bestehen. Diese Haarzellen enden in Stereozilien, einem Bündel aus Härchen (vgl. Weinzierl, 2008, S.42ff).

Ein Schallereignis erzeugt über Trommelfell und Gehörknöchelchen eine Schwingung am ovalen Fenster. Dadurch wird eine Wanderwelle auf der Basilarmembran ausgelöst. Je nach Wellenlänge der Frequenz des Schallereignisses entsteht die maximale Auslenkung der Wanderwelle an unterschiedlichen Orten auf der Basilarmembran. „Je tiefer die Frequenz ist, desto tiefer dringt die Wanderwelle in die Kochlea ein“ (Görne, 2015, S.114). Diese Eigenschaft des Gehörs wird als Ortstheorie des Hörens bezeichnet. Je nach Ort der maximalen Amplitude werden unterschiedliche Haarzellen auf der Basilarmembran angeregt, die Stereozilien an dieser Stelle geben einen elektrischen Impuls an den jeweiligen Hörnerv weiter. „Die Kochlea führt [also] eine Frequenzanalyse des Schalls durch“ (ebd., S.114). Die äußeren Haarzellen am jeweiligen Ort der Basilarmembran verstärken große und unterdrücken kleine Schwingungsamplituden. Neben dem Ortsprinzip wird über Impulssalven außerdem die Phase des Schallsignals kodiert, was als Salventheorie bezeichnet wird. Die Salventheorie ist vor allem bei tiefen Frequenzen unter 1 kHz relevant. Zwischen 1 kHz und 5 kHz wirken Salven- und Ortstheorie zusammen. Die Frequenzauflösung ist in diesem Bereich deshalb maximal. Frequenzen über 5 kHz werden vor allem durch das Ortsprinzip kodiert (vgl. ebd., S.115f).

## **2.2 Außenohr-Übertragungsfunktion**

Die spektrale Wahrnehmung eines Schallereignisses ist stark richtungsabhängig. Je nachdem, wo die Schallquelle relativ zum Ohr positioniert ist, entsteht eine charakteristische lineare Verzerrung des Schallsignals, da sich Direktschall und Schallreflexionen der Ohrmuschel am Außenohr überlagern und der Amplituden- und Phasenverlauf des Signals durch die dadurch entstehenden Kammfiltereffekte verändert wird. Verändert sich nun die relative Position der Schallquelle, so ändert sich auch, wie der Schall reflektiert wird und es entstehen andere Kammfilter. Diese richtungsabhängige lineare Verzerrung des Ursprungssignals ermöglicht die Ortung einer Schallquelle und wird mathematisch als Außenohr-Übertragungsfunktion bzw. Head Related Transfer Function (HRTF) im Frequenzbereich und Head Related Impulse Response (HRIR) im Zeitbereich beschrieben (vgl. Weinzierl, 2008, S. 90). Abbildung 2 zeigt die HRTF-Messung des linken Ohres einer Versuchsperson, 4 mm ohrkanaleinwärts bei frontalem Schalleinfall.

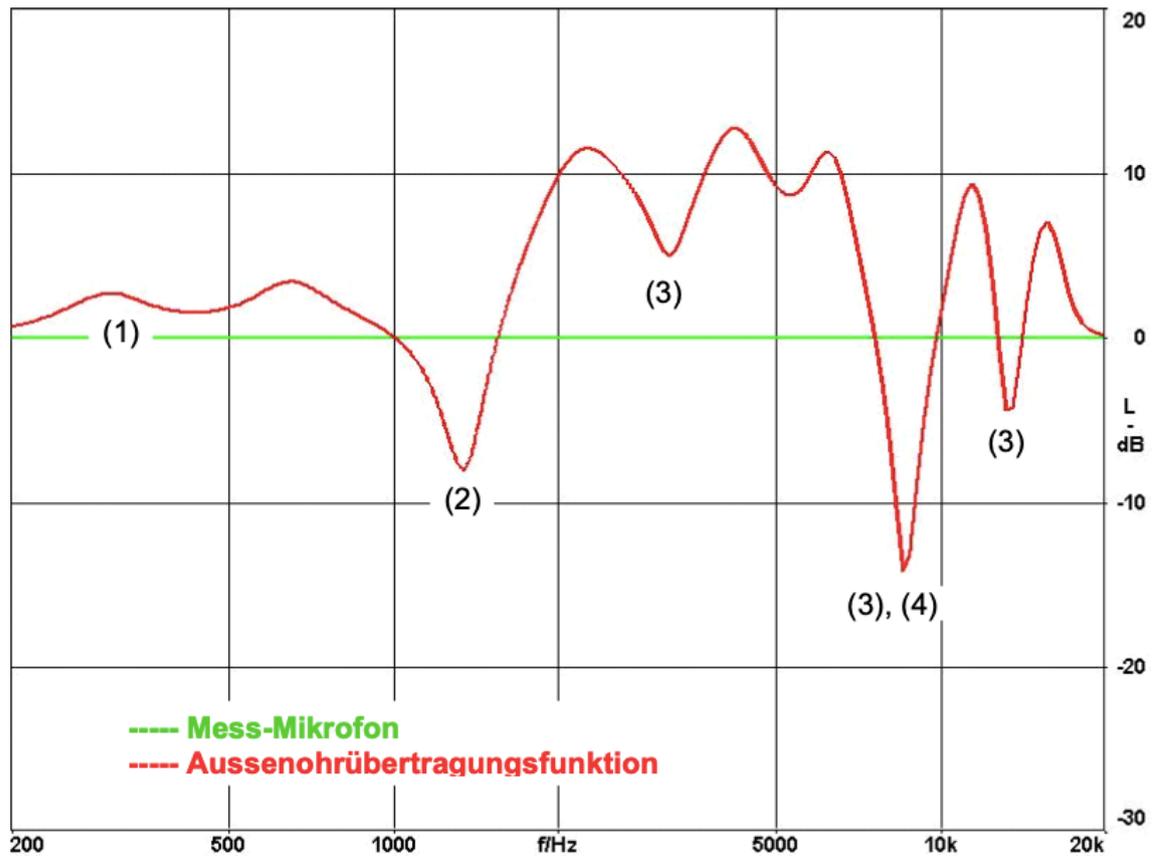


Abbildung 2: HRTF-Messung einer Versuchsperson (Daniel, 2007, S. 5)

Die in der Abbildung nummerierten, richtungsabhängigen Frequenzeinbrüche sind auf Reflexions- und Beugungseffekte am menschlichen Torso, Kopf und Außenohr zurückzuführen. Im Bereich unter 500 Hz liegt die Wellenlänge größer als die Abmessungen des Kopfes und die eintreffenden Schallwellen werden um den Kopf gebeugt (siehe Abbildung 2: Bereich 1). Im Bereich zwischen 1 kHz und 2 kHz entsteht ein Frequenzeinbruch, der durch Schallreflexionen an der Schulter und den daraus resultierenden Kammfiltereffekten entsteht (siehe Abbildung 2: Bereich 2). Die Einbrüche oberhalb von 3 kHz werden durch  $\lambda/4$ -Resonanzen innerhalb des Ohrkanals (siehe Abbildung 2: Bereich 3) und durch die Cavum Conchae, einer großen Vertiefung auf der Vorderseite der Ohrmuschel, hervorgerufen (siehe Abbildung 2: Bereich 4) (vgl. Ballou, 2008, S. 554). Für die Richtungsinformationen sind vor allem die linearen Verzerrungen oberhalb von 3 kHz wichtig (vgl. Daniel, 2007, S. 5).

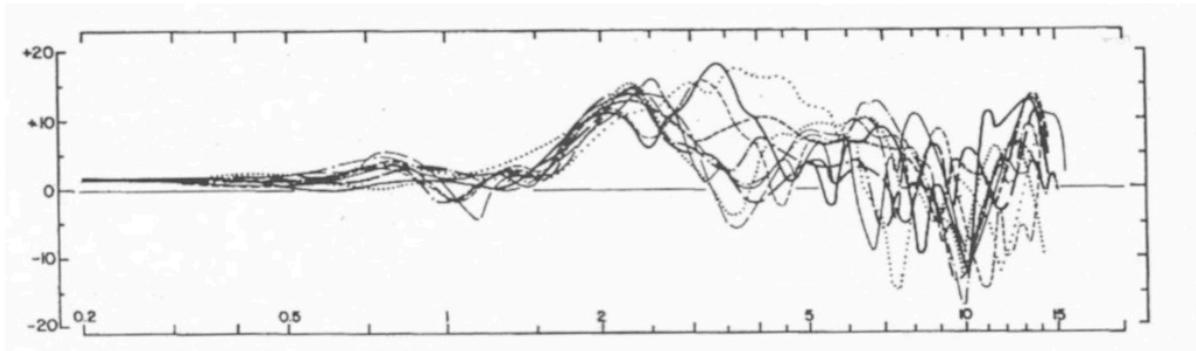


Abbildung 3: HRTF-Messung an zehn Versuchspersonen (Görne, 2015, S. 112)

Abbildung 3 zeigt die HRTF-Messungen von zehn Versuchspersonen, gemessen im linken Gehörgang. Die Verläufe der HRTF-Kurven der Versuchspersonen unterscheiden sich stark. Wie in Abschnitt 2.1 beschrieben, ist das Außenohr, speziell die Ohrmuschel jedes Menschen sehr individuell geformt. Je nach Form der Ohrmuschel wird der einfallende Schall anders gestreut. Das führt dazu, dass die HRTF jedes Menschen unterschiedlich ist und es umgekehrt keine universelle HRTF gibt, die auf die Anatomie der Ohrenpaare aller Menschen zutrifft. Eine solche universelle HRTF kann deshalb immer nur eine Mittelung aus verschiedenen HRTF-Messungen von mehreren Versuchspersonen darstellen (vgl. Görne, 2015, S. 113).

Die Messergebnisse variieren, je nach dem, an welchem Ort innerhalb des Ohrkanals die Messmikrofone platziert werden. Abbildung 4 zeigt die Ergebnisse dreier Messungen, bei denen das Mikrofon am Trommelfell, am offenen Ohrkanaleingang und am geblockten Ohrkanaleingang platziert wurden. Die Messungen wurden für jede Mikrofonposition an jeweils zwölf Versuchspersonen durchgeführt.

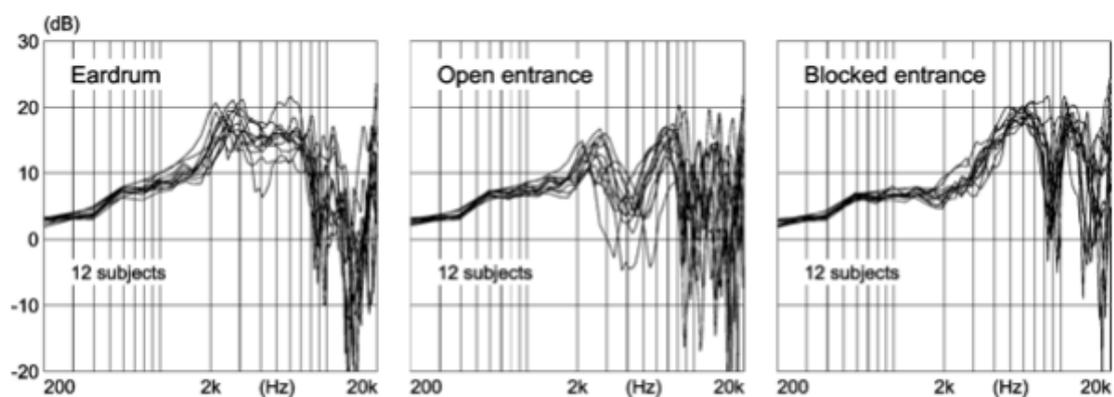


Abbildung 4: HRTF-Messung an unterschiedlichen Positionen im Ohrkanal (Weinzierl, 2008, S. 587)

Abbildung 4 veranschaulicht die Unterschiede zwischen den Messergebnissen an den drei Positionen. Weinzierl gibt zu bedenken, dass „sich dieser Einfluss in zahlreichen Untersuchungen als unab-

hängig von der Schalleinfallrichtung erwiesen [hat]“ (Weinzierl, 2008, S. 586). Die Unterschiede zwischen den drei Messpositionen sind deshalb weniger relevant. Wichtiger sind die Unterschiede zwischen den Messergebnissen der einzelnen Versuchspersonen an einer Messposition. Beim geblockten Ohrkanaleingang sind die Verläufe der Messkurven der Versuchspersonen einander am ähnlichsten. Deshalb wird eine HRTF-Messung typischerweise an dieser Position durchgeführt, da die Messung hier am wenigsten von der individuellen Ohrform beeinflusst wird und die Messergebnisse somit am universellsten auswertbar sind (vgl. ebd., S. 586f).

Abbildung 5 zeigt eine abschließende Übersicht über die genannten richtungsabhängigen und richtungsunabhängigen Komponenten, die die HRTF beeinflussen:

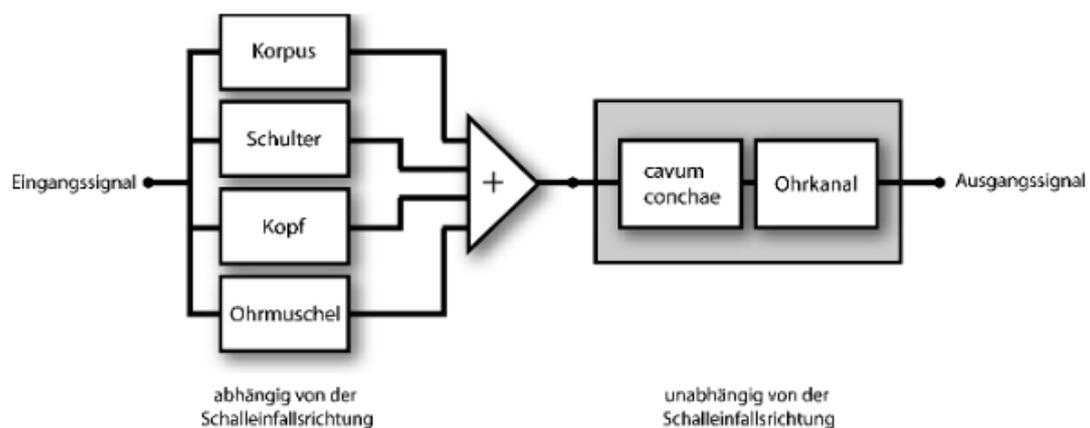


Abbildung 5: Komponenten der HRTF (Weinzierl, 2008, S. 586)

### 2.3 Lokalisation einer Schallquelle

Die Wahrnehmung und Ortung einer Schallquelle erfolgen neben dem Gehör auch durch andere Sinne, vor allem durch den Sehsinn, „man hört, was man sieht“ (Görne, 2015, S. 126). Auch der Tastsinn trägt zur Lokalisation bei.

Außerdem führt der Kopf ständig unbewusste kleine Bewegungen aus. Durch diese unwillkürlichen Peilbewegungen verändert sich die HRTF und das zentrale Nervensystem gewinnt weitere Richtungsinformationen über die Schallquelle. „Wegen der Trägheiten bei Kopfbewegungen kann nur dann ‚gepeilt‘ werden, wenn die Schallereignisse länger als ca. 200 ms andauern“ (Weinzierl, 2008, S. 88).

Der Knochenschall, der über den Schädelknochen übertragen wird, kann vernachlässigt werden, „denn die Hörschwelle für durch Luftschall erregten Knochenschall liegt mehr als 40 dB über der Hörschwelle für Luftschall“ (ebd., S. 89).

Zur Lokalisation einer Schallquelle tragen also weit mehr Faktoren, als nur die beiden Ohrsignale bei. „Hören ist [...] im Prinzip ein multimodales Phänomen“ (ebd., S. 89). Das ist vor allem bei der Anwendung von rein auditiven Medien zu berücksichtigen, da das Gehör hierbei der einzige Sinn

ist, der angesprochen werden kann. Da sich diese Arbeit ausschließlich auf Audioproduktionen bezieht, soll im nachfolgenden Abschnitt speziell auf die akustischen Faktoren des räumlichen Hörens eingegangen werden.

Zur Beschreibung der Position einer Schallquelle relativ zum Kopf der Empfängerin oder des Empfängers wird ein kopfbezogenes Koordinatensystem verwendet (siehe Abbildung 6).

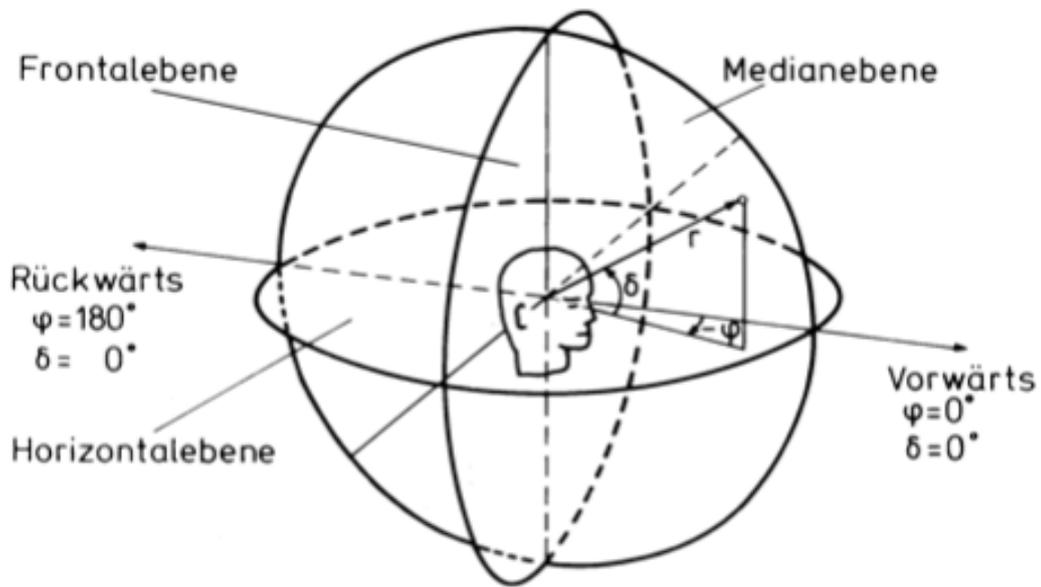


Abbildung 6: Kopfbezogenes Koordinatensystem (Weinzierl, 2008, S. 88)

Das Polardiagramm besteht aus Horizontal-, Frontal- und Medianebene. Die drei Ebenen sind orthogonal zueinander und haben ihren Ursprung in der Mitte des Kopfes. Die vertikale Auslenkung einer Schallquelle wird durch den Elevationswinkel  $\delta$  angegeben, die horizontale Auslenkung durch den Azimutwinkel  $\varphi$ . „Im Bereich der Tontechnik wird der Azimut  $\varphi$  oft statt im mathematisch positiven Sinne im Uhrzeigersinn gezählt“ (Weinzierl, 2008, S. 88), ein Azimutwinkel in der Tontechnik wird also mathematisch korrekt, als  $-\varphi$  notiert (siehe Abbildung 6). Der Abstand der Schallquelle wird durch die Entfernungs-Koordinate  $r$  angegeben (vgl. ebd., S. 88).

### 2.3.1 Lokalisation in der Horizontalebene

Wird eine Schallquelle seitlich zur Medianebene auf der Horizontalebene ausgelenkt, kommen an beiden Ohren unterschiedliche Signale an und das Schallereignis wird interaural wahrgenommen (siehe Abbildung 7), es entstehen interaurale Merkmale (engl. Cues): Einerseits treten interaurale Laufzeitdifferenzen (engl. Interaural Time Differences) (ITD) auf, da das Schallsignal an unterschiedlichen Zeitpunkten am jeweiligen Ohr ankommt. Andererseits entstehen interaurale Pegeldifferenzen (engl. Interaural Level Differences) (ILD), da das der Schallquelle abgewandte Ohr weniger Pegel bekommt, als das der Schallquelle zugewandte Ohr (vgl. Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 129f).

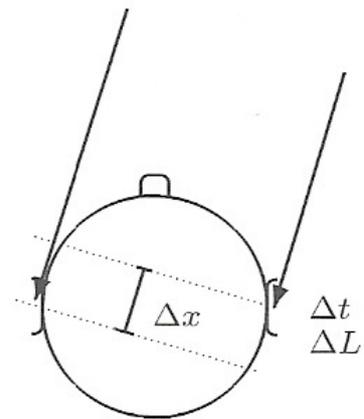


Abbildung 7: Seitlicher Schalleinfall (Görne, 2015, S. 126)

Bei frontalen Schallquellen liegt die Ortungsschärfe auf der Horizontalebene bei 2° bis 3°. Bei Signalen, die stark seitlich, also mit einem Azimutwinkel um 90° eintreffen, liegt die Ortungsschärfe bei 4° bis 5° (vgl. Friesecke, 2007, S. 134). Für die zeitliche Auflösung des Gehörs in der Horizontalebene gilt: „Ein Richtungswechsel von links nach rechts wird nach etwa 150 ms, von vorn nach hinten nach etwa 250 ms wahrgenommen“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 131). Bei einer bewegten Schallquelle liegt der kleinste wahrnehmbare Winkelunterschied, abhängig von der Geschwindigkeit der Schallquelle und der Signalform, zwischen 5° und 20° (vgl. Lindau & Weinzierl, 2008, S. 547).

#### 2.3.1.1 Interaurale Laufzeitdifferenzen

Bei seitlichem Schalleinfall entstehen unterschiedlich lange Schallwege von der Schallquelle zu den beiden Ohren. Abbildung 8 zeigt beispielhaft einige ITDs (siehe 2.3.1) in Abhängigkeit vom Azimutwinkel:

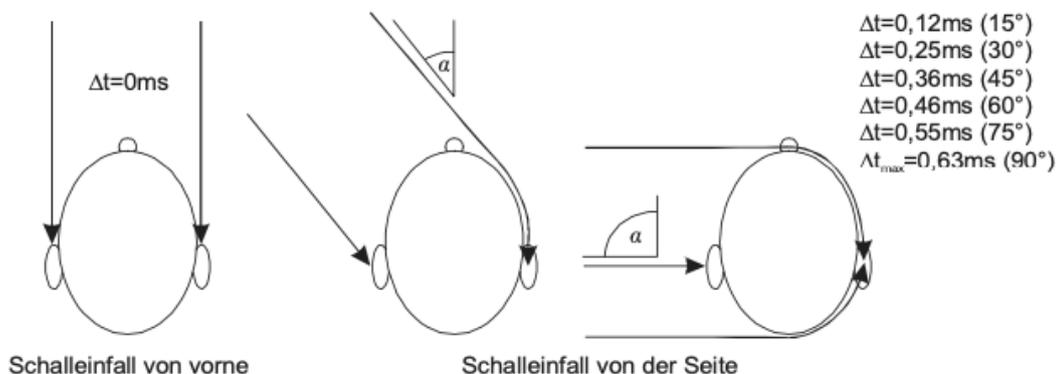


Abbildung 8: Lokalisation durch Laufzeitunterschiede (Friesecke, 2007, S. 135)

Bei frontalem Schalleinfall mit Azimut  $0^\circ$  beträgt die ITD 0 ms. Je größer der Azimutwinkel ist, desto größer ist die ITD. Die kleinste wahrnehmbare ITD liegt bei 0,01 ms und entspricht  $1^\circ$ . „Allerdings lässt sich erst ab etwa  $30\mu\text{s}$  eine tatsächlich neue Position einer Schallquelle sicher bestimmen“ (Friesecke, 2007, S. 134). Das entspricht einer Ortungsschärfe von  $2^\circ$  bis  $3^\circ$  und „einem Schallwegunterschied von nur 1 cm“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 129). Bei einem Schalleinfall mit Azimut  $90^\circ$  ist der Schallumweg und somit auch die ITD maximal und beträgt je nach Größe des Kopfes etwa 0,63 ms (siehe Abbildung 8). Dabei gibt es zwei mögliche Schallumwege: ein längerer Umweg führt über die Stirn, ein kürzerer über den Hinterkopf. Da beide Umwege eine andere ITD haben, kann das Gehirn die ITDs nicht eindeutig auswerten und es entsteht, im Vergleich zur frontalen Lokalisation, eine deutlich schlechtere Ortungsschärfe von etwa  $12^\circ$  bis  $18^\circ$  (vgl. Friesecke, 2007, S. 135).

### **2.3.1.2 Interaurale Pegeldifferenzen**

Bei seitlichem Schalleinfall entsteht ein Druckstau am Schall zugewandten Ohr und ein Schallschatten am Schall abgewandten Ohr. Dadurch werden ILDs (siehe 2.3.1) hervorgerufen. Eine Schallreflexion und die daraus resultierende Abschattung des abgewandten Ohres entsteht nur bei Signalen, deren Wellenlänge kleiner als der Kopfdurchmesser ist, dies gilt für Frequenzen ab etwa 2 kHz. Andernfalls wird der Schall um den Kopf gebeugt (vgl. Görne, 2015, S. 126f). Eine Beugung entsteht bei Frequenzen „unter 300 Hz“ (Friesecke, 2007, S. 135). Dabei sollte beachtet werden, dass die interauralen Pegeldifferenzen „gegenüber [den] interauralen Laufzeitdifferenzen [...] von geringerer Bedeutung für die Lokalisierung von Schallquellen [sind]“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 130).

### **2.3.1.3 Grenzen der horizontalen Lokalisation**

Tiefe Töne können besonders schlecht lokalisiert werden. Bei Frequenzen „unter 300Hz kann sich der Schall um den Kopf herum beugen und es sind demzufolge kaum noch Pegelunterschiede vorhanden“ (Friesecke, 2007, S. 135). Je tiefer die Frequenz ist, desto kleiner ist der Phasenwinkel zwischen den beiden Ohrsignalen bei gleichbleibendem Laufzeitunterschied. „Zu tiefen Frequenzen hin werden die Schwingungsverläufe an beiden Ohren immer ähnlicher“ (ebd., S. 135). Sowohl bei ILDs, als auch bei ITDs nimmt die Lokalisationsschärfe zu tiefen Frequenzen hin also ab.

Hohe Töne können schlecht über ITDs lokalisiert werden. Eine relativ hohe ITD von 0,61 ms entspricht einer Frequenz von etwa 1,63 kHz. Ist die Wellenlänge größer als der Umweg, kann ein Signal eindeutig lokalisiert werden. Bei hochfrequenten Signalen ist die Periodendauer kürzer als die ITD. Das führt dazu, dass die Ohrsignale phasengleich sind und mehrdeutige Richtungsinformationen liefern. Eine solche Fehlortung tritt bei sinusartigen, stehenden Tönen mit einer Signalfrequenz größer als 1,6 kHz auf (vgl. Görne, 2015, S. 126f). „Signalanteile oberhalb 1,6 kHz werden auf Grund von Zeitdifferenzen der Hüllkurven und nicht solchen der Feinstruktur dieser Signale lateralisiert“

(Weinzierl, 2008, S. 96). Oberhalb von 1,6 kHz wertet das Gehirn also nicht mehr den Schwingungsverlauf eines Signals, sondern nur noch dessen Lautstärkeverlauf aus.

Daraus lässt sich schließen, dass es sowohl bei sehr tiefen, als auch bei sehr hohen Frequenzen zu Fehlortungen kommen kann. Breitbandige Signale, die verschiedenen Frequenzen enthalten sind deshalb besser zu orten, als sinusartige Signale, die nur wenige Frequenzen enthalten. Je obertonreicher ein Signal ist, desto besser kann es lokalisiert werden.

Außerdem gilt, dass impulsartige bzw. zeitveränderliche Signale besser ortbar sind als stehende Töne bzw. gleichförmige Signale. Bei einem Impulssignal gibt es keine Schwingung, die sich wiederholt, die beiden Ohrsignale sind dadurch nie phasengleich und es wird eine exakte Lokalisation gewährleistet. Görne begründet die verbesserte Ortbarkeit eines Impulssignals gegenüber eines gleichförmigen Signals aus einer anderen Perspektive: „Evolutionsgeschichtlich betrachtet sind gleichförmige Signale weniger gefährlich als beispielsweise der knackende Ast“ (Görne, 2015, S. 127).

Die interaurale Hörwahrnehmung ist nicht die einzige Art, Schall zu lokalisieren. „Menschen können eine Schallquelle lokalisieren, selbst wenn eine einseitige Hörbehinderung besteht“ (Sengpiel, 2008). Selbst mit einem gesunden Gehör lässt sich eine Schallquelle nicht eindeutig orten: „eine Schallquelle schräg rechts vorne erzeugt die gleichen interauralen Signaldifferenzen wie eine Schallquelle schräg rechts hinten“ (Görne, 2015, S. 126).

ITDs und ILDs alleine ermöglichen also keine eindeutige Verortung einer Schallquelle. Häufig sind etwa Vorne-Hinten-Vertauschungen oder Oben-Unten-Vertauschungen. Abbildung 9 zeigt den so genannten „Cone of Confusion“:

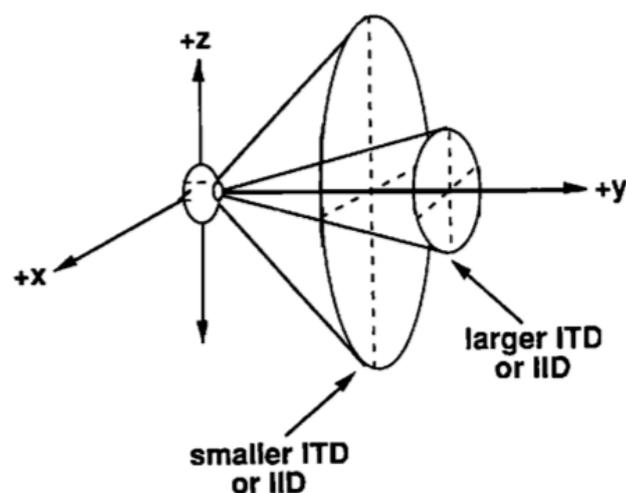


Abbildung 9: Cone of Confusion (Wenzel, Arruada, Kistler, & Wightman, 1993, S. 113)

Alle Schallquellen, die auf der Kegeloberfläche liegen, können nicht eindeutig durch ITDs und ILDs lokalisiert werden. Die Darstellung verdeutlicht außerdem, warum die Peilbewegungen des Kopfes (siehe 2.3) so wichtig für die Verortung einer Schallquelle sind: Wird der Kopf geneigt, dann ändert sich der Bereich der von der „Cone of Confusion“ abgedeckt wird (vgl. Wenzel, Arruada, Kistler, & Wightman, 1993, S. 112).

### 2.3.2 Lokalisation in der Medianebene

Die Hörwahrnehmung in der Medianebene erfolgt monaural, also einohrig. Sie bezieht sich auf die Höhe bzw. Elevation einer Schallquelle. Da der Abstand der beiden Ohren zu einer Schallquelle auf der Medianebene stets gleich ist, sind auch die beiden Ohrsignale sehr ähnlich. Die Richtungsinformation über einen Schalleinfall von vorne, oben oder hinten wird also monaural und nicht aus der Differenz der beiden Ohrsignal gewonnen (vgl. Weinzierl, 2008, S. 94). Bei der Ortung eines Schalleignisses auf der Medianebene „liegt [die Lokalisationsunschärfe] bei etwa 10°“ (Friesecke, 2007, S. 135).

1983 veröffentlichte Jens Blauert die Ergebnisse eines Experimentes, bei dem unterschiedliche Versuchspersonen mit mehreren schmalbandigen Rauschsignalen beschallt wurden. Die Beschallung erfolgte aus verschiedenen Elevationswinkeln und führte zur folgenden Schlussfolgerung: „Die Hörereignisrichtung hängt nicht von der Schalleinfallrichtung, sondern überwiegend von der Terzmittefrequenz ab“ (Weinzierl, 2008, S. 94).

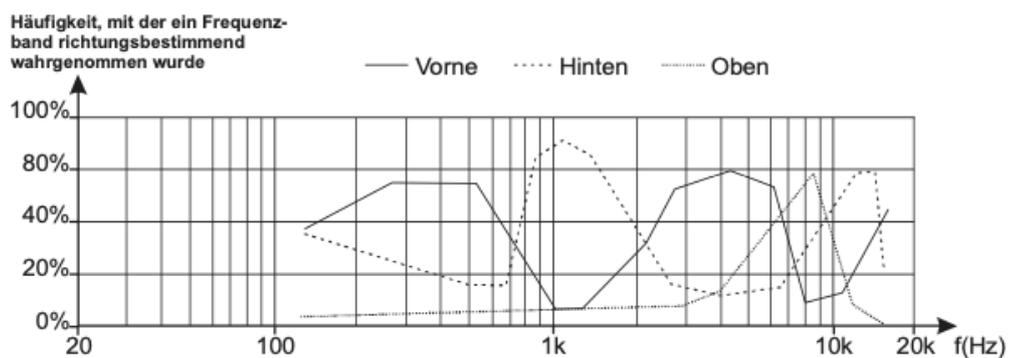


Abbildung 10: Richtungsbestimmende Bänder (Friesecke, 2007, S. 136)

Abbildung 10 veranschaulicht die Ergebnisse des Experimentes und zeigt die richtungsbestimmenden Bänder eines Signalspektrums: Ein Signal wird als von vorne kommend wahrgenommen, wenn die Frequenzanteile von 250 bis 500 Hz und von 2 bis 7 kHz erhöht sind. Sind die Frequenzen im Bereich bei 1 kHz und von 12 bis 16 kHz relativ laut, dann wird das Signal als von hinten kommend wahrgenommen. Ein Signal, das von oben kommt, hat eine Frequenzerhöhung bei 8 kHz.

Darüber hinaus entstehen durch Reflexionen an der Ohrmuschel Frequenzeinbrüche im Spektrum des Ohrsignals. Je nach Elevation der Schallquelle treten die Einbrüche bei unterschiedlichen Fre-

quenzen auf: für ein Signal auf Ohrhöhe bei 8 kHz, für ein Signal oberhalb der Ohrhöhe bei 10 bis 11 kHz und für ein Signal unterhalb der Ohrhöhe für 6 bis 7 kHz (vgl. Friesecke, 2007, S. 136). Abbildung 11 zeigt eine Übersicht der richtungsbestimmenden Bänder:

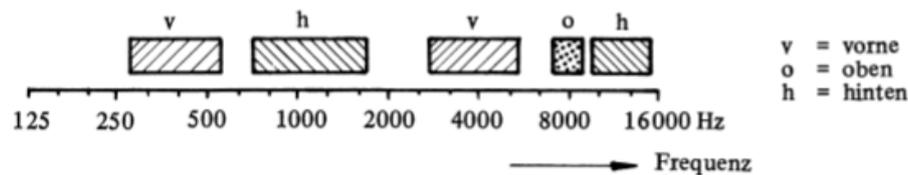


Abbildung 11: Veranschaulichung der richtungsbestimmenden Bänder (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 131)

### 2.3.3 Entfernungslokalisation

Um zu erkennen, wie weit eine Schallquelle entfernt ist, verarbeitet das Gehirn einerseits Amplituden- und Frequenzgang eines Signals und andererseits Informationen, die der Nachhall des Signals liefert. Eine Schallquelle erzeugt aber nur dann einen Nachhall, wenn sie sich in einem Raum befindet. Deshalb ist es „[nahezu unmöglich] eine gänzlich unbekannte Schallquelle im Freien in der Entfernung einzuschätzen“ (Friesecke, 2007, S. 143).

Für den Amplitudengang des Signals gilt, dass sich der Signalpegel pro Abstandsverdopplung um 6 dB verringert (vgl. Weinzierl, 2008, S. 99). Frequenzen über 1 kHz werden mit zunehmender Entfernung durch Luftabsorption gedämpft. Gleichzeitig tritt ein ähnlicher Effekt wie beim Nahbesprechungseffekt bei Druckgradientenmikrofonen (siehe Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 143ff) auf: „Die tiefen Frequenzen wirken in unmittelbarer Nähe einer Schallquelle stärker, als in großer Entfernung“ (Friesecke, 2007, S. 144).

Als „Erste Reflexionen“ werden die sechs Reflexionen der Schallquelle bezeichnet, die in einem rechteckigen Raum von den vier Seitenwänden, Decke und Boden erzeugt werden. Bei der Entfernungswahrnehmung wird der Zeitpunkt des Eintreffens des Direktschalls und der ersten Reflexionen verglichen. Befindet sich die Schallquelle relativ nah am Schallempfänger, dann ist der zeitliche Abstand zwischen Direktschall und den ersten Reflexionen relativ groß. Der Direktschall trifft deutlich früher als die ersten Reflexionen ein. Je weiter sich eine Schallquelle vom Schallempfänger oder -empfängerin entfernt, desto kleiner wird der zeitliche Abstand zwischen Direktschall und den ersten Reflexionen. Direktschall und erste Reflexionen treffen fast zeitgleich ein (vgl. ebd., S. 143).

Als Diffusschall wird der Nachhall bezeichnet, der nach den ersten Reflexionen das Ohr erreicht. Der Direktschallpegel nimmt mit zunehmender Entfernung ab, wohingegen „der Pegel des diffusen Schalls unabhängig von der Schallquellenentfernung konstant bleibt“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 133). Je größer also der Anteil des Diffusschalls ist, desto weiter ist die Schallquelle entfernt.

Die Entfernungswahrnehmung ist deutlich komplexer als die Lokalisation in der Horizontal- und Medianebene, da das Gehirn „mehrere Merkmale des Schallsignals, die in komplexer Weise auf die

Entfernungswahrnehmung einwirken“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 132) auswerten muss. Gerade deshalb sind die Hörerfahrung und die Bekanntheit eines Signals hierfür besonders wichtig (vgl. Weinzierl, 2008, S. 143).

#### **2.3.4 Lokalisation bei verzögerten Schallereignissen aus verschiedenen Richtungen**

Je nach Beschaffenheit des Raumes können Direktschall und Nachhall, vor allem die ersten Reflexionen sehr ähnlich sein. Alle Schallereignisse treffen unmittelbar hintereinander aus unterschiedlichen Richtungen auf das Ohr. Trotzdem kann die Schallquelle lokalisiert werden. Es gilt der Präzedenzeffekt bzw. das Gesetz der ersten Wellenfront: „Eine Schallquelle wird stets in der Richtung geortet, aus der die erste Wellenfront den Kopf erreicht“ (Görne, 2015, S. 128).

Der Präzedenzeffekt gilt ab einer Signalverzögerung von etwa 2 ms zwischen den Zeitpunkten, an denen Direktschall und Nachhall am Ohr eintreffen. Bei Werten unter 2 ms interpretiert das Gehör die Signalverzögerung als ITD (siehe 2.3.1.1). Die Lokalisationsrichtung des Signals verschiebt sich und es entsteht eine Phantomschallquelle. Ab einer Signalverzögerung von etwa 30 ms wird der Sekundärschall als Echo wahrgenommen.

Ab wann der Präzedenzeffekt eintritt und der Sekundärschall als Echo wahrgenommen wird, hängt stark von der Signalform ab. Ein kurzes impulshaftes Signal wird schon bei kurzen Verzögerungszeiten als Echo wahrgenommen. Bei einer statischen Klangfläche sind die Verzögerungszeiten zwischen Primär- und Sekundärschall weniger deutlich, deshalb liegt auch die Echschwelle deutlich höher (vgl. Friesecke, 2007, S. 139ff).

#### **2.3.5 Auditorische Signalverarbeitung**

Jedes Ohr übermittelt eine Folge von Nervenimpulsen, die das Gehirn verarbeitet. Theile beschreibt diese Signalverarbeitung anhand des Assoziationsmodells, das er 1980 aufstellte. Demnach verarbeitet das Gehör ein Ohrsignal auf zwei Ebenen: der Ortsassoziationsstufe und der Gestaltassoziationsstufe. Das Gehirn assoziiert die Ohrsignale dabei stets mit bereits gelernten Reizmustern und Erfahrungen. Obwohl beide Stufen zusammen die Eigenschaften des Hörereignisses bestimmen, findet die Verarbeitung der Information über Ort und Signal unabhängig voneinander statt (vgl. Theile, 1980, S. 23ff). Abbildung 12 zeigt den schematischen Aufbau des Lokalisationsmodells.

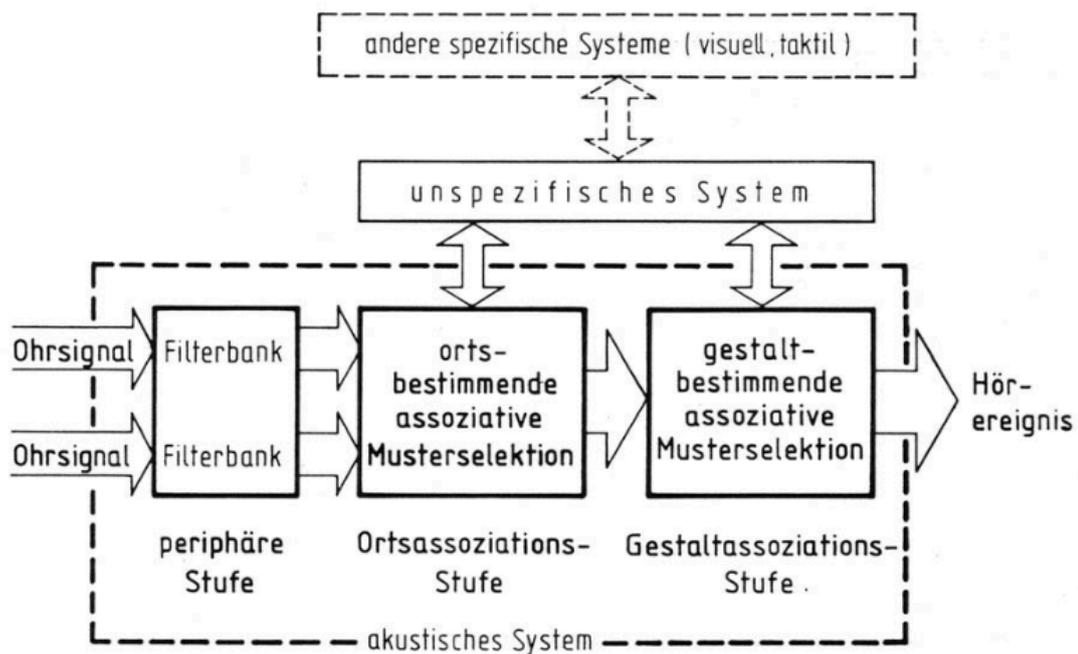


Abbildung 12: Funktionsprinzip des Assoziationsmodells (Theile, 1980, S.35)

Wie in den vorigen Abschnitten dieses Kapitels beschrieben, ermöglicht die Richtcharakteristik des Außenohrs die Lokalisation einer oder mehrerer Schallquellen. Diese richtungsgebenden Informationen werden in der Ortsassoziationsstufe ausgewertet. In der Gestaltassoziationsstufe verarbeitet das Gehirn die restlichen Eigenschaften wie die Klangfarbe des Signals. „Jeder der beiden Verarbeitungsprozesse geschieht auf dem Wege einer assoziativ gesteuerten Musterselektion: bestimmte Ohrsignale lösen nach der peripheren Verarbeitung in der ersten zentralen Stufe eine Ortsassoziation und in der höher- gelagerten Stufe eine Gestaltassoziation aus“ (Theile, 1980, S. 35). Im ersten Verarbeitungsschritt der Gestaltassoziationsstufe analysiert das Gehirn die Signalform eines Ohrsignals und vergleicht es mit dem anderen Ohrsignal. So können Merkmale und Muster von linkem und rechten Ohr zusammengeführt werden und die beiden separaten Ohrsignale werden einem Schallereignis zugeordnet. „Das Gehirn konzentriert sich im Moment des bewussten Zuhörens auf einen bestimmten Zeitabstand und Pegelunterschied der Musterfolgen zwischen linkem und rechten Ohr“ (Friesecke, 2007, S. 137). Auf diese Weise ist es möglich, dass sich die Hörerin oder der Hörer auf eine Schallquelle konzentrieren kann, obwohl im Alltag oft eine Vielzahl von Schallreizen gleichzeitig vorhanden sind. Dieses Phänomen wird als „Cocktailparty-Effekt“ bezeichnet. Dieser Effekt besagt, dass „ein Nutzsignal, das aus einer bestimmten Richtung eintrifft, von einem Störsignal, das aus einer anderen Richtung eintrifft, bei zweiohrigem Hören weniger stark verdeckt wird als bei einohrigem Hören“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 238). Dadurch ist es möglich, einem bestimmten Gespräch zu folgen und alle anderen Schallereignisse einer Geräuschkulisse, wie bei-

spielsweise auf einer Cocktailparty, auszublenzen. Fehlt das Signal eines Ohres, ist es nicht mehr möglich, die Musterfolgen beider Ohrsignale zu vergleichen. Das bewusste Zuhören ist deutlich schwieriger.

In einem weiteren Verarbeitungsschritt wird ein Schallereignis mit dem Klang bereits „erlernter“ Schallquellen verknüpft. „An dieser Stelle wird zum Beispiel unterschieden, welches Instrument man gerade hört, welcher Teil eines Wortes gerade gesprochen wird und, ob die Person bekannt ist, die das Wort spricht“ (Friesecke, 2007, S. 137).

Bei allen Schritten der auditorischen Signalverarbeitung werden die Ohrsignale stets mit anderen Sinneseindrücken in Zusammenhang gebracht. So vergleicht das Gehirn die Ohrsignale beispielsweise mit den Wahrnehmungen des visuellen und taktilen Systems (siehe Abbildung 12). Dieser Vorgang beeinflusst nicht nur die Lokalisation einer Schallquelle in der Ortsassoziationsstufe (siehe 2.3), sondern auch die allgemeine Wahrnehmung und Zuordnung der Ohrsignale zu einem Ereignis in der Gestaltassoziationsstufe (vgl. ebd., S. 137). Alle Sinneseindrücke zusammen erzeugen ein Bild der Umwelt, wobei die Informationen bestimmter Sinne je nach Situation mehr oder weniger Einfluss haben. Beim Hören können beispielsweise „auch visuelle Informationen [...] dominieren, man denke z.B. an den Bauchredner-Effekt“ (Weinzierl, 2008, S. 89).

### **3 Binauraltechnik**

Die Binauraltechnik ermöglicht die Wiedergabe einer dreidimensionalen Audiomischung über Kopfhörer. Dabei entsteht ein realistischer Höreindruck, der dem natürlichen Hören entspricht. Obwohl die Binauraltechnik bereits seit vielen Jahren Forschungsgegenstand ist, wurde sie bis heute nur in wenigen Audioproduktionen verwendet. In den letzten Jahren begünstigten allerdings einige aktuelle Entwicklungen ihre Anwendung in kommerziellen Medien: „In particular, the wide use of headphone listening, the sophistication and cheapness of signal processing, and the dramatic growth of virtual reality (VR) systems, have provided more of an impetus to integrate this technology into practical systems” (Rumsey, 2017, S. 524).

#### **3.1 Unterschiede zwischen Stereo- und Binauralmischungen**

Eine herkömmliche Stereoproduktion ist für die Wiedergabe auf zwei Lautsprechern optimiert. Da sich die Signale der beiden Lautsprecher überlagern, entsteht ein komplexes Schallfeld, das dem natürlichen räumlichen Hören entspricht, man spricht von raumbezogener Stereophonie (vgl. Görne, 2015, S. 130).

##### **3.1.1 Im-Kopf-Lokalisation**

Wird eine Stereoproduktion über Kopfhörer wiedergegeben, dann ändert sich die Wahrnehmung des Klangbildes. Das wird am deutlichsten, wenn beispielsweise ein Signal im Panorama extrem nach rechts gepannt ist. Bei der Lautsprecherwiedergabe wird das Signal zwar nur über die rechte Box wiedergegeben, kommt aber trotzdem an beiden Ohren an. Das Gehirn verarbeitet beide Ohrsignale und lokalisiert die Schallquelle rechts. Bei der Kopfhörerwiedergabe wird das Signal ausschließlich vom rechten Ohr gehört. Ein solcher Klangeindruck existiert beim natürlichen Hören nicht (vgl. Friesecke, 2007, S. 136). Wie in Abschnitt 2.3.5 beschrieben, vergleicht das Gehirn wahrgenommene Schallereignisse permanent mit bereits bekannten und „erlernten“ Klangeindrücken und ordnet diese einander zu. Tritt eine Schallquelle ausschließlich im rechten Ohrsignal auf, dann entsteht ein völlig unbekanntes Klangbild, das das Gehirn nicht verarbeiten kann. Infolgedessen wird die Schallquelle bei der Kopfhörerwiedergabe nicht im Raum, also außerhalb des Kopfes, sondern im Kopf zwischen den beiden Ohren lokalisiert. Man spricht deshalb von Im-Kopf-Lokalisation (vgl. Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 134).

Die räumliche Wahrnehmung einer Stereoproduktion bei der Kopfhörerwiedergabe ist eine völlig andere wie bei der raumbezogenen Stereophonie. Wie in Abschnitt 2.3 beschrieben, entstehen beim natürlichen Hören für jede Schalleinfallrichtung typische Signalfärbungen, die die Lokalisation einer Schallquelle ermöglichen. Diese richtungsgebenden Klangfärbungen sind bei der Kopfhörerwiedergabe nicht vorhanden. Eine Tiefenstaffelung der Elemente in der Mischung ist so nur eingeschränkt möglich (vgl. Görne, 2015, S. 131).

Zwar entsteht ein Stereopanorama im Kopf, zwischen den beiden Ohren, allerdings ist dieses Klangbild nicht mit dem Klangbild bei der Wiedergabe über Lautsprecher kompatibel. Bei der Lautsprecherwiedergabe befinden sich die Phantomschallquellen auf einer geraden Linie zwischen den beiden Lautsprechern. Bei der Verwendung von Kopfhörern und der daraus resultierenden Im-Kopf-Lokalisation ist diese Linie mittig in Richtung Schädeldecke gebogen. Phantomschallquellen in der Mitte des Stereopanoramas erscheinen also räumlich höher, als seitliche Phantomschallquellen. Auch bei der seitlichen Auslenkung von Phantomschallquellen gibt es Unterschiede zwischen der Lautsprecher- und Kopfhörerwiedergabe: „Während bei Lautsprecherwiedergabe der Hörereignisort bereits bei rund 15 dB Pegeldifferenz zwischen den Lautsprechersignalen ganz zur Seite der Lautsprecherbasis auswandert, wird bei Kopfhörerwiedergabe eine wesentlich größere Pegeldifferenz für eine Abbildung ganz seitlich benötigt“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 350).

### **3.1.2 Ziele der Binauraltechnik**

Angenommen ein Zuhörer besucht ein Orchesterkonzert in einer Konzerthalle. Er hört die einzelnen Instrumente und deren Reflexionen an den Wänden des Raumes. Dadurch verortet sich der Zuhörer selbst im Zuschauerraum vor der Bühne. Sein Gehirn lokalisiert die einzelnen Instrumente und erschafft ein Klangbild des Orchesters auf der Bühne. Alle Schallereignisse finden außerhalb des Kopfes und nicht im Kopf statt (vgl. Møller, 1992, S. 175f). Eine Binauralmischung soll diesen Klangeindruck reproduzieren: „The task of a binaural recording and playback system is to present the correct inputs to the hearing, that is to reproduce the eardrum signals correctly“ (Møller, 1992, S. 176). Am Beispiel des Orchesterkonzerts ist es also das Ziel von Binauraltechnik, den Klangeindruck genauso aufzunehmen und zu reproduzieren, wie ihn der Zuhörer beim Konzert wahrnimmt. Einer der ersten Versuche dies zu realisieren war die Kunstkopfaufnahme.

## **3.2 Kunstkopfstereofonie**

Die Idee bei der Kunstkopfstereofonie ist es, das Schallsignal genau an dem Punkt aufzunehmen, an dem es alle Richtungsinformationen zur räumlichen Ortung enthält. Bei der Originalkopfmikrofonie werden kleine Mikrofone in den Ohren des Zuhörers zu platzieren, sie wird in Abschnitt 3.2.6 genauer beschrieben. Alternativ verwendet man ein Kunstkopfmikrofon (engl.: „manikin“, „dummy head“). Der Kunstkopf ist eine künstliche Nachbildung des menschlichen Kopfes mit einem detailliert modellierten Gehör. Einige Kunstköpfe haben zusätzlich einen Torso, um die Schallbeeinflussung durch Schultern und Korpus eines Menschen zu simulieren. Einige Millimeter innerhalb der Ohrmuschel des Kunstkopfes sind zwei Mikrofone platziert, die die beiden Ohrsignale aufnehmen. Diese kopfbezogene Stereoaufnahme wird anschließend über Kopfhörer wiedergegeben und reproduziert genau das Klangbild, das ein realer Mensch am Aufnahmeort wahrgenommen hätte (vgl. Møller, 1992, S. 172).

### 3.2.1 Geschichte der Kunstkopfstereofonie

Die Kunstkopfstereofonie existiert bereits seit den 1930er-Jahren. Nach Jahren der Forschungsarbeit wurden in den 1970er-Jahren erste Kunstkopfaufnahmen mit dem *Neumann KU80* durchgeführt und die Kunstkopfstereofonie wurde für Medienproduktion eingesetzt. „Das erste im deutschen Rundfunk ausgestrahlte Hörspiel in Kunstkopfstereofonie war zur Funkausstellung 1973 in Berlin die RIAS/BR/WDR-Produktion ‚Demolition‘ nach dem Science-Fiktion-Roman von Alfred Bester ‚The Demolished Man‘, 1951“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 346). Mit dem *Neumann KU80* ließen sich vor allem Raumeindruck und Schallquellenentfernungen natürlich darstellen. Die Lokalisation von Schallquellen in der Horizontalebene funktionierte schlecht, außerdem kam es zu Vorne-Hinten-Vertauschungen (siehe 2.3.1.3). Diese Nachteile wurden beim *Neumann KU81*, dem Nachfolgermodell, das in den 1980er-Jahren auf den Markt kam, verbessert. Hierbei wurde die Form der Ohrmuschel detaillierter nachgebildet, wodurch sich die Darstellung der monauralen Hörwahrnehmung erheblich verbesserte und somit weniger häufiger Fehlortungen in der Medianebene entstanden (vgl. ebd., S. 354f). Außerdem wurden die Mikrofonensignale diffusfeldentzerrt (siehe 3.2.3). Bis heute ist der *Neumann KU100*, der Nachfolger des *KU81*, eines der Standardmikrofone für die Kunstkopfaufnahme.

### 3.2.2 Eigenschaften von Kunstkopfmikrofonen

Das Ziel der Kunstkopfstereofonie ist es, den aufgenommenen Klangeindruck möglichst universell zu reproduzieren. Das Kunstkopfmikrofon muss also so konstruiert sein, dass es der Anatomie möglichst vieler Menschen entspricht (siehe 2.2). Deshalb sind „die meisten Kunstkopfsysteme [...] ‚Durchschnittsköpfe‘, die sich an Mittelwerten aus anthropometrischen Datenbanken orientieren“ (Weinzierl, 2008, S. 587). Als Mikrofone werden üblicherweise zwei Sondenmikrofone verwendet. Für die Mikrofonposition innerhalb des künstlichen Ohrkanals gelten die gleichen Gesetzmäßigkeiten wie bei der HRTF-Messung (siehe 2.2). Deshalb werden die „Miniaturmikrofon[e] bündig einige Millimeter innerhalb des Ohrkanals eingesetzt“ (ebd., S. 586f).

### 3.2.3 Entzerrung des Kunstkopfmikrofons

Druckmikrofone können freifeld- oder diffusfeldentzerrt sein. „Für frontalen Schalleinfall, also bei 0°-Schalleinfallswinkel erhält man infolge des Druckstaus vor der Membran für hohe Frequenzen eine Anhebung in der Frequenzkurve“ (Sengpiel, 2009). Diese Anhebung wird bei freifeldentzerrten Mikrofonen durch entsprechende Filter ausgeglichen, der Verlauf der Frequenzkurve ist im Direktfeld bei frontalem Schalleinfall linear. Da die Höhenabsenkung auch bei diffusem Schalleinfall wirkt, wirkt der Klang des Diffusschalls bei der Aufnahme dumpfer. Bei einem diffusfeldentzerrten Mikrofon wird auf die Höhenabsenkung verzichtet: aufgrund des Druckstaus bei frontalem Schalleinfall gibt es im Freifeld eine starke Anhebung im Bereich von 8 kHz, dafür ist die Frequenzkurve im Diffusfeld linear (vgl. ebd.).

Ein freifeldentzerrtes Kunstkopfmikrofon hat bei frontaler Beschallung die gleiche Übertragungsfunktion wie ein freifeldentzerrtes Messmikrofon, ist aber fehlerhaft für alle anderen Schalleinfallrichtungen. Die Übertragungsfunktion eines diffusfeldentzerrten Kunstkopfmikrofons entspricht der eines diffusfeldentzerrten Messmikrofons im Diffusfeld. Bei Kunstkopfmikrofonen hat sich gezeigt, „daß das meßtechnisch klar definierte Kriterium ‚ebener Diffusfeldfrequenzgang‘ sehr gut mit der Eigenschaft ‚hohe Klangfarbentreue‘ korrespondiert“ (Theile, 1981, S. 151). Die Diffusfeldentzerrung entspricht einer Mittelung aller Schalleinfallrichtungen. Daniel et al schreiben, dass „die Anwendung eines diffusfeldentzerrten Kunstkopfes [...] etwas mehr Praxisrelevanz aufweist. Übliche Räume zeigen eine Vielzahl von Wandreflexionen und ähneln somit eher einem Diffusfeld als einem Freifeld“ (Daniel, 2007, S. 8). Durch die Diffusfeldentzerrung konnte die Lokalisationsschärfe und Klangfarbe der Kunstkopfaufnahme und vor allem die Kompatibilität zur Lautsprecherwiedergabe im Vergleich zu einem freifeldentzerrten Kunstkopfmikrofon erheblich verbessert werden (vgl. Theile, 1981, S. 151f) Bei der Wiedergabe der Aufnahme ist zu beachten, dass auch der Kopfhörer diffusfeldentzerrt sein sollte, dies wird in Abschnitt 3.3.3 genauer erläutert.

### 3.2.4 Nachträgliche Kunstkopfaufnahme einer raumbezogenen Aufnahme

Im vorherigen Abschnitt wurde stets davon ausgegangen, dass das Kunstkopfmikrofon direkt am Aufnahmeort verwendet wird. Grundsätzlich eignet sich ein Kunstkopfmikrofon aber auch, um eine raumbezogene Aufnahme (siehe 3.1) nachträglich in ein kopfbezogenes Signal zu wandeln. Dazu wird die raumbezogene Stereoaufnahme in einem Aufnahme- oder Regieraum über ein geeignetes Lautsprecherpaar abgespielt und mit dem Kunstkopfmikrofon aufgenommen. Die Qualität der Kunstkopfaufnahme wird dabei stark von der Räumlichkeit und der Wiedergabequalität des abgespielten Signals beeinflusst (vgl. Görne, 2015, S. 131). Weitere Einschränkungen werden in Abschnitt 3.2.7 beschrieben.

### 3.2.5 Beispiele für Kunstkopfmikrofone



Abbildung 13: Head Acoustics HMS III (links), Neumann KU 100 (rechts) (Weinzierl, 2008, S. 588)

Abbildung 13 zeigt beispielhaft zwei Kunstkopfmikrofone. Die Firma Neumann gehörte, wie in Abschnitt 3.2.1 beschrieben, zu den ersten Herstellern, die Kunstkopfmikrofone zum Einsatz für Medienproduktionen entwickelten. Der *Neumann KU 100* gehört zur dritten Kunstkopfgeneration der Firma und „wird überwiegend für Musik- und Sprachaufnahmen eingesetzt“ (Weinzierl, 2008, S. 587). Beim *HMS III* der Firma Head Acoustics ist nicht nur der Kopf, sondern auch ein Teil des Torsos modelliert, um dessen Einfluss auf die Richtungsinformation des Schalls zu simulieren (siehe 2.2).

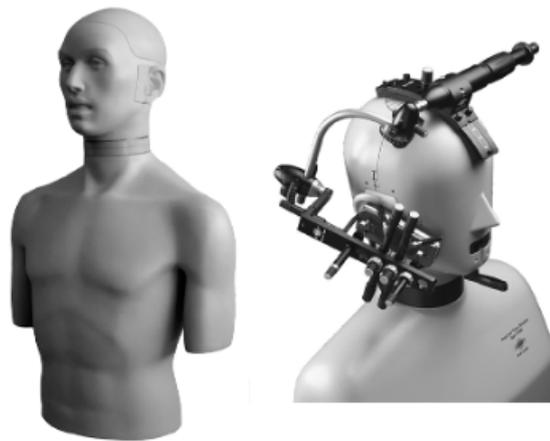


Abbildung 14: KEMAR KB 4004 (links), Brüel & Kjær HATS 4128 (rechts) (Weinzierl, 2008, S. 588)

Bei den beiden Kunstkopfmikrofonen in Abbildung 14 lassen sich noch genauere Aufnahmeeinstellungen festlegen: Das Außenohr des *KEMAR KB 4004* ist austauschbar. Der Hersteller bietet verschiedene Ohrtypen an, die an den Durchschnittsköpfen von Menschen unterschiedlichen Geschlechts und geographischer Herkunft angepasst sind. Das Mikrofon kann so schon vor der Aufnahme an die Physiognomie der Rezipientin oder des Rezipienten angepasst werden. Der Kunstkopf *HATS 4128* von Brüel & Kjær ist für messtechnische Aufnahmen optimiert. Neben einer Sprachsimulation ist es mit dem *HATS 4128* möglich, die Ohrkanal- und Trommelfellimpedanz zu simulieren. Durch eine Erweiterung mit speziellen Halterungen lässt sich beispielsweise die Nutzung von Mobiltelefonen audiomesstechnisch untersuchen (vgl. ebd., S. 587ff).

### 3.2.6 Originalkopfmikrofon

Ein Originalkopfmikrofon basiert auf dem gleichen Prinzip wie ein Kunstkopfmikrofon. Wie ein Kunstkopfmikrofon besteht auch dieses System aus zwei Druckempfängern, die als Miniaturmikrofone in den beiden Ohren platziert werden. Ein Originalkopfmikrofon wird also von einem Menschen getragen. Dessen Kopf dient als Trennkörper zwischen den beiden Mikrofonen. Die Mikrofone sind von einem Schaumstoff ummantelt und werden in den Eingang des Gehörgangs gesteckt. Für mehr Stabilität und einen festen Sitz haben Originalkopfmikrofone wie das *DPA 4560* zusätzlich einen Nackenbügel (vgl. DPA Microphones, o.D.). Bei einigen Modellen wird Mikrofon-

und Kopfhörerfunktion kombiniert (siehe 3.3.4): das *Sennheiser Ambeo Smart Headset* ist eigentlich ein In-Ear-Kopfhörer, bei dem außen ein Mikrofon angebracht ist (vgl. Sennheiser, o.D.).

Weinzierl schreibt zu den Originalkopfmikrofonen, dass „die räumliche Zuordnung von binaural aufgenommenen und wiedergegebenen Signalen mit dem eigenen Kopf in der Regel besser [gelingt] als über einen ‚fremden‘ Kunstkopf“ (Weinzierl, 2008, S. 587). Allerdings gilt dies nur dann, wenn die Aufnahme vom Rezipienten oder der Rezipientin selbst durchgeführt wurde. Wie in Abschnitt 3.2.2 beschrieben, sollte eine kopfbezogene Aufnahme möglichst universell sein, der Klangeindruck sollte für eine möglichst große Zahl von Hörerinnen und Hörern gleich sein. Für Aufnahmen mit einem Originalkopfmikrofon trifft dies nicht zu. Der Klangeindruck ist sehr inkonsistent, da er stark davon abhängt, wie sehr der Kopf des jeweiligen Mikrofonträgers oder -trägerin dem Kopf der Konsumentin oder des Konsumenten ähnelt.

Im Gegensatz zu einem Kunstkopfmikrofon kann ein Originalkopfmikrofon prinzipbedingt nicht auf einem Stativ platziert werden. Die Aufnahme wird „durch die medizinisch heikle Positionierung eines Mikrofons im ‚lebenden Objekt‘ und dessen unwillkürliche Bewegungen beeinträchtigt“ (ebd., S. 671). Die Qualität der Aufnahme ist also unmittelbar von der Kopfposition des Mikrofonträgers während der Aufnahme abhängig.

Im folgenden Abschnitt werden Einschränkungen in Bezug auf Kunstkopfmikrofone genannt. Sie treffen gleichsam auf Originalkopfmikrofone zu.

### **3.2.7 Einschränkungen bei der Kunstkopfstereofonie**

Eine der aus kommerzieller Sicht größten Einschränkungen bei der Kunstkopfstereofonie ist die transaurale Wiedergabe, die Wiedergabe der kopfbezogenen Aufnahmen auf Lautsprechern. Zwar konnte die Qualität durch die Diffusfeldentzerrung des Kunstkopfmikrofons verbessert werden (siehe 3.2.1), trotzdem lassen sich Klangfärbungen nicht vollständig vermeiden. Der Grund dafür ist, dass das Signal zweimal durch eine HRTF gefiltert wird: Zum einen durch die HRTF des Kunstkopfes während der Aufnahme und zum anderen durch die individuelle HRTF der Zuhörerinnen und Zuhörer bei der Lautsprecherwiedergabe. Die Konsumentin oder der Konsument muss eine Kunstkopfaufnahme also zwingend über Kopfhörer wiedergeben. Diese Tatsache erschwert die Vermarktung, da der Anbieter letztlich keine Kontrolle über das konsumentenseitige Wiedergabegerät hat (vgl. Schröder, 2018a).

Zwar ist ein Kunstkopfmikrofon so konstruiert, dass es der Anatomie möglichst vieler Menschen entspricht, trotzdem bleibt die fehlende Individualanpassung von Kunstkopfaufnahmen problematisch: „Der Kunstkopf ist zwar nach Durchschnittsmaßen gearbeitet, fußt aber in der Regel dennoch auf der Kopie eines bestimmten, typischen individuellen Kopfs, von dem die individuellen Kopfmaße eines Hörers mehr oder weniger abweichen“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 347). Bereits geringe Abweichungen führen zu großen Unterschieden bei der räumlichen Wahrnehmung für verschiedene Hörerinnen und Hörern.

Ein weitere Qualitätseinschränkung wird durch die statische Position des Kunstkopfes bei der Aufnahme hervorgerufen: Wie in Abschnitt 2.3 beschrieben, sind die unbewussten Peilbewegungen des Kopfes für die korrekte Lokalisation einer Schallquelle, vor allem in der Medianebene unerlässlich. Auch bei bewussten Kopfbewegungen wird die Schallquelle beim natürlichen Hören an einer festen Stelle im Raum verortet. Bei der Wiedergabe einer statischen Kunstkopfaufnahme wiederum dreht sich das Klangbild mit der Kopfbewegung mit. Dieser Klangeindruck widerspricht dem natürlichen Hören und führt zu Fehlortungen (vgl. Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 218).

Die Nachbearbeitung einer Kunstkopfaufnahme ist nur bedingt möglich, da beispielsweise bei einer Konzertaufnahme weder die Lautstärkeverhältnisse zwischen den Instrumenten und deren räumliche Balance, noch der Raumklang des Konzertsaals manipuliert werden können. „Die Abhängigkeit von den Gegebenheiten der Aufführung ist extrem groß und in den meisten Fällen unpraktikabel“ (ebd., S. 218). Die künstlerische Gestaltungsfreiheit bei der Postproduktion ist stark eingeschränkt. Dass eine nachträgliche Anpassung der Mischung unabdingbar ist, verdeutlicht das Beispiel eines gängigen Aufnahmeverfahrens für Orchesteraufnahmen: das Stützmikrofonverfahren. Hierbei werden zusätzlich zu einem stereofonen Hauptmikrofon mehrere Einzelmikrofone sehr nah an den Instrumenten platziert, da das Signal des Hauptmikrofons alleine in der Praxis meist keine befriedigende Darstellung des gesamten Klangkörpers liefert (vgl. Weinzierl, 2008, S. 603). Die Gründe hierfür sind vielfältig: Oftmals ist der Klangkörper nicht ausbalanciert oder es sollen künstlich einzelne Instrumente hervorgehoben werden. Der Klang eines akustisch ungeeigneten Saals kann nachträglich verbessert werden, indem mehr oder weniger Direktsignale hinzugemischt werden. Außerdem besteht die Möglichkeit, ein völlig neues räumliches Klangbild zu schaffen, unabhängig vom eigentlichen Aufnahmeort (vgl. Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 269). Die gleichen Gründe die gegen ein reines Hauptmikrofonverfahren sprechen, sprechen gegen ein reines Kunstkopfmikrofonverfahren. Die fehlende Manipulierbarkeit der Aufnahme macht die Kunstkopfstereofonie sehr unflexibel.

### **3.3 Kopfhörer**

Das Prinzip der Binauraltechnik basiert auf der Wiedergabe mit Kopfhörern. „Jedem Ohr [ist] ganz definiert ein elektroakustischer Wandler zugeordnet [...], der nur in vernachlässigbarem Maße auf das jeweils andere Ohr einwirkt, Gegebenheiten, die bei der üblichen Lautsprecherwiedergabe nicht realisierbar sind“ (ebd., S. 349). Nur dadurch wird gewährleistet, dass linkes und rechtes Signal jeweils nur ein Ohr erreichen. Die Eigenschaften und Funktionsweisen von Kopfhörern sind also ein wichtiger Bestandteil eines Binauralsystems und sollen im folgenden Abschnitt genauer beschrieben werden.

#### **3.3.1 Funktionsweise von Kopfhörern**

Kopfhörer unterscheiden sich prinzipiell nicht von Kleinlautsprechern. Abgesehen von der Position der Membran direkt am Ohr „entsprechen [die] Wandlerprinzipien und [die] elektroakustischen Ei-

genschaften [eines Kopfhörers] denen eines Lautsprechers“ (ebd., S. 211). Grundsätzlich kann eine Kopfhörermembran also den kompletten hörbaren Frequenzbereich von 20 Hz bis 20 kHz wiedergeben. Trotzdem wirken Kopfhörersignale weniger druckvoll als Lautsprechersignale, da der Kopfhörer keinen Körperschall erzeugen kann. Beim natürlichen Hören werden Frequenzen unter 150 Hz vom gesamten Körper und nicht nur von den Ohren wahrgenommen werden. Dieses Gefühl fehlt bei der Kopfhörerwiedergabe (vgl. Friesecke, 2007, S. 460).

### **3.3.2 Kopfhörerarten**

Das Audiosignal wird kabelgebunden oder kabellos, beispielsweise per Bluetooth-Standard vom Endgerät an den Kopfhörer übertragen. Je nachdem, wie stark die Kopfhörer die Hörerin oder den Hörer vom Umgebungsgeräuschpegel isolieren, spricht man von einem offenen, halboffenen oder geschlossenen Kopfhörer. Ein offener Kopfhörer ist schalldurchlässig, ein geschlossener Kopfhörer schließt das Ohr nahezu vollständig von äußerem Störschall ab, man spricht von passiver Geräuschisolierung (engl. Noise Isolating). Ein halboffener Kopfhörer ist eine Mischung aus beiden Varianten (vgl. ebd., S. 460).

Kopfhörer gibt es in drei Bauformen: ohraufliegende Kopfhörer, ohrumschließende Kopfhörer und In-Ear Kopfhörer. Die ersten beiden Bauformen werden durch Ohrmuschelhörer bzw. Bügelkopfhörer realisiert, deren Ohrpolster entweder auf dem Ohr liegen oder das gesamte Ohr umschließen. Ohrumschließende Kopfhörer bieten mehr Tragekomfort, können allerdings nicht so kompakt wie ohraufliegende Kopfhörer gebaut werden. In-Ear Kopfhörer werden in das Außenohr gesteckt. Deren Schallwandler ist entweder in einem Schaumstoff-Ohrhörer oder in einer Otoplastik, deren Form individuell an den Gehörgang der Hörerin oder des Hörers angepasst ist platziert (vgl. ebd., S. 460ff). Die Forschungsergebnisse zur Binauraltechnik beziehen sich meistens auf ohraufliegende oder ohrumschließende Bügelkopfhörer, laut Laumann hat das rein praktische Gründe: „Die im Endverbrauchermarkt mittlerweile weit verbreiteten Einsteckhörer sind hingegen in der Binauraltechnik bislang wenig beachtet worden und werfen messtechnische Herausforderungen auf“ (Laumann, 2016, S. 14), da sich die Testperson die Messmikrofone, ähnlich wie Originalkopfmikrofone (siehe 3.2.6) direkt in die Ohren, unter den Ohrmuschelhörer stecken muss. Bei In-Ear Kopfhörern ist dieser Versuchsaufbau nicht ohne weiteres möglich.

### **3.3.3 Entzerrung der Kopfhörer**

Eine raumbezogene Aufnahme wird bei der Kopfhörerwiedergabe im Kopf lokalisiert (siehe 3.1.1). Theile erklärt diesen Effekt anhand seines Assoziationsmodells für das natürliche Hören (siehe 2.3.5): Da das Außenohr bei der Kopfhörerwiedergabe keinen Einfluss auf die Ohrsignale hat, gibt es keine Informationen, die das Gehirn in der Ortsassoziationsstufe verarbeiten kann. „Die Außenohr-Übertragungsfunktion [wird] durch die Kopfhörer-Übertragungsfunktion ersetzt“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 350). Umgekehrt kann durch die Kopfhörerübertragungsfunktion eine bestimmte HRTF nachgebildet werden.

Bei einem freifeldentzerrten Kopfhörer entspricht die Entzerrung einer Annäherung an die HRTF eines Monosignals mit Schalleinfallrichtung von vorne ( $0^\circ$  Azimut und  $0^\circ$  Elevation) im freien Schallfeld. Dadurch sollen Klangfarbenfehler für frontale Schallereignisse verhindert werden. Allerdings können spektrale Verfärbungen für alle Schallanteile aus anderen Einfallrichtungen auftreten (vgl. Schröder, 2018b). Die Freifeldentzerrung ist besonders für stereofone Signale problematisch, da die erzeugten spektralen Ohrsignalmerkmale nicht zu den interauralen Ohrsignalmerkmalen passen. „Bei der Kopfhörerwiedergabe existiert keine bevorzugte Schalleinfallrichtung“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 350). Die Kopfhörerentzerrung muss sich deshalb auf ein richtungsneutrales Schallfeld beziehen: das Diffusfeld. Die Übertragungsfunktion eines diffusfeldentzerrten Kopfhörers entspricht Durchschnittswerten aus allen Richtungen im reflexionsreichen, diffusen Schallfeld und nicht der HRTF einer einzigen Richtung. Dazu werden mehrere HRTF-Messungen aus unterschiedlichen Schalleinfallrichtungen gemittelt (vgl. ebd., S. 350f). Abbildung 15 zeigt einen Vergleich der Freifeld- und Diffusfeld-HRTF:

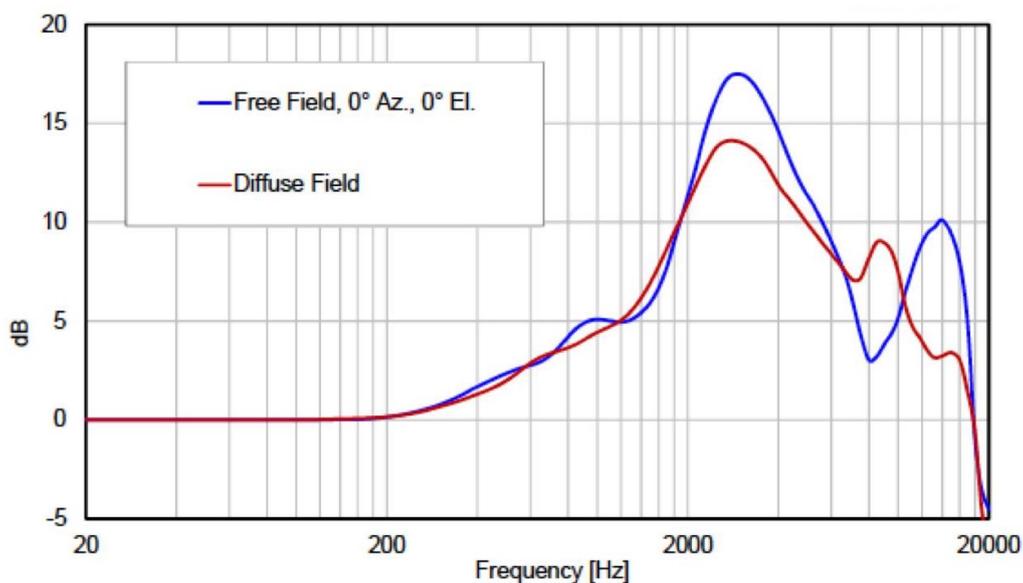


Abbildung 15: Freifeld-HRTF (blau) und Diffusfeld-HRTF (rot) (Schröder, 2018b)

Auch das Kunstkopfmikrofon sollte diffusfeldentzerrt sein (siehe 3.2.3). Die aufnahme- und wiedergabeseitige Diffusfeldentzerrung verhindert Klangfarbenfehler und sorgt für größtmögliche Klangneutralität. „Die Definition einer geeigneten Schnittstelle zwischen Aufnahme und Wiedergabe soll eine möglichst korrekte aber hörerenunabhängige Systemkette gewährleisten und das beliebige Austauschen von Komponenten auf beiden Seiten erlauben“ (Laumann, 2016, S. 15). Kunstkopfmikrofon und Kopfhörer bilden also eine Einheit. Prinzipiell sollten die gleichen spektralen Ohrsignalmerkmale vorliegen, egal ob sich der Zuhörer im Schallfeld befindet oder ob er über die Einheit aus Kunstkopfmikrofon und Kopfhörer beschallt wird (vgl. Daniel, 2007, S. 6).

Allerdings hat sich die Diffusfeldentzerrung als ein einheitlicher Bezugswert für Kopfhörer herstellerseitig nie vollständig durchgesetzt. „Kopfhörer weisen hochgradig individuelle Frequenzgänge auf, weshalb eine individuelle [sic!] Entzerrung für die Wiedergabe binauraler Signale unverzichtbar ist“ (Weinzierl, 2008, S. 674). Für die Entzerrung des Kopfhörers gibt es je nach System unterschiedliche Anwendungen. Eine Möglichkeit ist es, den Kopfhörer über die Messmikrofone in den Ohren aufzusetzen und so dessen Übertragungsfunktion zu bestimmen (siehe 3.3.2). Anhand der Messergebnisse wird ein Kompensationsfilter erstellt, der auf den individuellen Frequenzgang des Kopfhörers abgestimmt ist (vgl. Kane, 2019). Andere Systeme bieten eine Vorauswahl an Kompensationsfiltern, die auf die Frequenzgänge gängiger Kopfhörermodelle abgestimmt sind. Unter der Voraussetzung, dass die Nutzerinnen und Nutzer eines dieser Modelle verwendet, können sie so den passenden Kompensationsfilter auswählen (vgl. Waves, 2019).

### **3.3.4 Headset**

Ein Headset ist ein Kopfhörer mit integriertem Mikrofon. Auf dem Endverbrauchermarkt gibt es einige Headsets, die zum gleichzeitigen Hören und Sprechen bei Mobiltelefonie, Gaming oder Videokommunikation benutzt werden. Bei kabelgebundenen Headsets wird das Mikrofonsignal zusätzlich zum Kopfhörersignal über einen vierpoligen Klinkenstecker an das Endgerät übertragen. Drahtlose Headsets verwenden beispielsweise den Bluetooth-Standard zur Übertragung (vgl. NFON, o.D.).

Für Headsets bei der Interkom-Anwendung in der Luftfahrt ist das Mikrofon an einem Kopf- oder Nackenbügel befestigt. Aufgrund ihres geringen Gewichts werden häufig Elektretwandler eingesetzt. Je weiter seitlich das Mikrofon angebracht ist, desto stärker werden die Höhen angehoben, um eine möglichst hohe Sprachverständlichkeit zu gewährleisten (vgl. Weinzierl, 2008, S. 349). Bei kompakten In-Ear Kopfhörern ist das Mikrofon zur Telefonie oft am Kabel angebracht. Neuere Kopfhörermodelle, wie die drahtlosen In-Ear-Kopfhörer *Apple AirPods Pro*, besitzen drei Mikrofone pro Hörer: Ein nach innen gerichtetes Mikrofon analysiert das Ohrsignal, um die Entzerrung anzupassen, falls die Hörer nicht richtig im Gehörgang sitzen. Ein Mikrofon an der Außenseite dient zur Sprachaufnahme bei der Headset-Anwendung, ein weiteres Mikrofon wird für die aktive Geräuschunterdrückung verwendet (vgl. iFixit, 2019).

### **3.3.5 Aktive Geräuschunterdrückung**

Die aktive Geräuschunterdrückung (engl. Active Noise Cancelling) (ANC) ist eine Technik, die ursprünglich von Automobilherstellern genutzt wurde, um den Motorengeräuschpegel zu reduzieren. Bei der Kopfhörerwiedergabe kann ANC ebenfalls verwendet werden um Umgebungsgeräusche zu reduzieren. Wie in Abschnitt 3.3.2 erwähnt, trägt bei der passiven Geräuschisolierung lediglich die geschlossene Bauform des Kopfhörers zur Schallisolation bei. ANC-Kopfhörer besitzen zusätzlich ein oder mehrere Mikrofone, die innerhalb oder außerhalb der Hörmuschel platziert sind und die Außengeräusche an jedem Ohr aufnehmen. Die Polarität des aufgenommenen Signals wird durch

einen Mikroprozessor um 180° gedreht und zusätzlich zu den Umgebungsgeräuschen wiedergegeben, wodurch sich die beiden gegenphasigen Signale nahezu vollständig auslöschen. ANC eignet sich vor allem, um gleichförmige Signale in tiefen bis mittleren Frequenzen zu unterdrücken. Pilotinnen und Piloten verwenden seit den 1980er-Jahren ANC-Kopfhörer zur Kommunikation im Cockpit. In den letzten Jahren entwickelten immer mehr Hersteller ANC-Kopfhörer zur störungsfreien Musikwiedergabe für den Konsumerbereich (vgl. Teufel Blog, o.D.).

## 4 Binauralsynthese

### 4.1 Ideales System

Die Nachteile der Binauraltechnik mittels Kunstkopfstereofonie wurden in Abschnitt 3.2.7 genannt: die transaurale Wiedergabe einer Kunstkopfaufnahme ist stark eingeschränkt. Eine Kunstkopfaufnahme bezieht sich immer auf einen Durchschnittskopf, eine Anpassung an die individuelle Anatomie eines einzelnen Hörers oder Hörerin ist nur bedingt möglich. Die Aufnahme ist statisch. Eine nachträgliche Bearbeitung der Lautstärkeverhältnisse und der räumlichen Balance der einzelnen Elemente innerhalb einer Kunstkopfmischung ist nicht möglich.

Bei einem idealen binauralen Produktionssystem gäbe es diese Einschränkungen nicht. Es sollte kein spezielles Mikrofon für die Binauralaufnahme benötigt werden, vielmehr sollte es möglich sein, herkömmliche Mikrofonaufnahmen zu binauralisieren, um den Mehraufwand für die binaurale Mischung so gering wie möglich zu halten. Eine Nachbearbeitung der Aufnahme sollte möglich sein: Lautstärkeverhältnisse, Klangfarbe und Raumanteil der einzelnen Elemente sollten veränderbar sein, außerdem sollte es möglich sein, künstliche Räume zu erstellen. Ein ideales binaurales Produktionssystem sollte sich also in die Produktionsprozesse einer herkömmlichen Audioproduktion integrieren lassen. Die Wiedergabe sollte individualisierbar sein, sodass sich die Binauralmischung an die HRTF des jeweiligen Hörers oder Hörerin anpassen lässt. Für einen möglichst hohen Immersionsgrad (siehe 4.4.1) sollten sich die Ohrsignale dynamisch an die Kopfbewegung des Hörers oder der Hörerin anpassen.

Die eingangs genannten Einschränkungen bei der Kunstkopfstereofonie entstehen größtenteils dadurch, dass das Kunstkopfmikrofon die jeweiligen Ohrsignale direkt vor Ort aufnimmt. Bei der Binauralsynthese werden diese „Ohrsignale nicht direkt aufgenommen, sondern durch digitale Signalverarbeitung synthetisiert“ (Weinzierl, 2008, S. 671).

### 4.2 Prinzip der Binauralsynthese

Die Schallübertragung von einer Schallquelle zum Ohr entspricht einem linearen, zeitinvarianten System (engl. Linear Time Invariant System) (LTI-System). Ein LTI-System lässt sich im Zeitbereich durch eine Impulsantwort und im Frequenzbereich durch eine Übertragungsfunktion mathematisch beschreiben. Abbildung 16 zeigt eine schematische Darstellung eines LTI-Systems:



Abbildung 16: LTI-System (Görne, 2015, S. 136)

Bei der Binauralsynthese können durch Faltung des Signals der Schallquelle  $s(t)$  mit der Impulsantwort die beiden Ohrsignale  $g(t)$  synthetisiert und über einen Kopfhörer wiedergegeben werden (vgl. Weinzierl, 2008, S. 671).

#### 4.2.1 Faltung

„Das Ausgangssignal eines linearen Systems erhält man durch Faltung des Eingangssignals mit der Impulsantwort“ (Görne, 2015, S. 139). Auf die Binauralsynthese angewandt ist die Impulsantwort des Systems eine vollständige Darstellung der Übertragungseigenschaften eines Raumes. Für die diskreten Faltung gibt es zwei Methoden: die Implementierung im Zeitbereich und die Implementierung als schnelle Faltung.

Die erste Methode wird durch einen Finite Impulse Response (FIR) Filter realisiert. Impulsantwort und Eingangssignal werden Sample für Sample multipliziert und anschließend aufsummiert. Mit steigender Länge der Impulsantwort steigt die Anzahl der erforderlichen Rechenschritte bei einem FIR-Filter (vgl. Keinath & Tarnow, 2009, S. 12). Für die Binauralsynthese besser geeignet ist deshalb die zweite Methode, die schnelle Faltung. Sie basiert darauf, dass „die Faltung im Zeitbereich [...] einer Multiplikation im Frequenzbereich [entspricht]“ (Görne, 2015, S. 148). Abbildung 17 zeigt den Zusammenhang zwischen der Faltung im Zeitbereich und der schnellen Faltung:

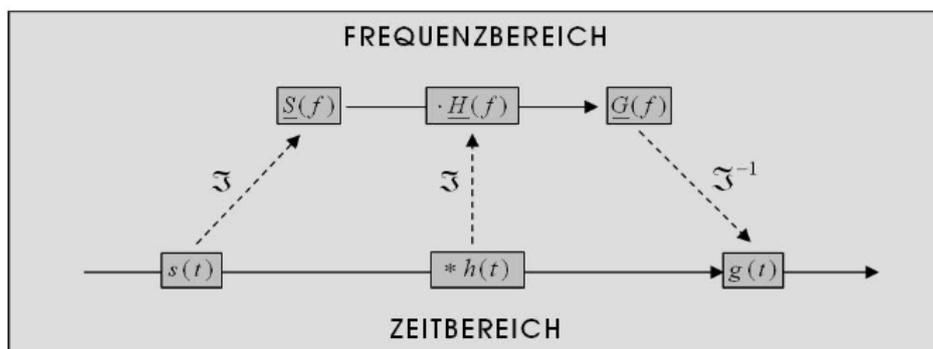


Abbildung 17: Faltung im Zeitbereich und schnelle Faltung (Keinath & Tarnow, 2009, S. 12)

Bei der schnellen Faltung werden zunächst die Impulsantwort und das Eingangssignal mit Hilfe der Fast Fourier Transformation (FFT) abschnittsweise vom Zeit- in den Frequenzbereich transformiert (siehe Abbildung 17: Eingangssignal  $s(t)$  zu  $S(f)$  und Impulsantwort  $h(t)$  zu  $H(f)$ ). Anschließend werden sie miteinander multipliziert. Das Ergebnis wird durch die inverse FFT in den Zeitbereich rückgewandelt (siehe Abbildung 17:  $G(f)$  zu  $g(t)$ ). Zwar muss das Signal zunächst in den Frequenzbereich transformiert werden, trotzdem ist der Rechenaufwand bei der schnellen Faltung deutlich geringer als bei der Faltung im Zeitbereich. Die schnelle Faltung eignet sich deshalb besonders für die Echtzeitverarbeitung eines Signals, wie sie etwa bei der dynamischen Binauralsynthese nötig ist (vgl. Görne, 2015, S. 140f).

## 4.3 Impulsantwort

Eine HRTF bzw. HRIR wird im reflexionsarmen Raum gemessen und beschreibt die Filterwirkung von Außenohr, Kopf und Torso, die für die Ortung einer Schallquelle verantwortlich ist (siehe 2.2). Sie kann als Impulsantwort zur Faltung mit einem Eingangssignal verwendet werden.

### 4.3.1 Binaurale Raumimpulsantwort

Wird eine solche Impulsantwort in einem geschlossenen Raum gemessen, dann spricht man von einer binauralen Raumimpulsantwort bzw. Binaural Room Impulse Response (BRIR) im Zeitbereich und einer Binaural Room Transfer Function (BRTF) im Frequenzbereich. Sie enthält neben dem Freifeldanteil auch spezifische Reflexionsmuster, die von der Raumbeschaffenheit und der Position von Schallquelle und -empfänger abhängig sind. Eine BRIR hat mindestens die Länge der Nachhallzeit des Raumes, sie ist somit deutlich länger als eine wenige Millisekunden lange HRIR, was sich im Rechenaufwand bei der Binauralsynthese widerspiegelt (siehe 4.2.1) (vgl. Weinzierl, 2008, S. 673). Typische Nachhallzeiten reichen von 0,3 s für ein Sprecherstudio, über 1,5 s für ein Opernhaus bis zu 3,0 s für eine große Kirche (vgl. Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 38).

### 4.3.2 Individualisierung

Wie schon in Abschnitt 3.2.2 beschrieben, sind die Maße eines Kunstkopfes so gewählt, dass sie eine durchschnittliche HRTF liefern, die der HRTF möglichst vieler Hörerinnen und Hörer entspricht. Zu große Abweichungen führen zu falschen interauralen Cues (siehe 2.3.1) und Fehlern bei der Lokalisation. Das gleiche gilt auch für die Binauralsynthese: die HRTF, mit der die Quellsignale gefaltet werden sollte möglichst genau der HRTF der Hörerin oder des Hörers entsprechen. Selbst wenn man die HRTF wie beim Kunstkopf nach statistischen Durchschnittswerten berechnet, kommt es für bestimmte Hörerinnen und Hörer immer noch zu Abweichungen. Diese Abweichungen führen zu klanglichen Änderungen und Fehllokalisationen. Um dies zu vermeiden ist es der einzige Ansatz, eine individuelle HRTF für jede Hörerin und jeden Hörer zu bestimmen (vgl. Lee & Kim, 2018, S. 2).

#### 4.3.2.1 Messbedingungen für eine HRTF-Messung

Für die korrekte Messung einer individuellen HRTF müssen folgende Bedingungen erfüllt sein: Die Messung muss unter Freifeldbedingungen stattfinden. Die Messung muss für verschiedene Schalleinfallrichtungen mit einer hohen Auflösung an Messpositionen stattfinden. Das Messmikrofon muss fest im geblockten Ohrkanal fixiert sein (siehe 2.2). Der Signal-Rausch-Abstand (engl. Signal-to-noise Ratio (SNR)) muss ausreichend groß sein. Bei einem Beispiel für ein Messaufbau steht die Versuchsperson auf einem Drehteller vor mehreren Lautsprechern, die bogenförmig entlang der Achse der Medianebene verteilt sind. Durch die Rotation der Versuchsperson und das Ansteuern einzelner Lautsprecher kann der komplette sphärische Bereich von Schallquellpositionen um die Versuchsperson abgedeckt werden (vgl. Møller, Sørensen, Hammershøi, & Jensen, 1995, S. 302ff).

Eine HRTF-Messung muss im reflexionsarmen Raum stattfinden, da die Übertragungsfunktion ansonsten zwangsläufig durch die Räumlichkeit beeinflusst wird (siehe 4.3.1). Doch auch unter Freifeldbedingungen können Reflexionen nicht vollständig ausgeschlossen werden, da es zu Schallreflexionen an den Lautsprechern, deren Befestigung und dem Drehteller kommen kann (vgl. Møller, Sørensen, Hammershøi, & Jensen, 1995, S. 305). Zwar gibt es Ansätze zur nachträglichen Eliminierung des Nachhalls aus dem Signal (siehe Warum, 2003, S. 28), trotzdem sollten Reflexionen weitestgehend vermieden werden.

Für die Messung können verschiedene Arten von Testsignalen, Miniaturmikrofonen und Lautsprecher mit unterschiedlichen Klangcharakteristiken verwendet werden, die ihrerseits das Ergebnis beeinflussen (siehe Lindau, 2006, S. 64ff). Das Messergebnis sollte zwingend „mit dem invertierten 0°/0°-Frequenzgang des Lautsprechers entzerrt werden“ (Weinzierl, 2008, S. 674), um es so vom Einfluss des Frequenzgangs des Lautsprechers zu bereinigen.

Da es sich bei der Versuchsperson um ein lebendes Objekt handelt, sind Kopfbewegungen und körpereigene Geräusche, wie das Atemgeräusch während des Messvorgangs nicht vermeidbar. Schon kleine Bewegungen des Kopfes führen dabei zu Messfehlern (vgl. Møller, Sørensen, Hammershøi, & Jensen, 1995, S. 306f).

Im Endverbrauchermarkt ist eine solche HRTF-Messung unter Laborbedingungen aus logistischen und wirtschaftlichen Gründen nicht realisierbar (vgl. Hwang, Park, & Park, 2008, S. 227). Deshalb gibt es alternative Ansätze für eine Individualanpassung, die sich nach Benutzerfreundlichkeit und Genauigkeit der Messung klassifizieren lassen.

#### **4.3.2.2 Auditive Messung**

Eine auditive Messung der individuellen HRTF unter den eingangs genannten Messbedingungen ist momentan noch recht kostspielig und damit nicht massentauglich. Die Anbieter solcher Messung zielen dabei ausschließlich auf den Audioliebhaber- und Profimarkt ab. Der *Smyth Realiser* dient der binauralen Raumsynthese (engl. Binaural Room Scanning (BRS)) einer Abhörregie im Tonstudio per dynamischer Binauralsynthese (vgl. Vaida, 2017). Im Gegensatz zu anderen BRS-Systemen, die den Referenzraum mit einem Kunstkopf vermessen (vgl. Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 1328f) ermöglicht es der *Realiser* den Nutzerinnen und Nutzern, eigene auditive Messung durchzuführen. Zwar ist es nicht das Ziel dieser Messungen, die individuelle HRTF zu bestimmen. Trotzdem soll das Messverfahren als ein Beispiel dafür dienen, wie einige der theoretischen Aspekte einer auditiven Messung (siehe 4.3.2.1) in einem massentauglichen Messverfahren angewendet werden können.

Durch die Messung der raumspezifischen Reflexionen und der Übertragungsfunktion der Lautsprecher ermöglicht es der *Realiser*, die eigene Abhörregie zu modellieren. Im Gegensatz zur HRTF-Messung sind die Voraussetzungen hierfür leichter zu erfüllen, da funktionsbedingt weder Freifeldbedingungen nötig sind, noch spezielle Messlautsprecher verwendet werden müssen. Der *Realiser*

modelliert den Klang der eigenen Abhörregie mit den eigenen Abhörmonitoren. Die Klangqualität der BRS steht also in unmittelbarer Abhängigkeit von der Klangqualität der echten Regie (vgl. Kane, 2019). Für die Messung muss sich die Nutzerin oder der Nutzer zwei Miniaturmikrofone in beide Gehörkanaleingänge stecken. Während des Messvorgangs spielt der *Realiser* drei Sinus-Sweeps an drei unterschiedlichen Kopfpositionen ab: mit Blick in Richtung Mitte, 30° nach links und 30° nach rechts. Die Messergebnisse werden als individuelle BRIR abgespeichert. Die Messung mit unterschiedlichen Kopfpositionen ermöglicht die dynamische Anpassung der BRIR per Headtracking im gemessenen Winkelbereich, dies wird in Abschnitt 4.4 genauer beschrieben. Allerdings sind die Messungen nicht fehlerfrei: Messfehler können dann entstehen, wenn die Nutzerinnen und Nutzer ihren Kopf während des Messvorgangs bewegen oder falsch ausrichten oder die Messmikrofone falsch im Ohr platziert wurden. Der Vorteil des *Smyth Realiser* ist es aber, dass der Messvorgang verhältnismäßig kurz ist und so oft wiederholt werden kann, bis die Binauralsynthese ein überzeugendes Ergebnis liefert (vgl. ebd.).

#### 4.3.2.3 Visuelle Messung und Erstellung eines virtuellen Kopfmodells

Die visuelle Messung hat gegenüber der auditiven Messung einige Vorteile, die sie sehr benutzerfreundlich macht: zur Messung wird lediglich ein Smartphone mit Kamera benötigt, die Messung kann in jedem beliebigen Raum, unabhängig von dessen akustischen Eigenschaften erfolgen. Messfehler, die durch eine falsche Platzierung der Messmikrofone oder Kopfbewegungen während des Messvorgangs entstehen können werden ausgeschlossen. Bei der visuellen Messung werden Torso, Kopf und Ohr per Foto oder Video vermessen (vgl. Genelec, 2019). Abbildung 18 zeigt die anthropometrischen Daten, die daraus gewonnen werden: 17 Maße für Kopf und Torso (Abbildung 18: links), 10 Maße für die Ohrmuschel (Abbildung 18: rechts):

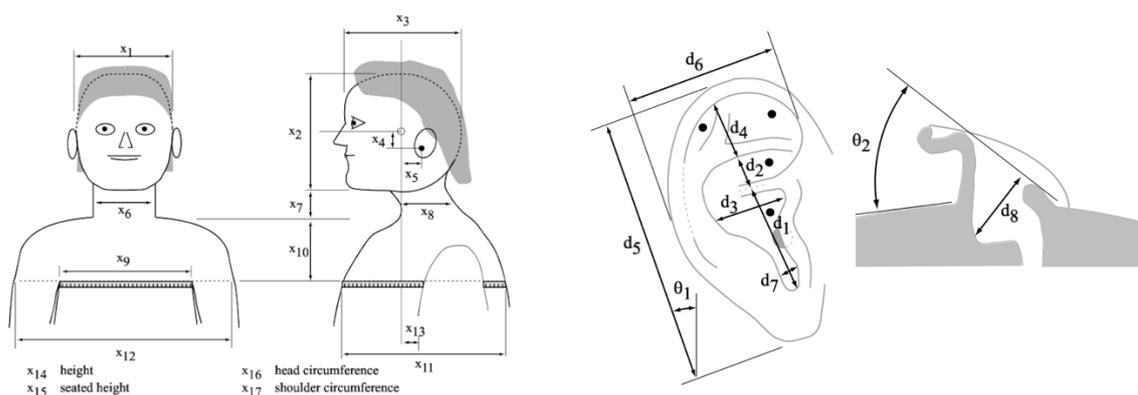


Abbildung 18: Anthropometrische Daten von Kopf, Torso und Ohr (Algazi, Duda, & Thompson, 2001, S. 2)

Anhand dieser Daten wird die HRTF berechnet. Die oben genannten Maße beziehen sich auf die CIPIC-Datenbank, ein frei zugängliche Sammlung von HRTFs des U.S. Davis CIPIC Interface Laboratory (vgl. Algazi, Duda, & Thompson, 2001, S. 1). Je nach Anbieter variieren die anthropometrischen Datensätze und die Art, wie diese Daten verarbeitet werden.

Die Firma Genelec bietet seit 2019 die Online-Software *Aural ID* an, mit der sich die individuelle HRTF anhand einer visuellen Messung erstellen lässt. Die Nutzerinnen und Nutzer müssen dazu mit ihren Smartphones Bilder und ein Video ihrer Ohren machen und hochladen. Auf den Bildern sollten die Ohrmuschel und ein Lineal als Referenzmaß zu sehen sein. Das Video sollte eine Rundum-Aufnahme des Torsos und eine Nahaufnahme der Ohrmuschel enthalten. Anhand dieser Daten erstellt *Aural ID* ein 3D-Modell des Kopfes und berechnet die HRTF der Nutzerin oder des Nutzers für unterschiedliche Schalleinfallrichtungen. Die Erstellung der individuellen HRTF mit *Aural ID* kostet aktuell 500,-€ und richtet sich an professionelle Tonschaffende und Audioliebhaber und -liebhaberinnen (vgl. Genelec, 2019).

#### **4.3.2.4 Visuelle Messung und Typisierung**

Die Erstellung eines 3D-Modells der Ohren, wie etwa bei *Aural ID* ist aufwendig. Bei der visuellen Messung und der anschließenden Typisierung wird auf die Berechnung der HRTF anhand eines 3D-Modells verzichtet. Stattdessen werden die Ergebnisse der visuellen Messung mit den Daten einer HRTF-Datenbank verglichen, um so die HRTF in der Datenbank zu finden, die am besten zu den individuellen Maßen von Kopf und Torso passt. Firmen, deren Angebot auf der Typisierung der Nutzerinnen und Nutzer basiert, erstellen entweder eigene Datenbanken oder verwenden frei zugänglichen Datenbanken wie die CIPIC-Datenbank (siehe 4.3.2.3), die LISTEN-HRTF-Datenbank (siehe Ircam, o.D.) oder die RIEC-Datenbank der Tohoku Universität (siehe Tohoku University, o.D.).

Bei der Typisierung lässt sich am Beispiel der CIPIC-Datenbank durch einen Vergleich der 27 Parameter der anthropometrischen Daten der Nutzerin oder des Nutzers und denen in einer Datenbank hinterlegten Daten die HRTF finden, bei der es die größte Übereinstimmung gibt. Lee und Kim erforschten die Verwendung eines künstlichen neuronalen Netzes, um so einen effizienteren Vergleich zu gewährleisten (vgl. Lee & Kim, 2018, S. 5f).

Sie erstellten außerdem eine Datenbank mit Bildern verschiedener Ohrformen und der zugehörigen HRTFs. Statt eines Vergleichs der anthropometrischen Daten experimentierten sie damit, die Bilder der Ohren der Nutzerin oder des Nutzers direkt mit den Bildern in der Datenbank zu vergleichen. Durch eine spezifische Bildbearbeitung und eine Mustererkennung und -analyse durch künstliche Intelligenz wurden die visuellen Merkmale der Bilder einander gegenübergestellt und die HRTF gefunden, die am besten zur Ohrform der Nutzerin oder des Nutzers passt. Dieses Verfahren macht die Typisierung noch effizienter, da die Erstellung der anthropometrischen Daten anhand der Messbilder so umgangen werden kann (vgl. ebd., S. 5f).

## 4.4 Dynamische Binauralsynthese

„Da die ermittelte Raumimpulsantwort nur für eine diskrete Kopfausrichtung gültig ist, wird der virtuelle Abhörraum bei Kopfrotationen einfach mitgedreht“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 362), so Theile über die statische Binauralwiedergabe. Bei der dynamischen Binauralsynthese wird dieses Problem vermieden, indem ein Binauralrenderer die Impulsantwort dynamisch an die momentane Kopfausrichtung des Hörers oder der Hörerin anpasst. Die perzeptive Plausibilität des Hörerlebnisses und die Lokalisation der Elemente in der Mischung werden dadurch erheblich verbessert, außerdem werden Fehler bei der Vorne-Hinten-Vertauschung (siehe 2.3.1.3) nahezu vollständig beseitigt (vgl. ebd., S. 347).

### 4.4.1 Virtual Reality

Im Forschungsfeld der dynamischen Binauralsynthese und den folgenden Abschnitten werden immer wieder Begriffe und Technologien aus dem Bereich Virtual Reality (VR) verwendet. Deshalb soll dieser Abschnitt eine kurze Übersicht über VR liefern.

„VR bezeichnet zunächst eine durch spezielle Hard- und Software erzeugte künstliche Wirklichkeit“ (Hammer, 2016). Dazu wird eine VR-Brille benötigt. Die VR-Brille isoliert die Nutzerinnen und Nutzer visuell und auditiv von der Umgebung und besteht aus zwei hochauflösenden Bildschirmen, ein Bildschirm für jedes Auge. Zur Audiowiedergabe sind entweder Kopfhörer in der VR-Brille integriert oder es werden externe Kopfhörer angeschlossen. Ein integriertes Trackingsystem erlaubt die Bestimmung der Kopfposition und -ausrichtung, die verschiedenen Funktionsweisen werden in Abschnitt 4.4.6 genauer erläutert. Durch unterschiedliche Controller ermöglichen viele VR-Systeme außerdem das Tracking von Finger-, Hand- und Armbewegungen. Dadurch können die Nutzerinnen und Nutzer mit der „virtuellen Wirklichkeit“ (ebd.) interagieren (vgl. Erl, 2020).

„Als Immersion wird das Gefühl des Eintauchens in eine virtuelle Welt bezeichnet. Im Gegensatz zu Erlebnissen auf dem Monitor [...] passieren die Dinge in VR um uns herum. Sie wirken nah und unmittelbar“ (ebd.). Das Ziel eines VR-Systems ist es, einen möglichst hohen Grad an Immersion, also eine möglichst realistische virtuelle Umgebung zu schaffen. Der Einsatz von VR ist sehr vielseitig: VR wird im Gaming-Bereich, für VR-Filme und für 360° Videos angewendet. VR erlaubt die virtuelle Begehung eines Museums, einer Sportveranstaltung oder eines Ausstellungsraumes. Darüber hinaus gibt es VR-Simulatoren zur Schulung etwa von Piloteninnen und Piloten, Ärztinnen und Ärzten und Astronautinnen und Astronauten (vgl. Hammer, 2016).

#### 4.4.2 Mögliche Kopfbewegungen

Die Bewegungsfreiheit eines Objektes im dreidimensionalen Raum wird im Zusammenhang mit VR durch die sechs Freiheitsgrade (engl. Six Degrees Of Freedom) beschrieben: die Positionskoordinaten des Objekts auf der x-,y- und z-Achse und die Ausrichtung des Objekts im Bezug zur x-,y- und z-Achse. Abbildung 19 zeigt die sechs Freiheitsgrade anhand der Kopfposition und -ausrichtung (vgl. Strickland, o.D.).

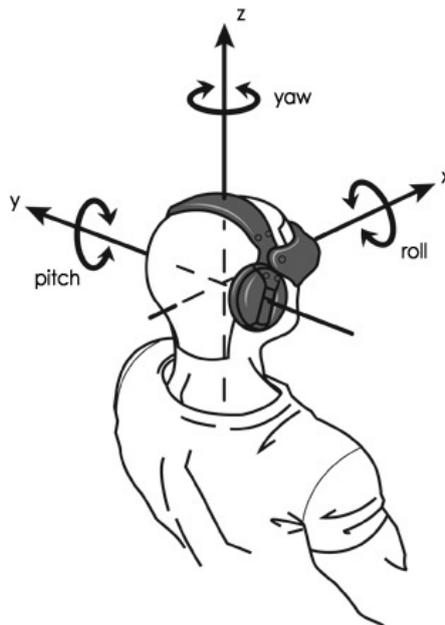


Abbildung 19: Kopfbewegungen im Bezug zur x-,y- und z-Achse (Strickland, o.D.)

Zur Lokalisation ist vor allem die Ausrichtung des Kopfes wichtig, also die horizontale Links- und Rechtsrotation des Kopfes um die z-Achse (engl. Yaw), die vertikale Neigung des Kopfes (Nicken) um die y-Achse (engl. Pitch) und das laterale Rollen des Kopfes um die x-Achse (engl. Roll) (siehe Abbildung 19) (vgl. Mackensen, 2004, S. 34). “Da durch die (horizontale) Rotation ILDs und ITDs am stärksten beeinflussbar sind, gilt diese Bewegung als wichtigste für die Lokalisation“ (Lindau, 2006, S. 43). Die Rotation des Kopfes alleine und die Kombination aus Rotation und Nicken sind die häufigsten Kopfbewegungen (vgl. ebd., S. 43).

#### 4.4.3 Prinzip der dynamischen Binauralsynthese

Die Position des Kopfes wird durch einen Headtracker ermittelt. Abbildung 20 zeigt das Prinzip der dynamischen Binauralsynthese mit einer BRIR-Datenbank:

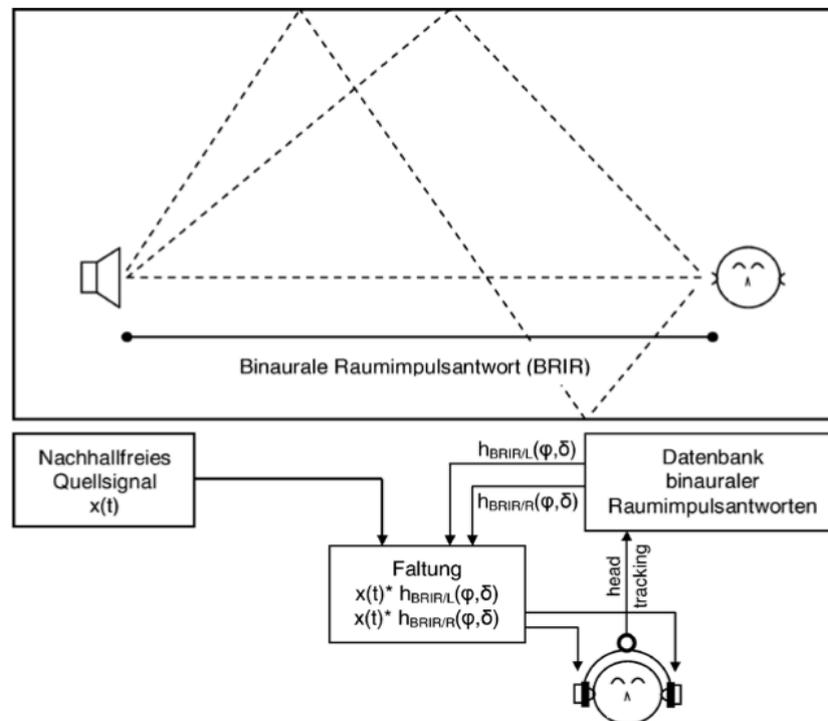


Abbildung 20: Dynamische Binauralsynthese (Weinzierl, 2008, S. 672)

Die Datenbank enthält unterschiedliche Impulsantworten, die alle möglichen Kopfpositionen repräsentiert und als SOFA-Dateien (engl. Spatially Oriented Format for Acoustics (SOFA)) gespeichert sind. Der Headtracker sendet permanent die Positionsdaten der Kopfausrichtung der Hörerin oder des Hörers an den Binauralrenderer. Dieser wählt die Impulsantwort aus, die der Kopfposition entspricht und verzögert die Ohrsignale so, dass die passenden ITDs entstehen. Laut Weinzierl sollte die Latenz zwischen der Kopfbewegung und der Nachführung der Impulsantwort für eine verzögerungsfreie Wiedergabe unter 60 ms liegen (vgl. Weinzierl, 2008, S. 677). Mackensen sagt, dass das Nachführen der Impulsantwort ab einer Latenz von 85 ms wahrnehmbar sei (vgl. Mackensen, 2004, S. 45). Die Echtzeitberechnung des Ausgangssignals ist also eine zwingende Voraussetzung für eine plausible Wiedergabe (siehe 4.2.1).

#### 4.4.4 Rasterauflösung des Datensatzes

Je mehr Impulsantworten unterschiedlicher Kopfausrichtungen vorhanden sind, desto feiner ist die Auflösung des horizontalen und vertikalen Rasters. Abbildung 21 zeigt eine Darstellung der Messpositionen von vorne (a) und im Profil (b):

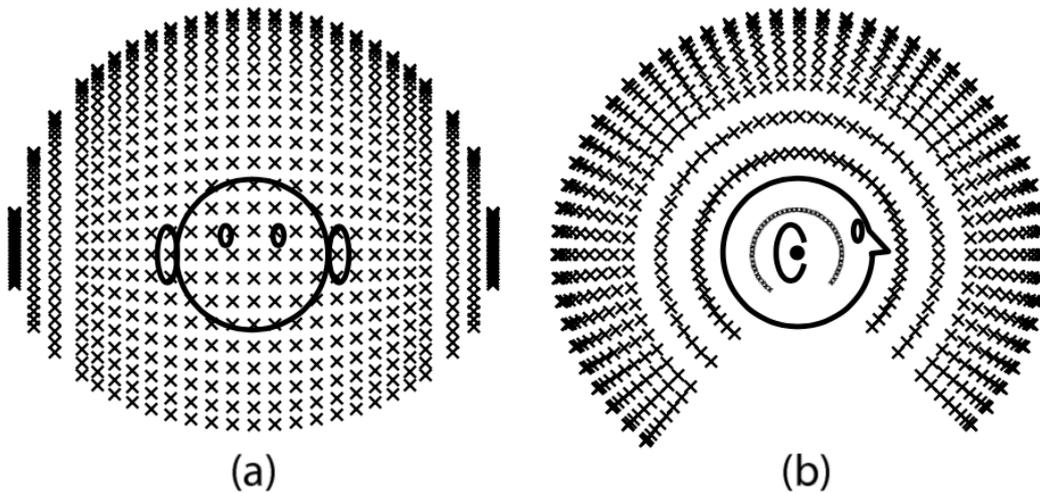


Abbildung 21: Raster eines Impulsantworten-Datensatzes (Algazi, Duda, & Thompson, 2001, S. 1)

Für Kopfpositionen, die nicht auf dem Raster liegen findet eine Interpolation zwischen den beiden nächstgelegenen Datensätzen und deren Verzögerungszeiten statt. Durch die Interpolation wird der Rechenaufwand erhöht, da die Faltung für zwei Datensätze gleichzeitig stattfindet. Der kleinste hörbare Winkelunterschied beim natürlichen Hören liegt zwischen  $5^\circ$  und  $20^\circ$  (siehe 2.3.1). Für HRTF-Messungen ist eine horizontale und vertikale Auflösung von  $2^\circ$  bis  $10^\circ$  üblich, bei der Binauralsynthese findet häufig ab einer Winkeländerung ab  $1^\circ$  eine Interpolation zwischen den zwei benachbarten HRTF-Datensätzen statt. „Da bei realen Räumen jedoch ständig Reflexionen aus allen Einfallsrichtungen vorliegen, können diese Daten nicht ohne weiteres auf den Fall einer Simulation anhand von BRIRs übertragen werden“ (Lindau & Weinzierl, 2008, S. 547). Lindau und Weinzierl empfehlen deshalb für BRIR-Datensätze  $2^\circ$  für die horizontale,  $1^\circ$  für die vertikale und  $3^\circ$  für die laterale Auflösung (vgl. ebd., S. 557).

#### 4.4.5 Trennung der Nachführung von frühem Hallanteil und Diffusschall

Wie in Abschnitt 2.3.3 beschrieben, bleibt der Pegel des Diffusschalls unabhängig von der Entfernung zur Schallquelle konstant, lediglich das Verhältnis zwischen Direkt- und Diffusschall beeinflusst die Entfernungswahrnehmung. Der Diffusschall selbst ist richtungsunabhängig. Deshalb reicht es, wenn nur der frühe Anteil der BRIR, also Direktschall und erste Reflexionen, an die Kopfposition der Hörerin oder des Hörers angepasst wird.

Abbildung 22 zeigt eine grafische Darstellung dieser Anpassung. Im linken Teil wird die komplette BRIR dynamisch ausgerichtet. Im rechten Teil werden nur Direktschall und erste Reflexionen nachgeführt, der Diffusschall wird für alle drei Kopfausrichtung aus derselben BRIR entnommen.

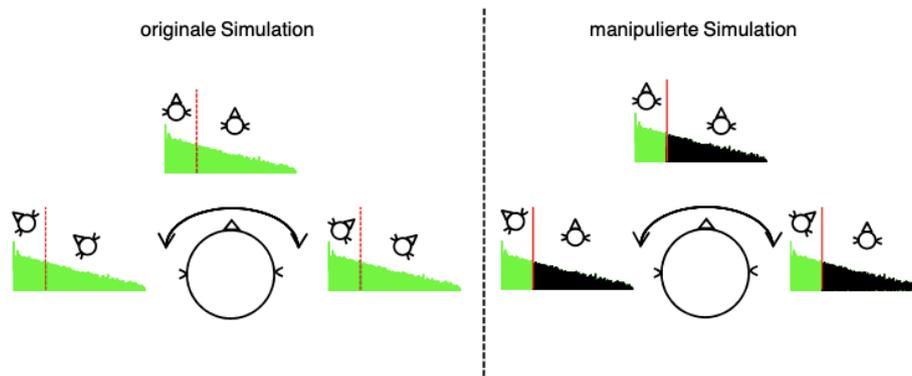


Abbildung 22: Trennung der Nachführung von frühem Hallanteil und Diffusschall (Kosanke, 2010, S. 52)

Je nach Raumvolumen verschiebt sich die Grenze zwischen frühem Anteil und Diffusschall. „Für kleine Räume erwies sich eine Beschränkung des dynamisch nachgeführten Anteils auf die ersten 40–80 ms als unkritisch [...], während für größere, Konzertsaal-ähnliche Räume eine Grenze von 130 ms ermittelt wurde“ (Weinzierl, 2008, S. 677).

#### 4.4.6 Headtracking-Systeme

Zur Umsetzung des Headtrackings gibt es unterschiedliche Technologien, die im Folgenden vorgestellt werden. Alle Systeme haben einen Sender, einen Empfänger und einen Prozessor, der die Kopfposition berechnet und an den Binauralrenderer weiterleitet (vgl. ebd., S. 677). Die Signale, die beim Binauralrenderer ankommen sind die Werte für Yaw, Pitch und Roll (siehe 4.4.2), die in Grad angegeben werden. Anhand dieser Werte kann der Binauralrenderer eine eindeutige Kopfausrichtung bestimmen (vgl. Strickland, o.D.). Als Sender- und Empfänger-Systeme zur dynamischen Binaural-synthese werden hauptsächlich Inertialsensoren und optische Sensoren und in selteneren Fällen elektromagnetische Sensoren verwendet. Die Latenz des Headtracking-Systems sollte klein genug sein, um eine verzögerungsfreie Nachführung der Impulsantwort zu gewährleisten (siehe 4.4.3). Außerdem sollte das System die Bewegungsfreiheit so wenig wie möglich einschränken.

##### 4.4.6.1 Elektromagnetische Sensoren

Bei einem elektromagnetischen Sensor bestehen Sender- und Empfänger aus jeweils drei orthogonal zueinander angeordneten Spulen. Der Sender erzeugt ein elektromagnetisches Feld, wodurch beim Empfänger Strom induziert wird. Jede der drei Spulen dient als Elektromagnet, sie können unabhängig voneinander aktiviert werden. Anhand der induzierten Spannung an den einzelnen Elektromagneten lässt sich die Ausrichtung des Empfängers in Relation zum Sender messen und berechnen.

Elektromagnetische Headtracking-Systeme sind relativ präzise und haben eine geringe Latenz. Sie erlauben also eine verzögerungsfreie Nachführung der Impulsantwort. Funktionsbedingt ist der Sensor aber sehr störungsanfällig gegenüber anderen ferromagnetischen Objekten, wie dem Magnetfeld anderer Gegenstände oder metallischen Gegenstände (vgl. Strickland, o.D.). Bei Systemen wie etwa dem *Polhemus FASTRAK* wird der Empfänger am Kopfhörerbügel oder einem Haarreif auf dem Kopf platziert. Ein Nachteil dieses Systems ist es, dass „die Befestigung eines Sensors am verfolgten Objekt erforderlich [ist], wodurch der Hörer, wie im vorliegenden Fall durch einen Haarreif und das Anschlusskabel, gestört wird. Die ideale Lösung wäre ein berührungsloses System“ (Laumann, 2016, S. 53). Die Bewegungsfreiheit der Hörerinnen und Hörer wird bei elektromagnetischen Sensoren eingeschränkt.

#### **4.4.6.2 Optische Sensoren**

Optische Sensoren lassen sich in markerbasierte und bildbasierte Systeme einteilen. Bei markerbasierten Systemen erleichtern optische Marker die Bildanalyse. In vielen Systemen werden Infrarot-LEDs als aktive Marker oder passive, Infrarot-Licht reflektieren Marker verwendet. Eine Kamera, die das Infrarot-Licht empfangen kann, dient als Empfänger. Das Infrarot-Licht wird als periodischer Puls aus unterschiedlichen LEDs gesendet. Die empfangenen Lichtimpulse werden an den Prozessor weitergeleitet, der daraus Position und Ausrichtung des Headtrackers erschließen kann. Bei einem bildbasierten System gibt es keine Marker. Eine Kamera, beispielsweise eine Webcam, filmt den Kopf der Nutzerin oder des Nutzers. Das Bild wird in Echtzeit von einem Grafikprozessor analysiert und etwa auf Ecken oder Kanten untersucht. Daraus kann der Prozessor die Kontur eines Objekts erschließen. Beim Headtracking untersucht der Prozessor das Bild auf die Kontur bestimmter Gesichtsmarkere, wie etwa die der Augen und kann so die Lage des Kopfes bestimmen (vgl. Mulder, Jansen, & van Rhijn, 2003, S. 4).

Ein optisches Trackingsystem hat sehr schnelle Datenraten und erlaubt die verzögerungsfreie Nachführung der Impulsantwort. Allerdings ist die Bildqualität und damit die Qualität des Headtrackers stark abhängig vom Kameramodell und der Belichtungssituation. Ein optischer Sensor ist sehr anfällig gegenüber extrem dunklem, aber auch extrem hellem Umgebungslicht. Problematisch ist außerdem, dass stets Sichtkontakt zwischen Sender und Empfänger vorhanden sein muss. Wird der Sichtkontakt unterbrochen, kommt es zu Fehlern bei der Bildanalyse (vgl. Strickland, o.D.). Deshalb ist die Bewegungsfreiheit der Nutzerinnen und Nutzer und der Abstand zwischen Sender und Empfänger oftmals beschränkt.

#### **4.4.6.3 Inertialsensoren**

Inertialsensoren gibt es in verschiedenen Ausführungen mit teils sehr komplexen Funktionsweisen. Für eine detaillierte Erklärung der Funktionsweisen wird im Folgenden deshalb an gegebener Stelle auf entsprechende Fachliteratur verwiesen. Dieser Abschnitt beschränkt sich auf die Beschreibung der grundlegenden Prinzipien und Eigenschaften von Inertialsensoren.

Inertialsensoren lassen sich in Beschleunigungssensoren (engl. Accelerometer) und Drehratensensoren (engl. Gyroscope) unterteilen. Beide zusammen fungieren als Lagesensor. Der Beschleunigungssensor misst die Beschleunigung in x-, y-, und z-Richtung (siehe 4.4.2). Prinzipiell besteht er aus einer gefederten Masse, die in einem Gehäuse aufgehängt ist. Bewegt sich das Gehäuse, dann verschiebt sich die Masse aufgrund ihrer Trägheit gegenüber dem Gehäuse und die Beschleunigung kann bestimmt werden (vgl. Theiss, 2011) (siehe Tränkler & Reindl, 2014, S. 543ff).

Ein Drehratensensor besteht aus einem Gyroskop bzw. Kreiselinstrument, das sich in einem kugelförmigen Rahmen frei drehen kann. Ein Gyroskop hat die Eigenschaften eines Kreisels: Wenn sich das Gyroskop nicht dreht, dann liegt es auf dem Boden. Rotiert das Gyroskop, dann steht es aufrecht und behält diese Lage bei, egal wie das Endgerät indem der Sensor verbaut ist ausgerichtet ist und dient so als konstante Referenz zur Bestimmung der Freiheitsgrade (vgl. Schanze, 2017) (siehe Tränkler & Reindl, 2014, S. 560ff).

Durch die Auswertung der Daten des Beschleunigungs- und Drehratensensors lässt sich beispielsweise die Ausrichtung eines Smartphones bestimmen. Dies wird vor allem für Navigationssysteme oder die automatische Hoch- oder Querformatausrichtung des Bildschirms genutzt. Im Gaming- und VR-Bereich (siehe 4.4.1) gibt es einige Ansätze, das Smartphone mithilfe des eingebauten Lagesensors als Headtracker zu verwenden. Beispielsweise wird das Smartphone am Kopfhörerbügel, auf dem Kopf platziert und überträgt die Daten des Lagesensors per Bluetooth oder WLAN an den Computer (vgl. GravityMsc, 2020). Inertialsensoren sind im Gegensatz zu elektromagnetischen oder optischen Sensoren nicht als Sender- und Empfänger-System aufgebaut und erlauben deshalb die volle Bewegungsfreiheit.

#### 4.4.6.4 Beispiele für Headtracking-Systeme

In aktuellen Systemen, die Headtracking zur dynamischen Binauralsynthese verwenden, werden oftmals mehrere Sensortypen kombiniert. Beim *Smyth Realiser* sind alle drei genannten Sensortypen eingebaut: ein elektromagnetischer Sensor, ein Inertialsensor und ein optischer Sensor. Laut Herstellerangaben wird ein Gyroskop als Hauptsensor genutzt, die anderen beiden Sensorarten dienen zur Unterstützung und zur Fehlerkorrektur dieses Inertialsensors. Mithilfe aller drei Sensortypen bestimmt der *Realiser* alle 5 ms die aktuelle horizontale Kopfausrichtung im Bereich von 30° links bis 30° rechts (vgl. Kane, 2019).

Die Firma Waves bietet Headtracking für ihr BRS-System *NX Virtual Mix Room* per Webcam, per Inertialsensor oder als Kombination aus beidem an. Bei reinem Webcam-Tracking können sich Nutzerinnen und Nutzer nur in einem Bereich von 60° und 1 bis 2 m Abstand bewegen. Mit einem zusätzlichen Inertialsensor, der als Headtracker auf dem Kopfhörerbügel montiert wird, lässt sich dieser Bereich auf 360° und 3 bis 6 m vergrößern (vgl. Waves, o.D.a).

Die Firma Apple hat Inertialsensoren zum Headtracking in den *Apple AirPods Pro* (siehe 3.3.4) verbaut. Die Anwendung wird in den Abschnitten 4.6.2 und 6.5.3 genauer beschrieben.

## 4.5 Verfahren zur Binauralsynthese

In Abschnitt 4.3 wurden zwei Arten von Impulsantworten beschrieben: die HRIR, die unter Freifeldbedingungen gemessen wird und die BRIR, die in einer Räumlichkeit gemessen wird. Die BRIR enthält im Gegensatz zur HRTF spezifische Reflexionsmuster, die den Raum charakterisieren, in dem das Schallereignis stattfindet. Sowohl HRIR, als auch BRIR können bei der Binauralsynthese als Impulsantwort dienen, mit der das Direktsignal gefaltet wird (siehe 4.2.1). Im Folgenden sollen drei solche Verfahren zur Binauralsynthese genauer beschrieben werden.

### 4.5.1 Faltung mit der HRTF

HRTF-Datensätze verschiedener Kunstkopfsysteme und Nutzerinnen und Nutzer sind mittlerweile weit verbreitet und öffentlich zugänglich (siehe 4.3.2.4). Wird ein Quellsignal mit einer HRTF gefaltet, dann wird das Schallereignis in der Richtung lokalisiert, die dieser HRTF entspricht. Diese Methode ist einfach, erfordert vergleichsweise wenig Speicherplatz und Rechenleistung und eignet sich deshalb speziell für Echtzeitanwendungen (siehe 4.2). Dies ist der Grund, warum sie kommerziell weit verbreitet ist (vgl. Lindau, 2006, S. 20ff). Bei einigen Verfahren kommen „sogar Raytracingverfahren zur Bestimmung von Erstreflexionen [sic!] niedriger Ordnung [...] zum Einsatz, so dass rudimentäre BRIRs in Echtzeit (!) erzeugt werden können“ (ebd., S. 23).

Ein Ambisonics-Binaural-Dekoder funktioniert nach demselben Prinzip. Ambisonics ist ein dreidimensionales Aufnahme- und Wiedergabeverfahren. Im Gegensatz zu anderen 3D-Audio-Formaten, wird die Mischung bei Ambisonics als dreidimensionales Klangfeld betrachtet, das erst durch den Ambisonics-Dekoder für eine dezidierte Anzahl an Wiedergabekanälen dekodiert wird. Die Bearbeitung und Postproduktion einer Ambisonics-Mischung erfolgt also lautsprecherunabhängig (siehe Abbildung 23) (vgl. Waves, 2017).

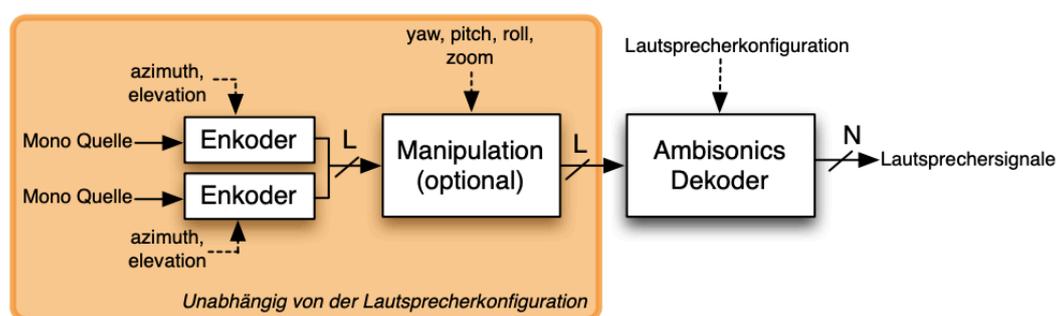


Abbildung 23: Ambisonics (Kronlachner, 2012)

Die Kanalanzahl wird durch die Ambisonics-Ordnung angegeben. Bei First-Order-Ambisonics sind es vier Kanäle. Bei Higher Order Ambisonics (HOA) steigt die Kanalanzahl auf bis zu 49 Kanälen für Sixth Order Ambisonics (vgl. Waves, 2017).

Für die Binauralisierung wird die Ambisonics-Mischung für eine rein virtuelle Lautsprecherkonfiguration dekodiert. Die Positionsdaten der virtuellen Lautsprecher werden an den Binaural-Dekoder übergeben. Dieser faltet das jeweilige Lautsprechersignale mit der HRTF, die zur jeweiligen Lautsprecherposition passt. Bei einer Konfiguration aus vier Lautsprechern ergeben sich insgesamt acht Faltungen, vier Faltung für jedes Ohrsignal. Die Summe der Ergebnisse liefert die beiden Ohrsignale (siehe Abbildung 24) (vgl. Zotter & Frank, 2019, S. 85).

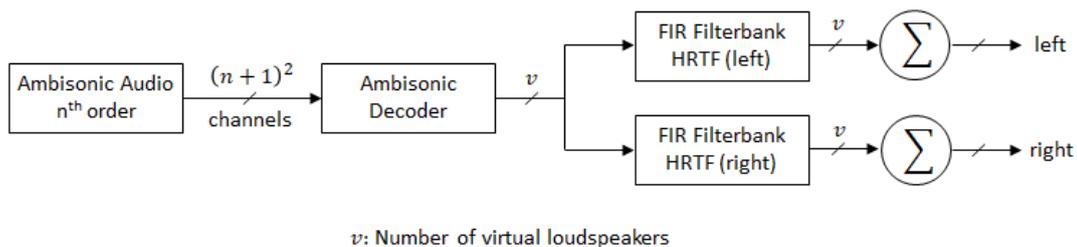


Abbildung 24: Ambisonics-Binaural-Dekoder (MathWorks, o.D.)

Die Optimierung der HRTF für diese Art der Binauralisierung, vor allem in Bezug auf interaurale Phasenunterschiede und den Einfluss der hohen Frequenzen auf die Klangfarbe, ist Gegenstand laufender Forschungen (siehe Schörkhuber, Zaunschirm, & Höldrich, 2018). Ein Vorteil dieser Methode ist die einfache Umsetzung der dynamischen Wiedergabe, da die Anpassung der Kopfausrichtung durch Rotation des Klangfeldes in die entsprechende x-, y- oder z-Richtung (siehe 4.4.2) umgesetzt werden kann (siehe Abbildung 23). Die Rotation erfolgt noch bevor die Mischung für eine Lautsprecherkonfiguration dekodiert wird, die Nachführung des Klangfeld ist also lautsprecherunabhängig. Dadurch wird vermieden, dass das Klangfeld abrupt von einem virtuellen Lautsprecher zum nächsten „springt“. Die Nachführung des Klangfelds ist gleichmäßig und stabil (vgl. Waves, 2017). Ein weiterer Vorteil ist es, dass die Impulsantwort nicht geändert werden muss, da sie entsprechend der virtuellen Lautsprecherposition gewählt und nicht permanent an die Kopfausrichtung angepasst wird. Deshalb wird „Ambisonics [...] als einfaches und flexibles Speicher- und Aufnahmeformat für interaktive 360°-Videos eingesetzt, z.B. in YouTube. Beim Drehen der Perspektive müssen nur die Werte der Ambisonics-Variablen geändert werden“ (Wittek & Theile, 2016). Nachteilig ist, dass es bei einer lautsprecherunabhängigen Mischung nicht möglich ist, eine Schallquelle durch einen dezidierten Lautsprecherkanal darzustellen. Die Lokalisationsschärfe eines Ambisonics-Binaural-Dekoders ist somit unmittelbar abhängig von der Ambisonics-Ordnung und der Anzahl der virtuellen Lautsprecherkanäle „da mit steigender HOA-Ordnung die räumliche Auflösung steigt und somit eine Quelle schärfer dargestellt werden kann“ (vgl. Balke, 2013, S. 46). Bei einer geringeren Ambisonics-Ordnung erscheint die Schallquelle somit unscharf.

#### 4.5.2 Faltung mit der BRIR realer Räume

Die Faltung des Signals mit der BRIR eines realen Raumes ist das Verfahren, das am realitätstreuesten klingt. Lindau schreibt darüber: „Eine Auralisation von solch perzeptiver Güte, dass sie an Grenzen der menschlichen Unterscheidungsfähigkeit von der Realität stößt, ist [...] nur mit der In-situ-Methodik möglich“ (Lindau, 2006, S. 21).

Für die Messung einer BRIR wird ein Lautsprecher verwendet, über den ein Messsignal im realen Raum abgespielt und mit einem Kunstkopfmikrofon aufgenommen wird. Dabei wird mit beweglichen Kunstkopfmesssystemen experimentiert, die eine Ausrichtung des Kunstkopfes für alle natürlichen Kopfbewegungen (siehe 4.4.2) ermöglichen. Je nach Genauigkeitsauflösung der BRIR-Messungen kann die Rasterauflösung durch Interpolation erhöht werden um so eine BRIR-Datenbank zu erstellen, die sich für die dynamische Binauralsynthese eignet (vgl. ebd., S. 21f). Dabei gilt es zu beachten, dass im Gegensatz zu HRIRs bei der dynamischen Nachführung von BRIRs besondere Herausforderungen entstehen, die in Abschnitt 4.4.4 beschrieben wurden.

Die gemessene BRIR sollte ausschließlich die Signalanteile enthalten, die die Raumakustik repräsentieren. Sie sollte möglichst frei von messtechnischen Verzerrungen sein, die durch den Einfluss des Frequenzgangs des Messmikrofons und -lautsprechers entstehen. Bei einer HRTF-Messung unter Freifeldbedingungen gilt dieselbe Bedingung, das HRTF-Messergebnis wird deshalb mit dem invertierten Frequenzgang des Lautsprechers entzerrt (siehe 4.3.2.1). Bei der BRIR-Messung ist das nicht möglich, da „die frequenzabhängige Richtcharakteristik des Lautsprechers praktisch untrennbar in die Raumimpulsantwort ‚einkodiert‘ [ist], da in der gemessenen BRIR keine Information über die Einfallsrichtung der einzelnen Raumreflexionen mehr enthalten ist“ (Weinzierl, 2008, S. 674). Als Lösungsansatz kann die Messung mit dem Diffusfeldfrequenzgang des Lautsprechers entzerrt werden, was den Fehler allerdings nicht vollständig kompensiert (vgl. ebd., S. 674).

Für raumakustische Messungen muss gewährleistet sein, dass alle Übertragungswege von der Quelle zum Empfänger angeregt werden. Deshalb werden Messlautsprecher verwendet, deren Abstrahlverhalten sich der omnidirektionalen Richtcharakteristik eines punktförmigen Kugelstrahlers annähert. Für die Binauralsynthese ist eine omnidirektionale Quelle ungeeignet, da „nahezu keine natürliche Quelle dieser Richtcharakteristik nahe kommt“ (Lindau, 2006, S. 66). Vorzugsweise sollte deshalb ein Lautsprecher gewählt werden, dessen Frequenzgang und Richtcharakteristik dem Objekt ähneln, das später binauralisiert werden soll. Für Sprache wurden spezielle Mundsimulatoren entwickelt (siehe Kob & Jers, 1999), allerdings sollte die Richtcharakteristik für die Binauralsynthese frei wählbar sein, um so das Abstrahlverhalten unterschiedlicher Instrumente zu simulieren. Ein Dodekaederlautsprecher ist ein Kugellautsprecher, dessen Treiber sich gruppenweise in Phase und Amplitude ansteuern lassen. Ein Ansatz zur Simulation unterschiedlicher Abstrahlverhalten ist es, durch die Überlagerung der ersten vier Basisfunktionen, Kugel und den drei Achten in x-, y- und z-Richtung, „invers zum Prinzip des Soundfieldmikrofons“ (Lindau, 2006, S. 69) verschiedene Richtcharakteristiken für den Messlautsprecher zu synthetisieren (siehe Misdariis, Nicolas, Warusfel, & Caussé,

2001). Das Abstrahlverhalten eines Instruments und die Abschattungseffekte, die durch den Spieler selbst entstehen, sind jedoch deutlich komplexer, als es sich mit einem solchen Lautsprechersystem darstellen lässt. Deshalb bleibt das Problem nicht vollständig gelöst und eine beliebige Schallquelle lässt sich nur annäherungsweise simulieren (vgl. Lindau, 2006, S. 69f).

Statt ein Direktsignal mit einer BRIR zu falten, können BRIRs für ein Auralisationsverfahren verwendet werden, dessen Funktionsweise der eines Ambisonics-Binaural-Dekoders (siehe 4.5.1) ähnelt, es basiert allerdings auf BRIRs und nicht auf HRTFs. Zunächst wird die BRIR eines realen Mehrkanalregieraum durch ein Kunstkopfmikrofon gemessen. Anhand der BRIR und der Positionsdaten der Lautsprecher des realen Regieraumes kann ein virtueller Regieraum konfiguriert werden. Wird nun eine 3D-Audiomischung abgespielt, erstellt ein Binaural-Renderer anhand des virtuellen Regieraumes die beiden Ohrsignale und simuliert so den realen Regieraum über Kopfhörer (vgl. Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 1243f). Der in Abschnitt 4.3.2.2 beschriebene *Smyth Realiser* funktioniert nach diesem Prinzip.

#### 4.5.3 Faltung mit der BRIR virtueller Räume

Bei der Faltung des Quellsignals mit der BRIR eines virtuellen Raumes wird die BRIR durch raumakustische Simulationsverfahren berechnet. Die Raumgeometrie wird dabei durch graphische und numerische Werte beschrieben, anhand derer ein 3D-Modell des Raumes berechnet wird (siehe Abbildung 25):

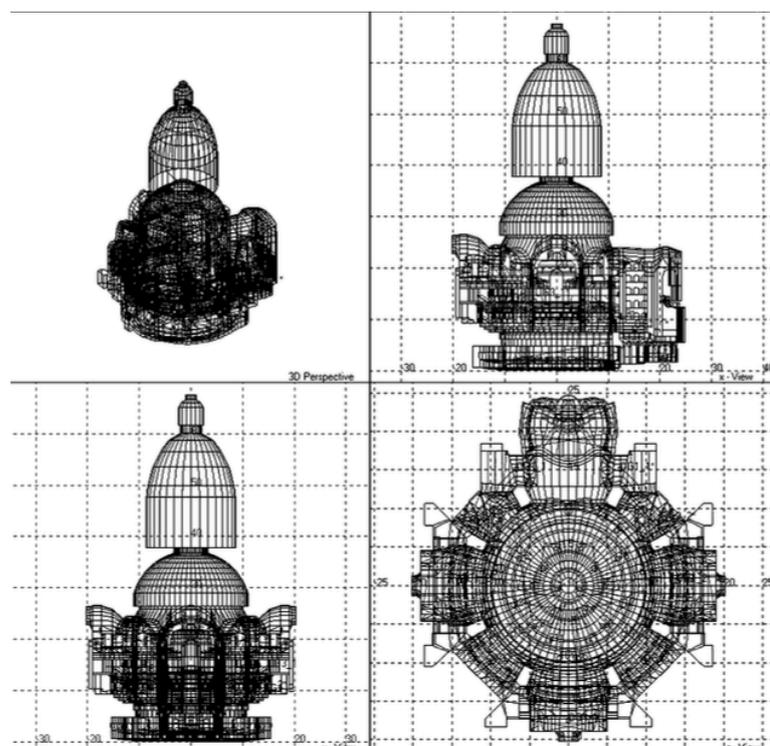


Abbildung 25: 3D-Modell der Frauenkirche Dresden (Weinzierl, 2008, S.243)

Um die Impulsantwort des Raumes zu berechnen gibt es unterschiedliche Methoden zur Simulation der Schallausbreitung. Typische Simulationsverfahren sind die Spiegelquellenmethode (engl. Image Source Model), die Strahlverfolgung (engl. Ray Tracing), Cone Tracing, Pyramid Tracing oder eine Kombination aus allen Verfahren (siehe Weinzierl, 2008, S. 244ff), die sich durch ihren Grad an Realitätstreue und dem benötigten Rechenaufwand unterscheiden. Zur Berechnung der BRIR werden die virtuellen Positionen von Quelle und Empfänger festgelegt. Die Schallstrahlen, die beim Empfänger eintreffen, werden entsprechend ihrer Einfallsrichtung mit der passenden HRTF gewichtet und liefern in Summe die gesuchte BRIR (vgl. ebd., S. 679).

Computersimulationen eignen sich besonders gut für die Berechnung von raumakustischen Daten, wie die Nachhallzeit oder der Seitenschallgrad. Im Vergleich zur Auralisation mithilfe von BRIR-Messungen aus realen Räumen gibt es bei der Auralisation mithilfe von Computersimulation jedoch deutlich hörbare Qualitätseinbußen. Dies liegt vor allem daran, dass geometrische Raumstrukturen in Computermodellen nur vereinfacht modelliert werden können. Hinzu kommt, dass Schall rein geometrisch und nicht als Welle betrachtet wird, da die meisten Computermodelle lediglich den Absorptionsgrad einer Fläche, nicht aber deren Beugungs- und Brechungsverhalten und die daraus resultierenden Interferenzeffekte von einfallenden und reflektierten Schallwellen berücksichtigen. Um solche wellentheoretischen Phänomene in die Simulation mit einzubeziehen müsste die Berechnung für das gesamte hörbare Frequenzspektrum von 20 Hz bis 20 kHz erfolgen. „Der dafür nötige Mess-, Modellierungs- und Rechenaufwand führt jedoch auf nicht mehr erlebbare Rechenzeiten, so dass die modale Berechnung von Schallfeldern in Räumen bisher auf niedrige Frequenzbereiche (<200Hz) und kleine Räume einfacher Geometrie beschränkt bleibt“ (Lindau, 2006, S. 21).

## **4.6 Reales System**

Im Kapitel 4 wurde die Binauralsynthese und die damit verbundenen technischen Herausforderungen beschrieben. Dabei haben sich drei Hauptfaktoren gezeigt, die die Qualität der Binauralsynthese beeinflussen: die Individualisierung, das Headtracking und die Art des Verfahrens zur Binauralsynthese. In Abschnitt 4.1 wurden die Gesichtspunkte eines idealen Produktionssystems genannt. „Die [...] Gesichtspunkte [beschreiben] zwar anspruchsvolle, mit modernen Verfahren der digitalen Signalverarbeitung jedoch durchaus lösbare Probleme, auch wenn die Ermittlung der für eine perzeptiv plausible Wiedergabe erforderlichen physikalischen Genauigkeit Gegenstand laufender Forschung ist“ (Weinzierl, 2008, S. 673). Es gibt also grundsätzlich Lösungsansätze, die alle in Abschnitt 4.1 genannten Anforderungen an ein ideales Produktionssystem erfüllen. Allerdings gibt es bisher kein universelles System, das alle Anforderungen in hinreichender Qualität vereint.

### **4.6.1 Vergleich realer Systeme**

Die folgenden Eigenschaftenmatrizen sollen einen Überblick über den aktuellen Stand der Technik liefern. Dazu wird eine Auswahl an aktuellen binauralen Produktionssystemen (siehe Tabelle 1), Binaural-Pannern (siehe Tabelle 2), Ambisonics-Binaural-Dekodern (siehe Tabelle 3) und binauralen Auralisations-Systemen (siehe Tabelle 4) verglichen. Die Auswahl der Werkzeuge erhebt keinen Anspruch auf Vollständigkeit. Die Zuordnung der Werkzeuge zu den einzelnen Matrizen erfolgte aufgrund ihrer Anwendung und nicht aufgrund ihrer technischen Funktionsweise.

Tabelle 1: Vergleich aktueller binauraler Produktionssysteme

	Anwendung	Individualisierung	Head-tracking	Verfahren zur Binauralsynthese	Kanäle	Kopfhörer-Entzerrung	Preis
New Audio Technologies <i>Spatial Audio Designer</i> (Plugin oder Hardware) <sup>1</sup>	3D-Audioproduktion	Nein	Nein	Faltung mit BRIR realer Räume (Tonstudios, Kinos)	Unbegrenzte Inputs, bis zu 330 Outputs	Grafischer Equalizer, Kompensationsfilter für gängige Modelle	89,-€ bis 1349,-€ (für Plugin)
Dear Reality <i>dearVR Pro</i> (Plugin) <sup>2</sup>	3D-Audioproduktion	Nein	Einbindung eines externen Head-trackers  OSC-Input Fähig	Faltung mit BRIR über 45 realer Räume (Tonstudios, Kinos, Bühnen, sonstige Räume)  Faltung mit der BRIR virtueller Räume	Unbegrenzte Inputs, bis zu 26 Outputs	Nein	310,-€
Sparta <i>Binauraliser</i> (Plugin) <sup>3</sup>	3D-Audioproduktion	Auswahl von HRTFs, Import eigener SOFA-Dateien	Einbindung eines externen Head-trackers  OSC-Input fähig	Faltung mit HRTF	Bis zu 64 Inputs, 2 Outputs	Nein	Kostenlos

<sup>1</sup> Vgl. New Audio Technology, o.D. und Latta, o.D.a

<sup>2</sup> Vgl. Dear VR, o.D. und Obst, 2017

<sup>3</sup> Vgl. McCormack, Delikaris-Manias, Politis, & Pulkki, 2020 und McCormack & Politis, 2019

Tabelle 2: Vergleich aktueller Binaural-Panner

	<b>Anwen- dung</b>	<b>Indivi- dualisie- rung</b>	<b>Head- tracking</b>	<b>Verfahren zur Binauralsynthese</b>	<b>Kanäle</b>	<b>Kopfhö- rer- Entzer- rung</b>	<b>Preis</b>
Sennheiser <i>Ambeo Orbit</i> (Plugin) <sup>4</sup>	Binaural- Panning	Nein	Nein	Faltung mit HRTF  Auswahl von Erstreflexionen verschiedener Oberflächen	-	Nein	Kosten- los
Apple Logic Pro <i>Binaural- Panner</i> (Plugin) <sup>5</sup>	Binaural- Panning	Nein	Nein	Faltung mit HRTF	-	Freifeld, Diffus- feld, Transau- rale Wieder- gabe	Kosten- los (in Apple Logic Pro inte- griert)

<sup>4</sup> Vgl. Sennheiser, o.D. und Schölzel, 2018

<sup>5</sup> Vgl. Apple, 2019 und White, 2019

Tabelle 3: Vergleich aktueller Ambisonics-Binaural-Dekoder

	Anwendung	Individualisierung	Head-tracking	Verfahren zur Binauralsynthese	Kanäle	Kopfhörer-Entzerrung	Preis
IEM <i>Binaural-Decoder</i> (Plugin) <sup>6</sup>	Ambisonics-Binaural-Dekoder	Nein	Nein	Faltung mit HRTF	Bis 7th Order Ambisonics (Input)	Kompensationsfilter für gängige Modelle	Kostenlos
Noisemakers <i>Ambi Head HD</i> (Plugin) <sup>7</sup>	Ambisonics-Binaural-Dekoder	Auswahl von HRTFs  Import eigener SOFA-Dateien	Mit <i>Ambi Eyes</i> (360° Videoplayer, der Tracking Daten übermittelt)	Faltung mit HRTF	Bis 1st Order Ambisonics (Input)	Nein	Kostenlos
S3A <i>VISR Binaural Renderer</i> (Plugin) <sup>8</sup>	Ambisonics-Binaural-Dekoder	Auswahl von HRTFs  Import eigener SOFA-Dateien	Einbindung eines externen Headtrackers.  OSC-Input fähig	Faltung mit HRTF oder BRIR	Bis Higher Order Ambisonics (Input)	Nein	Kostenlos

<sup>6</sup> Vgl. Rudrich, o.D.

<sup>7</sup> Vgl. Noisemakers, o.D.

<sup>8</sup> Vgl. S3A, 2019

Tabelle 4: Vergleich aktueller binauraler Auralisations-Systeme

	Anwendung	Individualisierung	Head-tracking	Verfahren zur Binauralsynthese	Kanäle	Kopfhörer-Entzerrung	Preis
Smyth <i>Realiser A16</i> (Hardware) <sup>9</sup>	Auralisation einer 3D-Abhör-Umgebung	Auditive Messung	Eigener Head-tracker (elektromagnetisch, optisch, inertial)	Individuelle Messung und Faltung mit der BRIR des eigenen Raumes  Faltung mit der BRIR realer Räume	Bis zu 16 Inputs und Outputs	Individuelle Messung des Kopfhörers	3552,-€
Waves <i>NX Virtual Mixing Room</i> (Plugin) <sup>10</sup>	Auralisation einer Surround Abhörumgebung	Angabe von Kopfumfang und Ohrabstand	Eigener Head-tracker (optisch, inertial)	Faltung mit der BRIR eines realen Raumes	Bis zu 5.1 Setup Output	Nein	158,-€ (mit Head-tracker)
Flux <i>Ircam HEar V3</i> (Plugin) <sup>11</sup>	Auralisation einer Surround Abhörumgebung	Nein	Nein	Faltung mit HRTF	Bis zu 10 Inputs und Outputs	Nein	89,-€
KLANG: technologies <i>KLANG: fabrik</i> (Hardware) <sup>12</sup>	3D-In-Ear-Monitoring für Musiker	Nein	Eigener Head-tracker (inertial)	Faltung mit HRTF	Bis zu 56 Inputs pro Mix, 16 Output-Mixe	Nein	4700,-€

<sup>9</sup> Vgl. Smyth Research, o.D. und Kane, 2019

<sup>10</sup> Vgl. Waves, o.D.b und Obst, 2016a

<sup>11</sup> Vgl. Flux Audio, o.D. und Obst, 2016b

<sup>12</sup> Vgl. Nötges, 2019 und Latta, o.D.b

#### 4.6.2 Diskussion

Die Optimierung einer anwenderfreundlichen Umsetzung der Individualisierung, des Headtrackings und des Verfahrens zur Binauralsynthese ist Gegenstand laufender Forschung (siehe 4.6). Die verschiedenen Systeme unterscheiden sich aktuell vor allem durch das jeweilige Verfahren, das für die Binauralsynthese verwendet wird (siehe 4.6.1). Verfahren, die Kunstkopf-HRTFs auf virtuelle Lautsprecherkonfigurationen herkömmlicher 3D-Audioformate wie Ambisonics oder Dolby Atmos rechnen, sind verhältnismäßig einfach umzusetzen und leicht in bestehende Produktionssysteme zu integrieren. Die Erstellung von BRIR-Datensätzen ist aufwendiger, deshalb sind Systeme, die BRIRs für die Faltung verwenden meist kostspieliger als andere Systeme. Zwar klingt die Faltung mit BRIRs sehr realitätsgetreu, allerdings ist die Anpassung der Raumparameter im Produktionsprozess eingeschränkter als bei der Verwendung von HRTFs und künstlichen Räumen (siehe 4.5.2).

Headtracking wird aktuell nur von wenigen Herstellern angeboten (siehe 4.6.1). Ist Headtracking möglich, dann sind zusätzlich Headtracker nötig, die am Kopfhörerbügel montiert werden müssen. Solche Headtracker mögen sich für die Anwendung in der Tonregie eignen, werden sich aber schon allein aus optischen Gründen nicht im Endverbrauchermarkt durchsetzen. Hinzu kommt, dass solche Headtracker bisher nicht für die kommerziell weit verbreiteten In-Ear Kopfhörer verwendet werden können (siehe 6.2), da bei In-Ear Kopfhörern der Kopfhörerbügel zur Befestigung des Headtrackers fehlt. Einen kompakten Headtracker will die Firma KLANG:technologies auf den Markt bringen: Der KLANG:vektor soll „entweder direkt im In-Ear-Monitor integriert oder an die Ohrstöpsel angeklebt“ (Nötges, 2019, S. 116) werden. Auch die Firma Apple bietet In-Ear Kopfhörer mit eingebautem Headtracker an. Seit einem Softwareupdate im Herbst 2020 können der in den *Apple AirPods Pro* (siehe 3.3.4) verbaute Beschleunigungssensor und Gyrosensor zur dynamischen Binauralsynthese verwendet werden (vgl. De Leuw, 2020).

Eine im System integrierte auditive Messung zur Individualisierung wird aktuell nur beim *Realiser* angeboten (siehe 4.6.1), wobei es sich hierbei um eine BRIR-Messung und nicht um eine HRTF-Messung handelt (siehe 4.3.2.2). Obwohl einige aktuelle In-Ear Kopfhörer Mikrofone integriert haben (siehe 3.3.4), die als Messmikrofone genutzt werden könnten, bleiben es der Messaufbau unter Freifeldbedingungen und die hohe Messdauer, die eine auditive Messung so aufwendig machen. Die Entwickler der Firma KLANG:technologies forschen derzeit an einem massentauglichen HRTF-Messverfahren, das laut Mitgründer und Marketingchef Pascal Dietrich maximal drei Minuten dauere und bei dem die Nutzerinnen und Nutzer selbst auf einem Drehteller im schalltoten Raum stehen sollen (vgl. Nötges, 2019, S. 114). Einige professionelle Tonschaffende mögen auf dieses Angebot zurückgreifen, für den reinen Musikkonsumenten oder -konsumentin wird eine solche Messung trotz der verkürzten Messdauer zu aufwendig und zu kostspielig bleiben. Visuelle Messungen, wie sie von der Firma Genelec angeboten werden, scheinen dabei anwenderfreundlicher zu sein, allerdings nur, sofern die Kosten dafür in Zukunft deutlich gesenkt werden können (siehe 4.3.2.3).

## 5 Binaurale Inhalte

In den vorangegangenen Kapiteln wurde aus technischer Sicht auf die Chancen und Herausforderungen der binauralen Audiotechnik eingegangen. Im folgenden Abschnitt soll binaurales Audio aus inhaltlicher und gestalterischer Sicht betrachtet werden. Die Inhalte auditiver Medien werden dabei aufgrund ihrer Gestaltungsmerkmale in drei Kategorien eingeteilt: Musik, Radio bzw. Podcasts und Hörspiele bzw. Hörbücher, wobei die Grenzen zwischen diesen drei Gruppen fließend sind. Es soll ausgearbeitet werden, welche Möglichkeiten zur Klanggestaltung die dynamische Binauralsynthese mit sich bringt und welche auditiven Inhalte davon profitieren könnten.

„Klanggestaltung bezeichnet die technische Bearbeitung von Audiosignalen mit dem vornehmlichen Ziel der Beeinflussung der Ausprägung klanglicher bzw. musikalischer Wahrnehmungsmerkmale“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 370). Dazu dienen technische Mittel zur Bearbeitung der Lautstärkeverhältnisse, der Panormaverteilung und der frequentiellen, dynamischen, räumlichen und harmonischen Klangeigenschaften (vgl. ebd., S. 370). Diese Arbeit beschränkt sich auf die Betrachtung der klanggestalterischen Maßnahmen durch die Panoramaverteilung und die dadurch beeinflusste Lokalisation eines oder mehrerer Elemente innerhalb der Binauralmischung und die räumlichen Klangeigenschaften. Die Studien, denen die nachfolgenden Umfrageergebnisse entnommen wurden, werden in Kapitel 6 genauer beschrieben.

### 5.1 Binaurale Musik

Schließt man anhand der Marktanteile von Musik, Radio bzw. Podcasts und Hörspielen auf das Interesse der Konsumentinnen und Konsumenten für die jeweilige Inhaltskategorie, dann wird deutlich, wie hoch die Nachfrage nach Musik auf dem Audiomarkt ist: 98% aller Nutzerinnen und Nutzer von Streamingdiensten streamen Musik (siehe 6.1). Im Schnitt hören die befragten Personen der „Studie zur Zukunft der Musik“ 21 Stunden und 30 Minuten Musik pro Woche (vgl. Clement, Lepthien, Kandziora, & Meyn, 2018a, S. 19). Daraus lässt sich schließen, dass die Musik aus Konsumentensicht zu den relevantesten Kategorien von Audioinhalten gehört. Im Folgenden soll anhand praktischer Beispiele geklärt werden, ob die binaurale Mischung von Musik das Hörerlebnis verbessern kann und welchen Mehrwert die Musik dadurch hätte.

#### 5.1.1 Gestaltungsmerkmale von Musik

Mit Musik ist die Aufnahme und Wiedergabe von Klängen gemeint, die von einem oder mehreren Musikschaaffenden in Form eines Konzerts oder einer Studioaufnahme erzeugt werden. Die Unterscheidung zwischen binauralen Studioproduktionen und binauralen Konzertmitschnitten im folgenden Abschnitt beruht auf Weinzierls Definition der unterschiedlichen Zielsetzungen bei der Klangbearbeitung von auditiven Inhalten: ein Konzertmitschnitt verfolgt ein illusionistisches Klangideal, „Aufnahmen orientieren sich klanglich am Aufführungsoriginal und zielen auf Natürlichkeit ab [...]. Audiobearbeitung wird im Rahmen des illusionistischen Konzepts [...] gezielt und ggf. deutlich zur

Überhöhung klanglicher Eigenschaften eingesetzt“ (Weinzierl, 2008, S. 779). Eine Studioproduktion dagegen verfolgt ein medial-autonomes Klangideal, „ein kreativer und innovativer Einsatz von Audiotreatmentmitteln ist hier nicht nur zulässig, sondern geradezu erforderlich. So nimmt populäre Musik zumeist nicht auf vorbestehende oder klanglich klar definierbare Aufführungssituationen Bezug“ (ebd., S. 779). Deswegen werden als Studioproduktionen vorwiegend Beispiele der Genres Pop, Rock und elektronischer Musik beschrieben, wohingegen Beispiele der Genres Klassik und Jazz eher dem Bereich Konzertmitschnitt zugeordnet werden.

## 5.1.2 Beispiele für binaurale Studioproduktionen

### 5.1.2.1 Studioproduktionen mit Kunstkopfstereofonie

Inspiziert von den Einsatzmöglichkeiten des neu erschienenen *Neumann KU80* (siehe 3.2.1) veröffentlichte der US-amerikanische Singer-/Songwriter Lou Reed 1978 mit *Street Hassle* (siehe Lou Reed, 1978) eines der ersten binauralen Studioalben. Die Basisaufnahmen entstanden bei einem Konzert in Ludwigshafen, das mit mehreren Kunstkopfmikrofonen aufgenommen wurde, die über dem Publikum schwebend platziert wurden. Die Kunstkopfaufnahmen klangen sehr indirekt und räumlich. Der Toningenieur Rod O’Brien sagte dazu:

“There was a certain percentage of that ambience, that binaural stuff, that we couldn’t go past. Because if we did, it started to sound a little soft and mushy – and soft and mushy wasn’t what Lou Reed did. Lou, definitely, even on some of his softest stuff, he was right there. He wanted you to hear it right in your face” (Love, o.D.)

Um einen weniger räumlichen, direkteren Klang zu erzeugen nahm Reed in einem New Yorker Studio nachträglich mehrere Instrumente im Overdub-Verfahren auf. Die Mikrofone wurden so im Bezug zu den jeweiligen Instrumenten, wie beispielsweise den Gitarren-Amps positioniert, dass deren Abstand und Ausrichtung den Pegel- und Laufzeitunterschieden der Kunstkopfaufnahme entsprach. Dieser aufwendige Prozess sollte gewährleisten, dass die Direktsignale und die Kunstkopfaufnahme gemischt werden können, ohne die räumliche Verteilung der Binauralaufnahme zu verfälschen (vgl. Love, o.D.). 1978 und 1979 veröffentlichte Reed zwei weitere binaurale Alben. Mit *Binaural* (siehe Pearl Jam, 2000) folgten Pearl Jam im Jahr 2000 Reeds Beispiel, Perfume Genius veröffentlichten 2007 das binaurale Studioalbum *No Shape* (siehe Perfume Genius, 2017). Im gleichen Jahr experimentieren auch Radiohead mit Kunstkopfaufnahmen (vgl. Hooke Audio, 2017).

Die Beschreibung des Aufnahmeprozesses zu *Street Hassle* bestätigt die in Abschnitt 3.2.7 genannten Einschränkungen einer Kunstkopfaufnahme und erklärt gleichzeitig, warum es bis heute nur vereinzelte Kunstkopf-Studioalben gibt: die Nachbearbeitung einer Kunstkopfaufnahme ist nur sehr eingeschränkt möglich. Der Musiker Damon Kurkowski sagt dazu: „fidelity to sound at large is not necessarily the same as fidelity to how we listen to music“ (Krukowski, 2014). Die Stärke eines Kunstkopfmikrofons ist die realitätsgetreue Reproduktion natürlicher Klänge, doch das ist nicht das, was die Konsumentinnen und Konsumenten von einem modernen Studioalbum im Bereich Pop- und

Rockmusik erwarten. Ernst Rothe, EMI Chef-Tonmeister, sagte dazu bereits in den 70er-Jahren: „the sound, the dummy head records, is not what listeners want to hear these days” (Krebs, 2016).

### 5.1.2.2 Studioproduktionen mit Binauralsynthese

Aus technischer Sicht bietet die Binauralisierung durch Binauralsynthese gegenüber der Kunstkopfstereofonie einige Vorteile (siehe 4.1). Auch aus gestalterischer Sicht liefert die Binauralsynthese neue Möglichkeiten für die Gestaltung einer Musikmischung. Gerade weil die Binauralsynthese ein aktueller Forschungsgegenstand ist (siehe 4.6) gibt es sehr wenige Studioproduktionen, die mit der Binauralsynthese explizit für eine binaurale Wiedergabe gemischt wurden. „Putting people in spaces hasn't really been part of the paradigm of producing music. So that's completely new” (Neal, 2016). Daraus lässt sich schließen, dass es noch keinerlei etablierte Hörgewohnheiten für 3D-Audioproduktionen im Bereich Musik gibt. Trotzdem sollen die folgenden Beispiele einen Eindruck darüber liefern, wie die Binauralsynthese bei aktuellen Studioproduktionen angewendet wird.

2016 veröffentlichten der Deep-House Produzent Robin Schulz und Sänger Nico Santos die Single *More Than A Friend* (siehe Robin Schulz feat. Nico Santos, 2017), die in Zusammenarbeit mit der Firma Sennheiser entstand. Die Single wurde binaural gemischt und demonstrierte die Anwendung der Sennheiser *Ambeo* 3D-Audio Produkte (siehe 4.6.1) im Zuge einer Werbekampagne (vgl. Wernke-Schmiesing, 2016).

2014 veröffentlichte die schwedische Rock- und Popband Mando Diao ihr Album *Aelita* (siehe Mando Diao, 2014) auf einer Blue-Ray-Disk als *Auro 3D* Mischung, ein 3D-Audio Format der Firma Auro Technologies zur Lautsprecherwiedergabe im Kino und Heimkino. Die *Auro 3D* Mischung wurde durch einen Binaural-Dekoder (siehe 4.5.1) binauralisiert (vgl. Auro Technologies, 2014). Carlos Barth, der Manager der Band, sagt zum Mehrwert der 3D-Mischung im Vergleich zur einer Stereomischung folgendes: „You build tension, you build unexpectedness. You can build up from, not just in volume and dynamics, but you can build up by direction and moving it forward to a focus” (Auro-3D, 2014). Dabei bezieht er sich auf die musikalische Steigerung innerhalb eines Arrangements eines Liedes und wie diese durch die Bewegung einzelner musikalischer Elemente im Panorama unterstützt werden kann.

2017 veröffentlichten Kraftwerk mit dem Album *3-D Der Katalog* (siehe Kraftwerk, 2017) ein Studioalbum, das für das 3D-Audioformat Dolby Atmos (siehe 6.5.2) und als Binauralmischung gemischt wurde. Das Album enthält die Lieder acht bereits veröffentlichter Studioalben als dreidimensionale Live-Versionen und wurde für den Grammy 2018 in der Kategorie Best Surround Sound Album nominiert (vgl. Klug, 2017). *3-D Der Katalog* wurde in Zusammenarbeit von Bandmitglied Fritz Hilpert und Tom Ammermann, dem Geschäftsführer von New Audio Technologies gemischt. Aus einem Mehrspurmitschnitt der Live-Konzerte erstellten sie drei verschiedene Mischungen: eine Dolby Atmos-Mischung für Heimkinos, eine Dolby Atmos-Mischung für große Kinosäle und eine

Binauralmischung. Für die Binauralsynthese wurde der *Spatial Audio Designer* (siehe 4.6.1) verwendet. Ammermann sagt über die Produktion:

„Die Musik von Kraftwerk [...] hat mir eine einmalige Gelegenheit gegeben, kreativ mit den Objektfahrten umzugehen. Ich spreche in diesem Zusammenhang auch gerne von Plausibilität. Damit meine ich, dass bei einer Mischung in 3D der Gesamteindruck homogen bleiben muss. Unrealistische Effekthascherei gilt es dabei tunlichst zu vermeiden“ (Stede, 2018).

Die, im Vergleich zu einer herkömmlichen Stereomischung, neuen Möglichkeiten zur Verteilung der Elemente in einer Binauralmischung sollten also den musikalischen Ausdruck unterstützen. „Effekthascherei“ bezeichnet in diesem Zusammenhang die ungewöhnliche, effektvolle und für die Zuhörenden möglichst auffällige Panoramaverteilung der Elemente in der Mischung.

### **5.1.2.3 8D-Audio**

Seit Herbst 2018 gibt es immer mehr „8D-Audio“-Versionen von bekannten Liedern. Die „8D-Audio“-Versionen werden vorwiegend auf der Videoplattform Youtube veröffentlicht und stoßen auf eine sehr positive Resonanz seitens der Youtube-Nutzerinnen und -Nutzer. „8D-Audio“ ist ein freigeählter Begriff, der keinen plausiblen Bezug zu acht „Dimensionen“ oder acht „Richtungen“ hat und scheint bewusst irreführend zu sein „um die Leute zum Klicken zu bringen“ (Peysner-Kreis, 2019). Für einen „8D-Audio“-Remix wird der Stereomaster eines Liedes binauralisiert. Der Binaural-Panner wird automatisiert, sodass eine für „8D-Audio“ typische Kreisbewegung auf der Horizontalebene (siehe 2.3) entsteht. Die Rotationsgeschwindigkeit kann variieren (vgl. Rieger, o.D.). Aus gestalterischer Sicht wirkt diese Form der Binauralmischung eher unmusikalisch und kann mit Ammermanns Begriff der „Effekthascherei“ assoziiert werden (siehe 5.1.2.2). Doch gerade, weil der Effekt sehr plakativ ist, scheint „8D-Audio“ ein hohes Maß an Akzeptanz bei den Internetnutzerinnen und -nutzern zu finden: „Der Hörer merkt schnell, dass hier eine Bewegung ist, die er so noch nie gehört hat“ (ebd.).

### **5.1.3 Beispiele für binaurale Konzertmitschnitten**

Ähnlich wie bei binauralen Studioproduktionen wurden die ersten binauralen Konzertmitschnitte als reinen Kunstkopfaufnahme realisiert (siehe 5.1.2.1). Die britische Rockband The Rolling Stones veröffentlichten 1991 das Live-Album *Flashpoint* (siehe The Rolling Stones, 1991), das 1989/90 auf der Steel Wheels/Urban Jungle Tour mit einem Kunstkopfmikrofon aufgenommen wurde (vgl. Fandom, o.D.).

Als Werbekampagne für ihre Produkte realisierten ab den 2010er Jahren verschiedene VR-Firmen weitere binaurale Konzertmitschnitte mit der Vision, „virtuelle“ Livekonzerte zu ermöglichen, die dem echten Konzerterlebnis in nichts nachstehen. David Chesky, Gründer des New Yorker Labels Chesky Records sagte dazu im Jahr 2016:

"In 10 years you could be living in Tahiti as a surfer and you'll go to your living room and say, 'Tonight should I see the Philharmonic? No? The Paris ballet? The Stones in Madison Square Garden?' And you're there" (Neal, 2016).

#### 5.1.3.1 Becks „Sound and Vision“

2013 führten der US-amerikanische Musiker Beck und ein 160-köpfiges Orchester im Rahmen einer Jubiläumsveranstaltung des Autoherstellers Lincoln eine Neuinterpretation von David Bowies *Sound and Vision* (siehe David Bowie, 1977) auf. Der Auftritt wurde als 360°-Video (siehe 4.4.1) mitgeschnitten (vgl. McMillan, 2013). Musikvideo-Regisseur Chris Milk und Designer Alan Scott entwickelten hierfür ein speziell angefertigtes Kunstkopfmikrofon: „a head with ears all around its circumference“ (Watercutter, 2013). Um den Kunstkopf sind acht Ohrmodelle verteilt, dadurch werden vier Kopfausrichtungen gleichzeitig aufgenommen (siehe Abbildung 26):



Abbildung 26: 360°-Kunstkopfmikrofon für Becks „Sound and Vision“ (Constine, 2014)

Mehrere solcher Kunstkopfmikrofone wurden zusammen mit mehreren 360°-Kameras auf der Bühne und im Zuschauerraum verteilt. Lincoln veröffentlichte das Video anschließend als 360°-Video im Internet und ergänzte das Video durch interaktive Elemente: so konnte das Internetpublikum auf einer Übersichtskarte die Position auswählen, von der aus das Konzert betrachtet werden sollte (siehe Abbildung 27: links unten). Die Verwendung einer VR-Brille (siehe 4.4.1) erlaubte außerdem die dynamische Anpassung des Videoausschnitts und des Tons an die aktuelle Kopfausrichtung der Zuschauerinnen und -zuschauer. Falls keine VR-Brille vorhanden war, wurde die dynamische Nach-

führung von Bild und Ton durch ein optionales Webcam-Headtracking (siehe 4.4.6.2) ermöglicht (siehe Abbildung 27: rechts oben) (vgl. Constine, 2014):

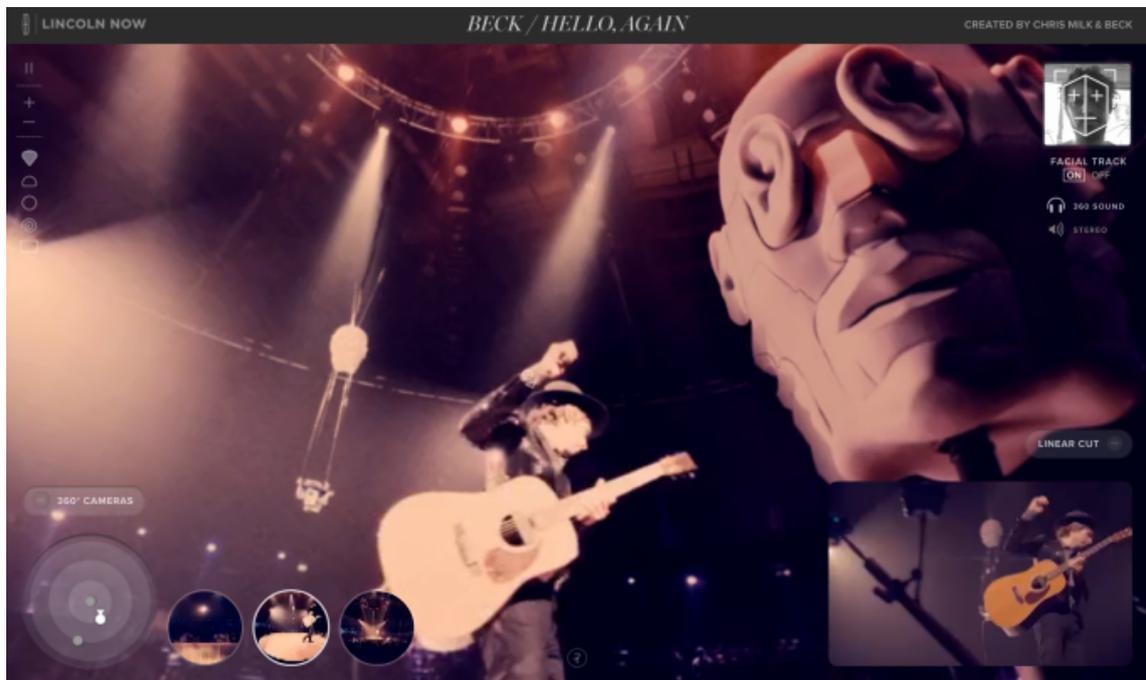


Abbildung 27: Screenshot des 360°-Videos zu Becks "Sound and Vision" (Constine, 2014)

### 5.1.3.2 BBC Proms

Die BBC Proms sind eine traditionelle Konzertreihe für klassische Musik, die seit 1895 stattfindet und seit 1942 in der Royal Albert Hall in London beheimatet ist. Im Jahr 2016 veröffentlichte die britische Rundfunkanstalt BBC zusätzlich zu den Stereomischungen erstmals binaurale Mischungen von einer Auswahl an Stücken, die bei den BBC Proms aufgeführt wurden. Die binauralen Mischungen werden auf der Website der BBC Proms veröffentlicht und seit 2017 auch live über das Internet gestreamt (vgl. Firth, 2019). Als Hauptmikrofon dient das 3D-Mikrofonarray *Schoeps ORTF-3D*, das über den Zuschauerraum hängt (siehe Abbildung 28):

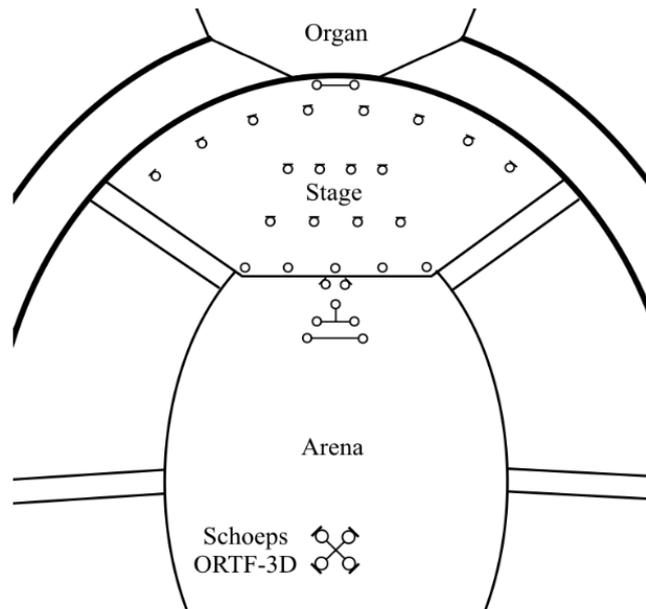


Abbildung 28: Mikrofonaufbau BBC Proms (Parnell & Pike, 2018, S. 2)

Die Direktmikrofone werden entsprechend ihrer Position auf der Bühne gepannt und ihre Laufzeiten an die Laufzeitverzögerungen auf dem Hauptmikrofonarray angepasst. Die Postproduktion findet zunächst unabhängig vom Ausgabeformat statt (vgl. Parnell, 2016). Für die Binauralsynthese wurde anfangs IRCAMs *Panoramix* verwendet, was Binauralpanner, Equalizer, Delay und Binaural-Renderer beinhaltet. Für die Binauralisierung wurden HRTFs des *Neumann KU 100* (siehe 3.2.1) verwendet (vgl. Parnell & Pike, 2018, S. 3). Seit 2018 verwendete die BBC ihren eigens entwickelten *BBC Renderer* für die Postproduktion. Der *BBC Renderer* ist individuell auf die Arbeitsabläufe der BBC-Toningenieurinnen und -ingenieure abgestimmt und erleichtert unter anderem die Erstellung unterschiedlicher Ausgabeformate und das Live-Rendering. Er erlaubt bis zu 128 Eingangskanäle und vereinfacht die Editierung der Signale in Gruppen (vgl. Firth, 2019).

### 5.1.3.3 VR-Konzerte

Samsung veröffentlichte 2014 über die App *NextVR* ein Konzert der britischen Pop- und Rockband Coldplay, das wie Becks *Sound and Vision* (siehe 5.1.3.1) den „virtuellen“ Konzertbesuch (siehe 4.4.1) ermöglichte, in diesem Fall allerdings ausschließlich mit der Samsung VR-Brille *GearVR* als Wiedergabegerät. Bei der *GearVR*-Brille wird das Smartphone in einer Halterung vor den Augen platziert und dient einerseits als Bildschirm und andererseits als Headtracker (siehe 4.4.6.3). Sowohl Video, als auch Audio werden dynamisch an die Kopfausrichtung der Nutzerinnen und Nutzer angepasst, außerdem kann zwischen der Position auf der Bühne oder im Publikumsraum gewechselt werden (vgl. Takahashi, 2014). Neben Samsungs *NextVR* gibt es mittlerweile einige andere Anbieter von VR-Konzerten, wie beispielsweise *MelodyVR* (siehe Magis, 2017a), *NoysVR* (siehe Magis,

2017b) oder *Amaze* (siehe Kümmeler, 2020). Wie genau die Konzertinhalte audioseitig aufgenommen und produziert werden wird von den Anbietern nicht verraten.

Obwohl es VR-Konzerte schon länger gibt, ist das Format bis jetzt nicht in der breiten Masse angekommen. Die Consumer Technology Association, ein Wirtschaftsverband der USA, kam nach einer Umfrage zu dem Schluss, dass die Menschen zwar an VR interessiert seien, VR aber eher mit dem Gaming-Sektor als mit Livemusik assoziiert werde. Es brauche mehr Inhalte im Bereich Film und Musik (vgl. DeLisa Coleman, 2017). Entgegen der eingangs beschriebenen Meinung von Chesky (siehe 5.1.3) sagt Jessica Brillhart, VR-Filmmacherin bei Google, dass es bei einem realen Konzert nicht nur um die Musik gehe, sondern auch um das soziale Erlebnis beim Konzertbesuch: „It be less about the music directly and more about the experience of this music in a unique space and at a unique time with people who you've never met“ (DeLisa Coleman, 2017).

#### **5.1.4 Mehrwert durch binaurale Audiotechnik**

In den vorangegangenen Abschnitten wurden einige Beispiele für binaurale Musik genannt. Anhand dieser Beispiele lassen sich folgende Eigenschaften binauraler Musik zusammenfassen:

Mithilfe einer binauralen Mischung kann der musikalische Ausdruck des jeweiligen Stücks unterstützt werden. So kann beispielsweise durch Objektfahrten das Arrangement des Stücks klarer werden. Durch Automationsfahrten kann ein Fokus auf bestimmte Elemente innerhalb des Panoramas gesetzt werden (siehe 5.1.2.2).

Die Stärke der Kunstkopfstereofonie ist die realitätsgetreue Abbildung eines natürlichen Klangbilds. Die Beispiele und Bewertungen von Studioproduktionen mit Kunstkopfstereofonie (siehe 5.1.2.1) haben gezeigt, dass dies nicht unbedingt den Erwartungen der Hörerinnen und Hörer an eine moderne Audioproduktion entspricht. Deswegen experimentierten Musiker wie Lou Reed damit, binauralisierte und nicht-binauralisierte Elemente in der Musikmischung zu kombinieren. Die nicht-binauralisierten Elemente werden im Kopf lokalisiert (siehe 3.1.1) und klingen dadurch druckvoller und direkter, so wie es die Hörerinnen und Hörer etwa vom Klang des Lead-Gesangs erwarten. Das gleiche Prinzip kann auf die Studioproduktionen mit Binauralsynthese übertragen werden.

Beim binauralen Konzertmitschnitt ist es ebenfalls die Kombination aus binauralisierten und nicht-binauralisierten Elementen, die einen Mehrwert der Binauralmischung gegenüber einer Stereomischung verspricht. So kann ein 3D-Mikrofonsetup, wie beispielsweise das *Schoeps ORTF-3D* (siehe 5.1.3.2) oder ein spezielles 3D-Kunstkopfmikrofon (siehe 5.1.3.1) als Hauptmikrofon dienen. Durch die Binauralisierung des Signals des Hauptmikrofons wird eine realistische Darstellung der Raumakustik des Konzertsaals erzeugt. Die Mono- und Stereo-Stützmikrofone an den einzelnen Instrumenten können entweder ebenfalls binauralisiert werden oder als nicht-binauralisierte Elemente für einen direkteren Klang sorgen, da sie dann wie bei einer herkömmlichen Stereomischung im Kopf lokalisiert werden (siehe 3.1.1).

## 5.2 Binaurales Radio und Podcasts

Zwar werden im Radio auch Musik oder Hörspiele wiedergegeben, trotzdem bezieht sich der Begriff Radio im Kontext dieser Arbeit ausschließlich auf die Übertragung von Sprache per Hörfunk. Als Podcast wird eine Sonderform von Radio bezeichnet. Er ist im Gegensatz zum Radio zeitunabhängig und wird als Audiodatei auf das Smartphone, den Computer oder ein anderes Wiedergabegerät heruntergeladen oder über das Internet gestreamt. Die Inhalte von Podcasts sind vielfältig und reichen „vom lockeren Gespräch am Küchentisch bis zum facettenreichen Hörspiel“ (BR, 2019). Sie werden von Privatpersonen, Prominenten, Unternehmen, Zeitungen und Radiosendern über Musikstreamingdienste oder eigene Mediatheken angeboten (vgl. ebd.).

Laut Online-Monitor hören 27% aller Befragten zumindest gelegentlich Podcasts und Webradio auf Abruf (vgl. Online-Audio-Monitor, 2019, S. 59). 17% der Befragten hören ein oder mehrmals im Monat Podcasts und Webradio (siehe 6.1).

### 5.2.1 Gestaltungsmerkmale von Radio und Podcasts

Wie eingangs erwähnt sagt die Klassifizierung Radio bzw. Podcast nichts über den redaktionellen Inhalt aus, da Radio bzw. Podcasts sowohl Musik, als auch Hörspiele enthalten können. Laut Online-Monitor sind die genutzten Inhalte von Radio bzw. Podcasts wie folgt verteilt: 70% Infosendungen, Wissens- und Lernbeiträge, 62% Musik, 55% Unterhaltung/Comedy, 54% Nachrichten, 44% Hörbücher/Hörspiele, 37% Regionales/Lokales, 35% Service, Ratgeber, Coaching, 19% Sportsendungen/Liveübertragungen (vgl. ebd., S. 68). Im Kontext dieser Arbeit werden lediglich die Inhalte betrachtet, bei denen Sprache als Hauptinhaltsträger dient, auch die folgenden Gestaltungsmerkmale beziehen sich auf Sprache.

Wie in Abschnitt 2.3.5 beschrieben, kann der auditive Sinneseindruck von anderen Sinneseindrücken, wie dem Sehsinn unterstützt werden. Gerade bei Sprache „kann das Sehen der Lippenbewegungen eines Sprechers zur Sprachverständlichkeit beitragen, auch wenn die Sprache alleine, etwa durch akustische Verdeckungseffekte oder Schwerhörigkeit des Zuhörers, keine ausreichende Information liefern würde“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 989). Dies erklärt, warum die Sprachverständlichkeit bei einem rein auditiven Medium wie Radio und Podcasts, bei dem der visuelle Sinneseindruck fehlt so wichtig ist. Folglich ist das Ziel bei der Wiedergabe von binauralem Radio und Podcasts, einen möglichst hohen Grad an Sprachverständlichkeit für alle Zielgruppen zu erreichen (vgl. ebd., S. 1247). Neben der Sprachverständlichkeit sind die Spracherkennung, also die Zuordnung einer Stimme zu einer bestimmten Person und die Lokalisation einer oder mehrerer Personen wichtige Gestaltungsmerkmale (vgl. Drullman & Bronkhorst, 2000, S. 2224).

## 5.2.2 Sprachverständlichkeit

Die Sprachverständlichkeit wird durch Silben-, Wort- und Satzverständlichkeit bestimmt. Dabei gilt, dass ein Wort eher verständlich ist, als eine Silbe, da eine Silbe aus dem Wortzusammenhang erschlossen werden kann. Gleichzeit ist ein Satz eher verständlich, als ein Wort, da ein Wort aus dem Satzzusammenhang erschlossen werden kann (vgl. Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 72).

Ein gängiges Vergleichsmaß für die Sprachverständlichkeit ist der so genannte Speech Reception Threshold (SRT), die minimale Verständlichkeitsschwelle. Der SRT wird in Dezibel angegeben und ist der SNR (siehe 4.3.2.1), bei dem 50% der gesprochenen Worte verständlich sind (vgl. Ben-Kiki, Ben-Asher, & Neoran, o.D.).

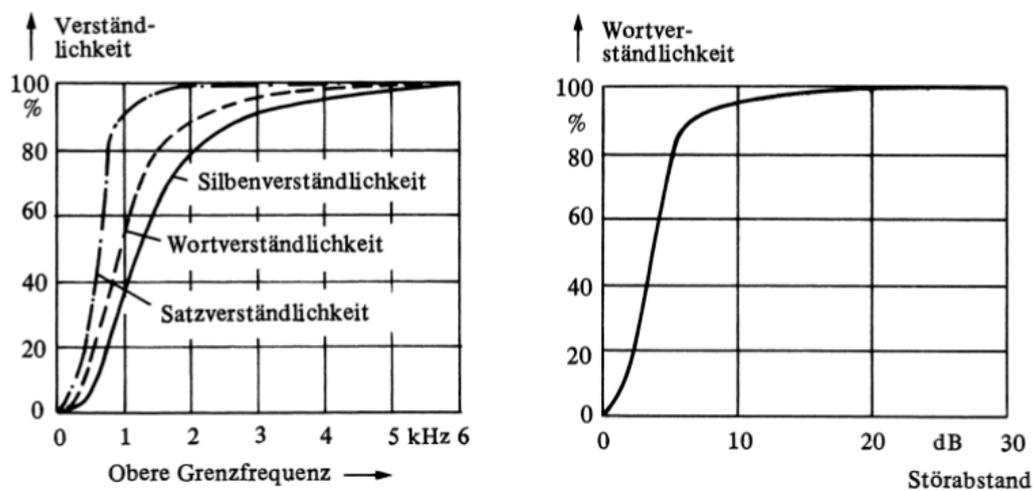


Abbildung 29: Sprachverständlichkeit (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 73)

Abbildung 29 (links) zeigt die Sprachverständlichkeit in Abhängigkeit von der oberen Grenzfrequenz. Für das Frequenzspektrum bis 4 kHz gilt: Je höher die obere Grenzfrequenz ist, desto verständlicher ist die Sprache. Ab etwa 4 kHz gibt es keine Verbesserung mehr, da sich die Maxima des Frequenzspektrums der menschlichen Sprache im Bereich von 100 Hz bis 5 kHz befinden (vgl. Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 69). Die meisten Anwendungen von Sprachübertragung wie Radio oder Telekommunikation nutzen deshalb bandbegrenzte Signale bis 4 kHz (vgl. Drullman & Bronkhorst, 2000, S. 2225).

Abbildung 29 (rechts) zeigt die Wortverständlichkeit im Bezug zum Störabstand. Bei einem Störabstand von 5 dB gilt eine Wortverständlichkeit von 60%. „Für Normalhörende liegt ein angemessener SNR bei etwa 6 dB, bei Schwerhörigen jedoch oft über 10 dB“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 1249). Ein niedriger SRT und damit eine niedrige Verständnisschwelle verbessern also nicht nur das Konsumerlebnis für Normalhörende, sondern dienen auch der barrierefreien Audiokommunikation für alle Zielgruppen (siehe ebd., S. 1247ff).

### 5.2.3 Mehrwert durch binaurale Audiotechnik

Die dynamische Binauralsynthese von Sprache kann den SRT senken und die Sprachverständlichkeit verbessern. Im folgenden Abschnitt wird zwischen einem und mehreren Sprechenden unterschieden.

#### 5.2.3.1 Dynamische Binauralsynthese von einem Sprechenden

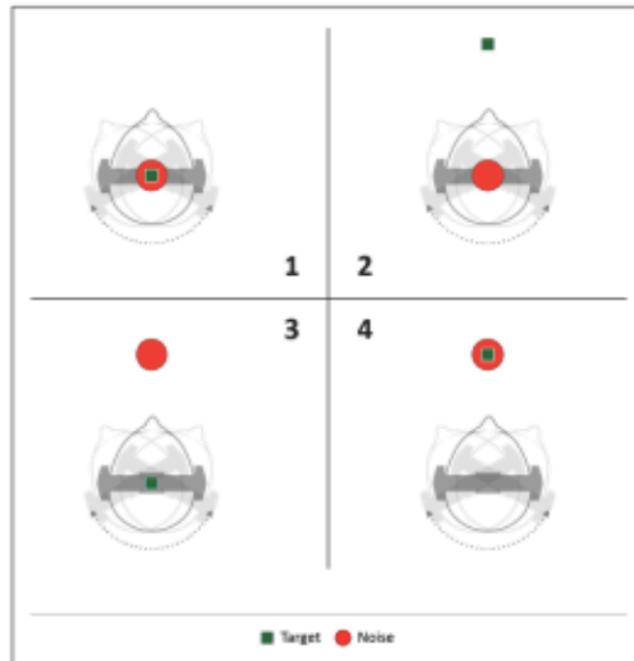


Abbildung 30: Dynamische Binauralsynthese von Sprache (grün) und Störsignal (rot) (Ben-Kiki, Ben-Asher, & Neoran, o.D., S. 2)

Ben-Kiki et al haben untersucht, welche Auswirkungen die dynamische Binauralsynthese von Sprache und Störgeräusch auf die Sprachverständlichkeit hat. Abbildung 30 zeigt die verschiedenen Tests. In Abbildung 30, 1 wurde weder Sprache (Abbildung 30: grünes Quadrat), noch Störsignal (Abbildung 30: roter Punkt) binauralisiert. In Abbildung 30, 2 - 4 wurde entweder nur die Sprache, nur das Störsignal oder beide Signale binauralisiert. Zur Binauralsynthese wurde die *Waves NX* Software (siehe 4.6.1) und ein *Waves NX* Headtracker (siehe 4.4.6.4) verwendet. Ben-Kiki et al konnten zeigen, dass der SRT um bis zu 8 dB verbessert werden kann, wenn nur die Sprache mit dynamischer Binauralsynthese externalisiert wird und das Störsignal im Kopf lokalisiert wird (siehe 3.1.1) (siehe Abbildung 30, 2). Werden Sprache und Störsignal externalisiert verbessert sich der SRT um 1 dB (siehe Abbildung 30, 4) (vgl. Ben-Kiki, Ben-Asher, & Neoran, o.D., S. 2).

#### 5.2.3.2 Dynamische Binauralsynthese von mehreren Sprechenden

Die Beeinflussung der Sprachverständlichkeit bei mehreren Sprechenden durch die Binauralsynthese von Sprache wurde hauptsächlich zur möglichen Anwendung von Binauraltechnik in der Telekommunikation untersucht. Die Ergebnisse sind gleichermaßen relevant für die Anwendung von Bin-

auralsynthese für Radio bzw. Podcasts, beispielsweise um die Verständlichkeit von mehreren Beteiligten in großen Diskussionsrunden zu verbessern. Generell gilt, dass die Sprachverständlichkeit nachlässt, je mehr Personen gleichzeitig sprechen (vgl. Drullman & Bronkhorst, 2000, S. 2234).

In Abschnitt 2.3.5 wurde der Cocktailparty-Effekt beschrieben, der es unter anderem erlaubt, einem Gespräch zu folgen, selbst wenn viele Gespräche gleichzeitig stattfinden. „Nachdem er Konversationen auf Partys monofon aufgenommen hatte und die Sprachverständlichkeit überraschend schlecht war“ (Dickreiter, Dittel, Hoeg, & Wöhr, 2014, S. 238), fand Colin Cherry bereits 1953 heraus, wie wichtig die stereofone Aufnahme und Wiedergabe für die Sprachverständlichkeit sind (siehe Cherry, 1953). Yost et al haben die Verständlichkeit von drei gleichzeitig sprechenden Personen beim normalen Hören, bei der monauralen Kopfhörerwiedergabe und für die binaurale Kopfhörerwiedergabe per Kunstkopfaufnahme verglichen. Die bandbegrenzten Sprachsignale wurden über ein halbkreisförmig angeordnetes Lautsprechersystem wiedergegeben. In der Mitte des Halbkreises wurde ein menschlicher Zuhörer, ein Mikrophon oder ein KEMAR-Kunstkopfmikrophon (siehe 3.2.5) platziert. Bei drei sprechenden Personen war die durchschnittliche Wortverständlichkeit beim normalen Hören und bei der binauralen Kopfhörerwiedergabe gleich. Sie lag bei etwa 40%. Bei der monauralen Kopfhörerwiedergabe lag sie beim gleichen Versuchsaufbau bei lediglich 18%. Aufgrund des Cocktailparty-Effekts, der sowohl beim normalen Hören, als auch bei der Kunstkopfaufnahme auftritt, lässt sich die Sprachverständlichkeit bei mehreren Sprechenden durch Binauralisierung verbessern (vgl. Yost, Dye, & Sheft, 1996, S. 1029ff). Die Wiedergabe bei diesen Versuchen war statisch. Ben-Kiki et al schreiben, dass sich durch Headtracking die Lokalisation und Sprachverständlichkeit bei mehreren Sprechenden aufgrund des Cocktailparty-Effekts weiter verbessert. Dies gelte sowohl für ein zweidimensionales, als auch für ein dreidimensionales dynamisches Klangfeld (vgl. Ben-Kiki, Ben-Asher, & Neoran, o.D., S. 3f).

Drullmann und Bronkhorst konnten zeigen, dass die Sprachverständlichkeit dann am höchsten ist, wenn das Sprachsignal aus den Richtungen  $-45^\circ$  oder  $45^\circ$  Azimut kommt. Je größer der Abstandswinkel zwischen den Sprechenden ist, desto besser ist die Sprachverständlichkeit. Drullmann und Bronkhorst zeigten außerdem, dass frontale Sprechende um  $0^\circ$  Azimut besser lokalisiert werden, als seitlich platzierte Sprechende (vgl. Drullman & Bronkhorst, 2000, S. 2234f). Wie in Abschnitt 2.3.1 beschrieben, ist die Lokalisationsschärfe bei frontalem Schalleinfall generell höher, als bei seitlichem Schalleinfall.

#### **5.2.4 Individualisierung für binaurales Radio und Podcasts**

In Abschnitt 4.3.2 wurde beschrieben, dass eine individuelle HRTF wesentlich zur Verbesserung der Lokalisierung einer Schallquelle beiträgt. Drullmann und Bronkhorst dagegen kamen zu dem Ergebnis, dass eine individuelle HRTF für bandbegrenzte Sprachsignale im Vergleich zu einer nicht-individuellen HRTF keine Verbesserung für die Sprachverständlichkeit, -erkennung und -lokalisierung bringt (vgl. ebd., S. 2231). Die nicht-individuelle HRTF wurde in einem unabhängigen Versuch als

Durchschnittswert aus der Typisierung von acht Personen (siehe 4.3.2.4) bestimmt (vgl. ebd., S. 2226). Begault und Wenzel kamen zum selben Ergebnis, ergänzen aber: „The evidence of individual differences found in the present study may suggest that some listeners were able to adapt more easily to the spectral cues of the nonindividualized HRTFs than others; or, put another way, perhaps their ears were more similar to SDOs” (Begault & Wenzel, 1993, S. 372), wobei “SDO” für die Person steht, deren HRTF als nicht-individuelle Durchschnitts-HRTF genutzt wurde (vgl. ebd., S. 362).

### **5.3 Binaurale Hörspiele**

Wie in Abschnitt 3.2.1 erwähnt war das Hörspiel „Demolition“ eine der ersten Audioproduktionen, die per Kunstkopfmikrofon aufgenommen und 1973 veröffentlicht wurde. Das zeigt, dass schon sehr früh die neuen gestalterischen Möglichkeiten eines binauralen Hörspiels erkannt wurden. Ein Hörspiel zeichnet sich dadurch aus, dass Sprache, Geräusch und Musik kombiniert werden (vgl. Maier, 2015, S. 41), für ein binaurales Hörspiel gelten folglich auch die Erkenntnisse aus den vorangegangenen Abschnitten 5.1 und 5.2.

Ein Hörspiel ist eine auditive Inszenierung einer meist fiktionalen Geschichte, deren Handlungsaussagen durch dramaturgische Mittel aus Theater und Film verstärkt werden. Als Feature wird ein Hörspiel mit non-fiktionalem dokumentarischem Inhalt bezeichnet. Bei einer Lesung liest eine Person im Monolog eine Geschichte aus einem Buch vor, was oftmals als Hörbuch bezeichnet wird. Andere Definitionen benutzen den Begriff Hörbuch als den Überbegriff für Hörspiel, Feature und Lesung und als Bezeichnung für das Medium, über das ein Verlag ein Hörspiel, ein Feature oder eine Lesung veröffentlicht (vgl. Gisselbrecht, o.D.). Im Kontext dieser Arbeit wird der Begriff Hörspiel als Überbegriff für alle genannten Unterkategorien verwendet und soll als Bezeichnung für jede Audioproduktionen stehen, bei der Klang und Sprache als zentrale Gestaltungselemente kombiniert werden.

#### **5.3.1 Gestaltungsmerkmale von Hörspielen**

Das Hörspiel besteht aus drei Elementen: Sprache, Geräusch und Musik. Der Monolog eines Sprechenden oder der Dialog mehrerer Sprechender wird durch den Einsatz von Sounddesign, also Geräuschen und Musik, unterlegt. Bei den meisten Hörspielen ist die Stimme das Hauptelement, führt durch die Handlung und dient als Hauptträger semantischer Information (vgl. Maier, 2015, S. 58ff). In einigen wenigen Hörspielen sind alle drei Elemente gleichberechtigt, die Stimme wird dann lautmalerisch eingesetzt und ist weniger semantischer Informationsträger, sondern dient mehr als „Instrumentarium und nutzt eine Form des Ausdrucks, der nicht durch das Wort bestimmt ist“ (ebd., S. 66).

Zwar ist das Geräusch als Gestaltungselement der Stimme untergeordnet und dient meistens lediglich der Unterstützung des gesprochenen Worts. Trotzdem lässt sich sagen, dass Geräusche „dem Hörer eine Vielzahl an präzisen Informationen vermitteln können, wofür sonst [...] viele Worte benötigt werden“ (ebd., S. 70). Als ein Beispiel dafür, welche inhaltlichen Informationen sich aus einem Geräusch ableiten lassen, soll der Klang mehrerer Schritte auf einem Kiesbett dienen: Das Geräusch

gibt einen Rückschluss auf die Schallquelle, die Schrittabfolge lässt auf die Anzahl der Personen schließen. Anhand des zeitlichen Abstands zwischen den Schritten lässt sich die Laufgeschwindigkeit der Person erahnen. Der Klang der Schritte liefert einen Rückschluss darüber, auf welchem Untergrund die Person läuft, wie dieser beschaffen ist, welches Wetter herrscht und an welchem Ort sich die Person befindet (vgl. Maier, 2015, S. 68ff).

Musik wird im Hörspiel grundsätzlich sparsamer eingesetzt, als beispielsweise beim Theater oder Film. Reine Hintergrundmusik wird selten verwendet. Wenn Musik verwendet wird, dann hat sie eine leitmotivische Handlungsfunktion, besitzt einen Symbolwert oder die musikalische Darbietung ist Teil der Handlung. Darüber hinaus hat Hörspielmusik oftmals eine Gliederungsfunktion und wird am Anfang und Ende eines Hörspiels und als Übergang zwischen den Szenen verwendet (vgl. Fischer, 1964, S. 118ff).

### **5.3.2 BBC „The Turning Forest“**

Die BBC VR-Produktion „The Turning Forest“ soll als ein Beispiel für ein binaurales Hörspiel dienen. Zwar handelt es sich hierbei nicht um eine rein auditive Produktion, allerdings wurde „The Turning Forest“ als Hörspiel geschrieben, das VR-Video (siehe 4.4.1) wurde erst nachträglich erstellt und basiert auf der Audioproduktion. Der Kurzfilm hatte seine Premiere auf dem Tribeca Film Festival 2016 (vgl. Watson, 2016).

Die Geschichte handelt von zwei Kindern, die im Wald spielen. Ein Kind rennt davon, das andere Kind bleibt im Wald und trifft auf ein riesiges Monster. Die Kreatur ist freundlich. Das Kind begleitet das Monster, reitet auf ihm und schwimmt auf dessen Rücken durch einen Fluss. Das Hörspiel beinhaltet alle drei Gestaltungselemente eines Hörspiels, also Sprache, Geräusche und Musik und wurde per dynamischer Binauralsynthese erstellt (vgl. Woodcock, et al., 2016, S. 3).

Die meisten Sprachaufnahmen fanden mit mehreren *Neumann RSM 191* Mikrofonen in einem Sprecherstudio statt. Für die Geräuschaufnahmen wurde ein Mikrofonarray aus mehreren *Sennheiser MKH 8050* in einem Wald aufgebaut. Die Mikrofone wurden kreisförmig in zwei Ebenen in 1,35 m und 2,45 m Höhe mit jeweils acht Mikrofonen aufgestellt, der Radius betrug 2,5 m. Für die tiefen Frequenzen wurden vier zusätzliche *Sennheiser MKH 20* Kugelmikrofone in der unteren Ebene platziert (siehe Abbildung 31).

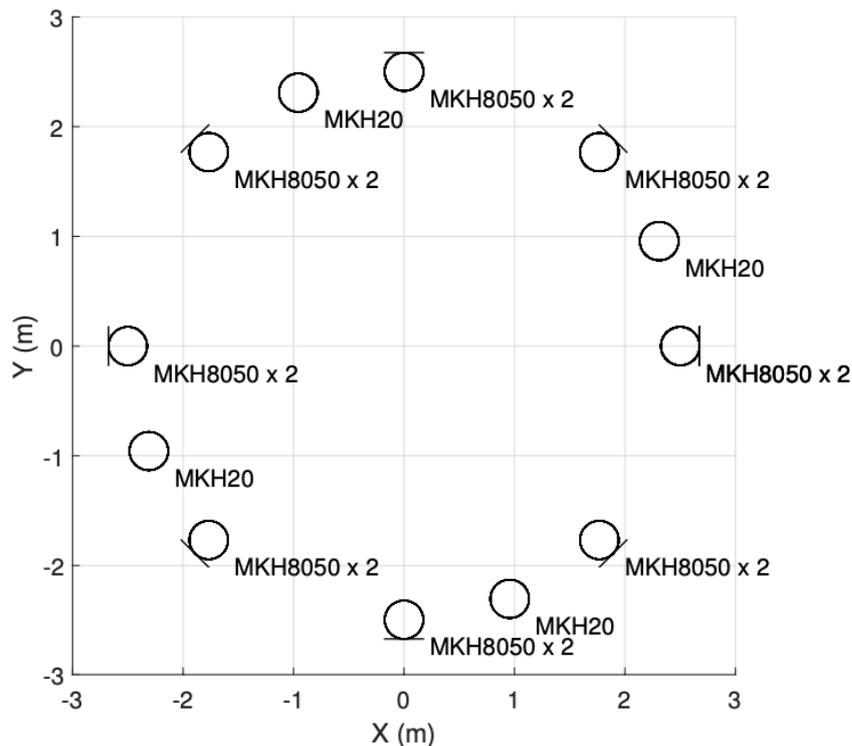


Abbildung 31: Mikrofonarray "The Turning Forest" (Woodcock, et al., 2016, S. 5)

Die Sprecher wurden für zusätzliche Sprachaufnahmen im Wald mit Lavalier-Mikrofonen ausgestattet. Die Geräusche, wie beispielsweise die Schritte der Kinder im Wald wurden mit einem *Neumann RSM 191* auf einer Ton-Angel aufgenommen (vgl. Woodcock, et al., 2016, S. 4).

Die Produktion wurde auf einem dreidimensionalen Lautsprecher-Setup gemischt, die Binauralsynthese der Objekte erfolgte mit einem eigens entwickelten Renderer (siehe 5.1.3.2) (vgl. ebd., S. 4). Anders als bei vielen herkömmlichen VR-Videos, erfolgt nicht nur die Videowiedergabe dynamisch, sondern auch das Audio passt sich an die Kopfausrichtung des Rezipienten oder der Rezipientin an. Das Headtracking erfolgt durch die VR-Brille (siehe 4.4.6.3). Bei der Premiere wurden die Gäste zusätzlich mit einem *SubPac* ausgestattet, das die tiefen Frequenzen unter 150 Hz, die ein Kopfhörer nur mangelhaft wiedergeben kann unterstützt (siehe 3.3.1) (vgl. Pike & Watson, 2016). „The Turning Forest“ lässt sich aktuell über die VR-Plattformen Samsung *GearVR* (siehe Oculus, 2017) oder Google *Daydream* (siehe Google Playstore, 2018) wiedergeben.

### 5.3.3 Mehrwert durch binaurale Audiotechnik

Wie in Abschnitt 5.3.1 beschrieben ist die Sprache das Hauptelement im Hörspiel, die Elemente Geräusch und Musik sind ihr untergeordnet. Wie die Sprachverständlichkeit und die Lokalisation eines oder mehrerer Sprechenden durch die dynamische Binauralsynthese verbessert werden kann wurde in Abschnitt 5.2 beschrieben. Welche Ansätze zur Mischung von binauraler Musik es gibt wurde in Abschnitt 5.1 gezeigt. Der folgende Abschnitt konzentriert sich deshalb auf die Wirkung von Geräuschen beim Sounddesign eines binauralen Hörspiels. Außerdem soll gezeigt werden, wie

die Kombination der drei Gestaltungselemente und die Binauralisierung eines Hörspiels die Dramaturgie der Handlung unterstützen können.

Die Binauralisierung erlaubt es, dass einige Geräusche und Klangeffekte noch eindrucksvoller wirken, als bei einem Stereo-Hörspiel. Der akustische Raum, indem sich Objekte platziert lassen, wird durch die Binauralisierung vergrößert. Catherine Robinson, Soundsupervisor bei der BBC, sagt dazu: „Writers should think in stereo. Think about things can be laid out between left and right. We now have the ability to lay things out 360 in the audio world“ (Collinson, 2016). Das wird am deutlichsten, wenn sich Objekte in der Mischung bewegen. So könnte beispielsweise der Klang eines Steinwurfes von rechts nach links, wie er in der BBC Produktion „Fire Rescue 360“ (siehe BBC Research & Development, 2016, ab 0:49 min) vorkommt oder der Klang eines Bienenschwarms, wie er in der BBC Produktion „The Resistance of Honey“ (siehe ebd., ab 2:05 min) vorkommt, in einer Binauralmischung noch realistischer wirken, als in einer Stereomischung. Die BBC Produktion „The Tragic Story of Betty Corrigall“ wurde binaural produziert. In einer Szene scheint sich die Sprecherin um den Kopf des Rezipienten zu bewegen und flüstert in dessen Ohr. Die Binauralisierung verdeutlicht die klangliche Wirkung der Bewegung um den Kopf und die scheinbar geringe Entfernung der Sprecherin zum Ohr der Rezipienten. Dadurch wird die unheimliche und düstere Atmosphäre, die das Hörspiel erzeugen will unterstützt (siehe ebd., ab 4:00 min).

Der, im Vergleich zur Stereoverision, vergrößerte akustische Raum erlaubt eine transparentere Mischung der einzelnen Elemente im Hörspiel. Bei „A Midsummer’s Night Dream“, einer BBC TV-Produktion mit binauralem Audio, bezieht sich diese Aussage speziell auf die Kombination aus orchesterlicher Musik und Sprache: „It helps capture the big bold sound of the orchestral score while letting the dialogue stand out with greater clarity“ (Moore, o.D.).

Ähnlich wie auch Rod O’Brien die Kunstkopfaufnahmen von Musik beurteilte (siehe 5.1.2.1), gibt Catherine Robinson zu bedenken, dass durch die gewonnene Räumlichkeit eines binauralen Hörspiels gleichzeitig etwas von der Intensität und Nähe einer Stereomischung verloren geht (vgl. Collinson, 2016). Eine Möglichkeit um dieses Problem zu lösen ist es, ausschließlich einzelne Elemente in der Mischung zu binauralisieren, wie es bereits in Abschnitt 5.2.3.1 in Bezug auf die Sprachverständlichkeit und in Abschnitt 5.1.4 in Bezug auf die binaurale Musikmischung beschrieben wurde. Bei einigen BBC Features, etwa in „Wireless Nights Series 5: Joujouka“ (siehe BBC Sounds, 2017, ab 1:22 min), werden ausschließlich die Umgebungsgeräusche binauralisiert, die Sprache nicht. Der Sprecher wird im Kopf lokalisiert (siehe 3.1.1). Dadurch wird eine klare Trennung zwischen der beschreibenden Erzählstimme und dem tatsächlichen Klangeindruck, der vor Ort aufgenommen wurde geschaffen (vgl. SCHOEPS Mikrofone, 2018).

## 6 Konsum binauraler Inhalte

Im folgenden Abschnitt soll binaurales Audio aus Sicht der Konsumentinnen und Konsumenten betrachtet werden. Die Ergebnisse unterschiedlicher Studien sollen ein bedarfsorientierter Eindruck darüber liefern, welchen Marktanteil die einzelnen Inhalte aktuell haben, über welche Wiedergabesysteme und Endgeräte sie abgespielt und wo sie konsumiert werden. Eine Bewertung der Studien soll die Ergebnisse in Bezug zu binauralem Audio setzen. Dadurch soll Aufschluss darüber gewonnen werden, wie das zukünftige Konsumverhalten bei einem breiteren Angebot von binauralem Audio aussehen könnte. Abschließend sollen drei aktuelle Technologien zum Konsum von binauralen Inhalten betrachtet werden.

Da es sich bei binauralem Audio um einen sehr aktuellen Forschungsgegenstand handelt (siehe 4.6), ließen sich keine aussagekräftigen Studien finden, die sich explizit auf dieses Thema beziehen. Stattdessen wurden drei themenverwandte Studien betrachtet, die sich mit unterschiedlichen Aspekten des Konsums von auditiven Medien im Allgemeinen beschäftigen.

Die erste Studie, der Online-Audio Monitor 2019 bezieht sich lediglich auf Audioinhalte, die online angeboten werden. Sie wurde von der Bayerischen Landeszentrale für neue Medien, dem Bundesverband Digitale Wirtschaft und dem Verband Privater Medien in Zusammenarbeit mit der Medienanstalt Berlin-Brandenburg, der Landesanstalt für Kommunikation Baden-Württemberg und der Landesanstalt für Medien NRW durchgeführt. Interviewt wurden 7554 Personen ab 14 Jahren per computergestützter telefonischer Befragung (CATI, Dual-Frame) mit mindestens 200 Personen pro Bundesland vom 06.05. bis 24.06.2019 (vgl. Online-Audio-Monitor, 2019, S. 110ff).

Die zweite Studie, die Bitkom Umfrage, bezieht sich nur auf Inhalte, die per Streaming-Dienst angeboten werden und besteht aus mehreren Teilmfragen, die durch den Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V. von Januar bis Juli 2019 per computergestützter telefonischer Befragung (CATI, Dual-Frame) durchgeführt wurden. Befragt wurden pro Teilmfrage durchschnittlich 1000 Personen per Zufallsstichprobe „über alle Altersgruppen hinweg“ (Bitkom, 2019, S. 27), ab 16 Jahren (vgl. ebd., S. 68ff).

Die dritte Studie, eine „Studie zur Zukunft der Musik“ der Universität Hamburg von 2018 liefert einen Eindruck über die Wiedergabesysteme, die am häufigsten zur Musikkwiedergabe genutzt werden. Befragt wurden 5140 Personen zwischen 16 und 70 Jahren im August 2018 per Online-Fragebogen über das Panel von Respondi. Die Auftraggeber der Studie sind der Bundesverband Musikindustrie e. V. (BVMI), der Bundesverband der Veranstaltungswirtschaft e.V. (bdv), die Gesellschaft für musikalische Aufführungs- und mechanische Vervielfältigungsrechte (GEMA), die Gesellschaft zur Verwertung von Leistungsschutzrechten (GVL), die Live-Musik-Kommission-Verband der Musikspielstätten in Deutschland e.V. (LIVEKOMM), der Verband der Deutschen Konzertdirektionen e.V. (VDKD) und der Verband unabhängiger Musikunternehmen e.V. (VUT) (vgl. Clement, Lepthien, Kandizora, & Meyn, 2018b, S. 1ff).

## 6.1 Welche Inhalte werden konsumiert?

Laut Online-Monitor 2019 nutzen 62% aller Befragten Online-Audio-Angebote (vgl. Online-Audio-Monitor, 2019, S. 4). Eine Umfrage zu den Inhalten ergab, dass 52% aller Befragten ein oder mehrmals im Monat Musik streamen, 41% Webradio, 17% Podcasts bzw. Webradio auf Abruf und 15% Hörbücher und Hörspiele hören (vgl. ebd., S. 14). Eine Umfrage des Bitkom ergab, dass 72% aller Befragten Musik im Internet streamen. 55% dieser Internetnutzerinnen und -nutzer streamen über Musikstreaming-Dienste (vgl. Bitkom, 2019, S. 26), davon streamen 98% Musik, 21% Podcast und 14% Hörbücher und Hörspiele (vgl. ebd., S. 30).

Eine weitere Umfrage des Online-Monitors 2019 ergab folgende Verteilung der redaktionellen Audio-Inhalte, die per Internet abgerufen werden: 86% aller Befragten hören Musik, 55% Nachrichten, 48% Dienste wie Wetter-, Verkehrs- oder Uhrzeitanzeigen, 44% regionale und lokale Informationen, 38% Informations-, Wissens- und Lernbeiträge, 35% Unterhaltung und Comedy, 21% Sportsendungen und -liveübertragungen, 20% Hörbücher und Hörspiele (vgl. Online-Audio-Monitor, 2019, S. 25).

Die Ergebnisse der Studien zeigen, dass Musik den mit Abstand größten Marktanteil hat. Radio, Podcasts und Hörspiele scheinen weniger relevant für die Konsumierenden zu sein. Zwar erfolgte die Befragung bei beiden Studien laut den Herausgebern über alle Altersgruppen hinweg (siehe 6), trotzdem sollte berücksichtigt werden, dass sich beide Studien ausschließlich auf online angebotene Inhalte beziehen. Eine gewisse Internetaffinität seitens der Befragten wird also vorausgesetzt. Daraus lässt sich schließen, dass die ältere Generation von Hörerinnen und Hörern in den vorliegenden Studien zu unterrepräsentiert ist, um die Ergebnisse auf das Verhalten aller Audiokonsumentinnen und -konsumenten zu übertragen. Da der Anteil von Musik als Audioinhalt jedoch bei allen Studien stark überwiegt, liefern die Ergebnisse trotzdem einen groben Eindruck über die Relevanz der einzelnen Inhaltskategorien.

## 6.2 Wie werden Inhalte konsumiert?

Binaurales Audio funktioniert nur über Kopfhörer. Deshalb ist es eine wichtige Frage, über welches Wiedergabesystem Audioinhalte aktuell konsumiert werden.

Laut der „Studie zur Zukunft der Musik“ nutzen 23% aller Befragten am häufigsten den im Endgerät integrierten Lautsprecher zu Musikwiedergabe, 22% hören Musik im Autoradio und 18% verbinden ihr mobiles Endgerät mit Kopfhörern. 18% hören Musik über eine Stereoanlage, 15% verbinden das mobile Endgerät mit einer Lautsprecherbox oder einer Stereoanlage und 4% verbinden ihr Smartphone mit den Autolautsprechern (vgl. Clement, Lepthien, Kandizora, & Meyn, 2018a, S. 10). Diskutabel ist im Kontext dieser Arbeit, dass sich die Umfrage lediglich auf die Wiedergabe von Musik bezieht und die Wiedergabe von Podcasts, Radio und Hörspielen hierbei außen vor gelassen wird. Die Wiedergabe per Kopfhörer ist eine zwingende Voraussetzung für den Konsum von binauralem Audio. Laut der Studie benutzt aktuell nur jeder fünfte Musikkonsumierende einen Kopfhörer. Ein

Großteil der Hörerinnen und Hörer muss sein Konsumverhalten also erst ändern und Kopfhörer verwenden, um binaurales Audio der Binauraltechnik entsprechend wiederzugeben.

Laut Bitkom Umfrage besitzen 71% aller Befragten einen Kopfhörer. In der Generation von 16 bis 49 Jahren besitzen im Schnitt 85% einen Kopfhörer, in der Generation von 50 bis 64 Jahren sind es 74% und erst ab 65 Jahren nimmt der Anteil der Kopfhörer besitzenden Personen mit 42% ab (vgl. Bitkom, 2019, S. 43). In der Altersgruppe von 16 bis 64 Jahren besitzen also mindestens drei von vier Personen einen Kopfhörer. Durch die weite Verbreitung von Kopfhörern erfüllt bereits ein Großteil der Konsumentinnen und Konsumenten die geräteseitigen Voraussetzungen zur binauralen Audiowiedergabe.

Bei der Bauform der Kopfhörer (siehe 3.3.2) zeichnet sich ein deutlicher Trend Richtung kabelgebundenen In-Ear-Hörer ab: 73% nutzen In-Ears mit Kabel, 28% Bügelkopfhörer mit Kabel, 22% In-Ears ohne Kabel und 13% Bügelkopfhörer ohne Kabel (vgl. ebd., S. 42f). Eine deutliche Mehrheit der Konsumentinnen und Konsumenten zieht also die Wiedergabe per In-Ear-Hörer der Wiedergabe per Bügelkopfhörer vor. Die Forschung im Bereich der Binauraltechnik sollte sich an diese Entwicklung anpassen und darf sich nicht auf die Untersuchung der Wiedergabe von binauralem Audio per Bügelkopfhörer beschränken (siehe 3.3.2). Hinsichtlich der Einschränkung der Wiedergabequalität seitens der Kopfhörer wäre eine Übersicht über die Verteilung der verwendeten Kopfhörermodelle aufschlussreich. Eine Umfrage über die Kopfhörermodelle, die Qualität oder das Preissegment der Kopfhörer ließ sich jedoch nicht finden.

### **6.3 Wo werden Inhalte konsumiert?**

Für die Gestaltung von binauralem Audio ist es relevant, in welchen Alltagssituationen Kopfhörer verwendet werden. Laut Bitkom Umfrage nutzen 51% der Befragten ihre Kopfhörer Zuhause, 46% beim Spaziergehen, 37% auf dem Fahrrad, 37% auf Zug- oder Flugzeugreisen, 30% beim Sport und 10% während der Arbeit (vgl. ebd., S. 43f). Wenn Kopfhörer im öffentlichen Raum oder im Straßenverkehr, beispielsweise auf dem Fahrrad verwendet werden, kann ein zu hoher Grad an Immersion (siehe 4.4.1), wie er von vielen binauralen Systemen und Inhalten angestrebt wird (siehe 4.1 und 5) problematisch werden: ein zu realistischer Klangeffekt im binauralen Hörspiel (siehe 5.3.3) der außerhalb des Kopfes lokalisiert wird könnte die Konsumentinnen und Konsumenten zu stark von der Umgebung ablenken. Nichtsdestotrotz werden Klangeffekte aber auch bei herkömmlichem Audio verwendet (siehe 5.3.3). Unabhängig von der Art des Audios kann ein Kopfhörer die Konsumentinnen und Konsumenten stark von der Außenwelt abschirmen (siehe 3.3.5). Ein gewisser Erfahrungswert der Konsumierenden im Zusammenhang mit der Kopfhörerwiedergabe im öffentlichen Raum oder im Straßenverkehr sollte also vorausgesetzt werden können.

## 6.4 Wo werden Inhalte angeboten?

Die Frage, wo Inhalte angeboten werden, bezieht sich auf die Veröffentlichung von Audio. Einerseits ist relevant, mit welchem Endgerät Audio wiedergegeben wird. Andererseits soll betrachtet werden, über welche Anbieter, Plattformen oder Apps Audio aktuell auf dem jeweiligen Endgerät gehört wird.

Mit 73% nutzen etwa drei von vier Befragten ein Smartphone als Endgerät zur Audiowiedergabe. 38% nutzen einen Laptop, 32% einen Desktop-PC, 28% ein Tablet, 27% ein Smartphone, das mit dem Autoradio verbunden ist, 18% einen Smart-TV, 13% einen Smart-Speaker (vgl. Online-Audio-Monitor, 2019, S. 21f). Wie bereits erwähnt, bezieht sich diese Studie ausschließlich auf Audioinhalte, die per Internet abgerufen werden (siehe 6). Folglich werden nur Internet-fähige Geräte gelistet. Eine Bitkom Umfrage zur Kopfhörerwiedergabe bezieht sich auf das Endgerät, mit dem die Kopfhörer verbunden werden: Mit 88% verbinden mehr als acht von zehn Befragten ihre Kopfhörer mit dem Smartphone. 31% nutzen die Kopfhörer mit der Musikanlage, 16% mit dem Laptop, 15% mit dem Fernseher, 10% mit dem MP3-Player, 8% mit dem Tablet und 7% mit dem Desktop-PC (vgl. Bitkom, 2019, S. 43). Beide Umfragen zeigen, dass das Smartphone das mit Abstand am meisten genutzte Endgerät ist.

Eine Bitkom Umfrage bezieht sich auf den Music Report 2019 und die Umsatzanteile beim Musikverkauf im Jahr 2018: 59% der Umsätze kommen aus dem digitalen Musikverkauf. Der Gesamtumsatz setzt sich zusammen aus 47% Musikstreaming, 25% physische Tonträger, 16% Aufführungs- und Synchronisationsrechte, 8% Musikdownloads und 4% sonstige digitale Erlöse (vgl. ebd., S. 25). Laut Bitkom streamen 72% der Internutzerinnen und -nutzer Musik (vgl. ebd., S. 25). 55% dieser Personen nutzen Musikstreaming-Dienste wie Spotify oder Deezer, 53% nutzen Video-Plattformen wie Youtube oder Vimeo und 33% streamen per Internetradio (vgl. ebd., S. 26).

Wegen ihrer hohen Nutzerzahl sind es also auch die Musikstreaming-Dienste und Video-Plattformen, die dynamische binaurale Audioinhalte erlauben sollten. Die Videoplattform Vimeo unterstützt aktuell nur Stereoformate. Die Betreiber der Videoplattform sagen: „Alle Quelldateien, die räumliches Hören enthalten, werden in 2 Kanäle umgewandelt und passen sich nicht den Bewegungen des Betrachters an. Wir haben die Absicht, in naher Zukunft auch räumliches Hören zu unterstützen“ (Vimeo, o.D.). Die Videoplattform Youtube unterstützt First-Order-Ambisonics Audioformate für 360 Grad Videos (vgl. Youtube, o.D.). Youtube verwendet einen eigenen Ambisonics-Binaural-Dekoder. Außerdem besteht die Möglichkeit, eine zusätzliche Stereospur hinzuzufügen. Diese „Head-Locked“ Stereospur ist statisch und wird nicht durch die Blickrichtung des Zuschauers verändert (vgl. Der Maschinenraum - Tonstudio Köln, 2018). Der Musikstreaming-Dienst Spotify unterstützt aktuell nur Stereoformate (vgl. Spotify, 2020). Der Musikstreaming-Dienst Deezer ermöglicht 3D-Audio für Premiumabonnenten und unterstützt *Sony 360 Reality Audio*, das im folgenden Abschnitt genauer erklärt werden soll (vgl. Deezer, 2020).

## 6.5 Aktuelle Technologien

In den vorangegangenen Abschnitten wurden drei Studien zum Konsum von Audioinhalten ausgewertet, dies führte zu folgenden Erkenntnissen: Musik ist die relevanteste der drei Inhaltskategorien. Der Musikkonsum erfolgt größtenteils online mit dem Smartphone über Musikstreaming-Dienste oder Videoplattformen. Das Wiedergabesystem ist nicht einheitlich und reicht vom integrierten Lautsprecher des Endgeräts, über einen Smartspeaker bis zu Kopfhörern. Im Folgenden sollen aktuelle Technologien beschrieben werden, die alle diese Erkenntnisse berücksichtigen und das binaurale Musikstreaming ermöglichen.

### 6.5.1 Sony 360 Reality Audio

*Sony 360 Reality Audio* ist eine Technologie zur dreidimensionalen Audiowiedergabe, die im Oktober 2019 vorgestellt wurde. *Sony 360 Reality Audio* basiert auf dem offenen Audiostandard MPEG-H des Fraunhofer IIS, der für Streaming- und Rundfunkanwendungen entwickelt wurde. Die Technologie ist geräteunabhängig und funktioniert mit den meisten 3D-fähigen Lautsprechern, wie beispielsweise dem *Amazon Echo Studio* oder der *Sennheiser Ambeo Soundbar*. Bei der Kopfhörerwiedergabe wird die Mischung binauralisiert (vgl. Fraunhofer IIS, o.D.).

Auf den Musikstreaming-Diensten Deezer, Nugs.net und Tidal werden spezielle *Sony 360 Reality Audio* Versionen angeboten. Über das Musiklabel Sony Music Entertainment ist Sony Rechteinhaber an vielen Musiktiteln. Aktuell stehen etwa hundert Alben als *Sony 360 Reality Audio* Versionen zur Verfügung, vorwiegend aus den 1970er- und 1980er-Jahren und dem Bereich der Popmusik. Über das Verfahren zur nachträglichen Binauralisierung (siehe 4.5) der Masterspuren dieser Alben gibt Sony keine Informationen. Musikmischungen, die speziell für *Sony 360 Reality Audio* liegen aktuell nicht vor (vgl. Bleisteiner, 2020).

Grundsätzlich kann jedes Kopfhörermodell zur Wiedergabe verwendet werden. Bei der Verwendung eines Sony Kopfhörers ist eine Individualisierung möglich (siehe 4.3.2): In der App *Sony Headphones Connect* wählen die Nutzerinnen und Nutzer aus einer Liste von Kopfhörermodellen ihren Kopfhörer aus. Durch eine visuelle Vermessung des Ohres mit der Smartphone-Kamera wird die individuelle Ohrform analysiert. Aufgrund dieser Daten wird eine geeignete HRTF ausgewählt (siehe 4.3.2.4) und die Mischung wird individualisiert. Die dynamische Musikwiedergabe ist mit *Sony 360 Reality Audio* nicht möglich, da aktuell kein Headtracking (siehe 4.4) unterstützt wird (vgl. Steele, 2019).

### 6.5.2 Dolby Atmos Music

*Dolby Atmos Music* wurden im Dezember 2019 vorgestellt und hat einige Gemeinsamkeiten mit *Sony 360 Reality Music*: Auch diese Technologie funktioniert sowohl mit 3D-fähigen Lautsprechern die Dolby Atmos fähig sind (siehe Dolby, o.D.), als auch bei der Kopfhörerwiedergabe. Wie *Sony 360 Reality Audio* ist auch *Dolby Atmos Music* im Musikstreaming-Dienst Tidal integriert. Zusätzlich kann *Dolby Atmos Music* über Amazon Music gestreamt werden (vgl. Bleisteiner, 2020).

Die Technologie basiert auf dem Dolby Atmos 3D-Audioformat, das bisher vor allem für bildbezogene 3D-Audio-Mischungen von Filmen und Serien verwendet wurde. Für das Musikangebot arbeitet Dolby mit den Labels Universal Music Group und Warner Music Group zusammen (vgl. Cohen, 2020). Da *Dolby Atmos Music* sehr neu ist, gibt es auch bisher nur ein sehr kleines Angebot an Liedern (siehe Dolby, o.D.). Giles Baker, Senior Vice President of Consumer Entertainment bei Dolby verspricht aber: „Along with those label partnerships, the format has been overwhelmingly embraced by some of music’s biggest names“ (Waniata, 2020) und bezieht sich dabei auf die zukünftige Zusammenarbeit mit Künstlern wie Elton John, Lizzo oder Coldplay (vgl. ebd.).

Zur Erstellung einer Dolby Atmos-Mischung stehen bis zu 128 Ausgangskanäle zur Verfügung. 10 Kanäle davon sind für das sogenannte „Audio Bed“ vorgesehen, eine 7.1.2 Surround Spur, die nicht gespannt werden kann. Zusätzlich können in der Mischung bis zu 118 Audio-Objekte hinzugefügt werden, die frei auf der Frontal-, Horizontal und Medianebene (siehe 2.3) verteilt werden können. Die Position der Audioobjekte wird in sogenannten Metadaten gespeichert. Der Dolby Atmos Renderer berechnet die 3D-Audio-Mischung je nachdem, welches Wiedergabesystem verwendet wird. Eine dynamische Wiedergabe und Individualisierung (siehe 4.4 und 4.3.2) bei der Kopfhörerwiedergabe von *Dolby Atmos Music* Inhalten über Tidal oder Amazon Music ist nicht möglich (vgl. Dolby Professionals, o.D.).

### **6.5.3 Apple Spatial Audio**

*Apple Spatial Audio* wurde mit einem Software-Update im September 2020 für Apple iPhones und iPads eingeführt. Zwar ist *Apple Spatial Audio* ausschließlich für das Streaming von Filmen und Serien ausgelegt und nicht für die 3D-Musikwiedergabe. Trotzdem soll diese Technologie nicht unerwähnt bleiben, da *Apple Spatial Audio* im Gegensatz zu *Sony 360 Reality Audio* und *Dolby Atmos Music* die dynamische Binauralwiedergabe per Headtracking (siehe 4.4) erlaubt (vgl. Bubeck, 2020). Die Technologie funktioniert bisher nur mit den Kopfhörern *Apple AirPods Pro* (siehe 3.3.4). Die drahtlosen In-Ear Kopfhörer enthalten einen Beschleunigungs- und Gyroskop-Sensor (siehe 4.4.6.3), mit denen sich die Position und Kopfausrichtung der Hörerin oder des Hörers bestimmen lässt (siehe 4.6.2). Das erlaubt *Apple Spatial Audio* die dynamische Binauralsynthese von 5.1-, 7.1- und Dolby Atmos-Mischungen (siehe 6.5.2). Neben der Kopfposition wird auch die Position des mobilen Endgerätes bestimmt, die 3D-Mischung wird relativ dazu ausgerichtet. „This means that even if you turn your head or reposition your device, dialogue will still be anchored to the actor on the screen“ (Parsons, 2020). Ob *Apple Spatial Audio* in Zukunft auch für die dynamische Binauralsynthese von rein auditiven Inhalten angewendet wird ist derzeit nicht bekannt (vgl. ebd.).

## 7 Fazit

Durch die Zugrundelegung der Eigenschaften des menschlichen Hörens und der Darlegung der verschiedenen Aspekte der binauralen Audiotechnik konnten aus technischer und gestalterischer Perspektive die Chancen und Herausforderungen von binauralem Audio aufgezeigt werden. Die Literaturarbeit lieferte durch diese Untersuchung folgende Ergebnisse.

### 7.1 Ergebnisse

Die HRTF jedes Menschen bestimmt, wo eine Schallquelle verortet wird. Sie entsteht durch die Filterwirkung von Außenohr, Kopf und Torso. Die Verwendung von Kopfhörern ermöglicht es, beide Ohren unabhängig voneinander zu beschallen. Wird ein Signal mit einer HRTF für linkes und rechtes Ohr gefaltet und über Kopfhörer wiedergegeben, dann wird es in der Richtung verortet, die ebendieser HRTF entspricht.

Ein Kunstkopfmikrofon simuliert die Filterwirkung von Außenohr, Kopf und Torso und erlaubt so die Wiedergabe von 3D-Audio über Kopfhörer. So entsteht ein Klangeindruck, der dem natürlichen Hören viel eher entspricht als der Klangeindruck, den die Kopfhörerwiedergabe einer Stereomischung liefert: Die Schallquelle wird außerhalb des Kopfes im Raum lokalisiert und nicht im Kopf wie bei einer Stereomischung. Jedoch ist die Kunstkopfaufnahme nicht-individuell, statisch und unflexibel in der Nachbearbeitung. Die dynamische Binauralsynthese ist für die Erstellung von binauralem Audio unter Betrachtung dieser Faktoren besser geeignet, so wird hierfür kein spezielles Mikrofonierungsverfahren benötigt, die Mischung und die einzelnen Elemente innerhalb der Mischung lassen sich entsprechend so nachbearbeiten, wie es auch bei Stereo-Audioproduktionen möglich ist. Die Technologie erlaubt außerdem die Individualisierung und die dynamische Wiedergabe der Binauralmischung. Neben diesen Chancen gibt es bei der Umsetzung der Binauralsynthese einige technische Herausforderungen die erklären, warum es bisher kein breites Angebot an binauralen Audioproduktionen gibt.

Eine Herausforderung ist die Individualisierung: Die Anatomie jedes Menschen ist unterschiedlich, deshalb ist auch die HRTF jedes Menschen individuell. Die Individualisierung der Binauralmischung sorgt für eine genauere Lokalisation eines Schallereignisses in der Mischung. Die Entwicklung eines anwenderfreundlichen und präzisen Verfahrens zur Bestimmung der individuellen HRTF ist ein wichtiger Aspekt der Binauralsynthese. Aus Anwendersicht ist die visuelle Messung im Gegensatz zu anderen Verfahren am einfachsten umzusetzen, weshalb sie von aktuellen Technologien wie *Sony 360 Reality Audio* verwendet wird.

Eine weitere Herausforderung ist die dynamische Wiedergabe einer Binauralmischung: Für die horizontale Lokalisation einer Schallquelle sind vor allem ITDs und ILDs verantwortlich. Durch unbewusste Peilbewegungen wird die Kopfposition ständig geändert, dadurch kann eine Schallquelle präziser verortet werden. Die dynamische Binauralsynthese erlaubt es, dass eine Schallquelle unabhängig von Kopfposition und -ausrichtung an einem festen Punkt im Raum verortet wird. Dafür ist ein

geeignetes Headtracking-System zur Bestimmung der Kopfausrichtung und -position nötig. Inertialsensoren sind dabei am anwenderfreundlichsten, da sie die Bewegungsfreiheit der Anwenderinnen und Anwender am wenigsten einschränken. Sie sind beispielsweise in VR-Brillen und in den *Apple AirPods Pro* eingebaut.

Das Verfahren, mit dem die Binauralmischung erstellt wird, wirft weitere technische Herausforderung auf: Psychoakustisch können breitbandige und impulshafte Signale besser lokalisiert werden, als gleichförmige und sinusartige Signale, die nur wenige Frequenzen enthalten. Für die vertikale Lokalisation sind Erhöhungen und Einbrüche im Frequenzspektrum des Signals verantwortlich. Die Entfernungslokalisation erfolgt sowohl durch den Amplituden- und Frequenzgang des Signals, als auch durch Informationen, die der Nachhall der Schallquelle in einem Raum liefert. Sowohl das Signal als auch dessen Nachhall beeinflussen also, wie gut die Lokalisation bei der Binauralwiedergabe funktioniert. Das Verfahren zur Binauralsynthese und die Art der Impulsantwort beeinflussen deshalb den Klang der Binauralmischung maßgeblich. Die Faltung mit der HRTF ist relativ einfach in bestehende 3D-Audio-Produktionssysteme zu integrieren. Die Faltung mit der BRIR eines realen Raumes klingt am realitätsgetreuesten. Problematisch ist hierbei, dass die Richtcharakteristik der Schallquelle bei der BRIR-Messung nachträglich nicht aus der Impulsantwort entfernt werden kann. Ein Ansatz ist es deshalb solche Messlautsprecher zu verwenden, deren Richtcharakteristik und Frequenzgang dem Objekt ähneln, das später binauralisiert werden soll. Die Faltung mit der BRIR eines virtuellen Raumes setzt die exakte Simulation der Raumgeometrie und -beschaffenheit voraus. Die Berechnung der Schallausbreitung im virtuellen Raum und aller daraus resultierenden wellentheoretischen Phänomene erfordert eine enorme Rechenleistung.

Aus gestalterischer Perspektive konnten einige Chancen aufgezeigt werden, die die binaurale Audio-technik für auditive Inhalte bietet. Die beschriebenen Beispiele zeigen, dass Muskschaffende und -produzierende das Potential, das eine 3D-Mischung von Studioproduktionen liefert, erkennen. Das Arrangement und damit der musikalische Ausdruck können durch eine passende Verteilung der Elemente innerhalb einer Mischung unterstützt werden. Nichtsdestotrotz gibt es bisher nur wenige Musikstudioproduktionen, die mit dynamischer Binauralsynthese erstellt wurden. Bei Konzertmitschnitten überzeugt vor allem der höhere Grad an Realismus, den eine Binauralmischung gegenüber einer Stereomischung verspricht. Besonders im VR-Bereich wird dabei die Chance gesehen, in Verbindung mit einer VR-Brille den virtuellen Konzertbesuch zu ermöglichen und so eine völlig neue Art zu schaffen, Live-Musik zu erleben. Vor allem die Kombination aus binauralisierten und nicht-binauralisierten Elementen innerhalb der Musikmischung schafft einen Mehrwert der Binauralmischung gegenüber einer Stereomischung: die Intensität einer Stereomischung lässt sich mit der vergrößerten räumlichen Tiefe und Umhüllung einer Binauralmischung vereinen.

Bei Radio und Podcasts ist der Mehrwert einer Binauralmischung gegenüber einer Stereomischung weniger im gestalterischen Sinne, als mehr in der Verbesserung der Sprachverständlichkeit und der Lokalisation bei mehreren Sprechenden zu finden.

Die 3D-Mischung eines binauralen Hörspiels bietet im Vergleich zu einer Stereomischung neue gestalterische Möglichkeiten in Bezug auf die Panoramaverteilung und den Einsatz von Klangeffekten. Der Grad an Immersion der Hörerinnen und Hörer kann deutlich verbessert werden. Die Mischung wird durch den vergrößerten akustischen Raum deutlich transparenter, vor allem dann, wenn mehrere Schallereignisse gleichzeitig stattfinden. Andererseits klingt eine rein binaurale Mischung weniger intensiv und indirekter als eine Stereomischung. Wie bei der Musikmischung können auch hier einzelne Elemente innerhalb der Mischung nicht-binauralisiert werden. Sie werden dann im Kopf lokalisiert und klingen druckvoller und direkter.

## 7.2 Diskussion und Ausblick

Diese Arbeit beschäftigt sich vorrangig mit der Produktion auditiver Medien. Kapitel 6 hat beschreibt den Konsum von auditiven Inhalten und vor allem bei der Veröffentlichung binauraler Inhalte werden einige Fragen aufgeworfen, die nicht abschließend beantwortet werden können: Die Plattformen, über die auditive Medien konsumiert werden sind vielfältig, was die prozentuelle Verteilung am Beispiel Abruf von Podcasts einer Umfrage des Online-Audio-Monitors zeigt: 51% der Befragten streamen Podcasts über Youtube oder andere Videoportale, 49% über Musikstreamingdienste, 46% über Websites oder Apps von Radiosendern, 39% über Audiotheken von Radiosendern, 35% über Websites anderer Anbieter, beispielsweise TV-Sender oder Zeitungen, 34% über andere Plattformen, beispielsweise Audible oder iTunes, 19% über spezielle Podcast-Apps und 17% über Radioplattformen (vgl. Online-Audio-Monitor, 2019, S. 65). Geht man davon aus, dass alle Podcasts als dynamische binaurale Audioproduktion zur Verfügung stünden, dann lege es dennoch an den Audio-Anbieter die Voraussetzungen für eine dynamische Wiedergabe schaffen. Wie in Abschnitt 6.4 beschrieben, gibt es einige Anbieter die bereits jetzt 3D-Audio-Formate ermöglichen. Trotzdem fehlen große Anbieter wie Spotify oder Vimeo. Geht man davon aus, dass alle Konsumierenden das SOFA-File ihrer eigenen HRTF besitzen, dann bliebe die Frage, wie die individuelle HRTF auf die fertige binaurale Audioproduktion gerechnet wird, die über einen Musikstreamingdienst gestreamt wird. *Sony 360 Reality Audio* löst dieses Problem, indem der Klang des Kopfhörers über eine App auf dem Smartphone individualisiert wird. Allerdings funktioniert die Individualisierung nur für Sony-Kopfhörern (siehe 6.5.1).

Zwar wurde in Abschnitt 5.2.3 beschrieben, dass der Einsatz von binauraler Audiotechnik die Sprachverständlichkeit bei einem oder mehreren Sprechenden verbessern kann, wie in Abschnitt 6.1 belegt besteht jedoch ein wesentlicher Teil der Inhalte von Radio und Podcasts aus Nachrichten-, Informations- und Unterhaltungsbeiträgen. Vernachlässigt man die verbesserte Barrierefreiheit für Schwerhörige, dann ist es eher fragwürdig, ob sich der Mehraufwand zur Erstellung einer binauralen Version lohnt, da es dem Großteil der Hörerschaft wohl kaum auf das auditive Klangerlebnis bei einer Nachrichtensendung ankommt. Vielmehr zählt bei einer Nachrichten- oder Informationssendung der reine verbale Inhalt des Beitrags. Andererseits ist es die erste Reaktion der meisten Höre-

rinnen und Hörer, die Lautstärke am Endgerät zu erhöhen, sobald die Sprache unverständlich ist. Wenn die Binauraltechnik die Sprachverständlichkeit verbessern kann, dann wäre diese Lautstärkeerhöhung nicht notwendig und das Gehör der Hörerschaft könnte geschont werden.

Die Arbeit hat gezeigt, dass es aus technischer Perspektive vor allem drei Faktoren gibt, die die Qualität der binauralen Mischung beeinflussen: die Individualisierung, die dynamische Nachführung der Mischung und das Verfahren, mit dem die binaurale Mischung erstellt wurde. Alle drei Faktoren beeinflussen wie realistisch die binaurale Mischung klingt. Beispielsweise gilt, dass die Schallquelle umso genauer lokalisiert werden kann, je besser die zur Binauralisierung verwendete HRTF zur individuellen HRTF der Konsumentin oder des Konsumenten passt. Verglichen mit einer Stereomischung ist es aus gestalterischer Sicht unter anderem dieser höhere Grad an Realismus, den der Einsatz von binauraler Audiotechnik verspricht. Dies wurde in Abschnitt 4.6.2 diskutiert. Allerdings gilt es zu berücksichtigen, dass es nicht das Ziel jedes existenten Formates ist, möglichst realistisch zu klingen. Bei einer Diskussionsrunde im Radio mag es wünschenswert sein, dass die Mischung den Hörerinnen und Hörern das Gefühl vermittelt, als säßen sie zusammen mit den Radiogästen in einem Raum. Bei einer modernen Popmusik-Produktion ist dies nicht der Fall. Eine Studioproduktion verfolgt ein medial-autonomes Klangideal. Es geht der Studioproduktion also gar nicht darum, einen realen Klangeindruck nachzustellen. Während des Produktionsprozesses soll eine autonome Klangwelt geschaffen werden, die den Ausdruck der jeweiligen Produktion unterstützt. So wäre es denkbar, dass sich dieses Klangideal an die technischen Einschränkungen bei der binauralen Wiedergabe anpasst. Bei einer binauralen Popmusik-Produktion ist es nicht zwingend notwendig, ein Schallereignis möglichst genau zu lokalisieren, um den Mehrwert gegenüber einer Stereoproduktion gestalterisch auszunutzen: der Eindruck einer breiten Streicherfläche kann erzeugt werden, ohne dass die einzelnen Streicher genau lokalisiert werden. Auffällige Panoramafahrten werden in einer Stereomischung eher selten verwendet und können deshalb auch bei einer binauralen Mischung sparsam eingesetzt werden. Der Eindruck einer großen Hallfahne nach einem Snaredrum-Schlag entsteht auch dann, wenn die binaurale Raumimpulsantwort nur eingeschränkt dem echten Raumklang entspricht. Löst man sich also vom Klangideal der möglichst genauen Lokalisation einer Schallquelle und der möglichst realistischen Nachstellung einer echten Räumlichkeit, dann können die Vorteile der binauralen Audiotechnik genutzt, auch wenn aktuell noch nicht alle technischen Herausforderungen gelöst sind. Unter diesem Hintergrund ist es fraglich, wie relevant es beispielsweise ist, die BRIR eines Raumes mithilfe eines Dodekaederlautsprecher (siehe 4.5.2) für möglichst viele Instrumente aufzunehmen. Eine Anpassung des Klangideals an die technischen Einschränkungen wäre nicht nur für binaurale Musik denkbar. Der gleiche Gedankengang kann auf das binaurale Hörspiel angewandt werden: Wann immer eine binaurale Audioproduktion eine neue Klangwelt schaffen soll, statt die Realität möglichst genau nachzustellen, dann sind die technischen Einschränkungen vernachlässigbar, denn die Gestaltung der Klangwelt kann sich an diese Einschränkungen anpassen und die Chancen der binauralen Audiotechnik mit dem jetzigen Stand der technischen Entwicklung nutzen.

Diese Literaturarbeit hat einen ersten Forschungsbeitrag zur theoretischen Betrachtung der technischen und gestalterischen Aspekte von binauralem Audio für unterschiedliche Inhalte geleistet. Weiterführende Forschungen könnten sich praktisch mit den genannten Aspekten auseinandersetzen. Denkbar wäre die Aufnahme und Produktion von mehreren binauralen Audiobeispielen für die verschiedenen Inhaltskategorien. Die Audiobeispiele sollten jeweils als Stereomischung und als individuelle und nicht-individuelle, dynamische und statische Binauralmischung, mit mehreren unterschiedlichen Verfahren zur Binauralsynthese erstellt werden. In einem Hörversuch mit diversen Probanden könnten die verschiedenen Versionen eines Audiobeispiels verglichen und so untersucht werden, wie groß der Einfluss der einzelnen Faktoren auf die Wiedergabequalität ist. Ein Vergleich der verschiedenen Inhaltskategorien könnte zeigen, ob die Einflussfaktoren aus Probandensicht für die einzelnen Inhalte mehr oder weniger relevant sind. Durch einen solchen Hörversuch könnten die Ergebnisse dieser Literaturarbeit praktisch überprüft werden.

## 8 Abbildungsverzeichnis

Abbildung 1: Aufbau des Ohrs .....	4
Abbildung 2: HRTF-Messung einer Versuchsperson.....	6
Abbildung 3: HRTF-Messung an zehn Versuchspersonen.....	7
Abbildung 4: HRTF-Messung an unterschiedlichen Positionen im Ohrkanal .....	7
Abbildung 5: Komponenten der HRTF .....	8
Abbildung 6: Kopfbezogenes Koordinatensystem .....	9
Abbildung 7: Seitlicher Schalleinfall.....	10
Abbildung 8: Lokalisation durch Laufzeitunterschiede.....	10
Abbildung 9: Cone of Confusion.....	12
Abbildung 10: Richtungsbestimmende Bänder .....	13
Abbildung 11: Veranschaulichung der richtungsbestimmenden Bänder.....	14
Abbildung 12: Funktionsprinzip des Assoziationsmodells .....	16
Abbildung 13: Head Acoustics HMS III (links), Neumann KU 100 (rechts) .....	21
Abbildung 14: KEMAR KB 4004 (links), Brüel & Kjær HATS 4128 (rechts) .....	22
Abbildung 15: Freifeld-HRTF (blau) und Diffusfeld-HRTF (rot) .....	26
Abbildung 16: LTI-System.....	29
Abbildung 17: Faltung im Zeitbereich und schnelle Faltung .....	30
Abbildung 18: Anthropometrische Daten von Kopf, Torso und Ohr .....	33
Abbildung 19: Kopfbewegungen im Bezug zur x-,y- und z-Achse.....	36
Abbildung 20: Dynamische Binauralsynthese.....	37
Abbildung 21: Raster eines Impulsantworten-Datensatzes .....	38
Abbildung 22: Trennung der Nachführung von frühem Hallanteil und Diffusschall.....	39
Abbildung 23: Ambisonics .....	42
Abbildung 24: Ambisonics-Binaural-Dekoder.....	43
Abbildung 25: 3D-Modell der Frauenkirche Dresden .....	45
Abbildung 26: 360°-Kunstkopfmikrofon für Becks „Sound and Vision“ .....	57
Abbildung 27: Screenshot des 360°-Videos zu Becks "Sound and Vision" .....	58
Abbildung 28: Mikrofonaufbau BBC Proms.....	59
Abbildung 29: Sprachverständlichkeit.....	62
Abbildung 30: Dynamische Binauralisierung von Sprache (grün) und Störsignal (rot) .....	63
Abbildung 31: Mikrofonarray "The Turning Forest" .....	67

## 9 Literaturverzeichnis

- Algazi, V. R., Duda, R. O., & Thompson, D. M. (2001). *The CIPIC HRTF Database*. Von U.C. Davis: [https://www.ece.ucdavis.edu/cipic/files/2015/04/cipic\\_CIPIC\\_HRTF\\_Database.pdf](https://www.ece.ucdavis.edu/cipic/files/2015/04/cipic_CIPIC_HRTF_Database.pdf) abgerufen am 4. Juni 2020.
- Anrews, M., & Chapman, B. (Regisseure). (2012). *Brave* [Kinofilm].
- Apple. (9. August 2019). *Logic Pro X: Binaural panning overview*. Von Apple Support: [https://support.apple.com/kb/PH24474?locale=en\\_US&viewlocale=en\\_US](https://support.apple.com/kb/PH24474?locale=en_US&viewlocale=en_US) abgerufen am 29. Juni 2020.
- Auro Technologies. (8. August 2014). *MANDO DIAO RELEASE WORLD'S FIRST POP ALBUM IN AURO-3D® IMMERSIVE SOUND*. Von Auro Technologies: <https://www.auro-3d.com/press/2014/08/mando-diao-release-worlds-first-pop-album-in-auro-3d-immersive-sound/> abgerufen am 24. Juli 2020.
- Auro-3D. (8. August 2014). *The mixing of Mando Diao's 'Aelita' in Auro-3D*. Von Youtube: <https://www.youtube.com/watch?v=1tIAAeKTKaE> abgerufen am 24. Juli 2020.
- Balke, S. (2013). *Wahrnehmungsbasierte Analyse von Higher-Order-Ambisonics-Signalen bei binauraler Synthese*. Hannover: Leibniz Universität.
- Ballou, G. (2008). *Handbook For Sound Engineers*. Amsterdam: Elsevier.
- Bartz, P. (15. April 2016). *Building an AHRS using the SparkFun "9DOF Razor IMU" or "9DOF Sensor Stick"*. Von GitHub: <https://github.com/Razor-AHRS/razor-9dof-ahrs/wiki/Tutorial> abgerufen am 10. Juni 2020.
- BBC Research & Development. (7. November 2016). *360 Video Showreel - BBC R&D*. Von Youtube: <https://www.youtube.com/watch?v=9mfKZjhtoLI&feature=youtu.be> abgerufen am 3. September 2020.
- BBC Sounds. (13. März 2017). *Wireless Nights - Series 5 - Joujouka*. Von BBC: <https://www.bbc.co.uk/sounds/play/b08hmd23> abgerufen am 5. September 2020.
- Begault, D. R., & Wenzel, E. M. (Juni 1993). Headphone Localization of Speech. *Human Factors*, S. 361-376.
- Ben-Kiki, A., Ben-Asher, M., & Neoran, I. (o.D.). *Improving Voice Intelligibility Using Waves Nx*. Waves Audio Research - Technical Paper.
- Bitkom. (2019). *Zukunft der Consumer Technology – 2019: Marktentwicklung, Trends, Mediennutzung, Technologien, Geschäftsmodelle*. Berlin: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V.
- Bleistainer, W. (21. Januar 2020). *Sony's 360 Reality Music VS Dolby Music*. Von Bayerischer Rundfunk Blog: <http://blog.br.de/highend/immersive-music/> abgerufen am 27. September 2020.

- BR. (27. Februar 2019). *Audio-Trend Podcast: Acht Dinge, die Sie über Podcasts wissen sollten*. Von BR: <https://www.br.de/radio/bayern2/podcast-faq-fragen-100.html> abgerufen am 10. Juli 2020.
- Bubeck, S. (18. September 2020). *Spatial Audio (3D-Audio) mit AirPods Pro nutzen: Anleitung, Apps und Inhalte*. Von Giga: <https://www.giga.de/artikel/spatial-audio-3d-audio-mit-airpods-pro-nutzen-anleitung-apps-und-inhalte/> abgerufen am 28. September 2020.
- Burton, T. (Regisseur). (1992). *Batman Returns* [Kinofilm].
- Cherry, E. C. (September 1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, S. 975-979.
- Clement, M., Lepthien, A., Kandziora, M. S., & Meyn, J. (2018b). *Ergebnisse der Studie zur Zukunft der Musik*. Hamburg: Universität Hamburg.
- Clement, M., Lepthien, A., Kandziora, M. S., & Meyn, J. (2018a). *Studie zur Zukunft der Musiknutzung 2018*. Hamburg: Universität Hamburg.
- Cohen, S. (3. September 2020). *What is Dolby Atmos Music, and how can you experience it?* Von Digitaltrends: <https://www.digitaltrends.com/home-theater/what-is-dolby-atmos-music-and-how-to-get-it/> abgerufen am 27. September 2020.
- Collinson, G. (1. Juni 2016). *360 Soundscapes: What You Need to Know about Recent Developments*. Von BBC Writers Room: <https://www.bbc.co.uk/blogs/writersroom/entries/474a4ea4-5a5e-4961-82c2-56f75c589d38> abgerufen am 3. September 2020.
- Constine, J. (13. April 2014). *How Oculus And 8 Fake Ears Could Morph Concerts Into Video Games*. Von Techcrunch: [https://techcrunch.com/2014/04/13/oculus-music/?gucounter=1&guce\\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\\_referrer\\_sig=AQAAANN-yvO8MEM8AEC4bxMynT\\_XkvoiX9TWzmWCiHFt\\_LV-XIsMBxt4CvMADuMHaUoQBU169fUjd6Rzfm5SYsCwVTAYfE4Ktt7SR3\\_q05twKFIMiGbnV9KjUef\\_AhrnwP](https://techcrunch.com/2014/04/13/oculus-music/?gucounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAANN-yvO8MEM8AEC4bxMynT_XkvoiX9TWzmWCiHFt_LV-XIsMBxt4CvMADuMHaUoQBU169fUjd6Rzfm5SYsCwVTAYfE4Ktt7SR3_q05twKFIMiGbnV9KjUef_AhrnwP) abgerufen am 30. Juli 2020.
- Daniel, P. e. (1 2007). *Kunstkopftechnik - Eine Bestandsaufnahme. Acustica Vol. 93*.
- David Bowie. (1977). *Sound And Vision. Auf: Low*. RCA Records.
- De Leuw, C. (24. Juni 2020). *Apple AirPods Pro mit Spatial Audio: Das bringt das Head-Tracking-Update*. Von Computerbild: <https://www.computerbild.de/artikel/avf-News-Audio-Apple-AirPods-Pro-Spatial-Audio-Head-Tracking-Update-26423787.html> abgerufen am 3. Juli 2020.
- Dear VR. (o.D.). *Dear VR Pro*. Von Dear VR: <https://www.dearvr.com/products/dearvr-pro> abgerufen am 29. Juni 2020
- Deezer. (29. Juni 2020). *360 by Deezer*. Von Deezer Support: <https://support.deezer.com/hc/en-gb/articles/360002137757-360-by-Deezer> abgerufen am 27. September 2020.

- DeLisa Coleman, L. (23. Februar 2017). *How The VR Concert Industry Is Boldly Jockeying For A Slice Of A Projected \$660M Pie*. Von Forbes:  
<https://www.forbes.com/sites/laurencoleman/2017/02/23/how-the-vr-concert-industry-is-boldly-jockeying-for-a-slice-of-a-projected-660m-pie/#19ed9ef10942> abgerufen am 8. September 2020
- Der Maschinenraum - Tonstudio Köln. (19. Juli 2018). *Die Verwendung von Spatial Audio in 360-Grad- und VR-Videos auf YouTube*. Von Firmenpresse:  
<https://www.firmenpresse.de/pressinfo1632542/die-verwendung-von-spatial-audio-in-360-grad-und-vr-videos-auf-youtube.html> abgerufen am 27. September 2020.
- Dickreiter, M., Dittel, V., Hoeg, W., & Wöhr, M. (2014). *Handbuch der Tonstudioteknik*. Berlin/Boston: Walter de Gruyter GmbH.
- Dolby. (o.D.). *Music in Dolby*. Von Dolby: <https://www.dolby.com/music/> abgerufen am 28. September 2020.
- Dolby Professionals. (o.D.). *Dolby Atmos for content creators*. Von Dolby Professionals:  
<https://professional.dolby.com/content-creation/dolby-atmos/2> abgerufen am 28. September 2020.
- DPA Microphones. (o.D.). *4560 CORE Binaurales Kopfbügelmikrofon*. Von DPA Microphones:  
<https://www.dpamicrophones.de/immersive/4560-binaurales-kopfbuegelmikrofon> abgerufen am 2. Mai 2020.
- Drullman, R., & Bronkhorst, A. W. (April 2000). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *The Journal of the Acoustical Society of America*, S. 2224-2235.
- El-Bawwab, R. (4. Januar 2015). *SURROUND-SOUND VON DAMALS BIS HEUTE (TEIL 1/2)*. Von AW-Insider: <http://av-insider.de/surround-geschichte-1> abgerufen am 2. Oktober 2020.
- El-Bawwab, R. (25. Juni 2015). *SURROUND-SOUND VON DAMALS BIS HEUTE (TEIL 2/2)*. Von AV-Insider: <http://av-insider.de/surround-geschichte-2> abgerufen am 2. Oktober 2020.
- Erl, J. (30. Mai 2020). *Virtual Reality für Einsteiger: Alles, was ihr wissen müsst*. Von Mixed:  
<https://mixed.de/virtual-reality-starter-guide/> abgerufen am 28. September 2020.
- Fandom. (o.D.). *Flashpoint: The Rolling Stones*. Von Fandom:  
[https://music.fandom.com/wiki/Flashpoint:The\\_Rolling\\_Stones](https://music.fandom.com/wiki/Flashpoint:The_Rolling_Stones) abgerufen am 29. Juli 2020.
- Firth, M. (18. Juli 2019). *Developing Tools for Live Binaural Production at the BBC Proms*. Von BBC: <https://www.bbc.co.uk/rd/blog/2019-07-proms-binaural> abgerufen am 31. Juli 2020.
- Fischer, E. K. (1964). *Das Hörspiel: Form und Funktion*. Stuttgart: Kröner Verlag.
- Flux Audio. (o.D.). *Ircam HEer V3*. Von Flux Audio: <https://www.flux.audio/project/ircam-hear-v3/> abgerufen am 3. Juli 2020.

- Fraunhofer IIS. (o.D.). *MPEG-H Audio*. Von Fraunhofer IIS:  
<https://www.iis.fraunhofer.de/de/ff/amm/rundfunk-streaming/mpegh.html> abgerufen am 27. September 2020.
- Friesecke, A. (2007). *Die Audio-Enzyklopädie: Ein Nachschlagewerk für Tontechniker*. München: K. G. Saur Verlag.
- Görne, T. (2015). *Tontechnik: Hören, Schallwandler, Impulsantwort und Faltung, digitale Signale, Mehrkanaltechnik, tontechnische Praxis*. München: Carl Hanser Verlag.
- Genelec. (Juni 2019). *Aural ID Operating Manual*. Von Genelec:  
[https://auralid.genelec.com/downloads/aural\\_id\\_user\\_manual-2019-05-14-v3.pdf](https://auralid.genelec.com/downloads/aural_id_user_manual-2019-05-14-v3.pdf)  
 abgerufen am 4. Juni 2020.
- Gisselbrecht, J. (o.D.). *Was ist ein Hörspiel?* . Von Hoerspiel:  
<http://www.hoerspiel.com/geschichte-zeitlauf-hoerspiel/definition-hoerspiel/> abgerufen am 10. Juli 2020.
- Google Playstore. (8. Mai 2018). *The Turning Forest*. Von Google Playstore:  
<https://play.google.com/store/apps/details?id=uk.co.bbc.turningforest> abgerufen am 8. September 2020.
- GravityMsc. (10. März 2020). *Smartphone Headtracking* . Von Github:  
<https://github.com/opentrack/opentrack/wiki/Smartphone-Headtracking> abgerufen am 13. Juni 2020.
- Hammer, P. (September 2016). *Virtual Reality: Die Erschaffung neuer Welten*. Von Zukunftsinstitut: <https://www.zukunftsinstitut.de/artikel/virtual-reality-die-erschaffung-neuer-welten/> abgerufen am 28. September 2020.
- Hooke Audio. (19. Juli 2017). *The Best Albums Recorded in Binaural Audio*. Von Hooke:  
<https://hookeaudio.com/blog/music/best-binaural-albums/> abgerufen am 18. Juli 2020.
- Hwang, S., Park, Y., & Park, Y.-s. (2008). *Modeling and customization of head-related transfer functions using principal component analysis*. Seoul, South Korea: International Conference on Control, Automation and Systems.
- Ifixit. (31. Oktober 2019). *Teardown der AirPods Pro*. Von Ifixit:  
<https://de.ifixit.com/Teardown/Teardown+der+AirPods+Pro/127551> abgerufen am 23. Mai 2020.
- Ircam. (o.D.). *LISTEN HRTF Database*. Von Ircam: <http://recherche.ircam.fr/equipes/salles/listen/> abgerufen am 8. Juni 2020.
- Kümmeler, L. (8. Juli 2020). *Amaze: VR-Videoplattform produziert Virtual-Reality-Konzerte*. Von Mixed: <https://mixed.de/amaze-vr-videoplattform-produziert-virtual-reality-konzerte/> abgerufen am 30. Juli 2020.

- Kane, M. (29. August 2019). *Realiser A16 Manual*. Von Smyth Research: [https://smyth-research.com/wp-content/uploads/2019/08/A16\\_manual\\_v091.pdf](https://smyth-research.com/wp-content/uploads/2019/08/A16_manual_v091.pdf) abgerufen am 29. Juni 2020.
- Keinath, D., & Tarnow, C. (2009). *Entwicklung und Erprobung einer Methode zur Integration binauraler Raumimpulsantworten bei der Mischung von Mehrspurproduktionen*. Detmold: Hochschule für Musik, Detmold.
- Klug, A. (4. Juni 2017). *Metal*. Von Kraftwerk - 3-D Der Katalog: <https://www.metal.de/reviews/kraftwerk-3-d-der-katalog-260093/> abgerufen am 24. Juni 2020.
- Kob, M., & Jers, H. (1999). *Directivity measurement of a singer*. Berlin: Collected Papers from the Joint Meeting "Berlin 99".
- Kosanke, L. (2010). *Untersuchung des Zusammenhangs physikalischer Maße zur Bestimmung der mixing time in Raumimpulsantworten und ihrem perzeptivem Äquivalent*. Berlin: Technische Universität Berlin.
- Krösmann, C., & Klöß, S. (25. Juni 2019). *Kopfhörer werden zum Status-Symbol*. Von Bitkom: <https://www.bitkom.org/Presse/Presseinformation/Kopfhoeerer-werden-zum-Status-Symbol> abgerufen am 2. Oktober 2020.
- Kraftwerk. (2017). 3-D Der Katalog. CD. Parlophone.
- Krebs, S. (3. August 2016). *The Failure of Binaural Stereo: German Sound Engineers and the Introduction of Dummy Head Microphones*. Von Kunstkopf stereophony: <https://binauralrecording.wordpress.com/2016/08/03/the-failure-of-binaural-stereo-german-sound-engineers-and-the-introduction-of-dummy-head-microphones/> abgerufen am 18. Juli 2020.
- Kronlachner, M. (2012). *Ambisonics Binaural Dekoder Implementation als Audio Plug-in mit Headtracking zur Schallfeldrotation*. Graz: Institut für Elektronische Musik und Akustik.
- Krukowski, D. (3. September 2014). *Back to Mono*. Von Pitchfork: <https://pitchfork.com/features/oped/9492-back-to-mono/> abgerufen am 18. Juli 2020.
- Kunz, A. (2019). Neue Dimension des Hörens. *Stereo*, 48-53.
- Latta, A. (o.D.b). *KLANG:fabrik 3D In-Ear-Monitoring für Bühne und Studio*. Von Sound And Recording: <https://www.soundandrecording.de/equipment/klangfabrik-3d-in-ear-monitoring-fuer-buehne-und-studio/> abgerufen am 29. Juni 2020.
- Latta, A. (o.D.a). *New Audio Technology Spatial Audio Designer im Test*. Von Sound And Recording: <https://www.soundandrecording.de/equipment/new-audio-technology-spatial-audio-designer-im-test/> abgerufen am 29. Juni 2020.
- Laumann, K. J. (2016). *Über Methoden zur Qualitätsbeurteilung von virtuellen Kopfhörern*. München: Technische Universität München.

- Lee, G. W., & Kim, H. K. (7. November 2018). Personalized HRTF Modeling Based on Deep Neural Network Using Anthropometric Measurements and Images of the Ear. *Applied Sciences*.
- Lindau, A. (2006). *Ein Instrument zur softwaregestützten Messung binauraler Raumimpulsantworten in mehreren Freiheitsgraden*. Berlin: Technische Universität Berlin.
- Lindau, A., & Weinzierl, S. (2008). *Zur Wahrnehmbarkeit von Diskretisierungen in der dynamischen Binauralsynthese*. Leipzig: Tonmeistertagung.
- Lou Reed. (1978). *Street Hassle*. CD. Arista.
- Love, D. (o.D.). *Babe, I'm On Fire: The Making Of Lou Reed's Street Hassle*. Von Damien Love: <https://damienlove.com/writing/babe-im-on-fire-the-making-of-lou-reeds-street-hassle/> abgerufen am 18. Juli 2020.
- Møller, H. (1992). Fundamentals of Binaural Technology. *Applied Acoustics Vol 36 (3/4)*, 171-218.
- Møller, H., Sørensen, M. F., Hammershøi, D., & Jensen, C. B. (Mai 1995). Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, S. 300-321.
- Mackensen, P. (2004). *Auditive Localization. Head movements, an additional cue in Localization*. Berlin: Technische Universität Berlin.
- Magis, M. (19. Juli 2017a). *Meilenstein für MelodyVR: Livemusik-VR-Portal gewinnt auch Sony Music*. Von VR Nerds: <https://www.vrnerds.de/meilenstein-fuer-melodyvr-livemusik-vr-portal-gewinnt-auch-sony-music/> abgerufen am 30. Juli 2020.
- Magis, M. (15. Dezember 2017b). *Noys VR – Konzert-App im Oculus Store für Rift erschienen*. Von VR Nerds: <https://www.vrnerds.de/noys-vr-konzert-app-im-oculus-store-fuer-rift-erschieden/> abgerufen am 30. Juli 2020.
- Maier, F. (2015). *Das Hörspiel: Eine technische Kunstform?* Würzburg: Julius-Maximilians-Universität.
- Mando Diao. (2014). *Aelita*. CD. Musica De La Santa.
- MathWorks. (o.D.). *Ambisonic Binaural Decoding*. Von MathWorks: <https://www.mathworks.com/help/audio/examples/ambisonic-binaural-decoding.html> abgerufen am 29. Juni 2020.
- McCormack, L., & Politis, A. (2019). SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods. *AES Conference on Immersive and Interactive Audio*. York: Audio Engineering Society.
- McCormack, L., Delikaris-Manias, S., Politis, A., & Pulkki, V. (6. März 2020). *Spatial Audio Real-time Applications (SPARTA)*. Von Aalto University. School of Electronical Engineering: [http://research.spa.aalto.fi/projects/sparta\\_vsts/plugins.html](http://research.spa.aalto.fi/projects/sparta_vsts/plugins.html) abgerufen am 27. Juli 2020.
- McMillan, G. (12. Februar 2013). *New, New, Electric Blue: Bowie's "Sound and Vision" Reinvented by Beck*. Von Wired: <https://www.wired.com/2013/02/new-new-electric-blue-bowies-sound-and-vision-reinvented-by-beck/> abgerufen am 29. Juli 2020.

- Misdariis, N., Nicolas, F., Warusfel, O., & Caussé, R. (2001). *Radiation Control on Multi-Loudspeaker Device : La Timée*. Paris: I.R.C.A.M - Centre Georges Pompidou.
- Moore, R. (o.D.). *A Midsummer Night's Dream: Immerse yourself in surround-sound Shakespeare*. Von BBC: <https://www.bbc.co.uk/programmes/articles/570p509ljD9gtdxqLzxJsJ4/a-midsummer-nights-dream-immersed-yourself-in-surround-sound-shakespeare> abgerufen am 3. September 2020.
- Mulder, J. D., Jansen, J., & van Rhijn, A. (2003). *An Affordable Optical Head Tracking System for Desktop VR/AR Systems*. Amsterdam: Center for Mathematics and Computer Science, CWI.
- Nötges, M. (Mai 2019). KLANG:technologies 3D In-Ear-Monitoring . *Tools4Music*, S. 112-116.
- Neal, M. (8. März 2016). *Why Are We Still Listening to Music in Two Dimensions?* Von Vice: [https://www.vice.com/en\\_us/article/pgka5z/why-are-we-still-listening-to-music-in-two-dimensions](https://www.vice.com/en_us/article/pgka5z/why-are-we-still-listening-to-music-in-two-dimensions) abgerufen am 24. Juli 2020.
- New Audio Technology. (o.D.). *Spatial Audio Designer - Plug-ins*. Von New Audio Technology: <https://newaudiotechnology.com/products/spatial-audio-designer/> abgerufen am 29. Juni 2020.
- NFON. (o.D.). *Headset*. Von NFON: <https://www.nfon.com/de/service/knowledgebase/knowledgebase-detail/headset> abgerufen am 23. Mai 2020.
- Noisemakers. (o.D.). *Ambi Pan & Ambi Head Manual*. Von Noisemakers: <https://www.noisemakers.fr/Downloads/AmbiManual.pdf> abgerufen am 27. Juli 2020.
- Obst, H. (15. Februar 2016b). *TEST: FLUX / IRCAM TOOLS HEAR V3*. Von Releasetime: <https://www.releasetime.de/test-flux-ircam-tools-hear-v3/> abgerufen am 3. Juli 2020.
- Obst, H. (10. Februar 2016a). *TEST: WAVES NX VIRTUAL MIX ROOM OVER HEADPHONES*. Von Releasetime: <https://www.releasetime.de/test-waves-nx-virtual-mix-room-over-headphones/> abgerufen am 29. Juni 2020.
- Obst, H. (28. Oktober 2017). *TEST: DEAR REALITY DEARVR MUSIC*. Von Releasetime: <https://www.releasetime.de/test-dear-reality-dearvr-music/> abgerufen am 29. Juni 2020.
- Oculus. (25. Mai 2017). *The Turning Forest*. Von Oculus: <https://www.oculus.com/experiences/gear-vr/1506375586042095/> abgerufen am 8. September 2020.
- Online-Audio-Monitor. (September 2019). *Online-Audio-Monitor 2019*. Von Online-Audio-Monitor: [https://www.online-audio-monitor.de/wp-content/uploads/2019-09-06\\_Bericht-OAM\\_2019\\_final.pdf](https://www.online-audio-monitor.de/wp-content/uploads/2019-09-06_Bericht-OAM_2019_final.pdf) abgerufen am 6. Juli 2020.
- Parnell, T. (2. September 2016). *Binaural Audio at the BBC Proms*. Von BBC: <https://www.bbc.co.uk/rd/blog/2016-09-binaural-proms> abgerufen am 31. Juli 2020.
- Parnell, T., & Pike, C. (2018). *An efficient method for producing binaural mixes of classical music from a primary stereo mix*. Milan: Audio Engineering Society.

- Parsons, T. (17. September 2020). *Apple spatial audio: what is it? How do you get it?* Von What Hifi: <https://www.whathifi.com/advice/what-is-apple-spatial-audio> abgerufen am 28. September 2020.
- Pearl Jam. (2000). Binaural. *CD*. Epic Records.
- Perfume Genius. (2017). *No Shape*. *CD*. Matador.
- Peysner-Kreis, I. (8. Mai 2019). *Wie ein Psychotrick den Super-Surround-Sound erzeugt*. Von Bayerischer Rundfunk: <https://www.br.de/puls/themen/netz/youtube-trend-8d-audio-100.html> abgerufen am 25. Juli 2020.
- Pierson, F. (Regisseur). (1976). *A Star is Born* [Kinofilm].
- Pike, C., & Watson, Z. (10. Mai 2016). *Virtual Reality Sound in The Turning Forest*. Von BBC Research and Development: <https://www.bbc.co.uk/rd/blog/2016-05-virtual-reality-sound-in-the-turning-forest> abgerufen am 8. September 2020.
- Rieger, M. (o.D.). *8D Audio – Die immersive Musik der Zukunft?!* Von VRTONUNG: <https://www.vrtonung.de/8d-audio/> abgerufen am 25. Juli 2020.
- Robin Schulz feat. Nico Santos. (2017). *More Than A Friend*. *Auf: Uncovered*. Tonspiel, Warner Music Group.
- Rudrich, D. (o.D.). *Plug-in Descriptions*. Von IEM Plug-in Suite: <https://plugins.iem.at/docs/pluginDescriptions/#binauraldecoder> abgerufen am 3. Juli 2020.
- Rumsey, F. (Juni 2017). Binaural audio and virtual acoustics. *The Journal of the Audio Engineering Society Vol. 65 (6)*, S. 524-528.
- S3A. (16. Oktober 2019). *VISR Production Suite User documentation*. Von S3A - Spatial Audio : <http://www.s3a-spatialaudio.org/wp-content/uploads/2019/10/userdoc-0.12.0.pdf> abgerufen am 27. Juli 2020.
- Schölzel, A. (16. Februar 2018). *Sennheiser AMBEO Orbit binaurales Panning-Plugin*. Von Delamar: <https://www.delamar.de/musiksoftware/audio-plugins/sennheiser-ambeco-orbit-44048/> abgerufen am 29. Juni 2020.
- Schörkhuber, C., Zaunschirm, M., & Höldrich, R. (2018). *Binaural Rendering of Ambisonic Signals via Magnitude Least Squares*. München: Deutsche Jahrestagung für Akustik.
- Schanze, R. (27. März 2017). *Gyroskop-Sensor im Hand: Was macht er? Wie funktioniert's? - Einfach erklärt*. Von Giga: <https://www.giga.de/apps/android/specials/gyroskop-sensor-im-handy-was-macht-er-wie-funktioniert-einfach-erklart/> abgerufen am 13. Juni 2020.
- SCHOEPS Mikrofone. (27. April 2018). *Tom Parnell (BBC R&D): Spatial Audio at the BBC*. Von Youtube: <https://www.youtube.com/watch?v=nSJoSBNDKMs> abgerufen am 3. September 2020.
- Schröder, J. (28. März 2018b). *Freifeldentzerrung und Diffusfeldentzerrung bei Kopfhörern*. Von Lowbeats: <https://www.lowbeats.de/freifeldentzerrung-und-diffusfeldentzerrung-bei-kopfoerern/> abgerufen am 18. Mai 2020.

- Schröder, J. (11. März 2018a). *Technik-Wiki: binaurale Aufnahmen*. Von Lowbeats: <https://www.lowbeats.de/technik-wiki-binaurale-aufnahmen/> abgerufen am 18. Mai 2020.
- Sengpiel, E. (April 2008). *Die Duplex-Theorie von Lord Rayleigh*. Von Sengpielaudio: <http://www.sengpielaudio.com/Duplex-Theorie.pdf> abgerufen am 10. April 2020.
- Sengpiel, E. (März 2009). *Druckstau bei Mikrofonen mit Kugelcharakteristik*. Von Sengpiel Audio: <http://www.sengpielaudio.com/DruckstauBeiMikrofonenMitKugelcharakteristik.pdf> abgerufen am 26. April 2020.
- Sennheiser. (o.D.). *AMBEO Orbit*. Von Sennheiser: <https://de-de.sennheiser.com/ambeo-orbit> abgerufen am 29. Juni 2020.
- Sennheiser. (o.D.). *Ambeo Smart Headset*. Von Sennheiser: <https://de-de.sennheiser.com/in-ear-kopfhoerer-ohrhoerer-3d-ambeo-smart-headset> abgerufen am 2. Mai 2020.
- Smyth Research. (o.D.). *A16 Realiser: The Future of Sound*. Von Smyth Research: <https://smyth-research.com> abgerufen am 29. Juni 2020.
- Spotify. (24. September 2020). *[iOS] Option to use "Spatial Audio" on iOS 14 (AirPods Pro)*. Von Spotify Community: <https://community.spotify.com/t5/Live-Ideas/iOS-Option-to-use-quot-Spatial-Audio-quot-on-iOS-14-AirPods-Pro/idi-p/5037421> abgerufen am 27. September 2020.
- Stede, C. (23. Januar 2018). *Interview Tom Ammermann – Mit Kraftwerk in die Atmos-Sphäre*. Von Mobile Fidelity Magazin: <https://www.mobilefidelity-magazin.de/interview-tom-ammermann-mit-kraftwerk-in-die-atmos-sphaere> abgerufen am 24. Juli 2020.
- Steele, B. (8. Januar 2019). *I want Sony's 360 Reality Audio to be the future of music*. Von Engadget: [https://www.engadget.com/2019-01-08-sony-360-reality-audio-hands-on.html?guccounter=1&guce\\_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce\\_referrer\\_sig=AQAAAD33hTovfFphy4PRc-RB93s3CNZahF9bFnmvMzoPbx9Yj97qP0-Oyd44ZgZqqLDIicQRGDohUypL4q0HiGNhSgSQgFp4WhioKO510](https://www.engadget.com/2019-01-08-sony-360-reality-audio-hands-on.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAD33hTovfFphy4PRc-RB93s3CNZahF9bFnmvMzoPbx9Yj97qP0-Oyd44ZgZqqLDIicQRGDohUypL4q0HiGNhSgSQgFp4WhioKO510) abgerufen am 27. September 2020.
- Strickland, J. (o.D.). *How Virtual Reality Gear Works: Virtual Reality Tracking Systems*. Von How Stuff Works: <https://electronics.howstuffworks.com/gadgets/other-gadgets/VR-gear6.htm> abgerufen am 9. Juni 2020.
- Takahashi, D. (10. Dezember 2014). *Coldplay and NextVR release virtual reality concert for Samsung Gear VR*. Von Venture Beat: <https://venturebeat.com/2014/12/10/coldplay-and-nextvr-release-virtual-reality-concert-for-samsung-gear-vr/> abgerufen am 30. Juli 2020.
- Teufel Blog. (o.D.). *Noise-Cancelling: Antischall im Kopfhörer*. Von Teufel: <https://blog.teufel.de/noise-cancelling-kopfhoerer-was-dahinter-steckt/> abgerufen am 18. Mai 2020.
- The Rolling Stones. (1991). *Flashpoint*. CD. Rollings Stones, Sony Music.

- Theile, G. (30. April 1980). *Über die Lokalisation im überlagerten Schallfeld*. Von Hauptmikrofon: [https://hauptmikrofon.de/theile/1980-1\\_Diss.\\_Theile\\_deutsch.pdf](https://hauptmikrofon.de/theile/1980-1_Diss._Theile_deutsch.pdf) abgerufen am 16. Mai 2020.
- Theile, G. (April 1981). Zur Kompatibilität von Kunstkopfsignalen mit intensitätsstereofonen Signalen bei Lautsprecherwiedergabe: Die Klangfarbe. *Rundfunktechnische Mitteilungen*, S. 146-154.
- Theiss, B. (22. August 2011). *So erkennt ein Smartphone jede Bewegung*. Von Connect: <https://www.connect.de/ratgeber/bewegungsmessung-von-smartphones-1169755.html> abgerufen am 13. Juni 2020.
- Tohoku University. (o.D.). *The RIEC HRTF Database*. Von RIEC Tohoku: <http://www.riec.tohoku.ac.jp/pub/hrtf/index.html> abgerufen am 8. Juni 2020.
- Tränkler, H.-R., & Reindl, L. (2014). *Sensortechnik: Handbuch für Praxis und Wissenschaft*. Berlin, Heidelberg: Springer-Verlag.
- Vaida, C. (November 2017). *Praxisreport: 3D Audio*. Von Professional Audio: <https://www.professional-audio.de/praxisreport-3d-audio/> abgerufen am 5. Juni 2020.
- Vimeo. (o.D.). *360-Grad-Videos hochladen*. Von Vimeo Hilfecenter: <https://vimeo.zendesk.com/hc/de/articles/115001877167-360-Grad-Videos-hochladen> abgerufen am 27. September 2020.
- Waniata, R. (13. Januar 2020). *Dolby Atmos is great for movies, but wait till you hear what it does for music*. Von Digitaltrends: <https://www.digitaltrends.com/home-theater/ces-2020-dolby-atmos-music-better-than-movies/> abgerufen am 28. September 2020.
- Warum, S. (2003). *Messung von Außenohrübertragungsfunktionen und direktionalen Raumimpulsantworten als Basis der Auralisation realer Räume*. Graz: Universität für Musik und darstellende Kunst in Graz.
- Watercutter, A. (19. Februar 2013). *Navigate Beck's Performance of 'Sound and Vision' as a 360-Degree Interactive Video*. Von Wired: <https://www.wired.com/2013/02/beck-360-degree-online-video/> abgerufen am 29. Juli 2020.
- Watson, Z. (11. Februar 2016). *A Virtual Reality Fairy Tale with Dynamic Binaural Sound is to be premiered at Tribeca Film Festival*. Von BBC Research and Development: <https://www.bbc.co.uk/rd/blog/2016-02-a-virtual-reality-fairy-tale> abgerufen am 8. September 2020.
- Waves. (10. Oktober 2017). *Ambisonics Explained: A Guide for Sound Engineers*. Von Waves: <https://www.waves.com/ambisonics-explained-guide-for-sound-engineers> abgerufen am 26. Juni 2020.
- Waves. (7. Oktober 2019). *Nx Headphone EQ Calibration – Supported Headphones*. Von Waves: <https://www.waves.com/support/nx-headphone-eq-calibration> abgerufen am 21. Juni 2020.

- Waves. (o.D.a). *Nx Head Tracker for Headphones: Overview*. Von Waves:  
<https://www.waves.com/hardware/nx-head-tracker#tab-overview> abgerufen am 14. Juni 2020.
- Waves. (o.D.b). *Nx Virtual Mix Room + Nx Head Tracker*. Von Waves:  
<https://www.waves.com/hardware/nx-virtual-mix-room-nx-head-tracker#tab-in-depth> abgerufen am 29. Juni 2020.
- Weinzierl, S. (2008). *Handbuch der Audiotechnik*. Berlin Heidelberg: Springer-Verlag.
- Wenzel, E. M., Arruada, M., Kistler, D. J., & Wightman, F. L. (Juli 1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, S. 111-123.
- Wernke-Schmiesing, S. (9. Juni 2016). *Musikvideo » Robin Schulz feat. Nico Santos - More Than A Friend*. Von Dance-Charts: <https://www.dance-charts.de/201606097320/robin-schulz-feat-nico-santos-more-than-a-friend?desk=1?mob=1?mob=1?mob=1> abgerufen am 24. Juli 2020.
- White, P. (Oktober 2019). *Binaural Panning In Logic Pro*. Von Sound on Sound:  
<https://www.soundonsound.com/techniques/binaural-panning-logic-pro> abgerufen am 29. Juni 2020.
- Wisse, E. (2018). Immersive Audio. *vdt magazin*, 3.
- Wittek, H., & Theile, G. (2016). *Die Anwendung eines stereofonen Mehrkanalverfahrens für 3D-Audio und VR*. Köln: 29th Tonmeistertagung.
- Woodcock, J., Pike, C., Melchior, F., Coleman, P., Franck, A., & Hilton, A. (2016). *Presenting the S3A objectbased audio drama dataset*. Manchester: University of Salford.
- Yost, W. A., Dye, J. R., & Sheft, S. (1996). A simulated “cocktail party” with up to three sound sources. *Perception & Psychophysics*, 1026–1036.
- Youtube. (o.D.). *Use spatial audio in 360-degree and VR videos*. Von Youtube Help:  
<https://support.google.com/youtube/answer/6395969?co=GENIE.Platform%3DDesktop&hl=en> abgerufen am 27. September 2020.
- Zotter, F., & Frank, M. (2019). *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Graz: Springer Verlag.

