

BACHELORARBEIT

IM STUDIENGANG AUDIOVISUELLE MEDIEN

3D AUDIO

WIE 3D AUDIO IM VERGLEICH ZU SURROUND SOUND DIE WAHRNEHMUNG UND DEN FOKUS EINES SPIELERS IN VIRTUAL REALITY-ANWENDUNGEN AUF BASIS VON KOPFHÖRERN BEEINFLUSSEN UND LENKEN KANN.

MIT AUSBLICK AUF 5D AUDIO.

VORGELEGT VON **CHRISTINA FUCHS** MARTIKEL-NR. 29952

AN DER HOCHSCHULE DER MEDIEN STUTTGART

EINGEREICHT AM 31.03.2018

ERSTPRÜFER: PROF. OLIVER CURDT

ZWEITPRÜFER: PROF. JENS-HELGE HERGESELL

ERKLÄRUNG

Hiermit versichere ich, Christina Fuchs, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: 3D Audio: Wie 3D Audio im Vergleich zu Surround Sound die Wahrnehmung und den Fokus eines Spielers in Virtual Reality Anwendungen auf Basis von Kopfhörern beeinflussen und lenken kann. Mit Ausblick auf 5D Audio 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.

Christina Fuchs

Stuttgart, den 31.03.2018

KURZFASSUNG

Die vorliegende Bachelorarbeit befasst sich mit der Technik des 3D Audio und Surround Sound sowie den damit verbundenen Möglichkeiten die Wahrnehmung, als auch den Fokus eines Spielers in Virtual Reality-Anwendungen mithilfe der Tongestaltung auf Kopfhörerbasis gezielt zu beeinflussen. Ziel dieser Arbeit ist die Analyse von Mechanismen der Wahrnehmungs- und Fokusmanipulation sowie ein Vergleich der Effektivität dieser Komponenten zwischen dem System des Surround Sound und des 3D Audio aus Sicht der Tongestaltung. Den Abschluss bildet ein Ausblick in die Arbeit mit 5D Audio. Diese Arbeit befasst sich mit der Thematik auf rein theoretischer Basis.

ABSTRACT

The present bachelor thesis deals with the technology of 3D Audio and Surround Sound as well as the associated possibilities of influencing the perception and focus of a player in virtual reality applications with the help of the earphone-based sound design. The work's aim is the analysis of perception and focus manipulation mechanisms as well as a comparison of the effectiveness of these components between the Surround Sound and 3D Audio systems with respect to the sound design point of view. The conclusion on this work is an outlook into the work with 5D Audio. This thesis deals with the topic on a purely theoretical basis.

INHALTSVERZEICHNIS

Abstract						
Ι		oildungsverzeichnis				
II	Abk	Abkürzungsverzeichnis				
1.	Einleitung					
2.	Hören					
	2.1.	Das menschliche Gehör	9			
	2.3.	3. Räumliches Hören				
		2.3.1. Lokalisation natürlicher Schallquellen	11			
		2.3.2. Interaurale Lokalisationmerkmale	12			
		2.3.3. Lokalisation in der Vertikalebene	14			
		2.3.4. Entfernungswahrnehmung	15			
		2.3.5. Cone of Confusion	16			
		2.3.6. HRTF und HRIR	17			
		2.3.7. Externe und interne Lokalisation	17			
		2.3.8. Phantomschallquellen	18			
3.	Surround Sound					
	3.1.	True Surround Sound	19			
	3.2.	Virtueller Surround Sound				
		3.2.1. Wellenfeldsynthese	21			
4.	3D A	O Audio				
	4.1.	Kanalbasiertier Sound	22			
	4.2.	Objektbasierter Sound				
	4.3.	Binauraltechnik				
		4.3.1. AB-Stereofonie	24			
		4.3.2. Kunstkopf-Stereofonie	25			
		4.3.3. Originalkopf-Stereofonie	26			
		4.3.4. Binauralsynthese	26			
5.	Virtual Reality					
	5.1.	Geschichte und Entwicklung	28			
	5.2.	. Virtual Reality Sound				
	5.3.	Kopfhörertechnik				

		5.3.1.	True und Virtual Surround Sound	30	
		5.3.2.	3D Audio	31	
6.	Psychologische Prozesse				
	6.1.	Wahrn	ehmung und Informationsverarbeitung	32	
		6.1.1.	Aufmerksamkeit	34	
		6.1.2.	Emotionen	36	
		6.1.4.	Propriozeption	37	
7.	Auditive Wahrnehmungs- und Fokusführung				
	7.1.		st		
		7.1.1.	Lautstärke	38	
	7.2.	Freque	nzen	41	
		7.2.1.	Frequenzselektion	41	
		7.2.2.	Frequenzmaskierung	42	
	7.3.	Emotio	onen	44	
		7.3.1.	Sound Effekte	44	
		7.3.2.	Musik	46	
		7.3.3.	Tempo	48	
		7.3.4.	Stimme	48	
8.	5D Audio				
	8.1.	Die Di	imensionen	54	
		8.2.1.	Tongestaltung in 4D	55	
		8.2.2.	Tongestaltung in 5D	55	
9.	Fazit			59	
III	Lite	raturve	rzeichnis	60	

ABBILDUNGSVERZEICHNIS

Abb. 2.1: Das menschliche Ohr, in schematischer Darstellung	9
Abb. 2.2: Die drei Lokalisationsebenen	11
Abb. 2.3: Schematisch dargestellte Horizontalebene und Laufzeitdifferenzen	12
Abb. 2.4: Blauert'sche Bänder/Richtungsbbestimmende Bänder	14
Abb. 2.5: Cone of Confusion	16
Abb. 3.1: Schematische Darstellung des Prinzips der Wellenfeldsynthese	21
Abb. 4.1: Schematische Darstellung des kanabasierten Systems	23
Abb. 4.2: Schematische Darstellung des objekbasierten Systems	23
Abb. 4.3: Neumann KU 100 Binaural Kunstkopf	25
Abb. 4.4: 360° Binaural Audio	25
Abb. 4.5: Schematische Darstellung des Prinzips der Binauraltechnik	27
Abb. 5.1: 7.1 True Surround Sound Headset mit mehreren Treibern	31
Abb. 5.2: 7.1 Virtual Surround Sound Headset mit einem Treiber	31
Abb. 6.1: Darstellung der Top-Down und Bottom-Up Verarbeitung	35
Abb. 7.1: Schematische Darstellung spektraler Maskierung	42
Abb. 7.2:Vereinfachte Darstellung von Maskierungen verschiedener Frequenze	n43
Abb. 7.3: Auszug aus dem Spiel Beat Saber, Hyperbolic Magnetism 2018	50

Ш

ABKÜRZUNGSVERZEICHNIS

1D Eindimensional
2D Zweidimensional
3D Dreisimensional
4D Vierdimensional
5D Fünfdimensional

Abb. Abbildung

BRIR engl. Binaural Room Response

Bspw. Beispielsweise Bzw. Beziehungsweise

cm Zentimeter dB Dezibel etc. et cetera

HRIR engl.: Head-Related Impulse Response HRTF engl.: Head-Related Transfer Function

Hz Hertz

IHL engl. Inside-Head Lokalization

ILD Interaurale Laufzeitdifferenz (engl.: ITD Interaural Time Difference)

IPD Interaurale Pegeldifferenz (engl.: ILD Interaural Level Difference)

m² Quadratmeter

MMN engl.: Mismatch-Negativity

ms Millisekunden

OHL engl.: Outside-Head Localization

ORTF franz.: Office de Radiodiffusion Télévision Française

OSS engl.: Optimal-Stereo-Signal

s. Kap. siehe Kapitel

SSA engl.: Stimulus Specific Adaptation

uU. unter Umständen

VR Virtuelle Realität (engl.: Virtual Reality)

WFS Wellenfeldsynthese

z.B. zum Beispiel

KAPITEL 1 **EINLEITUNG**

Die Entwicklung der Digitalen Medien schreitet mit einer enormen Geschwindigkeit voran und mir ihr auch die Entwicklung der Spielindustrie. Vor allem aber im Bereich des Virtual Reality (virtuelle Realität; kurz VR) möchte man die möglichst authentische Darstellung und Wahrnehmung der Wirklichkeit und ihrer physikalischen Eigenschaften in einem computergenerierten Umfeld darstellen. Es soll eine authentische Immersion und Interaktion des Spielers mit der virtuellen Umwelt erreicht werden (Wikipedia 2017c). Dieser Fortschritt erfordert dementsprechend der dem visuellen Reiz gleichwertigen Einbindung der auditiven Gestaltung der virtuellen Umwelten, was die Nachfrage an neuen und innovativen audiotechnischen Lösungen in Richtung Surround Sound, 3D Audio, als auch 5D Audio antreibt.

Immer wichtiger werden auch die psychischen Aspekte der Wahrnehmung und Interaktion des Spieles in und mit der virtuellen Umgebung, als auch die mit ihr verbundenen Möglichkeiten der Beeinflussung und Manipulation der Sinne, als auch die Fokusführung des Spielers auf auditiver Ebene um dessen Spielerlebnis einmalig zu gestalten.

KAPITEL 2 **HÖREN**

Um das menschliche Gehör während einer Interaktion in einer virtuellen Umgebung gezielt beeinflussen zu können, müssen die essentiellen Eigenschaften und Funktionsweisen des Gehörorgans, als auch dessen Wechselwirkung mit der Umwelt erörtert und bewusst gemacht werden.

Das Menschliche Gehör ist ein komplexes Organ das uns ermöglicht, unsere Umgebung akustisch wahrzunehmen und mit unserer Umwelt auf verschiedenen Ebenen interagieren zu können. Um sich aber auch in einer virtuellen Umgebung orientieren zu können und auch vollkommen einzutauchen (auch als Immersion bezeichnet) muss das Gehör, als auch das Gehirn, wichtige Informationen verarbeiten.

2.1. DAS MENSCHLICHE GEHÖR

Das menschliche Gehör lässt sich unterteilen in, das Außenohr mit der Ohrmuschel und dem äußeren Gehörgang, dem Mittelohr mit Trommelfell und Gehörknöchelchen (Hammer, Amboss, Steigbügel) und dem Innenohr mit der Cholera und dem Gleichgewichtsorgan.

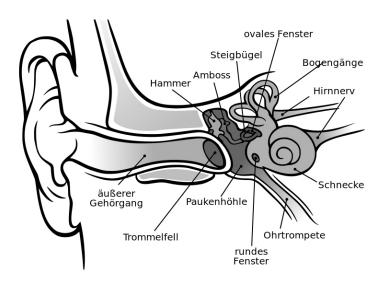


Abb. 2.1: Das menschliche Ohr, in schematischer Darstellung. Quelle: Storthmann 2013: 14-21

Der eintreffende Schall wird von der Ohrmuschel, der als richtungsabhängiger Schallreflektor dient, aufgefangen und gebündelt. Durch Überlagerung des Direktschalls und der Reflexion an der Ohrmuschel entstehen bestimmte Interferenzmuster, welche über den Gehörgang an das Trommelfell weitergeleitet werden. Diese spezifischen Verzerungen des Schalls sind bei der heutigen Entwicklung und Umsetzung der Wiedergabetechniken über Lautsprecher und Kopfhörer von enormer Wichtigkeit ist, um das Hörerlebnis in einer virtuellen Umgebung authentischer und realistischer zu gestalten.

Die Vibrationen des Trommelfells werden durch die drei Gehörknöchelchen an das Innenohr übertragen, wodurch die Schwingungen Aufgrund der Hebelwirkung der Gehörknöchelchen-Kette um das zwanzig-fache bis fünfzig-fache verstärkt wird. Diese verstärkten Schwingungen werden an das Innenohr weitergegeben, an die sogenannte Basilarmembran und das Cortische Organ, wo sie von den Haarzellen in bioelektrische Impulse umgewandelt werden. Diese werden danach von den Fasern des Hörnervs aufgenommen und zum Hörzentrum im Gehirn geleitet, wo die jeweiligen Impulse entschlüsselt, umgesetzt und interpretiert werden. So wird aus dem anfänglichen Schall Informationen (Gegenfurtner 2007: 23–28). Die Anatomie des menschlichen Gehörorgans ist ein wichtiger Aspekt, den man im Bereich gerade in der Binauraltechnik und der Gewinnung von HRTF-Daten, dazu mehr in Kap. 2.3.6. und 4.3.

2.2. SCHALLWAHRNEHMUNG

Das menschliche Hörorgan ist im Stande je nach physischem Alter und Zustand, Schallwellen mit Frequenzen im Bereich von etwa 16 Hz bis 20 kHz wahrzunehmen. Kopfhörer können diesen Bereich abdecken, was bei der Tongestaltung eine hohe Diversität bietet. Die untere Grenze der Hörfläche ist durch die Hörschwelle begrenzt. Sie beschreibt den niedrigsten Schalldruck, der einen Höreindruck hervorufen kann, welche stark Frequenzabhängig und individuell sehr unterschiedlich ist. Um etwa 4000 Hz reagiert das Ohr am empfindlichsten, in diesem Bereich befindet sich auch die menschliche Sprache, welche einer der stärksten Emotionsträger in der Tongestaltung darstellt. Nach oben hin wird die Hörfläche durch die Schmerzgrenze begrenzt, welche bei ca. 120 dB eine sichere Grenze definiert (Görne 2011: 120).

2.3. RÄUMLICHES HÖREN

Das räumliche Hören beruht auf der Wahrnehmung der räumlichen Lage und der Entfernung einer Schallquelle. Es spielt eine wichtige Rolle bei der Entwicklung neuer auditiver Systeme zur Darstellung einhüllender Klangerlebnisse. Gerade im Bereich der Spielindustrie und durch immer populärer werdende Virtual Reality, ist

das räumliche Hören von enormer Wichtigkeit; ohne diese Fähigkeit wären wir kaum in der Lage Schallereignisse in Ihrer Position, oder ihrer Entfernung einschätzen und lokalisieren zu können, was im Zeitalter des Virtual Reality und der Anwendung von Systemen in höheren Dimensionen immer mehr an Bedeutsamkeit gewinnt. Dieser Prozess ist jedoch von mehreren Faktoren und komplexen Vorgängen abhängig, die nachfolgend erläutert werden.

2.3.1. LOKALISATION NATÜRLICHER SCHALLQUELLEN

Die Lokalisation von Schallereignissen wird uns durch das binaurale Hören ermöglicht, aber auch durch das monaurale Hören, dem Hören mit nur einem Ohr. Binaurales Hören beschreibt das Hören mit beiden Ohren. Das ermöglicht uns das eintreffende Schallereignis in ihrer Richtung und Entfernung räumlich zu identifizieren und einschätzen zu können. Monaurales Hören bezieht sich dagegen auf das Hören mit nur einem Ohr, was die Lokalisation von Schallereignissen dennoch nicht ausschließt. Um ein Schallereignis lokalisieren zu können, behilft man sich mit den drei Lokalisationsebenen, auch Kopfbezogenes Koordinatensystem genannt.

Das kopfbezogene Koordinatensystem ist in drei ebenen aufgeteilt. Die Horizontaltebene (Abb. 2.2: links), auch Azimutebene oder Transversalebene genannt, welche in die obere und untere Hälfte unterteilt ist. Deren Hauptschalleinfallsrichtungen sind links, vorn, rechts und hinten. Die nächste ist die Medianebene (mittig), auch Sagittalebene genannt. Sie ist auf die Seiten links und rechts unterteilt, deren Hauptschalleinfallsrichtungen sich vorn, oben, hinten und unten befinden. Beim Hören in der Frontalebene, oder auch Coronalebene genannt, befinden sich die Hauptschalleinfallsrichtungen links, oben, rechts und unten. Unterteilt wird sie in die Flächen vorne und hinten (Sengpiel et al. 2012c).

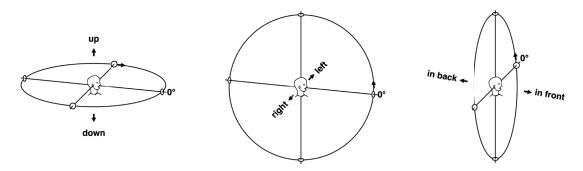


Abb. 2.2: Die drei Lokalisationsebenen v.l.n.r.: Horizontalebene, Medianebene, Frontalebene Quelle: Wikipedia: 2017a

Unsere räumliche Wahrnehmung ist für die Orientierung in der Horizontaltebene optimiert, da hier bei einem Schalleinfallswinkel von 0°-1°, die maximale Lokalisationsschärfe erreicht werden kann (Görne 2011: 126). Sie stellte bis vor noch wenigen Jahren die wichtigste Bezugsebene für die Darstellung von einhüllenden Klangwelten in der Virtual Reality Spielentwicklung dar. Aufgrund von Laufzeit- und Pegeldifferenzen sowie bestimmten Merkmalen des jeweiligen Reizes, ist es uns möglich eine Schallquelle in ihrer Position und Entfernung einzuschätzen.

2.3.2. INTERAURALE LOKALISATIONSMERKMALE

Die Lokalisierung in der Horizontalebene wird durch interaurale Zeit- und frequenzabhängige interaurale Pegeldifferenzen ermöglicht, sobald das Schallereignis aus einem von 0° oder 180° abweichenden Winkel auf das Gehör trifft.

INTERAURALE LAUFZEITDIFFERENZEN

Die Zeitdifferenz die der Schall benötigt, um beide Ohren in der Horizontalebene zu erreichen, bezeichnet man als *interaurale Laufzeitdifferenzen (ILD; engl.: ITD: Interaural Time Differences)*. Diese entstehen wenn der Schall nicht direkt von vorne oder hinten eintrifft, sondern in einem von 0° oder 180° abweichenden Einfallswinkel (Abb. 2.3). So wird bspw. ein von der linken Seite einfallender Schall von der linken Ohrmuschel früher erreicht als der rechten, da die rechte weiter entfernt ist und der Schall erst einen bestimmten Weg zum abgewandten Ohr zurücklegen muss. Dadurch ergibt sich die besagte Differenz. (Görne 2011: 126f).

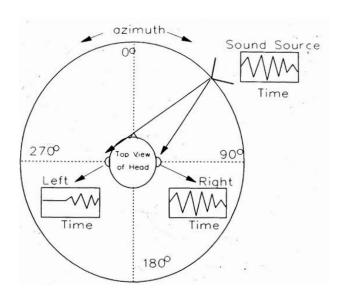


Abb. 2.3: Schematisch dargestellte Horizontalebene und Laufzeitdifferenzen Quelle: Heeger 2006

Keine Laufzeitunterschiede entstehen, wenn der Schall von vorne, der Kopfmitte, oder von hinten auftrifft (0° und 180°), da der Schallweg hierbei gleichlang ist. In den unteren Frequenzbereichen bis max. 1,6 kHz, kann das menschliche Gehör mithilfe der interauralen Phasendifferenz ein Schallereignis ausmachen. Geht man von einem reinen Sinuston aus der beide Ohren erreicht, kann es zu einer Fehllokalisation kommen, wenn das Signal keine Phasendifferenzen aufweist. Die Lokalisation wird einfacher, je größer die interauralen Phasendifferenzen im Signal sind, da sie dadurch einfacher zu unterscheiden sind. Hierbei dienen insbesondere die tieferen Frequenzen zur eindeutigen Identifizierung. Laufzeitdifferenzen besitzen eine frequenzspezifische Charakteristik. Laufzeitunterschiede bei höheren Frequenzen betragen lediglich 2/3 des niederfrequenten Werts (Neumeier 2016: 21). Es gibt eine maximale Grenzfrequenz von ca. 1,63 kHz, bei der Phasendifferenzen entstehen können, da die Schallwellenlänge in diesem Bereich noch größer als der Kopfdurchmesser ist und somit um den Kopf gebeugt wird. Erreichen diese den Kopfdurchmesser bzw. werden kleiner als dieser, kann kein eindeutiger Phasenunterschied mehr festgestellt werden. Ab diesem Zeitpunkt wird sich der Lokalisation durch Pegeldifferenzen bedient (Görne 2011: 126).

INTERAURALE PEGELDIFFERENZEN

Als Interaurale Pegeldifferenzen (IPD; engl.: ILD: Interaural Level Differences) oder auch Intensitätsunterschiede genannt, beschreibt man die Schalldruckdifferenz zwischen den Ohren. Bei dieser Art der Lokalisation in der Horizontalebene, wird die Schallquelle vom Kopf abgeschattet, wodurch der Schall das abgewandte Ohr mit einem geringeren Pegel erreicht, als das zugewandte Ohr. Am zugewandten Ohr entwickelt sich dementsprechend ein Druckstau, wohingegen am abgewandten Ohr ein Schallschatten entsteht. Diese Differenzen können jedoch nur entstehen, wenn die Schallwellenlänge kleiner als der Kopfdurchmesser ist. Bei höheren Frequenzen wie 5 kHz, beträgt die Pegeldifferenz ca. 12 dB, bei 10 kHz sind es um die 20 dB (Görne 2011: 128). Bei sehr niedrigen Frequenzen von unterhalb 300 Hz entstehen Aufgrund der Beugungserscheinungen des Schalls um den Kopf, so gut wie keine Pegeldifferenzen mehr. Das Gehör ist jedoch in der Lage, Pegeldifferenzen im gesamten für den Menschen hörbaren Frequenzbereich zu erkennen und in Richtungsinformationen umzusetzen (Dickreiter et al. 2014: 130). Die Audiowiedergabe in der Horizontalebene war bisher einer der wichtigsten Bezugspunkte im Hinblick auf die Wiedergabe im Bereich der Computerspiele in der Kopfhörertechnik. Die Surround-Systeme für die Kopfhörerwiedergabe in 5.1 und 7.1 sind auf diese Ebene ausgelegt und galten bisher als fester Standard.

2.3.3. LOKALISATION IN DER VERTIKALEBENE

In der Medianebene gibt es kaum interaurale Laufzeit- und Pegeldifferenzen an den Ohren, da die Ohren fast identisch beschallt werden, weshalb man auf die Auswertung des Signalspektrums angewiesen ist. Die Auslenkung des Hörereignisses wird durch die Gewichtung bestimmter Frequenzbänder bestimmt (Wegmann 2005: 8). Diese werden auch als Blauert'sche Bänder oder auch richtungsbestimmenden Frequenzbänder bekannt sind und ausschlaggebend für die Lokalisation in der Medianebene sind. Trifft ein Schallsignal auf das Ohr, durchgeht es eine Filterung Aufgrund der Beschaffenheit des Außenohres und des äußeren Gehörgangs. Dadurch entsteht ein richtungsabhängiger Filter, dessen spektrale Maxima und Minima von der Schalleinfallsrichtung abhängen. Je breitbandiger ein Signal ist, desto mehr spektrale Merkmale stehen dem Gehör zur Verfügung, die es auswerten kann. Mit dieser Information kann das Gehör die jeweilige Richtung einer Schallquelle einschätzen (Sengpiel et al. 2013).

Der Abbildung 2.4 nach lässt sich entnehmen, dass Frequenzen von ca. 250 Hz bis 500 Hz, sowie 3 kHz bis 4 kHz vorne lokalisiert werden. Bei ca. 1 kHz und 12 kHz vorzugsweise hinten und bei 8 kHz oben. Je bekannter ein Reizmuster ist, umso präziser fällt die Lokalisation aus. Wir drehen jedoch unseren Kopf meist, wenn wir ein Schallereignis nicht eindeutig ausmachen können um ein genaueren Eindruck zu bekommen. Sobald das der Fall ist, entstehen am Gehör interaurale Laufzeit- und Pegeldifferenzen, wodurch die Lokalisation durch die Richtungsbestimmenden Bänder nicht mehr primär ist.

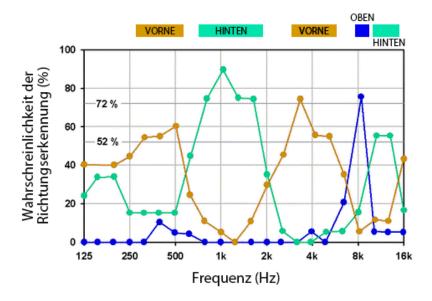


Abb. 2.4: Blauert'sche Bänder/Richtungsbestimmende Bänder Anlehung an Sengpiel et al. 2012b

2.3.4. ENTFERNUNGSWAHRNEHMUNG

Die auditive Entfernungswahrnehmung im offenen Raum, basiert in erster Linie auf den Erfahrungswerten des Hörers zu einem Schallereignis; sprich die Erfahrung des Gehörs, deren Übung, als auch die Bekanntheit der Schallquelle bzw. des Reizmusters. Befindet sich ein Schallereignis in unmittelbarer Nähe, sind die interauralen Pegeldifferenzen an den Ohren groß (etwa. 4 – 20 dB) und können demnach gut Informationen zur Distanzeinschätzung liefern. Je weiter weg sich eine Schallquelle befindet, desto schwieriger ist es für das Gehör dessen Entfernung korrekt einschätzen zu können. Der Pegel nimmt im Freien 6 dB mit der Verdopplung der Entfernung ab. Neben der Lautheit einer Schallquelle, bedient sich das Gehör zwei weiterer Schätzmechanismen wie der Tonhöhe; da mit einer zunehmenden Entfernung hohe Frequenzen eher absorbiert werden, klingen diese Schallquellen dumpfer, als auch der Bewegungsparallaxe, bei der durch die Bewegung des Kopfes sich nahe Schallquellen schneller in Gegenrichtung bewegen zu scheinen. Somit ist die Lautstärke der Schallereignis mitunter wichtig für die Entfernungswahrnehmung (Freie Universität Berlin 2010: 34f).

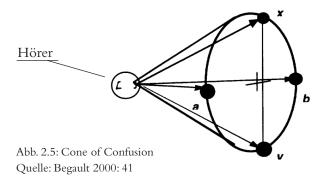
Kurz zusammengefasst lässt sich sagen, dass diese Differenzen von der Art des Schallsignals abhängen. Unter 1,6 kHz werden interaurale Laufzeitdifferenzen ausgewertet, um ein Schallereignis zu lokalisieren. Oberhalb 1,6 kHz wertet das Gehör primär Pegeldifferenzen aus. Die mitunter wichtigsten Schallsignale, die wir im täglichen Leben erfahren wie Sprache, weisen Frequenzen oberhalb 1,6 kHz auf (Triebsch 2010: 12). Die Lokalisation durch Zeit-, Phasen- und Pegeldifferenzen werden im englischen auch als Sound Localization Cues bezeichnet, diese Faktoren sind Ausschlaggebend bei der Gewinnung von HRTF-Daten für bspw. Anwendungen im Bereich der Surround als auch der 3D Audiotechnik. Neben all diesen Verfahren, die wir unterbewusst anwenden, kann das Gehirn jedoch auch über weitere Merkmale wie die Klangfarbe und weitere Reize wie beispielsweise Sehsinn, Tastsinn, die Bekanntheit des Reizmusters, aber auch durch weitere Kopfbewegungen, einen genaueren Eindruck über das Hörereignis gewinnen. All diese Faktoren müssen bei der Entwicklung auditiver Wiedergabesysteme; ob Lautsprecher oder Kopfhörer, als auch Softwareanwendungen beachtet und miteinbezogen werden. Ohne eine genaue Analyse kann es spätestens bei der Wiedergabe zu einem ungenauen Abbild des Klangbilds durch Fehllokalisationen oder Verzerrungen von Klangereignissen kommen.

COCKTAIL-PARTY-EFFEKT

Die Fähigkeit unseres Gehörs verschiedene Signale von einander zu trennen oder ein bestimmtes Signal herauszuhören geht ebenfalls auf das Binaurale Hören, also das Hören mit beiden Ohren zurück. Ein Beispiel hierfür bietet der sogenannte Cocktail-Party-Effekt. Dieser Effekt beschreibt die Fähigkeit des Gehörs, sich beispielsweise während einer Unterhaltung in einem lauten Umfeld, voll und ganz auf das Gespräch zu konzentrieren und alle unnötigen Umgebungsgeräusche ausblenden zu können (Stifelman 1994: 1). Oder aber, den eigenen Namen in einem lauten Raum herauszuhören, auch wenn man im jeweiligen Moment in eine Unterhaltung involviert ist (Gerrig et al. 2008). Für die Aufnahme von binauralen Ohrsignalen gibt es vier bekannte Verfahren, die folgend erläutert werden.

2.3.5. CONE OF CONFUSION

Bei interauralen Zeit- und Pegeldifferenzen, kann es zu Mehrdeutigkeiten in der vertikalen (x-v) und horizontalen (a-b) Ebene kommen, geht man davon aus, dass der Kopf eine perfekte Sphäre ist. In der natürlichen Umgebung nimmt das gehör mehrere Töne und Geräusche wahr gleichzeitig wahr (Natur). Ein sogenannter *Cone of Confusion*, (Abb.2.5) entsteht dann, wenn mehrere Schallereignisse mit der selben interauralen Zeit- und Pegeldifferenzen auf das Gehör treffen, sodass es zu Schwierigkeiten bei der Lokalisation des Schallereignisses kommen kann; der Mensch kann die einzelnen Schallquellen nicht mehr von einander trennen. Gelöst werden kann dies durch eine leichte Kopfbewegung (Moore 2012: 262). Dieses Problem trat of bei virtuellem Klangwiedergaben über Kopfhörer auf, wenn das Soundmaterial nicht genug Distanz/Raum hatte, wie das bei natürlichen Hören der Fall ist, sondern sozusagen im Kopf lokalisiert wurden (s. Kap. 2.3.7). Dadurch kam oftmals zu einer Verwechslung von Schallquellen die sich hinter oder vor dem Hörer befanden (Goldstein 2004:447).



2.3.6. HRTF UND HRIR

Die Spektrale Filterung, die am Eingang des Ohres auf dem Weg zum Trommelfell stattfindet, bezeichnet mal als Außenohrübertragungsfunktion (engl.: HRTF: Head-Related Transfer Function) oder auch kopfbezogene Übertragungsfunktion (Möser 2009: 633). Diese Filterung oder auch Verzerrung ist Abhängig von der Anatomie und Beschaffenheit eines menschlichen Körpers, bzw. dessen Kopf, Torso, Schultern und Ohren. Trifft ein Schall auf das Außenohr, erfährt der Schall Beugungen und Reflexionen durch die Ohrmuschel (Pinna), aber auch den zuvor genannten anatomischen Gegebenheiten. Diese von der Einfallsrichtung abhängende spektrale Filterung erkennt das Gehör und kann dementsprechend die Richtung einer Schallquelle lokalisieren. So gesehen hat jeder Mensch ein ähnliches, jedoch nie selbes HRTF-Muster, da sich die Anatomie eines jeden Ohrs unterscheidet. (Sengpiel et al. 2012a). Somit beschreibt das HRTF die Übertragungsfunktion im Frequenzbereich. Die entsprechende Impulsantwort dazu im Zeitbereich wird als kopfbezogene Impulsantwort (HRIR; engl.: Head-Related Impulse Response) bezeichnet und beschreibt das Verhältnis zwischen dem Eingangs- und Ausgangssignal einer Schallquelle am Ohreingang und ist das direkte Signal von der Schallquelle bis zum Hörer (ohne Nachhall). Die Messungen der HRIRs finden unter reflexionsarmen Bedingungen statt, sodass ist die Information der Impulsantwort nur durch die Einflüsse des menschlichen Körpers bestimmt werden, wie durch den Kopf und des Torsos, weshalb es als kopfbezogene Impulsantwort bezeichnet wird. Diese Messungen werden durch Methoden wie der Kunstkopfstereofonie erzielt, beschrieben in Kap. 4.3. Als weitere Option gibt es Datenbanken, die vielerlei HRTFs und/ oder HRIRs anbieten, die zuvor mit verschiedenen Kontrollpersonen aufgenommen wurden, als auch künstliche Kopf- und Torso Simulationen (Kayer 2009: 2).

Das Wissen über die HRTFs ist besonders im Hinblick auf die Weiterentwicklung auf dem Gebiet der Kunstkopfaufnahmen und bei der Computersynthese von binauralen Signalen für Virtual Reality Anwendungen von großer Wichtigkeit. Durch die Gewinnung von HRTF-Daten ist es möglich ein einzigartiges und Hörer-spezifisches Klangbild zu erstellen und diese in späteren Kopfhörerdarbietungen in verschiedenen virtuellen Anwendungen einzusetzen.

2.3.7. EXTERNE UND INTERNE LOKALISATION

Die externe Lokalisation, oder Außer-Kopf-Lokalisation (engl.: OHL: Outside-Head Localization) beschreibt die Ortung eines Schallereignisses außerhalb des Kopfes. Durch die Auswertung der Informationen der eintreffenden Signale, kann das Gehör den Schallquellenursprung analysieren und dementsprechend Informationen zur Richtung des Schallereignisses darbieten.

Die interne Im-Kopf-Lokalisation oder auch Lateralization (engl.: IHL: Inside-Head Localization) kommt vor allem dann auf, wenn ein Signal über Kopfhörer übertragen wird. Das führt dazu, dass ein Klangereignis nicht räumlich wahrgenommen wird. Das eintreffende Signal wird nicht nur auf interaurale Laufzeit- und Pegeldifferenzen analysiert, sondern gleicht auch das Reizmuster des Schallereignisses mit schon bekannten und erfahrenen Reizmustern ab. Sind die Ohrsignale nah am Ohr, annähernd gleich, sodass ein fast einheitliches Hörereignis entsteht und oder ist das Reizmuster unbekannt, ist es für das Gehör nicht möglich diese korrekt auszuwerten wodurch es zur Im-Kopf-Lokalisation kommt, das Hörereignis findet also im Kopf statt. Die Signale werden auf einer Linie zwischen beiden Ohren im Kopf lokalisiert, was die Tiefenstaffelung des Signals nicht mehr ermöglicht, da hier die Informationen der "Tiefe" nicht vorhanden sind. Kommt ein Signal mit der gleichen Laufzeit an beiden Ohren an, wird der Schall als Phantomschallquelle frontal aus der Mitte der Medianebene kommend wahrgenommen (Friesecke 2014: 135).

2.3.8. PHANTOMSCHALLQUELLEN

Als Phantomschallquelle wird eine Lokalisation beschrieben, dessen Schallereignis aus einer Richtung kommt, in der sich kein Schallwandler (z.B. Lautsprecher) befindet. Beim Natürlichen Hören stimmt das Hörereignis normalerweise mit der tatsächlichen Schallquelle überein, sodass Phantomschallquellen beim natürlichen Hören praktisch nicht auftreten. Eine Phantomschallquelle entsteht jedoch dann, wenn ein Schallsignal über zwei oder mehrere Lautsprecher in einer Stereo- oder Surround-Anordnung wiedergegeben wird. Alle kanalgebundenen Lautsprechersysteme basieren auf der Nachproduktion von Phantomschallquellen. Diese Phantomschallquellen verhalten sich jedoch nicht wie reale Schallquellen. Bewegt man den Kopf, bewegen sich auch die Phantomschallquellen mit. Sie wandern also mit der Zuhörerposition. Gibt man zwei identische Signale über zwei Lautsprecher wieder, werden diese aus der Mitte der Anordnung geortet. Verzögert man eines dieser Signale, wandert die Phantomschallquelle Richtung des schnelleren Signals (Görne 2011: 129f). Wie schon unter 2.3.7. erläutert, können Phantomschallquellen auch "im Kopf" auftreten, was sich in einer Im-Kopf-Lokalisation bemerkbar macht. Bei der Kopfhörerwiedergabe wird jedoch eine größere Pegeldifferenz als bei der Lautsprecherwiedergabe benötigt, um das Signal zu einer Seite auslenken zu können (Dickreiter et al. 2014: 350).

KAPITEL 3 SURROUND SOUND

Die Technik des Surround Sounds beschreibt eine Stereofonie, die auf dem Einsatz mehrerer Lautsprecher basiert. Im Vergleich zur Stereoanordnung wird die Anordnung im Surround um mindestens zwei weitere Lautsprechern erweitert. Diese sind um den Hörer herum platziert, sodass das Schallereignis aus mehreren Richtungen wahrgenommen werden kann.

3.1. TRUE SURROUND SOUND

Es gibt vielerlei Ansätze und Ausführungen zu Surround Sound Systemen. Der grundlegende Aufbau eines Surround-Systems besteht aus zwei Stereo Lautsprechern, die links und rechts vor dem Hörer platziert werden, als auch zwei Surround-Lautsprechern, die sich links und rechts hinter dem Hörer befinden. Dieses System wird auch als Quadrofonie (4.0) bezeichnet (Breh 1973). Ein wirksamer Surround Sound wurde versucht durch die Bildung von Phantomschallquellen mittels Pan-Pots 1 über vier separate Kanäle zu erzeugen, welches eine Verbesserung im Vergleich zum Stereobild darbot. 1975 brachte die Firma Dolby Laboratories das analoge Format Dolby Stereo auf den Markt, das hauptsächlich im Kino eingesetzt wurde. Später entwickelte Dolby Laboratories ein Surround-System für den Privatgebrauch, das Dolby Surround. Damit stellten sie die Weichen der Surround-Wiedergabesysteme für zukünftigen den Heimbereich, und bereitete mit ihrem Stereo-System den Weg für weitere Surround Sound-Wiedergabeformate für verschiedene Anwendungen, wie wir sie heute kennen. Langfristige Anwendung, vor allem im Heimkino und bei Computerspielen (Kopfhörertechnik) fanden die Surround-Formate 5.1 und 7.1, die jeweils zu den kanalbasierten Systemen zählen (nachfolgend erklärt). Beide Formate sind heute die am weitesten verbreiteten Surround-Verfahren für die Anwendung im Privatbereich. Es gibt 5 Hauptkanäle, welche über die L, C, R, Ls und Rs Lautsprecher wiedergegeben werden und dem sechsten Tieftonkanal (engl.: LFE: Low Frequency Enhancement), der über den Subwoofer wiedergegeben wird. Die 5 Hauptkanäle sind gleichwertig und haben einen Frequenzumfang von ca. 20 Hz - 20 kHz. Der Tieftöner

Auch Panoramaregler genannt, durch die die Lautstärkeverteilung eines Audiosignals geregelt wird.

hingegen gibt Frequenzen ca. unterhalb 120 Hz wieder. 7.1 wird durch zwei weitere diskrete Kanäle erweitert, die über Lautsprecher wiedergegeben werden. Die herkömmlichen Surround Sound-Systeme agieren auf der horizontalen, der Azimut-Ebene (zweidimensional; 2D). Es fehlt demnach die Möglichkeit, die Klanginformationen in der vertikalen Ebene, sprich der Höhe und Tiefe unterzubringen.

Bei der Wiedergabe mit Lautsprechern gibt es nur den kleinen Bereich der sich *Sweet-Spot* nennt, in dem die optimale Klangqualität durch die Wiedergabe von Klangereignissen über Lautsprecher-Systeme gewährleistet werden kann. Dieser ist im Normalfall mit einem Öffnungswinkel von 60° zu den Lautsprechern und demnach in der Mitte der Basis. Außerhalb dieses Spots kann es zu weiteren Verzerrungen und Fehllokalisationen kommen. Dieser Sweet-Spot fällt jedoch bei Kopfhöreranwendungen vollkommen weg.

Heutzutage gibt es eine Vielzahl an Systemen, die surroundspezifische Signale auch über eine Kopfhörerdarbietung authentisch und mit herausragender Qualität abbilden können. Sie alle lehnen sich an die vorangegangene Technik der Wiedergabe über Lautsprechersysteme an. Die Nachfrage nach derartigen technischen Lösungen für die Darbietung über Kopfhörer ist Aufgrund der weiter steigenden Popularität des Virtual Reality sehr groß geworden und entwickelt sich dementsprechend rasant fort.

3.2. VIRTUELLER SURROUND SOUND

Oft ist es im Privatbereich aus platztechnischen Gründen nicht möglich, eine Surround-Anlage mit 6 oder gar mehr Lautsprechern aufzustellen. Hier möchte die Idee des virtuellen Surrounds Abhilfe schaffen. Die Technik des virtuellen Surround Sounds greift auf ein System mit zwei bis drei Lautsprechern zurück, um einen realitätsnahen Raumklang zu erzeugen. Dem Gehör soll vorgegaukelt werden, dass ein Klangereignis, welches eigentlich aus einem 2.0 oder 2.1 Aufbau kommt, als ein 5.1 oder höheres Wiedergabesystem wahrgenommen wird. Dabei müssen jedoch ebenfalls interaurale Laufzeit- und Pegeldifferenzen beachtet und mitgerechnet werden, die uns die Lokalisation erst ermöglichen. Das wird mithilfe von der Außenohr-Übertragungsfunktion (HRTF, s. Kap. 3.2.), oder auch der Nutzung von Reflexionen im Raum erreicht.

Diese Technik lässt sich auf ähnliche Weise auch auf die Wiedergabe über Kopfhörersysteme anwenden, die im Gegensatz zu True Surround Sound-Kopfhörern die Wiedergabe nur auf zwei Lautsprecher setzt. Neben einem erschwinglicheren Preis bietet diese Variante einen kraftvolleren und dynamischeren Klang Aufgrund der Größe der eingebauten Treiber; siehe dazu Kap. 5.3.1.

3.2.1. WELLENFELDSYNTHESE

Die Technik des virtuellen Surrounds greift auf die Vorgehensweise der *Wellenfeldsynthese (WFS)* zurück, welche auf dem Huygens-Prinzip beruht. Demnach kann jede Wellenfront in eine Folge von einzelnen sogenannten Elementarwellen zerlegt werden. Im Umkehrschluss kann durch Überlagerung (beliebig vieler) einzelner Wellen, den Elementarwellen, jedes Wellenfeld synthetisiert werden – jeder Punkt einer Wellenfront, kann als Ausgangspunkt einer neuen Elementarwelle betrachtet werden (siehe Abb. 3.1).

"Einzeln steuerbare Lautsprechermembranen, [...] werden dazu exakt in dem Moment ausgelenkt, zu dem die Wellenfront einer virtuellen Schallquelle ihren Raumpunkt durchlaufen würde. Aus einer ausreichend großen Zahl solcher Elementarwellen entsteht dann eine gemeinsame Wellenfront, die sich physikalisch nicht von einer realen Wellenfront unterscheidet." (Oellers, 2017)

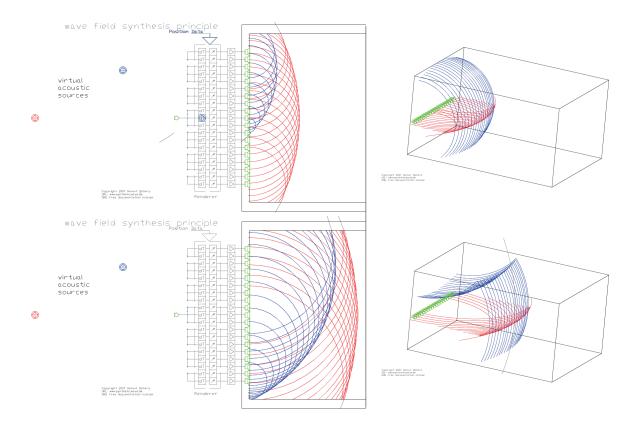


Abb. 3.1: Schematische Darstellung des Prinzips der Wellenfeldsynthese.

L: horizontale Ansicht; R: räumliche Ansicht.

Quelle: Oellers, 2017

KAPITEL 4 3D AUDIO

Mit dem Konzept des dreidimensionalen (3D) Audio verfolgt man eine natürliche und daraus hervorgehende realistischere räumliche Wiedergabe. Somit hat man sich zum Ziel gesetzt, die räumliche Wiedergabe in die Höhe und Tiefe zu erweitern, wie man es aus dem realen Leben gewohnt ist. Nach Hu et al. ist die Definition des 3D Audio weitgehend akzeptiert, sofern es durch Charakteristiken wie das Lokalisieren von Klangereignissen aus einer beliebigen Richtung in einem dreidimensionalen Raum, die Realisation der Distanzwahrnehmung, als auch ein verbessertes Gefühl der klangakustischen Szenerie beschrieben werden kann (2012: 82)

"In einer alternativen Anwendung zusätzlicher Kanäle wird die Horizontalebene verlassen. Lautsprecheranordnungen in der Ebene oberhalb des Hörers vervollständigen den räumlichen Gestaltungsbereich und ermöglichen in bestimmten Grenzen den Aufbau eines dreidimensionalen Klangbilds" (Theile 2014: 311f).

Wir nehmen unsere Umgebung akustisch nicht nur von vorne, hinten und seitlich auf, sondern auch über und unter uns. Durch die bisherigen Surround-Systeme ist es nach wie vor möglich Klänge im Raum zu platzieren, jedoch sind diese in den meisten Fällen Kanalgebunden (nachfolgend erklärt) und fungieren ausschließlich in der horizontalen Ebene. Das Konzept des 3D Audio ermöglicht Klänge frei im dreidimensionalen Raum zu positionieren, ohne dabei an eine bestimmte Lautsprecherkonstellation oder – anzahl gebunden zu sein. Somit hat man nahezu unbegrenzte Möglichkeiten bei der Klanggestaltung um die Illusion der Immersion mit einer virtuellen Umwelt zu verstärken und noch nie dagewesene Klangräume zu erschaffen.

Es gibt jedoch auch einige Hybrid-Systeme, die beide Verfahren; kanalals auch objektbasiert, miteinander verbinden und daraus ihren Nutzen ziehen.

4.1. KANALBASIERTER SOUND

Kanalbasierter Sound bedeutet, dass der Toningenieur schon bei der Mischung im Studio, die Informationen der Spuren fest über die jeweiligen Kanäle, die bestimmte Schallwandler ansprechen, den Lautsprechern zuweist. Demnach muss im Vorfeld

logischerweise bekannt sein, auf welchem Wiedergabesystem das Projekt abgespielt wird (5.1, 7.1 etc.). Diese zuvor genannten Kanäle werden auch als diskrete Kanäle bezeichnet, da sie nur einem bestimmten Lautsprecher zugewiesen sind (s. Abb.4.1.). Bei der Wiedergabe in einer neuen Umgebung muss die selbe Lautsprecheranordnung gewählt werden, sodass die korrekte Konfiguration des Ingenieurs wiedergegeben werden kann.

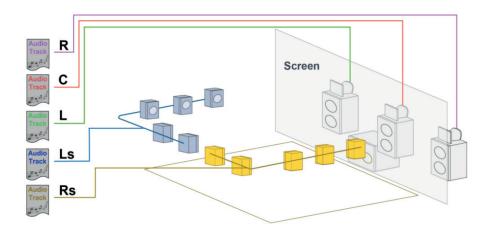


Abb. 4.1: Schematische Darstellung des kanabasierten Systems Quelle: Maile 2016: 7

4.2. OBJEKTBASIERTER SOUND

Ein Vorteil der objektbasierten zur kanalbasierten Variante ist, dass man hierbei eine spezifische Schallquelle einem bestimmten Objekt zuweisen kann und diese an gewünschten Koordinaten, also frei im dreidimensionalen Hörraum positionieren kann. Das jeweilige Element enthält Metadaten wie die Position im Raum, Lautstärke, Verzögerung, aber auch die Flugbahn und Beschaffenheit der Schallquelle (diffus, punktuell) (Schuller 2016).

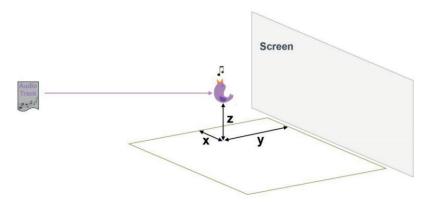


Abb. 4.2: Schematische Darstellung des objekbasierten Systems Quelle: Maile 2016: 8

Dem Nutzer wird es mit Hilfe der Metadaten möglich, einzelne Elemente in ihrer Lautstärke zu verändern, wie beispielsweise einen Dialog, aber auch eine gesamte Szene zu rotieren. Die Möglichkeit eine Szene während einer Interaktion rotieren zu können, ist vor allem bei VR-Anwendungen im Zusammenhang mit Head-Mounted-Displays oder auch Head-Tracking-Systemen wichtig, um eine Szene mit den Bewegungen des Nutzers synchronisieren zu können (Beermann 2017: 31). Bei der Ausgabe rendert ein Audioprozessor den Sound und berechnet welches Klangobjekt mit welcher Lautstärke auf jeder einzelnen Box wiedergegeben werden muss. Dies ist unabhängig von der Anzahl der Lautsprecher, wodurch sich Klangobjekte im Gegensatz zu kanalbasierten Systemen, präziser platzieren lassen können.

Dadurch lässt sich auch das bisherige Problem der optimalen Abhörposition; des Sweet-Spots, beseitigen (Weinzierl 2014: 93).

4.3. BINAURALTECHNIK

Die Binauraltechnik beschäftigt sich mit der Erzeugung und Wiedergabe von hörgetreuen Ohrsignalen. So hat sie sich innerhalb der letzten Jahre zu einem wesentlichen Bestandteil der Forschung im auditiven Bereich entwickelt. Ziel ist es, eine möglichst authentische akustische Simulation bzw. Reproduktion einer realen Umgebung, wie sie am menschlichen Trommelfell ankommt. Dazu müssen einige schallfeldbeeinflussende Komponenten wie der Eintrittswinkel, Laufzeit-, Pegel- und Spektraldifferenzen am Ohr, als auch die Ausbreitung im virtuellen Raum beachtet werden (Lentz 2008: vi).

"Ein häufig benutzter Ansatz zur hörgetreuen Übertragung basiert auf der folgenden Annahme: Wenn die Schallsignale an den Trommelfellen bei der Wiedergabe genau denjenigen entsprechen, die sich bei der Aufnahmesituation an den Trommelfellen ergäben, dann werden die Zuhörer auch genau das gleiche hören, was sie in der Aufnahmesituation hören würden." (Blauert 1996)

4.3.1. AB-STEREOFONIE

Unter die AB-Stereophonie fallen die beiden Stereoaufnahme-verfahren ORTF (Office de Radiodiffusion Télévision Française, Radio France) und OSS (Optimal-Stereo-Signal; oder auch Jecklin-Verfahren). Beide haben eine ähnliche Funktions- und Wirkungsweise.

Das ORTF ist eine Kombination aus der AB- und XY Aufnahmetechnik. Zwei Mikrofone werden in einem Abstand von genau 17 cm (durchschnittlicher Kopfdurchmesser und einem Öffnungswinkel von 110° angeordnet. Es berücksichtigt Laufzeitund Pegelunterschiede.

Das OSS ist dem Aufbau des ORTF sehr ähnlich mit der Ausnahme, dass zwischen beiden Mikrofonen eine Scheibe (ca. 30 cm) platziert wird. Das hat den Grund, dass beim natürlichen Hören die Pegeldifferenzen Aufgrund der Schallbeugung um den Kopf, frequenzabhängig sind. Durch Einsatz dieser schallabsorbierender Scheibe wird dieser Effekt erzielt, was einen höheren Stereo-Effekt bietet, als das ORTF-Verfahren (Sengpiel et al. 2014).

4.3.2. KUNSTKOPF-STEREOFONIE

Die Kunstkopfstereofonie beabsichtigt mit ihrem Verfahren die Ohrsignale (HRTFs; s. Kap. 2.3.6), wie sie am menschlichen Trommelfell ankommen so aufzunehmen, dass sie bei der Reproduktion so hörgetreu wie möglich wiedergegeben werden können. Wichtig sind die Messungen der HRTFs für die Erstellung von dreidimensionalem Klang, als auch für die Tongestaltung im Surround, was seine Anwendung im Bereich Virtual Reality mittlerweile unverzichtbar macht.

Für die Gewinnung dieser Daten wird ein Kunstkopf mit einer Nachbildung beider Außenohren, deren äußeren Gehörgängen und zwei nagelgroßen Messmikrofonen, die als Trommelfell dienen, verwendet. Anstatt ein Schallfelld künstlich nachzubilden, werden die Ohrsignale mittels des Kunstkopfes an einer bestimmten Position aufgenommen. Die Wiedergabe erfolgt über Kopfhörer. Im besten Fall stimmen die aufgenommen Daten mit den Ohrsignalen des Hörers überein. (Raffaseder 2010: 166). Man kann dieses Verfahren auch mithilfe eines Menschen durchführen Jedoch sind die HRTFs bei jedem Menschen verschieden, weshalb der Kunstkopf viel mehr als ein Durchschnitt angesehen wird.



Abb. 4.3: Neumann KU 100 Binaural Kunstkopf

Quelle: Neumann (2017)



Abb. 4.4: "360° binaural audio", entwickelt von Chris Milk

Quelle: 360cameraonline

4.3.3. ORIGINALKOPF-STEREOFONIE

Statt einen Kunstkopf zu verwenden, lassen sich die kleinen Messmikrofone auch in die Ohren eines Menschen legen. Sie ähneln den Kopfhörers eines Walkmans, mit dem Unterschied, dass sie die Ohrsignale aufnehmen, statt Klang abzugeben. Das Verfahren ist mir dem des Kunstkopfes zu vergleichen.

4.3.4. BINAURALSYNTHESE

Die Schallübertragung bei der binauralen Synthese ist von einer Schallquelle im Raum bis vor das menschliche Trommelfell weitgehend linear und unabhängig von zeitlichen Verschiebungen. Es kann im Zeitbereich für eine Impulsantwort, als auch im Frequenzbereich durch eine Übertragungsfunktion beschrieben werden. Binaurale Signale nachzubilden wird mittels einer Faltung (Konvolution ²) erreicht. Dabei wird durch das vor Ort aufgenommene nachhallfreie Quellsignal, mit zwei zuvor aufgenommenen oder auch simulierten kopfbezogenen binauralen Raumimpulsantworten (engl.: Binaural Room Responses, BRIRs) synthetisiert, sodass man ein binaurales Signalpaar für beide Ohren erhält (Litfin 2014: 49). Hierbei können folglich zwei Verfahren unterschieden werden. Die binauralen Raumimpulsantworten berücksichtigen im Gegensatz zur kopfbezogenen Impulsantwort (s. Kap.: 2.3.6), die von der Schallquelle ausgehenden ersten Reflexionen und den Nachhall im Raum, sodass alle Ausbreitungswege des Schalls bis zu den Ohren aufgefasst werden können. Wichtig wird dieses Verfahren in Anwendung von 3D Audio, um die erstellten virtuellen Umgebung auch akustisch realistisch und natürlich gestalten zu können.

STATISCHE UND DYNAMISCHE BINAURALSYNTHESE

Bei der statischen Binauralsynthese wird die Kopfposition/-rotation des Hörers bei der Wiedergabe nicht getracked (verfolgt). Heißt, wenn sich dieser bewegt, bewegen sich sämtliche Schallquellen mit der Kopfbewegung, was so in der Natur nicht vorkommt.

Im Gegensatz zur statischen Binauralsynthese, wird bei der dynamischen die Kopfbewegung durch einen angebrachten Head-Tracker erfasst. Diese Daten werden mithilfe eines Algorithmus mit den dazu passenden kopfbezogenen binauralen Raumimpulsantworten gefaltet und erlauben es, den Schallquellen unabhängig von der Kopfbewegung, deren Position, relativ zum Hörer, im simulierten Raum beizubehalten (Schlenstedt 2015: 13). Hinzukommend wird bei dieser Simulation ebenso

² Ein math. Operator, der für zwei Funktionen eine dritte Funktion ausgibt

die Kopfgröße des Hörers berücksichtigt, da der individuelle Abstand beider Ohren und die daraus resultierenden interauralen Laufzeitdifferenzen, die Lokalisation einer Schallquelle grundlegend beeinflussen.

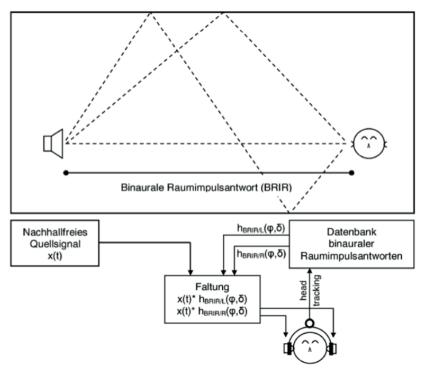


Abb. 4.5: Schematische Darstellung des Prinzips der Binauraltechnik Quelle: Weinzierl 2008: 672

Das Verfahren der Binauralsynthese ermöglicht demnach eine kopfhörerbasierte realitätsnahe Wiedergabe dreidimensionaler Schallquellen, zur akustischen Reproduktion bzw. Simulation eines realen oder virtuellen Raumes. Dabei müssen die spontanen Kopfbewegungen des Hörers in Echtzeit erfasst und miteinkalkuliert werden, um einen realitätsnahen Eindruck des Raumes gewährleisten zu können, damit nicht der Eindruck des Mitwanderns der Schallquellen entsteht.

KAPITEL 5 VIRTUAL REALITY

Der Begriff Virtual Reality (VR, dt.: Virtuelle Realität) wurde in den 1990ern geprägt und wird unter anderem als

"[...] eine neuartige Benutzeroberfläche, in der die Benutzer innerhalb einer simulierten Realität handeln [...] und sich im Idealfall so wie in ihrer bekannten Umgebung verhalten." (Bill 2009: 6)

"[...] werden Techniken verstanden, die es erlauben, einen Menschen unmittelbar in eine Computergenerierte Welt zu integrieren. Als die Mensch-Maschine-Schnittstelle der Zukunft angesehen, sprechen Techniken der virtuellen Realität mehrere Sinne des Menschen zugleich an." (Preißer 2010: 2)

beschrieben. Mit VR wurden zunächst primär Computer, kopfbezogene Anzeigegeräte (engl.: HDM; Head Mounted Displays), als auch bewegungssensitive Datenhandschuhe verbunden. (Pietschann 2017: 40). Es wird nicht selten der Begriff Immersion benutzt, welches als "Das Eintauchen von jemandem oder etwas in eine Flüssigkeit" oder auch "Tiefe mentale Mitwirkung in etwas (z.B. eine Aktivität)" beschrieben wird. (frei übersetzt aus dem englischen; Oxford Dictionaries 2018).

5.1. GESCHICHTE UND ENTWICKLUNG

Die Idee des Virtual Reality begann in etwa um 1838 mit er Erfindung des *Stereoskops* durch den englischen Wissenschaftler und Erfinder Charles Wheatstone. (Virtual Reality Society 2017). Das Stereoskop sollte den Eindruck der Tiefe durch zwei fast identische Bilder verschaffen, die vom linken und rechten Auge jeweils aus verschieden Blickwinkeln betrachtet werden. Der Begriff Virtual Reality wurde jedoch erst ab den 1990ern Jahren populär. Schon um 1920 erschien der erste elektromechanische Flugsimulator, der sogenannte *Link Trainer* von Edwin Link. Dieser Flugsimulator war noch ein einfaches Metallkonstrukt mit einer moderaten Einrichtung und einem kleinen Motor darin, womit das Ruder und die Lenksäule gesteuert werden konnten. Dadurch war es 500.000 Piloten im zweiten Weltkriegs erstmals möglich, anfänglich

durch diese Simulatoren zu lernen und ihre Fähigkeiten auszubauen (Plunkett 2011).

1962 entwickelte und patentierte Morton Heilig das Sensorama, welches neben dem visuellen Reiz auch erstmals den Gehör-, Tast- und Geruchssinn miteinbezog. Das war sozusagen die erste VR-Maschine. Es beinhaltete Stereo-Boxen, einen stereoskopischen 3D-Monitor, Ventilatoren, Geruchserzeuger, als auch eine vibrierende Sitzgelegenheit (Turi 2014).

5.2. VIRTUAL REALITY SOUND

"[...] Leute werden sagen, dass die Bilder eines Computerspiels durch hochqualitatives Audiodesign aufgewertet werden, jedoch wird Audiodesign durch hochwertige Bilder nicht besser; sie lassen es sogar schlechter klingen. [...]" (frei übersetzt aus dem englischen; Begault 2000: 10f).

Die Wiedergabe des Virtual Reality Sounds für VR Anwendungen wird heutzutage hauptsächlich über Kopfhörer dargeboten, damit der Sound sozusagen direkt ins Ohr geleitet wird und keine weiteren Verzerrungen durch Kopf, Torso, oder auch Reflexionen im Raum erfährt, wie das bei der Wiedergabe über ein Lautsprechersystem der Fall wäre. Durch das Verfahren der Kunstkopfstereofonie, oder auch der Binauralsynthese ist es möglich, eine glaubwürdige Klanggestaltung zu erzeugen. Wo noch vor wenigen Jahren Surround Sound den Standard für VR Anwendungen gebildet hat, rückt sich nach und nach die Gestaltung mittels 3D Audio in den Vordergrund. Um eine virtuelle Umgebung glaubhaft zu machen ist nicht nur der visuelle Eindruck wichtig, sondern auch die klangliche Darstellung.

5.3. KOPFHÖRERTECHNIK

Die heutige Technik der Schallwandlung bietet uns durch verschiedene Bauformen und Systeme vielerlei Möglichkeiten, das gewünschte Schallsignal direkt am oder im Ohr zu erzeugen, sodass der gesamte vom Menschen hörbare Frequenzbereich von ca. 20 Hz bis 20 kHz, wiedergegeben werden kann. Das Problem, dass sich jedoch bei der Wiedergabe von Signalen über Kopfhörersysteme bisher stellte ist die Im-Kopf-Lokalisation (s. Kap.: 2.3.7). Lokalisiert wird das Hörereignis je nach Signal und Ort der Schallquelle, im Kopf bzw. in der Mitte des Kopfes da das Signal direkt am Ohr anliegt. Ein Vorteil der Wiedergabe über Kopfhörer ist, dass jedem Ohr ein separates Signal zugeführt werden kann, wodurch es nicht zum *Crosstalk (XT; dt.: Übersprechen)* kommen kann und diese ebenfalls, je nach Modell (geschlossen, halboffen, offen,

In-Ear), eine akustische Isolation von der externen Umgebung bieten und dadurch eine höhere Immersion mit der virtuellen Umgebung ermöglichen.

5.3.1. TRUE UND VIRTUAL SURROUND SOUND

Durch die Wiedergabe mit einem Lautsprechersystem in der Heimanwendung kann man nie vollkommen sicher sein, wie der Sound am Ende im jeweiligen Raum klingen wird. Abhängig von der Anlage, dem Raum und der Position der Lautsprecher, kann die Qualität des Endproduktes unter Umständen leiden, da die Schallquellen vom Lautsprecher bis zum Gehörkanal einen bestimmten Weg zurücklegen müssen und dadurch gewisse Verzerrungen erfahren, wodurch das jeweilige Schallsignal nicht immer wie gewünscht abgebildet werden kann. Eine adäquate Lösung bieten hierbei die Surround Sound-Kopfhörersysteme. Grob kann man zwischen dem *True Surround Sound* (nachfolgend True SS), der sozusagen als echter Surround Sound bezeichnet wird und dem *Virtual Surround Sound* (nachfolgend Virtual SS) unterscheiden.

Beim True SS befinden sich mehrere kleine diskrete Lautsprecher pro Ohrseite in der Kopfhörermuschel, die alle einem bestimmten Kanal zugewiesen sind und das Ohr in unterschiedlichen Winkeln beschallen. Durch die verschiedenen Schalleinfallswinkel wird das Problem der Im-Kopf-Lokalisation umgangen, da der Schall ausschließlich an der Ohrmuschel verzerrt wird, bevor er in den Ohrkanal weitergeleitet wird. Diese Technik basiert grundlegend auf den Surround-Lautsprechersystemen; je nach System; 5.1 oder 7.1, werden dementsprechend 6 bis 10 Lautsprecher für die Wiedergabe verwendet (Scheeden 2010).

Virtueller SS beruht dagegen auf einem Stereo-Lautsprechersystem. Je nach System wird ein 5.1 oder 7.1 Mix durch einen komplexe Signalverarbeitungsalgorithmen in einen Stereo-Mix umgerechnet (Temporal 2017). Durch die Anwendung von HRT-Fs (s. Kap. 2.3.7.) kann der Effekt eines mehrkanaligen Sounds aus einem Stereo-Lautsprecherpaar simuliert werden. Theoretisch bieten True SS Kopfhörer aufgrund der für jeden Kanal fest zugeordneten Treiber eine genauere Abbildung des erstellten Sound Designs, was vor allem bei Multiplayer Shooter von Wichtigkeit sein kann, wo die exakte und optimale Platzierung der Schallquelle den Unterschied zwischen töten und getötet werden bedeuten kann. Eine derartig genaue Platzierung ist mit dem System des Virtual SS in der Form nicht möglich. Dafür sind Virtual SS jedoch in der Lage, einen kraftvolleren und dynamischeren Sound wiederzugeben, da deren eingebauten Lautsprecher im Vergleich größer sind (Scheeden 2010).



Abb. 5.1: 7.1 True Surround Sound Headset mit mehreren Treibern Treiber Ø: front 30mm, center 30mm, rear 23mm, side 27mm, sub 40mm

Quelle: G.SKILL 2016



Abb. 5.2: 7.1 Virtual Surround Sound Headset mit einem Treiber Treiber Ø: 40mm Quelle: Gearbeast 2017

5.3.2. 3D AUDIO

Im Vergleich zu den kanalbasierten Stereo und Surround-Varianten, sind 3D Audio Systeme objektbasiert und haben keine feste Anzahl an Kanälen, sodass die jeweiligen Klangereignisse beliebig im dreidimensionalen Raum platziert werden können. Das bedeutet, dass die Kopfhörer das Klangereignis genau dort wiedergeben können, wo sich die Schallquelle auch tatsächlich im Raum bzw. im virtuellen Raum befindet und muss wie bei den vorherig genannten Systemen, keine Näherung der Schallquelle aus zwei Kanälen bestimmen, was zu Ungenauigkeiten der Schallquellenposition führt. Das bietet schon den Vorteil, dass man gestalterisch nicht mehr ausschließlich auf die horizontalen Ebene begrenzt ist. Die für 3D Wiedergabe vorgesehenen Kopfhörer beinhalten ebenfalls mehrere Lautsprecher pro Kopfhörermuschel und zudem ein Head-Tracking-System (oft integriert) um die Kopfposition in Echtzeit zu aktualisieren und die Positionen der Schallquellen dementsprechend neu zu berechnen und anzupassen. Die Firma Audeze bedient sich bei ihrem Mobius Modell zudem eines integrierten HRTF Kalibrierungssensors, der die spezifischen HRTFs eines Hörers misst, sobald die Kopfhörer aufgesetzt werden um so ein auf den Hörer bestmöglich abgestimmtes dreidimensionales Klangbild zu berechnen (Audeze 2017).

KAPITEL 6 PSYCHOLOGISCHE PROZESSE

Es ist wichtig zu wissen, wie der Mensch seine Umwelt wahrnimmt und wie diese Wahrnehmungsmechanismen funktionieren, damit man diese, als auch den Fokus später bei der Anwendung in einer virtuellen Umgebung gezielt beurteilen und beeinflussen kann. Die auditive Wahrnehmung, als auch weitere Sinne, werden während einer Interaktion mit der virtuellen Umgebung unterbewusst als auch bewusst stark beansprucht. Verschiedene Arten von Computerspielen benötigen demnach verschiedene Fähigkeiten und Fertigkeiten, die der Spieler mitbringen oder erlernen muss, um den Anforderungen eines Spiels gewachsen zu sein. Darunter fallen unter anderem das Reaktionsvermögen, räumliches Vorstellungsvermögen, die Kombinations- als auch Konzentrationsfähigkeit, Auge-Hand-Koordination, rechnerisches und strategisches Denken (Schönthaler 1998: 24).

6.1. WAHRNEHMUNG UND INFORMATIONSVERARBEITUNG

Unter dem Begriff der Wahrnehmung versteht sich die Fähigkeit, die Umwelt, aber auch die Geschehnisse innerhalb des eigenen Körpers über die Sinnesorgane zu erfassen, also die Aufnahme von Reizen, die im weiteren Verlauf durch das Gehirn analysiert, dekodiert und interpretiert werden. Das macht es dem Menschen möglich, sich in der Umwelt behaupten und anpassen zu können.

Der visuelle Sinn stellt nach wie vor die wichtigste Informationsquelle bei der Wahrnehmung von virtuellen Welten dar, wodurch die auditive Gestaltung dieser lange als zweitrangig betrachtet wurde. Durch die Notwendigkeit dem Nutzer das Gefühl der vollständigen Immersion in die virtuelle Welt zu ermöglichen, muss die auditive Wahrnehmung gleichermaßen miteinbezogen werden.

Die meisten Reize die wir aufnehmen, als auch die Funktionen und Abläufe des Körpers laufen unterbewusst ab und werden demnach unterbewusst verarbeitet. Kann beispielsweise ein eintreffendes Schallereignis im dreidimensionalen Raum nicht auf Anhieb lokalisiert werden, wird of Position des Kopfes unbewusst angepasst um einen besseren Richtungseindruck der eintreffenden Schallquelle zu erhalten. Das Bewusstsein integriert die Signale der verschiedenen Sinnesreize, transferiert diese in Langzeitspeichersystem, welche auf weiterem Weg kognitive "Filterprozesse" durchlaufen

(Maderthaner 2017: 101f). Als mitunter die wichtigsten Funktionen des Bewusstseins beschreiben Mandler (1979: 78)

"[...] die Prüfung potenzieller Handlungsmöglichkeiten und die Bewertung der situativen Gegebenheiten [...]".

Nach Solso (2005: 150) ist das Bewusstsein

"[...] der hauptsächliche Prozess [...], mit dessen Hilfe sich das Nervensystem an neuartige, herausfordernde und informative Ereignisse der Welt anpasst" [...].

Befindet sich ein Anwender in einer für ihn vollkommen neuen virtuellen Umgebung, kann er nach einer bestimmten Eingewöhnungsphase, Anhand des für das Spiel spezifischen akustischen Designs, sprich die Codes und Muster, differenzieren, welche Klänge und Geräusche für bestimmte Ereignisse angelegt wurden und kann dementsprechend eine Handlung kalkulieren und ausführen. Diese stattfindenden Prozesse lassen sich nach Freud (1923: 12) in drei Kategorien aufteilen: bewusste, vorbewusste und unbewusste Vorgänge, die er als topografisches Modell bezeichnete.

Diese Prozesse umfassen alle unterbewussten Inhalte und Abläufe, die dem Bewusstsein nicht direkt zugänglich sind, wie beispielsweise der automatisierte Vorgang der Orientierung des Körpers im Raum oder auch den unbewussten Einfluss von auditiven Reizen auf die Psyche und den Körper.

Vorbewusste Prozesse beschreiben Vorgänge, die zu einem bestimmten Zeitpunk nicht bewusst sind und können im Gegensatz zu den unbewussten Vorgängen durch aktives Suchen wieder ins Gedächtnis gerufen werden. Die vorbewussten Prozesse umfassen beispielsweise Erinnerungen von Wahrnehmungen, Erlebnisse und Personen (vgl. Stangl 2018a). Der Spieler lernt durch vorbewusste Vorgänge während des Spiels automatisch die Codes oder Muster der visuellen und auditiven Gestaltung, wie ein Ton oder Geräusch, das zu einem spezifischen Ergebnis führt, und kann diese ohne nachzudenken anwenden oder aktiv ins Gedächtnis rufen, falls sich diese "Automation" noch nicht verfestigt hat.

Bewusste Prozesse umfassen Wahrnehmungen, Vorstellungen und Gedanken; also alle Zustände, die vom jeweiligen Menschen im Moment erlebt werden. Hierbei liegt es am Menschen selber, seinen aktiven Fokus auf eine bestimmte Sache zu legen (vgl. Stangl 2018a). Ein aktiver (Filter-)Prozess der bewussten Wahrnehmung ist die Aufmerksamkeit. Diese ist neben der Wahrnehmung an weiteren psychischen Prozes-

sen wie dem Denken, Emotionen, Motivation, Entscheidungen oder auch Handlungen beteiligt (Roll 2017: 25).

Ekman (2009) definierte vier Bereiche, die einen großen Einfluss auf die auditive Wahrnehmung ausüben: Aufmerksamkeit, Emotionen, Multimodalität und Propriozeption.

6.1.1. AUFMERKSAMKEIT

Die Aufmerksamkeit stellt einen grundlegenden Faktoren im Hinblick auf die Klanggestaltung einer virtuellen Welt dar. Sie beschreibt die selektive Ausrichtung des Wahrnehmens, Denkens und Vorstellens und ist auf bestimmte Erlebnisinhalte gerichtet. Objekte, deren aufgrund eigener Ziele, Achtung geschenkt wird, ist als aktive oder auch zielgesteuerte Aufmerksamkeit bekannt. Man kann diese als eine Art Scheinwerfer betrachten, der wichtige oder gewünschte Informationen beleuchtet, jedoch Reize die außerhalb des Kegels liegen, vorerst nicht weiterverarbeitet oder bewusst ausblendet (Rauh et al. 201: 8); (s. Kap.: 2.3.4.). So wird beispielsweise bei der Verteidigung einer Flagge oder eines Platzes innerhalb eines *First-Person-Shooters* ³, die auditive Aufmerksamkeit gezielt auf nähernde Schritte oder motorisierte Fahrzeuge gelegt, um dessen Richtung und Distanz ggf. rechtzeitig erfassen und dementsprechend handeln zu können.

Es ist also die bewusste und eigenmächtige Lenkung der Aufmerksamkeit durch Interessen, Vorstellungen, Bedürfnissen und Gedanken. Dagegen versteht sich die passive oder reiz-induzierte Aufmerksamkeit als Reize aus der Umgebung, die die Aufmerksamkeit durch deren Auffälligkeit oder Unbekanntheit, unabhängig von den Zielen des Wahrnehmenden auf sich ziehen (Gerrig et al. 2008: 141).

Es gibt hierzu Verarbeitungsprozesse, auf deren Weise ein Reiz wirksam wird. Diese werden als *Top-Down* und *Bottom-Up* bezeichnet. Die Bottom-Up Verarbeitung ist ein reiz-gesteuerter Vorgang; eine datengeleitete Informationsaufnahme. Sie beschreibt die Aufnahme von äußeren Sinnesreizen in das Wahrnehmungssystem, dessen Weiterleitung und die Gewinnung von Informationen. So löst bspw. ein niederfrequenter Ton (äußerer Reiz) in einem Horrorspiel negative Emotionen wie Nervosität oder Angst aus. Das führt wiederum zu einer zu einer Erkenntnis (gefährliche Situation) und Handlungsanweisung (Flucht; Verteidigung) des Gehirns. Demnach werden die Informationen von außen nach innen prozessiert (Goldstein 2007: 7-12).

Auch als Ego-Shooter bezeichnet wird für die Kategorie der Computerspiele verwendet, in der der Spieler aus der Egoperspektive des Spielcharakters heras agiert

Der Top-Down Prozess wird auch als wissensbasierte Verarbeitung bezeichnet, da er die Wahrnehmung durch kognitive Prozesse beeinflusst. Die Interpretation der reinkommenden Reize wird vom im Gedächtnis zur Verfügung stehenden Wissen beeinflusst, wie der Lernerfahrung (Motive, Erwartungen, Kontext), willentliche Handlungen (Entscheidungen), aber auch Emotionen (Huestegge 2018). Die Top-Down Verarbeitung ist während des gesamten Wahrnehmungsprozesses aktiv, denn nur dadurch kann die Aufnahmebereitschaft einer Person gegenüber bestimmten Reizen durch etwa Erwartungen, Wissen und Emotionen beeinflusst werden, was demnach zu einer selektiven Wahrnehmung führt (Gerrig et al. 2008: 141ff), (s. Abb.: 6.1).

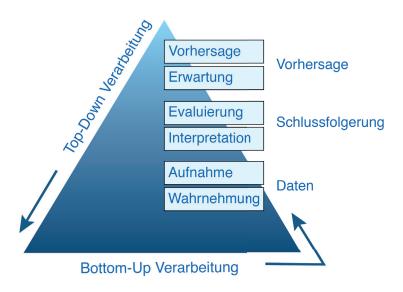


Abb. 6.1: Darstellung der Top-Down und Bottom-Up Verarbeitung Anlehnung an Hunt 2017

Die Aufmerksamkeitsprozesse können jedoch auch durch eine hierarchische Aufteilung einer Klangszene beeinflusst werden, in der die Prozesse der Aufmerksamkeitsauswahl Schrittweise den Zugang zu komplexeren Details eröffnen, während der Wahrnehmende die Hierarchie durchläuft. So fällt beispielsweise die Konzentration auf eine bestimmten Klangquelle in der Umgebung (Bsp.: Musik wird gespielt); dies eröffnet den Fokus für mehr Details, wie ein spezifisches Instrument (Klavier). Und wenn die Aufmerksamkeit auf jenes Instrument fixiert ist, werden einzelne Details (Klang, Anschlag) erst bewusst (Ekman 2009: 3f).

6.1.2. EMOTIONEN

Emotionen organisieren und aktiveren die Handlung eines Spielers und haben dementsprechend eine große Auswirkung auf Weise, wie die virtuelle Umgebung wahrgenommen wird, die Einstellung gegenüber dem Spiel selbst, als auch wird Einfluss auf die Handlungsgeschwindigkeit ausgeübt (Isbister et al. 2001: 188).

Emotionen und Aufmerksamkeit sind eng miteinander verbunden, sodass Emotionszustände sowohl den Inhalt des Bewusstseins, als auch die Leistung, sprich darunter auch die Aufmerksamkeit bzw. den Fokus beeinflussen (Faucher et al. 2002). So wird die sensorische Reizverarbeitung nicht nur durch die Bottom-Up oder Top-Down Prozesse (s. Kap. 6.1.1.) bestimmt, sondern es konnte gezeigt werden, dass sich die Stimmung auf die Art und Weise, in der die eingehenden Information erfasst und verarbeitet werden, auswirkt. Der emotionale Zustand eines Spielers innerhalb einer virtuellen Interaktion, ist mir der Präsenz in einer virtuellen Umgebung verbunden und kann demnach einen enormen Einfluss auf die Fokusführung dessen und weiterführenden Handlungen haben (Vanlessen et al. 2012).

Emotionen während einer Interaktion in einer virtuellen Welt hervorzurufen und den Spieler somit auch auf dieser Ebene mit einzubinden, ist einer der wichtigsten Elemente der Spielentwicklung und -gestaltung. Emotionen binden den Anwender an das Erlebte und sind ein Beispiel für einen spontanen Entscheidungsprozess, der fokusführend fungiert. Gerade im Bereich der Tongestaltung kann dies ein mächtiges Werkzeug sein, denn das Gehör lässt sich nicht bzw. nicht so schnell verschließen, wie das Sichtfeld.

Forschungsergebnisse legen nahe, dass die menschliche Wahrnehmung von Emotionen ihren Höhepunkt erreicht, wenn sie Schallwellenreizen ausgesetzt ist. Demnach haben Klänge nicht nur das Potential die Wahrnehmung einer visuellen Szene zu beeinflussen, sondern auch die Immersion innerhalb einer virtueller Umgebung, die Tiefe als auch einen emotionalen Eindruck zu erzeugen (Grimshaw et al. 2015: 78f). So kann sich ein starker emotionaler Zustand des Spielers auf dessen Wahrnehmung auswirken, indem es das Gedächtnis beeinflusst, welches sich wiederum auf die Assoziationen und Interpretationen der vom Spieler aufgenommenen auditiven Reize auswirken kann. Emotionen können ihren Ursprung in vielerlei Situationen und Reize haben und durch äußere (Lob, Kränkung), als auch innere (Gedanken, stimmungsverändernde Substanzen) Umstände, hervorgerufen werden (Ekman 2009: 3).

6.1.3. MULTIMODALITÄT

Die multimodale oder auch multisensorische Wahrnehmung beschreibt den Einfluss von Informationen weiterer Sinne auf die Interpretation eines Klangs im Bezug zu einem bestimmten Ereignis. Im Zusammenhang mit Virtual Reality ist es demnach vor allem der visuelle, als auch haptische Sinn, die ihren Einfluss auf die Informationsinterpretation der auditiven Wahrnehmung ausüben.

6.1.4. PROPRIOZEPTION

Propriozeption bezieht sich auf das Wahrnehmen von Klängen die innerhalb des menschlichen Körpers entstehen, oder welche durch den Körper hervorgerufene wie Herzklopfen, Kauen oder Schlucken (Grimshaw et al. 2015: 80).

KAPITEL 7 AUDITIVE WAHRNEHMUNGS- UND FOKUSFÜHRUNG

"Das Auge führt den Menschen in die Welt, das Ohr führt die Welt in den Menschen ein." (Lorenz Oken, 1779 – 1851)

In den vorangegangenen Kapiteln wurden wichtige Merkmale der Funktion des Gehörs, als auch der Psychoakustik des Menschen erläutert, die bei der Manipulation von Wahrnehmung und Fokus für Klänge als auch der auditiven Informationsaufnahme und -verarbeitung in einer virtuellen Umgebung, tragende Elemente bilden. Welche Faktoren gibt es also unabhängig der verwendeten Wiedergabeverfahren, um den Fokus eines Spielers in einer virtuellen Umgebung auf eine bestimmte Sache oder Situation zu führen, mit nahezu unendlichen Möglichkeiten?

7.1. KONTRAST

Kontraste zwischen Objekten und/oder Ereignissen in einer virtuellen Umgebung zu erschaffen, um damit einen Fokuspunkt zu kreieren bzw. deren Bedeutsamkeit herauszuheben, ist ein elementares Hilfsmittel. Das Gehirn kann sich nicht gleichzeitig zwei oder mehreren neuen komplexen Aufgaben widmen, da das Bewusstsein zu jedem Zeitpunkt ausschließlich einen Inhalt bearbeiten kann; es springt sozusagen zwischen diesen Aufgaben hin und her. Wenn gewisse Aufgaben schon mehrmals ausgeübt wurden und einen flachen Komplexitätsgrad haben, ist die Fähigkeit des Multitasking begrenzt möglich (vgl. Stangl 2018b). Deshalb muss der Informationsfluss von vorn herein reguliert werden bzw. muss der Fokus auf bestimmte Ereignisse geführt werden. Je größer der Kontrast ist, desto höher fällt Aufmerksamkeitsfokus aus und das Gehirn wird sich der wichtigeren Information widmen.

Um einen Kontrast mithilfe der Tongestaltung zu kreieren, gibt es einige Möglichkeiten, die folgend beschrieben werden.

7.1.1. LAUTSTÄRKE

Der einfachste Weg einen auditiven Kontrast zu erzielen ist mittels einer Amplitudenveränderung. Die Empfindung von Lautstärke folgt dem Energiegehalt eines Signals, denn ein schmales und impulsartiges Signal wird als leiser empfunden als eines das brei-

ter ist, jedoch die selbe Amplitude aufweist (Görne 2017: 57). So kann die Aufmerksamkeit des Spielers durch das Hervorheben eines Objektes oder einer Situation durch eine deutliche Lautstärkenvariation zum Umgebungsklang gezielt in dessen Richtung beeinflusst werden. Beispielsweise durch das starke Zufallen einer Tür, eine Explosion bzw. einen Knall, wird ein Kontrast zur umliegenden Klangumgebung hergestellt, wodurch eine Richtungslokalisation der Schallquelle erfolgen kann. Weiter kann ein steigender Lautstärkenverlauf als Indikator für ein sich näherndes Objekt oder Ereignis eingesetzt werden. Große Objekte werden häufig lauter eingeschätzt, als kleine (Görne 2011: 123); so kann eine steigende Lautstärke zur Annahme eines schnell nähernden oder auch großen Objektes führen, was im Aufmerken des Spielers resultiert.

Plötzlich auftretende Klangereignisse können ebenfalls in einer Orientierungsreaktion resultieren, da durch ihre kurze Zeit einen hohen Kontrast zur Umgebung schaffen. Ebenso wie neue oder komplexe Reize, sich widersprechende oder Signalreize, wodurch eine gesteigerte Aufmerksamkeit hervorrufen wird. Das führt unter anderem zu einer erhöhten Aufnahmebereitschaft für auditive, als auch visuelle Reize. Solche impulsartigen Ereignisse können unter anderem eine Abwehr- oder Panikreaktion hervorrufen (Görne 2017: 48), was den Spieler auf der emotionalen und psychischen Ebene in das virtuelle Geschehen miteinbezieht und dadurch den Fokus führt, als auch dessen handlungsorientierte Entscheidungen beeinflussen oder beschleunigen kann.

Die Lautheit kann auf eine bspw. (große) Wucht oder Stärke eins Objekts oder Ereignisses hinweisen, was auf eine drohende Gefahr schließen lassen kann. Aufgrund der menschlichen Evolutionsgeschichte löst dies eine Reihe von wehrhaften physiologischen, kognitiven, emotionalen, als auch verhaltensbezogenen Reaktionen aus, was in dieser Form im Hinblick auf sich bewegende Objekte in anderweitige Richtungen, nicht beschreiben lässt. Somit kann die Aufmerksamkeit gezielt auf dieses Ereignis geführt werden (Neuhoff 2016: 1).

STILLE ALS KONTRAST

Um die andere Seite der Lautstärkenanwendung miteinzubeziehen, lässt sich auch durch Stille ein hoher Kontrast erzeugen. Nach Balázs (1930: 159) hat Stille nur dort eine Bedeutung, wo es auch laut sein könnte. In den meisten Fällen wird die Stille als spannungsaufbauendes Gestaltungsmittel eingesetzt, wie z.B. im VR Survival-Horror ⁴-Game The Doors of Silence - The Prologue (Balckbite Interactive 2017). Der Spieler läuft durch einen Flur bis zu einem dunklen, bruchreifen Zimmer, in dem eine personenähnliche Puppe vor einem Monitor sitzt. Plötzlich durchbricht ein Rauschen die Stille. Der Monitor beginnt in Raum, worin keine anderen Geräusche zu hören sind abgesehen von den eigenen Schritten, an, ein Signal abzugeben, so-

dass der Fokus des Spielers instant darauf gezogen wird. Dieser Kontrast dient dazu, den Spieler an den Tisch zu führen um einen wichtigen Gegenstand einzusammeln.

Demzufolge wird sich jedes lautere Klangereignis aus diesem Stilleteppich der umgebenden klanglichen Atmosphäre herausheben und eine automatische Richtungsführung hervorrufen.

7.1.2. TONHÖHE

Die Tonhöhe oder auch Pitch, wird mitunter durch die Frequenz bestimmt, durch deren Variation ebenfalls ein klanglicher Kontrast geschaffen werden kann. Die Frequenzen werden in drei Gruppen geteilt: *low* (bass; dt.: tief/ Tiefen), *midrange* (dt.: Mitten) und *high* (treble; dt.: hoch/Höhen). Bass-kräftige Klänge werden mit Macht, Stärke und Größe verbunden und demnach als bedeutungsvoll, kompliziert oder auch schwer assoziiert. So wird ein niederfrequentes Grölen (z.B. eines Monsters), oder schwere sich nähernde Schritte eher in den Aufmerksamkeitsfokus rücken und als Gefahr interpretiert, als im Vergleich zu einer leichten Version, sollten diese gemeinsam auftreten.

Frequenzen im Mittenbereich geben den Klängen ihre Energie. Das menschliche Gehör ist für die mittleren Frequenzen von etwa 1-5 kHz am empfindlichsten; somit kann hier dem Hörer ein Maximum an Informationen vermittelt werden (Görne 2011: 115). Auch die Sprache ist in etwa dem selben Frequenzrahmen verankert.

Hochfrequente Klänge sind ein starkes Hilfsmittel um einen Kontrast zu erzielen. Sie ermöglichen uns Klänge und Geräusche zum einen klar wahrzunehmen (Carlson 2018), geben ihnen zum anderen aber auch eine Form von Präsenz. Sie können beispielsweise als akustische Führung und Informationsgebung eingesetzt werden, wie das gerne für Waffenfadenkreuze in futuristischen Science Fiction Spielen (Bsp.: *Archangel; Skydance Interactive* 2017) verwendet wird. Befindet sich ein Ziel im Fadenkreuz und wird fixiert, ertönt oft ein hoher telemetrischer Klang der den Spieler wissen lässt, dass das Ziel für einen Augenblick fest im Visier ist, woraufhin es zu einer weiterführenden Handlung kommt. Zum anderen können diese hochfrequenten Töne und Klänge auch als warnendes Signal eingesetzt werden, wie das bei Sirenen der Fall ist, die ob über lang oder kurz gehalten, aus der restlichen Klangwelt herausstechen und dem Spieler eine Gefahr signalisieren.

"Klangobjekte, die einen gesellschaftlich definierten kommunika-

Zählt zu einem "Computerspiel-genre, bei dem der Spieler im Rahmen einer Horror-Geschichte Angriffe von unheimlichen Kreaturen überleben und Rätsel lösen muss." Wikipedia 2017b

tiven Gehalt haben, heißen umgangssprachlich 'Signale'. Der kommunikative Gehalt von Signalen ist im Allgemeinen eine Handlungsaufforderung, die häufig mit einem Warnhinweis verbunden ist. Klanglich haben die meisten Signale eine sehr einfache Gestalt. Meist bestehen sie nur aus einem oder zwei Sinustönen [...] oder sie sind harmonisch strukturiert." (Flückiger 2010: 159).

Aber auch durch den gezielten Einsatz von Filtern kann man eine indirekte Fokusführung erreichen. Durch bspw. der Anwendung eines Tiefpassfilters auf alle Umgebungsgeräusche, im Falle einer Verletzung des Spielcharakters um den kritischen Zustand dessen zu verdeutlichen. Dies führt dazu, dass sich der Spieler auf die schnelle Heilung seines Avatars konzentrieren muss um weiterspielen zu können (Ekman 2014: 6).

7.2. FREQUENZEN

7.2.1. FREQUENZSELEKTION

Eine weitere Möglichkeit den Fokus zu führen, ist mittels einer Frequenzselektion. Studien von Slee et al. (2015), Dick et al. (2017) und Kaya et al.
(2017) zeigten, das sich der auditive Kortex auf bestimmte Frequenzen "einpegelt" bzw. sich adaptiert, wenn diese vom Gehirn als wichtig erachtet werden,
und legt dementsprechend den Aufmerksamkeitsfokus gezielt auf die jeweiligen Frequenzen und maskiert diejenigen, die außerhalb der Zielfrequenz liegen.

Ertönt ein bestimmter Ton für eine im Spiel wichtige Aktion, passt sich das Gehör der Frequenz dieses Tones sozusagen an. Auf der neuronalen Ebene wird die Frequenz des Ton dann mit einer wichtigen Information assoziiert. In Farpoint (Sony Interactive Entertainment 2017) wurde ein spezifischer Ton für den Alarm für das Überhitzen einer Waffe verwendet. Sobald die Überhitzung eintritt, bekommt der Spieler immer den selben Ton zu hören, was ihm signalisiert, dass er die Waffe abkühlen lassen, bzw. diese durch eine andere ersetzen muss, sodass er sich weiterhin verteidigen kann, da ansonsten im schlechtesten Fall der Tod durch den Gegner eintritt. Um auf diese Information automatisch und unterbewusst reagieren zu können muss das jeweilige Muster jedoch erst erlernt werden. Das menschliche Gehirn strebt stets nach Kohärenz, Ordnung und Struktur; einem Muster nachdem es Voraussagen auf Ereignisse treffen kann. Ein Beispiel für ein erlerntes Muster ist ein Gewitter; wir hören Donner und schließen daraus, dass es bald anfangen wird zu regnen. Je öfter dem Spieler ein bestimmtes Signal/ein Ton oder Geräusch in Verbindung

mit einer bestimmten Aktion bewusst gemacht wird, desto schneller prägt sich diese Verbindung ein. Das neuronale System verknüpft wiederholte Reize mit einer bestimmten Reaktion; und umso schneller fällt die Handlungsreaktion auf das jeweilige Signal bzw. Situation aus. Das Finden und die Identifikation von Mustern dient dem Überleben, es ist evolutionsbedingt verankert (Beitman 2009: 256). Das Heraushören eines Raubtiers im Wald bricht das Muster aus der klanglich homogenen Waldszenerie und resultiert in der Zunahme des Aufmerksamkeitsfokus auf die Bedrohung, und demnach dem eigenen Schutz.

7.2.2. FREQUENZMASKIERUNG

Bei der Frequenz- oder Spektralmaskierung handelt es sich um die Verdeckung spektraler Komponenten ähnlicher bzw. gleicher Frequenzen. Demnach kann beim Erklingen zweier Töne ähnlicher Frequenzen, jedoch mit unterschiedlicher Lautstärke, der leisere Ton nur wahrgenommen werden, wenn sich seine Frequenz ausreichend von der des lauteren Tons unterscheidet.

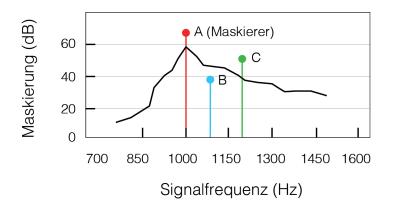


Abb. 7.1.: Schematische Darstellung spektraler Maskierung Anlehnung an Seeber 2006

Abbildung 7.1. zeigt vereinfacht den Maskierungseffekt von einem Ton A auf Ton B. Ton B befindet sich noch unter der Hörschwelle, oder auch Mithörschwelle (schwarz) von Ton A und wird somit verdeckt, er ist also nicht hörbar oder nur sehr schlecht hörbar. Ton C befindet sich ebenfalls im selben Frequenzbereich, hat jedoch einen höheren Lautstärkepegel als die Hörschwelle und kann somit wahrgenommen werden.

Die Maskierung betrifft demnach auch umliegende Frequenzen, die sich unter der Schwelle der Hörbarkeit befinden. In Abbildung 7.2. ist deutlich zu sehen, in wie weit die Maskierung bestimmter Frequenzen durch vorangegangene Frequenzen mit unterschiedlichen Lautstärkepegeln auswirkt.

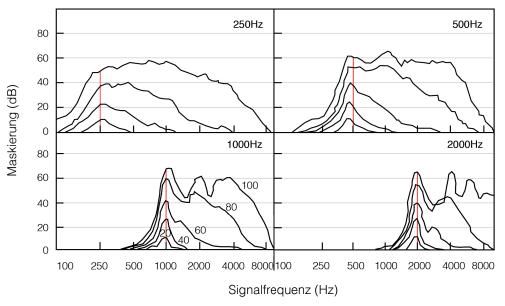


Abb. 7.2.:Vereinfachte Darstellung von Maskierungen verschiedener Frequenzen Anlehnung an Ehmer (1959: 1117)

So kann ein 250 Hz Ton von 75 dB, Frequenzen bis zu 4 kHz maskieren. Die Maskierungsbereich ist von der Frequenz der Töne anhängig. Hohe Töne haben einen größeren Maskierungsbereich, der über 4 kHz betragen kann, bei tieferen Tönen liegt dieser dagegen bei unter 100 Hz (Lipinski 2017).

Durch eine Maskierung kann der Mix eines Computerspiels undefiniert klingen und einzelne wichtige Elemente im Spiel unhörbar werden, wodurch der Fokus auf bestimmte Elemente beeinträchtigt werden können. Um dies zu umgehen gibt es einige Möglichkeiten: bei einer schon vorhandenen Maskierung von zwei Elementen, kann eines der beiden Elementen in eine Ton- und damit eine andere Frequenzlage transportieren, wenn es dem Sinn und der Gestaltung des Elements nicht negativ zugute kommt. Der Einsatz eines Equalizers (EQ) kann ebenfalls Abhilfe schaffen, indem man beim "wichtigeren" Objekt die gewünschten Frequenzen anhebt, beim jeweils anderen jedoch genau diese absenkt. Durch das Panning beide Elemente im Surround- oder 3D-Raum lässt sich eine Maskierung wirksam umgehen. Ein weiteres Problem können zu langen Hallfahnen darstellen. Einzeln mögen sie gut klingen, jedoch können sie mit darauffolgenden Töne interferieren, was die klangliche Separation der Klangelemente erschwert (Weiss, 2013).

7.3. EMOTIONEN

Virtual Reality stützt sich stark auf die richtige Auswahl spezifischer Wahrnehmungsmerkmale um Emotionen zu aktivieren. Demnach können positive Klangumgebungen zum psychologischen, als auch physiologischen Wohlbefinden des Menschen beitragen, während negative Klangumgebungen Stress, Angst oder Furcht erzeugen können, die ebenfalls physiologisch auf den Körper und damit die darauffolgende Handlung des Spielers auswirken (Chung et al. 2017: 1).

In vielen Spielgenres beruht ein emotionales Erlebnis oft auf dem Aufbau einer Spannung und negativen Emotionen während einer bestimmten Herausforderung wie einem Kampf, gefolgt durch positive Emotionen, wenn diese Herausforderung vorüber ist, bzw. man diese gemeistert hat. Diese Zustände können sich wie folgt, unterschiedlich auf den Aufmerksamkeitsfokus auswirken.

7.3.1. SOUND EFFEKTE

Durch Vanlessen et al. (2012) wurde nachgewiesen, dass sich der Aufmerksamkeitsbereich durch das Erzeugen von positiven Emotionen wie Freude, Spaß, Liebe, Interesse, Hoffnungen etc., erweitert. Dieser erweiterte Aufmerksamkeitszustand führt zu einer offenen, kreativen und flexiblen Denkweise. Es findet jedoch keine spezifisch unterbewusste Fokussierung auf ein Objekt oder Ereignis statt, wie das bei negativen Emotionen der Fall ist, da sich der Aufmerksamkeitsbereich unter Stress, Angst, Furcht etc. verkleinert und der Fokus auf einzelne Geschehnisse oder Objekte gelenkt wird, sodass eine schnelle Reaktion auf die Gefahr hin gewährleistet werden kann.

Das Erleben und Erfahren von negativen Emotionen wie Furcht, Angst, Traurigkeit, Stress oder auch Anspannung, können gezielt gefördert werden, indem man den Spieler durch *priming cues* ⁵ "vorbereitet". So aktiviert der Prime ein Assoziationsfeld, mit dem die darauffolgenden Information in Verbindung gebracht werden (Gramatke 2014). Innerhalb eines Survival Horror Spiels bspw., kann durch bestimmte Sound Effekte das ständige Denken und das Gefühl über die bedrohliche und unheimliche Situation des Spiels begünstigt werden. Nach diesem Priming, sprich der Vorbereitung, ist es wahrscheinlicher, dass der Spieler bestimmte Reize in der vom Spielentwickler gewünschten Weise beurteilt, wie die Angst, wenn im Spiel plötzlich eine bedrohende Kreatur durch ein dämonisches Grölen bemerkbar macht (Toprac et al. 2010).

Ist der Spieler einer direkten Gefahr ausgesetzt, wird die Aufmerksamkeit

Vorbereitug. Als Priming wird ein erster Reizn (Prime) durch die Sinne bezeichnet, der die Interpretation bzw. Reaktion auf nachfolgende Reize maßgeblich beeinflusst (Gramatke 2014)

unterbewusst auf das Gefahrenobjekt oder -ereignis fokussiert, defensive Systeme rufen eine schnelle reflexive Handlung hervor, wie die *Kampf-oder-Flucht-Reaktion* (Mobbs et al. 2015: 1). Kann die Gefahr jedoch noch nicht identifiziert werden, befindet sich der Organismus in einem alarmbereiten Zustand, sodass die Aufmerksamkeit auf nichts spezifisches gerichtet ist. Somit ist der Spieler stets wachsam und richtet den Fokus instant auf die Gefahr, sobald sich dessen Position erfassen lässt.

Untersuchungen durch Wolfson und Case (2000) zeigen, dass sich gewisse Klangereignisse auf den Körper auswirken, wie durch eine Erhöhung des Pulses, was sich wiederum im Aufmerksamkeitsfokus bemerkbar macht. Es wird angenommen, dass physiologische Veränderungen wie eine erhöhte Herzfrequenz unterbewusst mit Angst, Furcht oder an Anspannung verbunden werden (Toprac 2010).

STILLE ALS EMOTIONSTRÄGER

Eine weitere Möglichkeit der Fokusführung besteht im gezielten Einsatz von Stille. Die Stille an sich muss keine negativen Emotionen oder Gefühle hervorbringen, das hängt ganz davon ab, in welchem Genres diese eingesetzt wird. Meist findet sich diese jedoch in altbekannten Horror-Genres- Sie wirkt erdrückend, schafft Anspannung und ist ein Vorbote für das Unheimliche.

"Die bedrohliche Dimension der Stille resultiert aus der potentiellen Gefahr der Totenstille: Nur was tot ist, ist still." (Flückiger 2002: 233)

Hiermit wären wir wieder bei Anspannung, Angst und Furcht. Eine derartige Stille wird meist als Vorahnung eines negativ bevorstehenden Ereignisses interpretiert und somit ebenfalls mit vielerlei negativen Emotionen oder auch eigenen Erfahrungen verbunden. Wataru Hachisako, der Audio Director des VR Spiels *Resident Evil 7: Biohazard (Capcom 2017)* erläuterte in einem Interview wie wichtige es ist, Geräusche des täglichen Lebens, die man sonst unterbewusst wahrnimmt ins Spielt einzubauen, da es dem Spieler eine Art von Routine, Komfort und Sicherheit vermittelt um im Kontrast zu diesem die Stille einzusetzen:

"Es ist gut, das Gefühl des täglichen Lebens zu etablieren, um es dem Spieler zu ermöglichen, sich an Aspekte des Spiels zu gewöhnen [...]. Es hilft, das Eintauchen zu vertiefen und gibt Spielern eine erhöhte Sensibilität für alles, was von dieser Norm abweichen könnte. Dies ist eine perfekte Gelegenheit, die Stille zu nutzen und dem Spieler all die Sicherheit und den Komfort zu entziehen, die er hat." (frei übersetzt aus dem englischen; Wong, 2017).

7.3.2. MUSIK

Musik ist mit der Sprache das vielleicht am aussagekräftigste Hilfsmittel, das eine große Vielfalt an Emotionen hervorrufen kann. Sie wird dazu verwendet, die Umgebung oder auch ein Geschehen, eine Handlung auf der auditiven Ebene zu unterstützen. Musik ist ein bedeutender Faktor im Hinblick auf die Immersion mit virtuellen Welt und wird meistens ohne Gesang eingesetzt, um die kognitive Informationsaufnahme nicht zu überlasten und den Spieler nicht abzulenken (Ganiti et al. 2018: 18). Viele Studien legen nahe, dass das übliche Ziel der Musik, die Beeinflussung von Emotionen ist (Juslin et al. 2008: 559) und hilft zumindest im Bereich der Computerspiele die Handlungsperformance und den darauf bezogenen Aufmerksamkeitsfokus zu führen. Die Musik übt demnach einen wesentlichen Einfluss auf die ausgeführte Handlung aus. Sie wird nicht nur in Form von Hintergrund eingebunden, sondern auch interaktiv handlungsbedingt eingesetzt werden oder kann gar als zentrales Element eines Spiels dienen.

POSITIVE IMPLIKATIONEN DURCH MUSIK

Grundsätzlich ist man der Meinung, dass Musik, die den Spieler in ein positives Erlebnis taucht und in diesem dementsprechende Emotionen erweckt und vermittelt, sich positiv auf diesen auszuwirken kann, wie durch eine Reduktion des Aggressionsniveaus oder auch das Ablenken von Frustrationsgefühlen, wenn sich der Stressfaktor in Grenzen hält (Lawrence 2012: 18). Wie auch schon unter Kap. 7.3.1. erläutert, erweitert sich der Aufmerksamkeitsbereich unter positiven Emotionen.

Unter bspw. musikalisch harmonischen Strukturen mit einer geringen Taktzahl pro Minute, wie die des Herzschlages eines ruhenden Menschen von etwa 60 - 80 Schlägen pro Minute (Treasure 2011: 137), tendieren Spieler dazu, Handlungen langsamer und entspannter auszuführen. Musikalische Variablen beeinflussen also das menschliche Verhalten. Natürlich lässt sich auch positiver Stress durch ein schnelles Tempo und einem erhöhten Rhythmus erzeugen, was zu einer gesteigerten Fokussierung auf die zu bearbeitende Aufgabe führt, demnach aber auch zu einem engeren Aufmerksamkeitsbereich, und unter anderem mit einer gesteigerten Motivation einhergeht (BioRics 2018).

NEGATIVE IMPLIKATIONEN DURCH MUSIK

Eine stärkere Aufmerksamkeitswirkung zeigen jedoch durch Musik negativ induzierte Emotionen; diese Art von Musik wird in der Regel in Verbindung mit bevorstehenden Gefahrsituationen eingesetzt, um diese anzukündigen oder zu unterstützen. Ein Beispiel hierfür liefert das VR-Spiel Primordian (Stonepunk Studios 2018). Zu Beginn des Spiels erkundet man mit seinem Charakter einen Wald, der durch eine seichte, jedoch epische Musik begleitet wird, eine Kombination aus Streichern, weichen "Engles"-Chören und Synth-Pad-Sounds, dessen atmosphärischer Klang zum allgemeinen Erkunden der neuen Welt einlädt. Je näher man jedoch einem bestimmten Tempel kommt, desto mysteriöser und unheilvoller wird die begleitende Musik und bringt eine angespannte Stimmung mit sich. Im Tempel angekommen, verstärkt sich die erdrückende Stimmung durch die Musik, eine dunkle Stimme ruft aus der Richtung, in der sich zwei Waffen befinden. Bisher war die Musik sehr homogen und gleichmäßig. Direkt nach dem Griff zu den Waffen, erhält die Musik einen Energieaufschub, wodurch dem Spieler eine enorme Stärke seines Spielcharakters vermittelt wird. Einen Augenblick später erscheint der erste Gegner, womit instant große und dunkle Trommeln, weitere Streicher, Bläser und männliche Stimmen einsetzen; das Tempo wird angezogen. Der Spieler macht sich kampfbereit, der Aufmerksamkeitsfokus steigt und fixiert sich auf dem gefährlich schnell auf den Charakter zukommenden Objekt. Während solch einem Ereignis erfolgt eine Ausschüttung von Adrenalin und Cortisol, das wiederum, wie in Kap. 7.3.1 erläutert, in einer evolutionsbedingten Kampf-oder-Flucht-Reaktion und so auch mit erhöhter Aufmerksamkeit auf das Gefahrenobjekt um schneller auf diese reagieren zu können (Müsseler et al. 2017: 206).

ORIENTIERUNG MITHILFE VON MUSIK

Durch Musik kann aber auch eine akustische Landkarte erstellt werden. In dem für verschiedene Szenen/Gebiete eine für jene charakteristische Klanggestaltung erstellt wird, gibt man dem Spieler die Möglichkeit, sich neben der visuellen, auch der akustischen Merkmale der virtuellen Welt zum Zweck der Orientierung zu bedienen (Ardito et al. 2007: 202f). So kann der Fokus vom Spieler spezifisch der Landkarte nach erfolgen.

7.3.3. TEMPO

Das Tempo von akustischen Ereignissen nehmen ebenfalls Einfluss auf Emotionen, Aufmerksamkeit, als auch Handlung. Wenn wir Sound Effekte als Beispiel nehmen, geht von schnellen Tönen eine Dringlichkeit aus, eine Zeitknappheit. Ein Explosionsobjekt wird oft als langsamer Ton, der, je näher das Objekt der Explosion kommt, exponentiell schneller wird. Die Interpretationen eines positiven oder negativen Eindrucks ist jedoch stark vom Genre des Spiels Abhängig. Nichtsdestotrotz, führen sie den Aufmerksamkeitsfokus durch ihre Auffälligkeit und unterbewusste Gedrungenheit auf das jeweilige Objekt oder Geschehen. Selbiges lässt sich auf Musik anwenden. Ein Musikstück mit einem erhöhten Tempo ruft Emotionen mit hoher Energie wie Wut oder Freude auf. Das wiederum, wie schon in voran gegangenen Kapiteln beschrieben, wirkt sich auf den Aufmerksamkeitsfokus aus, in dem der Aufmerksamkeitsbereich kleiner und fokussierter wird, wodurch die Handlung tendenziell schneller und reflexhafter erfolgt (Lawrence 2012: 18).

7.3.4. STIMME

Wie schon erwähnt, gehört die Stimme mit der Musik zu einem der stärksten Auslöser für Emotionen, da sie diese, als auch weitere persönliche innere Zustände nach außen trägt. Sie kann auf verschiedenen Ebenen wie der physiologischen, der artikulatorischen, als auch akustischen, zum Ausdruck gebracht werden (Johar 2016: 11). Die Stimme stellt damit den wichtigsten zwischenmenschlichen Informationsträger dar.

So erkennen wir sofort an der jeweiligen Sprechweise eines jeden, ob diese/r bspw. froh, traurig, ängstlich oder wütend ist. Für den VR Action Shooter ⁶ Robo Recall (Epic Games 2017) hat der Sound Ingenieur Aaron Brown einen kurzen Simulationstest durchgeführt, um das Sound Design des Roboters zu testen; ohne jegliche Musik, Atmosphäre oder Stimmen. Hierbei hat er den Roboter, der später im Spiel den Gegner darstellen wird, auseinander genommen. Sprich geschlagen, angeschossen, als auch dessen Extremitäten entrissen. Er hat dabei festgestellt, wie wenig Empathie man diesem Charakter entgegenbringt, wenn dieser keinerlei Gefühle, wie in diesem Fall, Schmerz, während der Prozedur ausdrückt, wie das in der finalen Version des Spiels der Fall ist. Man baut also keinerlei emotionale Verbindung zum Gegner, als auch der Situation auf, wenn dieser seine Emotionen nicht kommunizieren kann (Brown 2017). Das Tempo oder auch die Gedrungenheit einer Stimme kann wie bei der Musik, den Spieler ebenfalls zu einer schnelleren Handlungsabwicklung leiten.

STIMME ALS WEG- UND RICHTUNGSWEISER

Der Mensch reagiert sensitiv auf die menschliche Stimme im Vergleich zu nicht-menschlichen Geräuschen (Varvatsoulias 2014: 71). Das kann man sich dementsprechend

⁶ Gehört zu den First-Person-Shootern und bildet eine Unterkategorie

als fokusführendes Element zunutze machen. Eine vertraute Stimme in einem Gewirr von Geräuschen, Musik, als auch weiteren Stimmen, ist schneller auszumachen. So kann man für die für den Spielcharakter wichtige Stimmen weiterer Charaktere sensibilisiert werden, sodass deren akustische Informationen schneller erfasst werden können. Insbesondere zieht eine Stimme dann Aufmerksamkeit auf sich, wenn durch deren Lautstärke einen Kontrast zur umgebenden klanglichen Umwelt entsteht, als auch ein Signal damit gesetzt wird, wie ein Hilferuf, Schrei, oder Weinen. So wird die Stimme zum Wegweiser, wenn man den jeweiligen Charakter nicht sieht, sondern nur die akustische Information erhält. Je nachdem welche Emotionen diese Stimme auslöst, erfolgt eine darauf bezogene Handlung dementsprechend schneller (Wut, Angst) oder langsamer (Freude). Aber auch durch konkrete Ansagen der Richtung, wie das im VR Spiel Archangel (Skydance Interactive 2017) der Fall ist, wird dem Spieler die Richtung des Feindes mitgeteilt, sodass dieser instant reagieren kann.

STIMME ALS INFORMATIONSSTAUS

Um eine Information über ein bestimmtes Ereignis oder einen Status dem Spieler mitzuteilen, wird neben Sound Effekten auch die Stimme verwendet. Oft ist es unter einer Vielzahl von Sound Effekten, die dem Spieler direkte Informationen zur Handlung geben schwierig, noch weitere Sounds einzubinden, die sich, wie auch alle anderen, in irgend einer Form abheben sollen um dem Spieler eine wichtige Information mitzuteilen. Da ist es unter Umständen sinnvoll, eine gesprochene Sequenz einzubringen.

Vielerlei Spiele, gerade verschiedene Arten Shootern geben dem Spieler durch eine sprechende Stimme Informationen wie bspw. die erfolgreiche Vernichtung eines Zieles, das Nachladen einer Waffen, den Energiegehalt eines Schutzschildes oder auch ankommende Gefahren bzw. Beschuss. Dies gibt dem Spieler einen Anhaltspunkt oder auch einen direkten Anweisung für die folgende Handlung.

ZUSAMMENFASSUNG

Je nach Genre eines Spiels, haben die Komponenten auch eine andere Wirkungsweise und somit ein anderes Ziel im Hinblick auf den Spieler und das virtuelle Erlebnis. Alle und viele weitere Komponenten lassen sich miteinander kombinieren um einen hohen Fokus des Spielers zu ermöglichen. Die Aufmerksamkeit stützt sich jedoch nicht ausschließlich auf das Gehör, sondern wird in Verbindung mit weiteren Sinnen, wie primär dem Sehen, verstärkt. Bonneel et al. haben gezeigt, dass durch das Erhöhen der akustischen Details, die visuelle Detailgenauigkeit zunahm. Das bringt uns wieder zur multimodalen Wahrnehmung (s. Kap.: 6.1.3) und dem damit einher-

gehenden gegenseitigen Einfluss der Sinne zueinander. Andererseits konnte nachgewiesen werden, dass Klänge das Potential haben, einen gewissen Teil der Aufmerksamkeit des Spielers von visuellen Reizen abzulenken (Cowan et al. 2015: 1209).

7.4. ANWENDUNG AUF SURROUND SOUND UND 3D AUDIO BEI KOPFHÖRERWIEDERGABE

Die Wahl eines passendes Wiedergabesystems richtet sich stets nach den Anforderungen des Spiels und kann abhängig vom System, Auswirkungen auf die Wahrnehmung und die Qualität des Aufmerksamkeitsfokus mit sich bringen.

Für manche Spiele wäre es für die Spielweise mehr hinderlich als hilfreich das Audio in die Höhe und Tiefe durch 3D zu erweitern. Nehmen wir als Beispiel *Beat Saber (Hyperbolic Magnetism 2018)*; es gleicht der Spielserie *Guitar Hero (RedOctane 2006)*, in der der Spieler versucht Original-Musikstücke während dem Hören, authentisch mit den Gitarren-Controllern nachzuspielen. Im Spiel Beat Saber verfügt man über zwei farblich unterschiedliche Controller, die virtuell zwei Laserschwerter darstellen, mit denen zum Takt die farblich korrekten Blöcke erfasst werden müssen (s. Abb. 7.3.).



Abb. 7.3.: Auszug aus dem Spiel Beat Saber, Hyperbolic Magnetism 2018 Quelle: LIV 2018

Diese kommen auf einer horizontalen Ebene von vorne auf den Spieler zu, weshalb hier ein Sound Design im Surround mehr Sinn macht, als in 3D, was dem Spieler in ungünstigsten Fall den Fokus auf das Geschehen direkt vor diesem nehmen könnte.

Auf der anderen Seite macht es bei vielerlei Anwendungen in VR mittlerweile mehr Sinn und erhöht die Immersion, mit dem Einsatz von 3D Audio. Einen wichtigen Unterschied kann dies bspw. bei sogenannten *Open-World-Spielen* 7 machen. Die Spieler sind an keinen spezifischen Handlungsablauf gebunden, sondern können sich frei in der virtuellen Welt bewegen und selbst entscheiden, welche Handlung als nächs-

⁷ In Open-World-Spielen haben Spieler von Beginn an ungehinderte Bewegungsfreiheit und besitzen demnach die freie Wahl der zeitlichen und örtlichen Erkundung der virtuellen Welt.

te getätigt wird und in welche Richtung sich der Spielcharakter bewegen soll. Bei diesen Spielen ist eine Erweiterung durch 3D notwendig, um eine bestmögliche Immersion erzielen zu können, da die Immersion nicht nur das Gefühl einer physischen und räumlichen Umgebung (dem Gefühl, tatsächlich "da zu sein") beinhaltet, sondern auch eine persönliche Bewertung der Attraktivität und der Natürlichkeit bzw. Glaubwürdigkeit der virtuellen Umgebung, als auch deren Inhalt miteinbezieht (Baños et al. 2004). Demnach lässt sich sagen, je realer eine Umgebung dargestellt wird, desto höher die Immersion und desto höher fällt das emotionale Involvieren des Spieler aus.

Bei der Arbeit in Surround oder 3D gibt es einige wichtige Unterschiede, die man bei der Tongestaltung nicht vergessen darf. Die Gestaltung für die klassischen 5.1 und 7.1 Surround-Formate, die bei der Kopfhörerdarbietung Gang und Gebe sind, sind auf die horizontalen Ebene (2D) limitiert; das schränkt die akustische Gestaltung einer virtuellen Welt ein, sofern man einen hohen Grad an Natürlichkeit und Realismus erzielen möchte. 3D Audio wird mittlerweile vermehrt und mit Nachdruck im Bereich des VR angewendet. Da hierbei alle drei Lokalisationsebenen miteinbezogen werden wodurch ein höherer Realitätseindruck vermittelt werden kann. Nichtsdestotrotz lassen sich die gestalterischen Mittel für ein gutes Game Sound Design auf beide Systeme anwenden. Im Bezug auf einen Kontrast durch Lautstärke wird die Lautheit eines Ereignisses am besten auf der horizontalen Ebene von 90° rechts bzw. 270° links wahrgenommen, in der auch die Lokalisation von Schallereignissen am höchsten ist.

Wie gut so ein Kontrast erzielt, als auch der Fokus unter Surround geführt werden kann, zeigt der VR First-Person Shooter Battlefield 1 (EA DICE 2016). Man befindet sich im Szenario des ersten Weltkriegs. Alles was in diesem Spiel passiert, wurde konkret im Sound Design verankert und gestaffelt, es liefert einen präzisen Eindruck, aus welcher Richtung ein bestimmtes Schallereignis kommt. Das Surround-Panorama erlaubt das wandern von Klangereignissen, wenn der Spielcharakter bewegt wird. Auch wenn sich das Sound Design auf Surround bezieht und somit nur in der horizontalen Ebene agiert, fällt das im Spiel selbst nicht auf. Das Geschehen findet meist am Boden statt, sodass fast alle Geräusche ebenfalls auf der selben Höhe wie der Spielcharakter wahrzunehmen sind und sich der Spieler größtenteils auf diesen Bereich konzentriert. Einen deutlichen Lautstärkekontrast zur umliegenden klanglichen Umgebung kann man hier immer dann feststellen, wenn bspw. ein Schuss abgefeuert wird, sich andere Spieler oder Fahrzeuge nähern. Ebenfalls lässt sich aufgrund der verwendeten Frequenz und Lautstärken sehr gut unterscheiden, ob es sich um ein schweres oder kleines Geschütz handelt. Ist der Spielcharakter verletzt, dumpfen alle Geräusche bis auf die des Charakter ab, sodass eine Fokussierung auf den Charakter selbst stattfindet. Einen klanglichen Kontrast durch Lautstärke in Surround oder 3D lässt sich demnach ohne Probleme erzielen. Die Schwierigkeit liegt mehr darin, diesen Kontrast richtig zu setzen und damit den Fokus gezielt führen zu können. Durch die fehlende Möglichkeit Höhen und Tiefen im Surround korrekt abzubilden, kann es zu akustischen Ungenauigkeiten der Position eines Schallereignisses im Spiel selbst führen. Schallquellen die sich über oder unter dem Spielcharakter befinden, können nicht gezielt als solche dargestellt werden, sondern befinden sich stets auf oder etwas oberhalb/ unterhalb der horizontalen Ebene, die Position derer ist eher ungenau. Somit kann es zur Missinterpretation bzw. Misslokalisation bezüglich der Richtung und Distanz der Schallereignisse kommen. Finden Handlungen also auch ober- oder unterhalb des Spielcharakters statt, macht es Sinn 3D Audio zu verwenden, um das gesamte sphärische Klangereignis ausnutzen zu können, denn im Surround-Format können diese schlichtweg nicht korrekt abgebildet werden. Das Sound Design des Open-WorldVR Spiels Horizon Zero Dawn (Guerrilla Games 2017) bietet dem Spieler genau das; ein sphärisches und detailreiches und sich änderndes Klangerlebnis. Dadurch, dass in dieser virtuellen Welt Kreaturen sowohl hoch im Himmel, als auch unter der Erde aktiv sind, haben sich die Sound Designer die Technik des 3D Audio zunutze gemacht. Das ermöglicht dem Spieler die Dimensionen des Spiels besser zu erfassen.

"[...] Der Unterschied eines Erlebnisses mit einem Multi-Channel Set Up wie 5.1, 7.1 zu 3D Audio ist definitiv die Nutzung einer Höhen- und Tiefenebene. In einem flachen Surround-System (wie 5.1, 7.1) können die Höheninformationen nicht so klar und deutlich vermitteln, wie in einem 3D System. [...]" (Frei übersetzt aus dem englischen; PlayStation Europe 2017)

Durch diesen höheren Realitätsbezug erhöht sich wiederum die Immersion in die virtuelle Welt und präzisiert die Lokalisation von Klangobjekten. Dadurch, dass bei beiden Systemen alle hörbaren Frequenzen wiedergegeben werden können, kann sowohl im Surround, als auch in 3D ein hervorragender Kontrast durch die Anwendung von Tonhöhe erzielt werden, es dies nicht vom angewendeten System selbst abhängig ist. Einen weiteren Vorteil bietet 3D anhand der Möglichkeit, Audioobjekte frei im dreidimensionalen Raum platzieren zu können, als auch der erweiterten Dimension. Durch diese Aspekte gibt es mehr Platz und Möglichkeiten, Geräusche und Klänge sinnvoll im Raum zu verteilen, was zum einen die Gefahren einer Frequenzmaskierung minimiert und somit die Fokusführung hindern bzw. verfälschen, zum anderen auch den auditiven Raumeindruck verbessern kann.

Im Hinblick auf Emotionen haben Baños et al. (2004) gezeigt, dass diese eine entscheidende Rolle bzgl. des subjektiven Urteilsvermögens, den automatischen Reaktionen, das Beeinflussen des Lernvorgangs, als auch die Generelle Verständnis und Reaktion auf unsere Umgebung. Die Präsenz in einer virtuellen Umgebung ist also nicht nur vom Gefühl für das räumliche bzw. der Immersion; dem Gefühl sich "im Spiel" zu befinden abhängig, sondern auch der persönlichen und subjektiven Beurteilung und Erfahrung des Nutzers, als auch die Natürlichkeit bzw. der Realismus des Spiels. Je natürlicher und realistischer eine virtuelle Welt erschaffen werden kann, desto größer ist die emotionale Bindung des Spielers an diese. Emotionen lassen sich ohne zu zweifeln in beiden Systemen außerordentlich gut vermitteln, als auch im Spieler selbst erwecken, jedoch hat 3D in Sachen Natürlichkeit und Realismus einige geringfügige Vorteile. Bisher wurden bei typischen Spielen in 5.1 oder 7.1 die Vertonung von Kreaturen so vorgenommen, dass deren Geräusche sozusagen in der Mitte oder zwischen den Füßen der Kreatur platziert wurden. Sprich alle Geräusche, Schritte, Gröhlen, Klatschen kamen von einem Punkt, statt sie den spezifischen Stellen der tatsächlichen Schallquellen zuzuordnen. Im Surround macht das keinen sehr großen Unterschied, da alles auf einer Ebene im Stereo-Panorama verteilt wird. Mit der Verwendung von 3D Audio wird diese Verteilung jedoch möglich und deren Verwendung sinnvoll, sodass man die Klangereignisse den eindeutigen Stellen des Schallquellenursprungs zuweisen kann und sollte (Smurdon 2015).

KAPITEL 8 **5D AUDIO**

Durch die immer populärer werdende Technologie des Virtual Reality und des steigenden visuell realistischeren Inhalts, möchte man auch abgesehen vom Visuellen, das höchste Maß an Realität erzielen. Dazu wird auch im Bereich des Audio viel in Verbindung mit höheren Dimensionen experimentiert und gearbeitet. Lange Zeit war es der Standard, bspw. Musik oder auch Computerspiele in 2D anzulegen und zu erfahren. Mit dem Aufleben des VR und der Möglichkeiten sich frei im dreidimensionalen Raum zu bewegen, wollte man dieses Konzept auch erneut in der Audiobranche aufgreifen. Die Technik des 3D Audio ist seit Jahren auf dem Vormarsch und wird vermehrt auch für Computerspiele; vor allem im Bereich des VR eingesetzt.

8.1. DIE DIMENSIONEN

1D beschreibt die Länge (x-Achse) und kann als einfache Linie dargestellt werden. 2D ist die Breite (y-Achse) und eine Erweiterung zur Länge. Sie stellt eine Fläche dar und damit einen Teil eines dreidimensionalen Raumes. 3D erweitert wiederum durch ihre Höhe (z-Achse) die Länge und Breite und gibt allen Objekten ein Gefühl von Raum und eine Querschnitt. Es ist eine Momentaufnahme der vierten Dimension. 4D ist die Dimension der Zeit. Diese regelt die Eigenschaften aller bekannten Materie an einem bestimmten Zeitpunkt im Raum und ist ein kleiner Teil aller möglichen Resultate bzw. Ausgänge auf dem fünfdimensionalen Raum. 5D ist ein Konzept und beschreibt die Vorstellung von weiteren möglichen Welten, die sich etwas von unserer eigenen Welt unterscheidet. Es gäbe uns theoretisch die Möglichkeit, Unterschiede zwischen unserer Welt und den möglichen Parallelwelten. Die fünfte Dimension ist die Dimension der Möglichkeiten/Entscheidungen, in der wir Aufgrund unserer Entscheidungen jeden Tag existieren. Sie beinhaltet demnach alle möglichen Resultate für alle Ereignisse, die ein Spieler in einer virtuellen Welt treffen kann. 5D ist eine Momentaufnahme aller möglichen Resultate (Brown 2017), (Williams 2014).

8.2. DIE DIMENSIONEN IN DER TONGESTALTUNG

Wie lassen sich die Dimensionen auf die Tongestaltung anwenden?

3D Audio ist uns nun als Technologie und Gestaltungsmittel bekannt, mit der man einen dreidimensionalen Raumklang erzeugen kann, in dem man die gewünschten

Audioobjekte frei platzieren kann um uns ein Klangerlebnis zu bieten, wie wir es vom Natürlichen Hören her gewohnt sind; vorzugsweise wird dies über die Darstellung von Kopfhörern erreicht.

8.2.1. TONGESTALTUNG IN 4D

Paul Oomen et al. (2018) haben mit ihrem 4DSound Projekt klangliche Gestaltung um eine Dimension erweitert und mit eigens dafür konzipierten Programmen lässt sich jeder einzelne Sound steuern und beliebig im Raum und Zeit platzieren. Anwendung findet dieses System jedoch nicht über Kopfhörer, sondern über ein Setting von 57 Lautsprechern. Dabei wird eine Fläche von 16 m² mit 16 Säulen beschallt, in der sich 48 Satellitenlautsprecher, innerhalb einer Säule 3, befinden, als auch insgesamt 9 Tieftöner unterhalb der Tanzfläche. Der Sound dieses Systems ist vollkommen omnidirektional und erlaubt es, Phantomschallquellen unabhängig der Position des Zuhörers im Raum zu bewegen. Mit ihrem Setup erreichen die Entwickler, dass die Lautsprecher während der Wiedergabe nahezu verschwinden. Der Sound kann dem Zuhörer unendlich weit weg, oder sehr nahe kommen, es bewegt sich oberhalb, unterhalt, als auch um den Zuhörer herum. Man hört nicht mehr die Lautsprechern im Raum, sondern mehr den gesamten Raum selbst.

8.2.2. TONGESTALTUNG IN 5D

5D Audio ist dagegen mehr ein Konzept und wird, wie schon erwähnt, als Dimension der Wahl bezeichnet und wendet sich an die Komponenten der fünften Dimension, um die Faktoren dieser, die Entscheidung, zu beeinflussen. Brown (2017) wendet es im Bereich des Audio an indem er hervorhebt, dass gerade interaktive virtuelle Umgebungen dem Nutzer vielerlei Auswahl bietet und Entscheidungen abverlangt, ob bewusst oder unbewusst. Damit der Spieler die Reize versteht die man ihm aussetzt, muss man ihm eine grundlegende Erfahrung bieten. Brown erforschte, wie man die Spielerfahrung eines Nutzers durch eine höhere Dimension des Audio verbessern kann, woraus sich folgende Kriterien ergeben haben:

DAS BEEINFLUSSEN DER AUSWAHL/ENTSCHEIDUNG

Durch das Arbeiten mithilfe von Kontrast durch Lautstärke und Stille, sprich den Aufmerksamkeitsfokus auf eine bestimmte Auswahl lenken. Je höher der Kontrast, desto höher die Aufmerksamkeit. Weiteres durch Frequenzselektion- und Maskierung.

VERBESSERUNG DES MIXES

Eine Verbesserung des Mixes kann ebenfalls durch die Nutzung eines hohen Dynamikumfangs erreicht werden. Es sollte genug Headroom für wichtige Klänge und Geräusche gelassen werden. Dynamik, Frequenz und Stille können als Kontrast für bestimmte Schlüsselmomente verwendet werden, damit sich diese abheben, so geben sie dem Mix eine klare Struktur. Ereignisreiche Geräuschkulissen sollten sinnvoll gestaffelt und verteilt werden, sodass sich diese nicht gegenseitig Maskieren und unnötig verwirren. Mithilfe von selektiven Side-Chain Duckings, auf der Basis von Lautstärke als auch spezifischer Frequenzbänder kann ebenfalls zu der Ausgeglichenheit, Klarheit und Dynamik eines Mixes beitragen. Mithilfe von Frequenzanalysen lassen sich bspw. bei Stimmen, die für die Stimme spezifischen Frequenzen auslesen, sodass wenn diese aufspricht, anderweitige Geräuschkulissen in genau diesen Frequenzen leicht gesenkt werden können, sodass die Stimme nicht verdeckt wird und somit mehr Platz hat. Automatisches Ducking kann uU. eine negative Auswirkung auf die Immersion im Spiel haben und diese beeinträchtigen, da es grob und ungenau ist. Klangereignisse die im direkten Zusammenhang mit einer Visualisierung zusammenhängen sollten selektiv angestimmt werden, da der visuelle Fokus meist auf das gerichtet ist, was im jeweiligen Moment wichtig ist (Brown 2017).

VERBESSERUNG DER IMMERSION

Bei der Verbesserung der Immersion helfen bestimme stetig und unterbewusst ablaufende psychische Prozesse wie die *Mismatch-Negativity (MMN)*. Sie wird auch als *Ereignis-korreliertes Potential (EKP)* beschrieben und lässt sich nur im Rahmen akustischer Reize beobachten. Die MMN kann die Immersion innerhalb 150 ms brechen und tritt dann auf, wenn Klänge/Töne aus einer homogenen Reizfolge abweichen (Boeckmann et al. 2018). Es kann bspw. durch das Brechen von akustischen Mustern erfolgen; wird ein Muster gebrochen, fällt das meist direkt auf; ob unbewusst oder bewusst. Man hat das Gefühl, dass etwas nicht ganz richtig erscheint. Durch inkorrekte physische Darstellung von visuellen Gegebenheiten wie Schritte, als auch Stille bei physischer Interaktion/Aktivität. Wenn solche Muster gebrochen werden, verringert sich der Eindruck der Immersion des Spielers. MMN sollte man sich zunutze machen und Muster bzw. Regeln für die Tongestaltung erstellen um sich an diesen zu orientieren.

Einen weiteren Aspekt bildet die Stimulus Specific Adaptation (SSA; dt.: Stimulus-spezifische Adaption). Sie beschreibt die Eigenschaft von hörbahneigenen Neuronen, identische sich wiederholende Reize (bspw. ein Muster) mit einer verringerten neuronalen Aktivität zu repräsentieren, behält jedoch die Fähigkeit bei, auf

neue Reize mit voller Aktivität zu reagieren (von der Behrens 2009: 1) wodurch der Aufmerksamkeitsfokus in Kraft tritt. Setzt man sich gewisse Regeln/Muster für die auditiven Gestaltung fest, sollte man sich an diese halten. Wenn solch ein Muster versehentlich durchbrochen wird, kann dies ebenfalls zu einem Bruch der Immersion führen, was bei der SSA innerhalb 30 ms geschieht, oder auch den Spielerfokus fälschlich beeinflussen. Spieler können jedoch oft keine genaue Aussage darüber treffen was sie tatsächlich stört, es fühlt sich für sie lediglich "falsch" an (Brown 2017).

ORGANISCHES GAMEPLAY

Ein organisches Gameplay kann ebenfalls durch die Tongestaltung unterstützt werden. Muster und Wiederholungen trainieren den Spieler unterbewusst, sodass keine Text eingeblendet werden müssen, man kann also durch die Tongestaltung hinweg mit dem Spieler kommunizieren und ihm Anweisungen geben. Auditive Muster und Wiederholungen können einen Spieler ebenfalls trainieren. Ein Beispiel dafür stellt Inside (Playdead 2016). In diesem Spiel lernt der Spieler durch visuelle und auditive Cues, was es sicher ist bestimmte Tätigkeiten ausführen und wann nicht. In einem Abschnitt muss der Spielcharakter eine Brücke überqueren, auf dem in bestimmten Abständen Objekte wie Container stehen, oder Objekte die er durch die Betätigung eines Hebels aufrichten muss. Im Hintergrund sieht und hört man in gleichmäßigen zeitlichen Abständen eine Schockwelle, die sich bis zur Brücke ausweitet. Der Spieler kann seinen Weg über die Brücke immer nur dann sicher fortführen, wenn die Schockwelle nicht aktiv ist und muss sich während dieser, hinter den Objekten verstecken. Ohne jegliche textliche oder stimmliche Anweisungen, erlernt der Spieler durch solche sich wiederholenden Muster auf eine natürliche Weise, wie bestimmte Aufgaben zu bewältigen sind (Brown 2017). Dieser Effekt basiert auf einer der einfachsten Lernregeln im Bezug auf das Zustandekommen des Lernens in neuronalen Netzwerken, der Hebb'schen Lernregel. Diese lautet:

"Je häufiger ein Neuron A gleichzeitig mit Neuron B aktiv ist umso bevorzugter werden die beiden Neuronen aufeinander reagieren ("what fires together that wires together")" (Bauer 2018).

Es verbindet demnach die visuellen und auditiven Gegebenheiten, die immer zusammen auftreten.

TRIGGERN VON EMOTIONEN

Wie auch schon in Kap. 6.1.2. beschrieben ist es wichtig den Spieler auch auf der emotionalen Ebene miteinzubeziehen, denn Emotionen beeinflussen und verändern den Fokus, als auch die Wahrnehmung eines Ereignisses. Eine starke positive emotionale Reaktion eines Menschen kann in weniger als einer Sekunde registriert werden. Da wir für den Frequenzbereich von ca. 1-5 kHz am empfindlichsten sind, sollte man laut Brown, Schreckensmomente in genau diesem Frequenzbereich anwenden, somit haben diese eine erhöhte Auswirkung auf den Spieler haben – jedoch sollte man es mit der Anwendung dieser nicht übertreiben, da sich der Spieler daran gewöhnt oder es uU. sogar als anstrengend empfinden könnte. Ebenfalls befindet sich die Stimme im Bereich von etwa 2-3 kHz, welche zu den stärksten Auslöser für Emotionen gilt (s. Kap. 7.3.4).

KAPITEL 9 **FAZIT**

Wie wichtig eine gute Tongestaltung ist, zeigt sich in allen möglichen Audiovisuellen Medien, nicht nur in Virtual Reality-Anwendungen. Es spielt eine große Rolle, in wie weit man gerade bei interaktiven Spielen den Nutzer miteinbezieht und ihn mit einer positiven Erfahrung zur erneuten Nutzung der Anwendung führen kann.

Sowohl durch Surround Sound als auch 3D Audio lassen sich die Wahrnehmung als auch der Fokus mithilfe verschiedener Komponenten der Tongestaltung für virtuelle Umgebungen hervorragend beeinflussen und führen. Die Verwendung eines jeweiligen Formats hängt ausschließlich von den Anforderungen des Spiels ab und kann je nach Spielformat Vor- und Nachteile mit sich bringen.

Nichtsdestotrotz bietet 3D Audio einige Vorteile gegenüber den bisherigen kopfhörerbasierten Surround Sound-Systemen, wie einen höheren Eindruck an Natürlichkeit und Realismus, das Positionieren einzelner Audio-Objekte im freien Raum, oder die Erweiterung in die Höhe und Tiefe und damit die Darstellung eines sphärischen Klangbildes. 3D Audio hat seit dem Aufkommen des Virtual Reality einen neuen Aufschwung erhalten und sich rasant entwickelt. Der Bereich des Virtual Reality ist ein extrem schnell wachsender Markt und die Nachfrage nach neuen Systemen die den Eindruck der Immersion erhöhen steigen. In den kommenden Jahren wird sich 3D Audio mit hoher Wahrscheinlichkeit noch stärker nach vorne rücken und nicht nur für den Einsatz von VR Anwendungen populär werden wird. Mit dem Hintergrund der hohen realistischen und natürlichen Wiedergabe von akustischen Inhalten in allerlei Bereichen und nahezu unendlich vielen Gestaltungsmöglichkeiten, macht die Technologie des 3D Audio zu einem starken Gestaltungsmittel.

||| LITERATURVERZEICHNIS

- 360cameraonline (2016); Why Combine A 360 Camera With Binaural Microphone Head; online verfügbar unter: https://360cameraonline.com/360-camera-binaural-microphone/; erstellt: 04.01.2016; Zugriff: 09.01.2018
- Ardito, C.; Costabile M. F.; De Angeli A.; Pittarello F.; (2006): Navigation help in 3D worlds: some empirical evidences on use of sound; in: Multimedia Tools and Applications, Vol. 33, Ausg. 3, S.201-216; Hrsg. Springer US; online: DOI 10.1007/s11042-006-0060-0
- Audeze (2018): *Audeze Mobius*; online: https://www.audeze.com/products/mobius-series/mobius-headphone; Zugriff: 07.03.2018
- Baños, R.M.; Botella, C.; Alcañiz, M.; Liaño, V.; Guerrero, B.; Rey, B. (2004): *Immersion and Emotion: Their Impact on the Sense of Presence*, in: CyberPsychology & Behaviour; Vol. 7, Ausg. 6, S.734-741, 2004; Mary Ann Liebert; online: https://pdfs.semanticscholar.org/fb4d/cbd818e5839f025a6b-c247b3bc5632be502f.pdf
- Balázs, Bela (1930): Schriften zum Film; Band 2: Der Geist des Films, Artikel und Aufsätze 1926 - 1931; Henschelverlag Kunst und Gesellschaft Berlin
- Bauer, Martin (2018): *Hebb'sche Lernregel*; Hrsg. uni-protokolle.de; online: http://www.uni-protokolle.de/Lexikon/Hebb%27sche_Lernregel.html; Zugriff: 21.03.2018
- Beermann, Nils (2017): Ein Vergleich von Virtual Reality Ton auf Basis von Kopfhörern und Lautsprechern; Bachelorarbeit, Hochschule der Medien Stuttgart; online: https://www.hdm-stuttgart.de/~curdt/ Beermann.pdf
- Begault, Durand R. (2000); 3-D Sound for Virtual Reality and Multimedia; Hrsg. NASA Ames Research Centeer; National Aeronautics and Space Administration; online: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20010044352.pdf
- Beitman, Bernard D. (2009): *Brains Seek Patterns in Coincidences*, in: Psychiatric Annals, Vol. 39, Ausg. 5;, S.255-261; Hrsg. Charles Slack; online: https://www.drjudithorloff.com/main/wp-content/uploads/2017/09/Psychiatric-Annals-Brains-Seek-Patterns.pdf

- BioRICS (2018): *Positive and Negative Stress*; online: https://www.biorics.com/wp-content/uploads/2017/03/background-info-positive-negative-stress.pdf
- Blauert, Jens (1996): Zum gegenwärtigen Stand der Binauraltechnik, Tonmeistertagung 1996; Hrsg K.G. Saur Verlag; online: http://www.vdtshop.de/WebRoot/Store8/Shops/17637291/50B-F/9F3A/08B4/3F28/6E00/C0A8/28BA/4A6D/leseprobe_vdt_tmt1996_96003_abstract_zum_gegenwaertigen_stand_der_binauraltechnik.pdf
- Boeckmann, Martin; Rahne, Torsten; von Specht, Hellmut (2018): Mismatch Negativity (MMN) zum Nachweis auditorischer Objekte; Hrsg. Universitätsklinikum Regensburg; online: https://www.ukr.de/imperia/md/content/kliniken-institute/halsnasen-ohren-heilkunde/audiologie/itgboeckmann.pdf
- Bonneel, Nicolas; Suied, Clara; Viaud-Delmon, Isabelle; Drettakis George (2010): Bimodal perception of audio-visual material properties for virtual environments; ACM Transactions on Applied Perception, Association for Computing Machinery, 2010; Vol. 7, Ausg. 1, S.1-16; online: https://hal.inria.fr/inria-00333266/document
- Breh, Karl (2017): *Was ist Quadrophonie?*; Hrsg. Deutsches Hifi-Museum Wiesbaden; online: http://www.hifimuseum.de/1929.html; erstellt am 05.1973; Zugriff: 29.12.2017
- Brill, Manfred (2009): Virtuelle Realität; Hrsg. Günther; Prof. Dr. O.; Karl, Prof. Dr. W.; Lienhart, Prof. Dr. R.; Zeppenfeld, Prof. Dr. K.; Springer-Verlag Berlin Heidelberg
- Brown, Aaron (2017): 5D Audio in VR Expanding 3D Audio to Enhance VR Experience Using Higher Dimensions of Sound; Austin Game Conference 2017; online: http://www.aaronbrownsound.com/5d-audio-in-vr-expanding-3d-audio-to-enhance-vr-experiences-using-higher-dimensions-of-sound/; Zugriff: 04.03.2018
- Carlsson, Sven E. (2018): Nine Components of Sound; online: http://filmsound.org/articles/ninecomponents/9components.htm#Pitch; Zugriff: 13.02.2018

- Chung, Andy; To, Wai Ming; Schulte-Fortkamp, B. (2017): *Virtual Reality and Sonic Perception*,; online: https://www.researchgate.net/publication/319624702_Virtual_Reality_and_Sonic_Perception
- Dick, Frederic K.; Lehet, Matt I.; Callaghan, Martina E; Keller, Tim A.; Sereno, Martin I.; Holt Lori L. (2017): Extensive Tonotopic Mapping across Auditory Cortex Is Recapitulated by Spectrally Directed Attention and Systematically Related to Cortical Myeloarchitecture, in: The Journal of Neuroscienc; Vol. 37, Ausg. 50, S.12187–12201; 13. Dez. 2017; online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5729191/pdf/zns12187.pdf
- Ehmer, R. H. (1959): *Masking patterns of tones*, in: Journal of the Acoustical Society of America, Vol. 31, 1115-1120; online: https://doi.org/10.1121/1.1907836
- Ekman, Inger (2009): Modelling the Emotional Listener: Making Psychological Processes Audible; Center for Knowledge and Innovation Research, Helsinki School of Economics, Finland; online: https://www.researchgate.net/publication/233406211_Modelling_the_Emotional_Listener_Making_Psychological_Processes_Audible
- El-Bawwab, Rami (2017): Surround Sound von Damals bis heute (Teil 1/2); Hg. v. av-insider.de; online: http://av-insider.de/surround-geschichte-1; erstellt am 04.01.2015; Zugriff: 29.12.2017
- Freie Universität Berlin (2010): Auditive Wahrnehmung; online: http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/allg_neuro/studium/allgpsy_4.pdf?1286344518
- Freud, Signmund (1923): *Das ich und das Es*; Internationaler Psychoanalytischer Verlag Ges. m. b. H. Wien 1923; online: https://archive.org/details/Freud_1923_Das_Ich_und_das_Es_k
- Friesecke, Andrea (2014)s: Die Audio-Enzyklopädie -Ein Nachschlagewerkt für Tontechniker; Verlag Walter de Gruyter, 2. Aufl., 19.08.2014
- Ganiti, Aikaterini; Politopoulos, Nikolaos; Tsiatsos, Thrasyvoulos (2018): *The Impact of Background Music on an Active Video Game*, in: Interactive Mobile Communication, Technologies and Learning, S.18-28; Springer International Publishing AG; online: https://doi.org/10.1007/978-3-319-75175-7_3
- Gearbeast (2017): Somic G910 7.1 Surround Sound USB Gaming Headset BLACK; online: https://des.gbtcdn.com/uploads/pdm-desc-pic/Electron-ic/image/2016/05/26/1464255073326736.jpg; Zugriff: 30.01.2018

- Gegenfurtner, Prof. Karl R (2007): *Wahrnehmung*; Hrsg. Universität Giessen; online: http://www. allpsych.uni-giessen.de/karl/teach/Wahrnehmung/ Wahr-10-audit.pdf
- Gerrig, Richard J.; Zimbardo, Philip G. (2008): Psychologie; Pearson Deutschland GmbH; 18. Auflage 2008
- Görne, Thomas (2011): *Tontechnik*; Hrsg. Prof. Dr. Ulrich Schmidt, Hochschule für angewandte Wissenschaften Hamburg; Carl Hanser Verlag München 2011, 3. Aufl.
- Görne, Thomas (2017): Sounddesign: Klang, Wahrnehmung, Emotion; Hrsg. Dr. Ulrich Schmidt, Hochschule für angewandte Wissenschaften Hamburg; Carl Hanser Verlag München 2017
- Goldstein, E. B. (2007): Einführung in die Wahrnehmung, in Wahrnehmungspsychologie Der Grundkurs; Hrsg. Gegenfurtner, Karl; 7. Aufl.; Springer-Verlag Berlin Heidelberg; online: http://www.springer.com/cda/content/document/cda_downloaddocument/9783827417664-c1.pdf?SG-WID=0-0-45-869657-p173758620
- Gramatke, Carsten (2014): *Der Priming Effekt Teil I;* Hrsg. NPL-Zentrum Berlin; online: https://nlp-zentrum-berlin.de/infothek/nlp-psychologie-blog/item/der-priming-effekt; Zugriff: 23.02.18
- Grimshaw, Mark; Garner Tom (2015): Sonic Virtuality
 Sound As Emergent Perception; Oxford University
 Press
- G.Skill (2016): *Ripjaws SR91*0; online: https://www.gskill.com/en/product/ripjaws-sr910---real-7-1-headset; Zugriff: 30.01.2018
- Heeger, Prof. David (2017): Perception Lecture Notes
 Auditory Pathways and Sound Localization; Hrsg.
 NYU/CNS enter for Natural Science, 2006; online:
 http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/localization/localization.html;
 Zugriff: 21.12.2017
- Hofmann, Jan; Szczypula, Oliver (2008): Game Sound Audiotechnische Aspekte in Computerspielen; VDM Verlag Dr. Müller AG & Co. KG, Saarbrücken 2008
- Hu, Ruimin; Dong, Shi; Wang, Heng; Zhang, Maosheng; Wang, Song; Li, Dengshi (2012): Perceptual Characteristiv and Compression Research in 3D Audio Technology, in: From Sounds to Music and Emotions Hrsg. Aramaki et al. S.82-98; Springer-Verlag Berlin Heidelberg 2013

- Huestegge, Lynn (2018): *Top-down-Verarbeitung*; Hrsg. Markus Antonius Wirtz, Hogrefe AG Schweiz; online: https://m.portal.hogrefe.com/dorsch/top-down-verarbeitung/; Zugriff: 05.02.2018
- Hunt, Dr. Maria (2017): *Top-Down and Bottom-Up Funcioning*; Hrsg. MeWe Mindfullness; online: https://www.mewemindfulness.com/topic/top-down-and-bottom-up-functioning/;Zugriff: 28.01.2018
- Isbister, Katherine; Schaffer, Noah (2011): Game Usability Advancing the Player Experience; CRC Press Taylor & Francis Group
- Johar, Swati (2016): Psychology of Voice, in: Emotion, Affect and Personality in Speech; S.9-15; Springer International Publishing; online: DOI 10.1007/978-3-319-28047-9_2
- Juslin, Patrik N.; Västfjäll, Daniel (2008): Emotional responses to music: The need to consider underlying mechanisms, in Behavioral and Brain Sciences 559-621; Vol. 31, Ausg. 6, Dez. 2008; Camebridge University Press; online: doi:10.1017/S0140525X08005293
- Kaya, Emine Merve; Elhilali, Mounya (2017): *Modelling Auditory Attention*, Hrsg. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, Vol. 372 (1714)
- Kinchla, R. A.; Wolfe, J. M. (1979): The order of visual processing: "Top-down", "bottom-up," or "middle-out"; in: Attention, Perception, & Psychophysics Vol. 25, Ausg. 3; Hrsg. Springer Nature (1979); online: https://link.springer.com/article/10.3758/ BF03202991
- Lentz, Tobias (2008); *Binaural Technology for Virtual Reality*; Hrsg. Janina Fels und Michael Vorländer, RWTH Aachen; Aachener Beiträge zur Technischen Akustik, Band. 6, Logos Verlag Berlin; online: http://publications.rwth-aachen.de/record/51299/files/Lentz_Tobias.pdf
- Lipinski, Klaus (2017): Maskierung, in: ITWissen. info; Hrsg. DATACOM Buchverlag GmbH; online: https://www.itwissen.info/Maskierung-masking. html; Zugriff: 19.02.2018
- Litfin, Dominik (2014): Binauralsynthese von Mehrspurproduktionen auf mobilen Endgeräten; Bachelorarbeit, Hochschule der Medien Stuttgart; online: https://www.hdm-stuttgart.de/~curdt/Litfin.pdf
- LIV (2018): PERFECT Combo in Beat Saber! (Hard Mode) in Mixed Reality; Channel: LIV; online: https://www.youtube.com/watch?v=_E1d-qLnNxs; Zugriff: 11.03.2018

- MacGateway (2012): The History of Sound Cards and Computer Game Music; Hrsg. MacGateway; online: http://macgateway.com/featured-articles/sound-card-history/#1; zuletzt aktualisiert am 13.07.2012; Zugriff: 10.12.2017
- Maderthaner, Prof. Dr. Rainer (2017): *Psychologie*; 2. Aufl.; Facultas Verlags- und Buchhandels AG 2017
- Maile, Benedikt; Schiffner, Daniel (2016): Konzeption und praktische Umsetzung einer 3D-Audio-Live-Produktion Untersuchung und Entwicklung mit den Schwerpunkten im technischen Design und dreidimensionaler Live-Mischungen im Kontext populärer Musik; Masterarbeit, Hochschule der Medien Stuttgart; online: https://www.hdm-stuttgart.de/~curdt/Maile_Schiffner.pdf
- Mandler, George (1979): Denken und Fühlen Zur Psychologie emotionaler und kognitiver Prozesse; Junfermann Verlag 1979
- Mobbs, Dean; Hagan, Cindy C.; Dalgleish, Tim; Silston, Brian; Prévost, Charlotte (2015): The ecology of human fear: survival optimization and the nervous system, in Frontiers in Neuroscience 18.03.2015; online: doi.org/10.3389/fnins.2015.00055
- Moeser, Michael (2009); *Messtechnik der Akustik*; Hrsg. Michael Möser; Springer Verlag; Aufl. 2009
- Moore, C. J. Brian (2012); An Introduction to the Psychology of Hearing; Hrsg. University of Cambridge; Emerald Group Publishing Ltd; 6. Aufl. 2012
- Müsseler, Jochen; Rieger, Martina (2017): *Allgemeine Psychologie*; Hrsg. Jochen Müsseler; Springer-Verlag Berlin Heidelberg 2017
- Neuhoff, John G. (2016): Looming Sounds are perceived as faster than Receding Sounds in Cognitive Research: Principles and Implications; online: https://link.springer.com/content/pdf/10.1186/s41235-016-0017-4.pdf; Springer Open Choice
- Neumann, Georg (2017); *Dummy Head KU 100*; online: http://www.neumann.com/?cid=ku100_description&id=current_microphones&lang=en; Zugriff: 09.01.2018
- Neumeier, Sibylle (2016): Entwicklung der nutzerfreundlichen Individualisuerung von Binauraltechnik; Bachelorarbeit, Hochschule der Medien Stuttgart; online: https://www.hdm-stuttgart.de/~curdt/ Neumeier.pdf
- Oellers, Helmus (2017): *Wellenfeldsynthese*; Hrsg. SyntheticWaveAudio.de; online verfügbar unter: http://www.syntheticwave.de/Wellenfeldsynthese. htm; Zugriff: 30.12.2017

- Oomen, Paul; Holleman, Poul; van Weelden, Luc; Breed, Salvador (2018): 4DSound: A New Dimension In Sound; online: http://www.4dsound.net/about/#about-overview; Zugriff: 22.03.2018
- Oxford Dictionaries (2018): *Immersion*; Hrsg. Oxford University Press; online: https://en.oxforddictionaries.com/definition/immersion; Zugriff: 21.01.2018
- Pietschmann, Daniel (2017): Erleben virtueller Welten; Verlag Werner Hülsbusch; 2. Aufl. 2017
- PlayStation Europe (2017): Horizon Zero Dawn | 3D Audio Dev Diary Platinum Wireless Headset | PS4; online: https://www.youtube.com/watch?v=n8FAGDax5_g; Zugriff: 26.03.2018
- Plunkett, Luke (2011): *Meet The World's First Flight Simulator*; Hrsg. Gizmodo Media Group; online: https://kotaku.com/5792892/meet-the-worlds-first-flight-simulator, Zugriff: 04.12.2017
- Preißer, Alexander (2010): Virtuelle Realität; Hochschule Rhein Main; online: https://www.cs.hs-rm.de/~linn/fachsem0910/preisser/Virtuelle_Realitaet.pdf
- Raffaseder, Hannes (2010): *Audiodesign*; Hrsg. Prof. Dr. Ulrich Schmidt; Hochschule für angewandte Wissenschaften Hamburg; Carl Hanser Verlag München 2010, 2. Aufl.
- Rauh, Andreas; Müller, Jörn; Nießeler, Andreas (2016): Aufmerksamkeit - Neue humanwissenschaftliche Perspektiven; Hrsg. Andreas Rauh; Transcript Verlag Bielefeld 2016
- Roll, Julia (2017): Kommunikation im öffentlichen Raum Aufmerksamkeit und Mediennutzung; Springer VS; Springer Fachmedien Wiesbaden GmbH
- Rothbucher, Martin; Veprek, Kajetan; Paukner, Philipp; Habigt, Tim; Diepold, Klaus (2013); Comparison of head-related impulse response measurement approaches; Hrsg. The Journal of the Acoustical Society of America; Acoustical Society of America, Vol. 134; online: http://asa.scitation.org/doi/10.1121/1.4813592
- Schäfer, Thomas (1998): Soundkarten in der Praxis, Hrsg. Technische Universität Chemnitz; online https://www.tu-chemnitz.de/informatik/RA/ news/stack/kompendium/vortraege_98/sound2/ firstpag.html; Zugriff: 02.12.2017
- Schedeen, Jesse (2010): Surround Sound Headphones Explained; in: Imagine Games Network; Hrsg. Ziff Davis; online: http://www.ign.com/articles/2010/08/02/surround-sound-headphones-explained; Zugriff: 29.01.2018

- Schlenstedt, Gunar (2015): Zur Qualität von binauraler Wiedergabe mit transauralen Wandlern; Masterarbeit; Technische Universität Berlin; online: https://www2.ak.tu-berlin.de/~akgroup/ak_pub/abschlussarbeiten/2015/SchlenstedtGunar_MasA.pdf
- Schönthaler, Markus (1998): Computerspiele versus Spieler; Hamburg, Diplomica Verlag GmbH
- Schuller, Christian (2016): 3D Sound mit Dolby Atmos Der aktuelle Stand der Dinge Teil 1; Hrsg. Heimkinoraum; online verfügbar unter: https://www.heimkinoraum.de/tests/3d-sound-mit-dolby-atmos-der-aktuelle-stand-der-dinge-teil-1-70; Zugriff: 01.02.2018
- Seeber, Bernhard (2006): Psychoakustische Maskierung; online: http://www.bseeber.de/itg_page/maskierung.html; Zugriff: 17.02.2018
- Sengpiel, Eberhardt; Sengpiel Alexander (2012a): Kopfbezogene Übertragungsfunktion HRTF; online: http://www.sengpielaudio.com/KopfbezogeneUebertragungsfunktionHRTEpdf; zuletzt geändert: 19.05.2012; Zugriff: 30.12.2017
- Sengpiel, Eberhardt; Sengpiel Alexander (2012b): *Richtungsbestimmende Bänder*; online: http://www.sengpielaudio.com/BlauertBaender01C.gif; Zugriff: 04.01.2018
- Sengpiel, Eberhardt; Sengpiel Alexander (2012c): Richtungshören als Lokalisation beim natürlichen Hören in den drei Ebenen; online: http://www.sengpielaudio.com/RichtungshoerenInDreiEbenen.pdf; zuletzt geändert: 20.06.2012; Zugriff: 15.12.2017
- Sengpiel, Eberhardt; Sengpiel Alexander (2013): Die Bedeutung der Blauertschen Bänder für die Tonaufnahme; online: http://www.sengpielaudio.com/DieBedeutungDerBlauertschenBaender.pdf; zuletzt geändert: 22.03.2013; Zugriff: 04.01.2018
- Sengpiel, Eberhardt; Sengpiel Alexander (2014): Jecklin-Scheibe - eine Stereo-Mikrofonanordnung mit Trennkörper; online: http://www.sengpielaudio.com/Jecklinscheibe.pdf; zuletzt geändert: 12.04.2014; Zugriff: 09.01.2018
- Slee, Sean J.; David, Stephen V. (2015): Rapid Task-Related Plasticity of Spectrotemporal Receptive Fields in the Auditory Midbrain, in: The Journal of Neuroscience, Vol. 35, Ausg. 38, 23. Sep. 2015; online: http://www.jneurosci.org/content/35/38/13090.full.pdf
- Smurdon, Tom (2015): Oculus Connect 2: 3D Audio: Designing Sounds for VR; Hrsg. Oculus auf YouTube, online: https://www.youtube.com/watch?v=IAwFN9sFcso; Zugruff: 19.03.2918

- Solso, Robert L. (2005): *Kognitive Psychologie*; Springer Verlag Heidelberg 2005
- Stangl, W. (2018a): Bewusstseinsebenen; Lexikon für Psychologie und Pädagogik; online: http://lexikon.stangl.eu/8954/bewusstseinsebenen-ebenen-bewusstsein/; Zugriff: 25.01.2018
- Stangl, W. (2018b): Multitasking eine Illusion; Lexikon für Psychologie und Pädagogik; online: http://paedagogik-news.stangl.eu/multitasking-eine-illusion/; Zugriff: 09.02.2018
- Stifelman, Lisa J. (1994); The Cocktail Party Effect in Auditory Interfaces: A Study of Simultaneous Presentation, MIT Media Laboratory Technical Report, University of Memphis; online: http://citeseerx.ist.psu.edu/viewdoc/download?-doi=10.1.1.2.9918&rep=rep1&type=pdf
- Strohtmann, Christiane (2013): Physiologische Grundlagen des Hörens, In: Phänomene der Wahrnehmungspsychologie, Hrsg. Folkwang Universität der Künst, Online verfügbar unter http://icemwww.folkwang-uni.de/icem-web/wp-content/uploads/2013/12/2.-Physiologische-Grundlagen-des-H%C3%B6rens.pdf
- Temporal (2017): *Headphone Virtual Surround Sound; in: linustechtips*; Hrsg. Linus Media Group; online verfügbar unter: https://linustechtips.com/main/topic/718841-headphone-virtual-surround-sound/; Zugriff: 30.01.2018
- Theile, Günther (2014): *Tonaufnahme und Tonwiedergabe*; in: Dickreiter et al. (2014) Handbuch der Tonstudiotechnik; Hrsg. Dickreiter, Michael; Dittel, Volker; Hoeg, Wolfgang; Wöhr, Martin; Band 1, 8. Aufl.; De Gruyter Verlag
- Toprac, Paul; Abdel-Meguid, Ahmed (2010): Causing Fear, Suspense, and Anxiety using Sound Design in Computer Games, in: Game Sound Technology and Player Interaction: Concepts and Developments; Hrsg. Mark Grimshaw; Information Science Reference 2010; S. 176-191; online: DOI: 10.1080/09084280802636413
- Treasure, Julian (2011): *Sound Business*; Hrsg. Julian Treasure; Management Books 2000 Ltd; 2 Auflg. 2011
- Triebsch, Isabell (2010): Lokalisation von Schallquellen in der Horizontalebene im virtuellen Surround; Masterarbeit, Hochschule der Medien Stuttgart; online: https://www.hdm-stuttgart.de/~curdt/Triebsch.pdf

- Turi, Jon (2014): The Sights and Scents of the Sensorama Simulator; Hg. v. Oath Tech Network Aol Tech; online: https://www.engadget.com/2014/02/16/ morton-heiligs-sensorama-simulator/; Zugriff: 07.12.2017
- Ubbens, Theo (2015): Surround Sound und 3D Audio, Hrsg. InfoTipService GmbH; online: https://kompendium.infotip.de/3d-audio.html, erstellt am 02.10.2015, Zugriff: 30.12.2017
- Vanlessen, Naomi; Rossi, Valentina; Raedt, Rudi De; Pourtois, Gilles (2012): Positive emotion broadens attention focus through decreased position-specific spatial encoding in early visual cortex: Evidence from ERPs, in: Cognitive, Affective & Behavioral Neuroscienc; Hg. v. The Psychonomic Society; Vol. 13, Ausg. 1; 2013; online: DOI 10.3758/s13415-012-0130-x
- Varvatsoulias, George (2014): Voice-Sensitive Areas in the Brain, in: Psychological Thought; Hrsg. PsychoOpen Publishing; Vol. 7, Ausg. 1, 66–79; online: doi:10.5964/psyct.v7i1.98
- Virtual Reality Society (2018): *History of Virtual Reality*; online: https://www.vrs.org.uk/virtual-reality/history.html; Zugriff: 19.01.2018
- Von der Behrens, Wolfger (2009): Repräsentation akustischer Abweichungen durch Neurone und lokale Feldpotenziale im auditorischen Kortex der wachen Ratte; Dissertation, Johann Wolfgang Goethe-Universität 2009; online: https://core.ac.uk/download/ pdf/14520182.pdf
- Wegmann, Dominik (2005); Zu Unterschieden in der Hörereigniswahrnehmung bei Wellenfeldsynthese und Stereofonie im Vergleich zum natürlichen Hören; Diplomarbeit, Institut für Rundfunktechnik München; online: https://www.irt.de/IRT/publikationen/Diplomarbeiten/DA_Wegmann.pdf
- Weinzierl, Stefan (2008): *Handbuch der Audiotechnik*; Hrsg. Prof. Dr. Stefan vWeinzierl; Springer-Verlag Berlin Heidelberg 2008
- Weinzierl, Stefan (2014): Virtuelle Realität fürs Ohr; in Technology Rewiev Dez. 2014; Hrsg. Heise Zeitschriften Verlag 2014; online: https://www.ak.tu-berlin.de/fileadmin/a0135/downloads/stefan_weinzierl_aktuelles/Virtuelle_Realitaet_fuers_Ohr_Technology_Review_2014_12.pdf
- Weiis, Matthew (2013): 5 Dynamic Ways to Think About Masking, in ProAudioFiles; online: https://theproaudiofiles.com/masking-dynamics-frequency/; Zugriff: 20.02.2018

- Wikipedia (2017a): Lokalisation (Akustik); Hrsg. Die freie Enzyklopädie Wikipedia; online: https://de.wikipedia.org/wiki/Lokalisation_(Akustik), zuletzt aktualisiert am 23.07.2017, Zugriff 02.12.2017
- Wikipedia (2017b): *Survival Horror*; Hrsg. Die freie Enzyklopädie Wikipedia; online: https://de.wikipedia.org/wiki/Survival_Horror; zuletzt aktualisiert am 16.05.2017, Zugriff 23.02.2018
- Wikipedia (2017c): *Virtuelle Realität*; Hrsg. Die freie Enzyklopädie Wikipedia; online: https://de.wikipedia.org/wiki/Virtuelle_Realität, zuletzt aktualisiert am 30.11.2017; Zugriff 05.12.2017
- Williams, Matt (2014): A Universe of 10 Dimensions; Hrsg. Phys.org; online: https://phys.org/news/2014-12-universe-dimensions.html; Zugriff: 18.03.2018
- Wolfson, S; Case, G. (2000): The effects of sound and colour on responses to a computer game, in: Interacting with Computers, Vol. 13, Ausg. 2, S.183-192
- Wong, Kevin (2017): The Scariest Thing in 'Resident Evil 7' Is the Silence; Hrsg. Vice Media, LLC.'s; online: https://motherboard.vice.com/en_us/article/535zb5/the-scariest-thing-in-resident-evil-7-is-the-silence; Zugriff: 25.02.2018