



TECHNISCHE  
UNIVERSITÄT  
WIEN  
Vienna University of Technology

## **DIPLOMARBEIT**

Aural Multiplicities

**ausgeführt zum Zwecke der Erlangung des akademischen Grades  
eines Diplom-Ingenieurs  
unter der Leitung**

**Thomas Lorenz**

E264

Institut für Kunst und Gestaltung

**eingereicht an der Technischen Universität Wien**

Fakultät für Architektur und Raumplanung

von

**Clemens Hausch**

9626612

Wien, am

**eigenhändige Unterschrift**



# Abstract

## Aural Multiplicities

*Wie Schall Raum erzeugt und digitale Technologien diesen verändern.*

Schall ist von Natur aus ein körperliches Phänomen und somit immer räumlich. Er kann als Funktion der Veränderung der physischen Welt über die Zeit verstanden werden. Er ist also eine räumlich bestimmte, aber auch eine räumlich bestimmende Kraft: eine räumlich bestimmte Kraft, weil die physische Konstellation unserer materiellen Umgebung ihn formt und leitet; eine räumlich bestimmende Kraft weil er selbst wiederum den Zustand der Welt verändern kann: Physisch gesehen kann Schall Materie zum Schwingen bringen, also in Bewegung setzen. Psychisch gesehen kann er auf uns Menschen vielfältig Einfluss nehmen, uns leiten und unsere Handlungen beeinflussen. Er berührt uns und wir können mit ihm berühren. So stellt er eine körperliche Verbindung zu unserer Umwelt her.

Diese Arbeit stellt die These auf, dass Schall selbst Raum schaffen kann. Unsere körperliche Umgebung ist eine akustische, und wir sind in ihr sowohl Sender als auch Empfänger. Es besteht also eine aurale Rückkopplungsschleife zwischen uns und unserem Umraum. Elektronische Modifikatoren, die unsere Wahrnehmung unserer räumlichen Umgebung verändern, haben schon lange verschiedene Bereiche unserer persönlichen und sozialen Räume durchdrungen. Immersion in solch „verformten“ Raum ist aber nicht nur bedingt durch die Glaubwürdigkeit der Modifikatoren, sondern gleichermaßen durch deren Kapazität auf uns zu reagieren. Wir selbst müssen Modifikator dieses veränderten Raumes werden, genauso wie wir Modifikatoren unserer physischen Welt sind.

Durch die Herstellung einer Rückkopplungsschleife zwischen uns und diesen neuen Räumen können die virtuellen Veränderungen des Raumes, die sich auf unsere Wahrnehmung beschränken, tatsächliche Veränderungen werden. Aufgrund aktueller Technologien kann anhand der Schallwellen eines Ereignisses oft nicht unterschieden werden, ob es sich um eine Simulation oder eine Aufnahme handelt, oder ob das Wahrgenommene auf physischen Grundlagen beruht. Ganz unabhängig von der Ursache können sie uns berühren, leiten und unsere weiteren Handlungen beeinflussen. Der Effekt ist der gleiche, und zieht emotionale, intellektuelle und letztlich physische Folgen nach sich, wenn das Individuum beginnt, sich nach dieser neuen Räumlichkeit zu verhalten. Der Raum wird erweitert und verändert. Die Unmittelbarkeit des Hörens bedingt, dass wir uns ihm nicht entziehen können.

Teil 1 der schriftlichen Arbeit befasst sich mit den theoretischen Grundlagen für ein künstlerisches Projekt, das im Rahmen dieser Arbeit entstanden ist und in Teil 2 beschrieben ist. Kapitel 1.1 erklärt die physikalischen Grundlagen von Schall und die Funktionsweise räumlichen Hörens beim Menschen. Kapitel 1.2 untersucht bisherige Beschreibungen unserer akustischen Umgebung durch die Kulturwissenschaften der Sound Studies. Kapitel 1.3 führt in den Begriff der auralen Architektur ein und erläutert einerseits, wie unsere materielle Welt Schall formt und andererseits, wie Schall selbst Raum erzeugt. Kapitel 1.4 schließlich befasst sich mit dem Begriff der Virtualität und behandelt die Möglichkeiten und Konsequenzen eines elektronisch modifizierten auralen Raumes.

Teil 2 beschreibt das Projekt „*Sonic Manifolds*“ das, aufbauend aus den Erkenntnissen aus Teil 1, den Versuch unternimmt, virtuelle aurale Räume mittels einer selbst entworfenen computergesteuerten Rauminstallation zu erzeugen. Indem eine eine Rückkopplungsschleife zwischen physischem Hier und künstlichem Dort hergestellt wird kann der virtuelle Raum ertastet werden. Da die Bewegungen des Besucher im Realraum eine Antwort in der künstlichen Umgebung erhalten entsteht erst die geforderte aurale Immersion. Dynamik und Kontrast sind die bestimmenden Elemente dieser Umgebung, und Bewegung und Innehalten sind die Methoden, mit denen sie erkundet werden kann. Das Ohr ist das Organ, welches uns leitet. Am Spiel mit verschiedenen Ordnungen und Brüchen in den Konfigurationen soll so der Kern dessen approximiert werden, woraus auraler Raum besteht.





# Abstract

## Aural Multiplicities

*How sound engenders space and digital technologies transform it.*

Sound is a bodily phenomenon by nature, and thus always spatial. It can be understood as a function of change of the physical world over time. It is a spatially determined, but also spatially determining force: spatially determined because it is formed and directed by the physical configuration of our material world; spatially determining because sound itself can change the state of our world: viewed from the perspective of physics it can set matter into vibration and ultimately effectuate movement. Psychologically it can exert influence on us humans in a variety of ways, and can guide us and affect our actions. It touches us, and we can touch through it. Thus it represents a bodily connection to the world around us.

With this thesis i want to make the case that sound itself is able to engender space. Our physical environment is also an acoustic one, in which we are both transmitter and receiver. Thus exists an aural feedback loop between us and our surrounding. Electronic „modifiers“ that change our sensory perception of the world have penetrated various spheres of our personal and social space for a long time now. Immersion in such a „altered“ space is not just presupposed by the credibility of the modifications that take place, but just as much by the capacity of the new space to react to us. We ourselves must become modifiers of this altered space, just as we are modifiers of our physical world.

By establishing an aural feedback loop between us and these new spaces (akin to the one between us and the traditional physical world) the virtual alterations that are confined to our sensory perception can become an actual alteration. With todays technologic capabilities it is often impossible to deduct from the soundwaves of an auditory event if they originate in a simulation, a recording or an actual physical event. In either case they can touch us, guide us and affect our actions. The effect will be the same, and will entail emotional, intellectual and finally physical consequences, when the individual starts to behave under the influence of the new spatiality. Space is expanded and altered. The immediacy of hearing brings about that we can not elude it.

Part 1 of the written thesis treats the theoretical basis for an art project that was conducted in the course of this thesis and which is documented in part 2. Chapter 1.1 explains the physical basics of sound and the perceptual workings of human spatial hearing. Chapter 1.2 investigates the previous attempts of scholars of social science to describe our spatial acoustic environment. Chapter 1.3 builds upon this fundament and elaborates how our material world shapes the sounds it contains, and vice versa, how sound itself shapes our spatial environment through physical, emotional and social means. Finally Chapter 1.5 sheds light on the possibilities and consequences of an electronically modified aural space.

Part 2 describes the project “*Sonic Manifolds*”, which was conceived of the knowledge documented in part 1. It attempts to create aural spaces by means of a self-designed computer-controlled sound installation. By establishing a feedback loop between the physical *Here* and the artificial *There* virtual space becomes physically palpable for visitors. Since the visitor’s movements in the actual space receive an answer in the artificial simulation true aural immersion can take place. Dynamics and contrast are the defining elements of this environment, and movement and pause are the means to explore it. The ear is the organ that leads us. Through the interplay between diverse collocations and discontinuities in the configuration of the virtual space the core of what constitutes aural space shall be approximated.



## Danksagungen

An dieser Stelle möchte ich mich bei den Leuten bedanken, die sowohl intellektuell als auch als auch persönlich zur Genesis meiner Arbeit beigetragen haben.

Mein größter Dank gilt meinen Eltern Dorith und Erich Hausch für ihre unermessliche Geduld und ihre unablässige Unterstützung, die ich all die Jahre über geniessen durfte.

Weiters danke ich vor allem meinem Betreuer Thomas Lorenz für die ausdauernde Begleitung dieser Arbeit, Christine Hohenbüchler und dem gesamten Institut E264 für den langjährigen Beistand und Nikolai Kirisits und Dörte Kuhlmann für entscheidende Hinweise, die die vorliegende Arbeit beeinflusst haben.

Insbesondere danke ich auch Erich Hausch für sein lexikalisches Wissen und seine tiefgreifenden Erklärungen naturwissenschaftlicher und musikalischer Prinzipien, sowie Gernot Schlager und Leopold Paul Lehner für ihre grosse Anteilnahme und ihren prüfenden Blick bei der Lektüre der Erstfassungen dieser Arbeit.

Weiters möchte ich noch einigen Personen danken, die im Laufe der Jahre ihr Wissen mit mir geteilt und so die Werdung des Projektes beeinflusst haben. Unter diesen sind Florian Widhalm, Georg Nebehay, Florian Gruber, Ulrich Kurt Kühn, Nils Peters, Jan Schacher, Hannes Franks und Alessandro Ludovico. Jörg Wollman möchte ich dafür danken, dass er mich so lange ich mich erinnern kann stets sowohl künstlerisch als auch geistig förderte und inspirierte und Bertram Haller für seine unermessliche Großzügigkeit und auch so manch bezahlte Zeche.

Zu guter Letzt äußere ich meinen Dank an meine Leonie Schlager – ohne ohne sie hätte ich es nicht geschafft.



## **Inhalt**

### **Teil 1**

<b>Einleitung</b>	<b>13</b>
<b>1. Räumlich Hören</b>	<b>23</b>
<b>2. Sound Studies</b>	<b>41</b>
<b>3. Aurale Architektur</b>	<b>63</b>
<b>4. Aurale Multiplizitäten</b>	<b>85</b>

### **Teil 2**

<b>1. Übersicht</b>	<b>109</b>
<b>2. Videotracking</b>	<b>113</b>
<b>3. Datenauswertung</b>	<b>117</b>
<b>4. Szenen</b>	<b>121</b>
<b>5. Klangerzeugung</b>	<b>125</b>
<b>6. Raumklang</b>	<b>131</b>
<b>7. Raumsimulation</b>	<b>145</b>
<b>8. Perspektiven</b>	<b>161</b>
<b>Quellenverzeichnis</b>	<b>169</b>



# **Teil 1**

# **Aural**

# **Multiplicities**





## Einleitung

Unsere Welt ist durchdrungen von Klang. Wirkliche Stille existiert nirgends. Vielleicht klatscht der Regen ans Fenster, oder es brummt ein Transformator am Strassenrand; oder eine Fliege summt um den Frühstückstisch; ein lautes Piepen eines Lastwagens, der in eine Parklücke schiebt; ein Vogel, der ein Solo trällert um seine Sippe zu wecken; der Presslufthammer auf der Baustelle gegenüber; das Meeresrauschen; das Singen der Geleise, das vom baldigen Eintreffen des Zuges kündigt; das Gedudel von Kaufhausmusik; eine Explosion; Kindergeschrei am Ende des Schultages; Rauschen, Klirren, Knattern, Gurgeln, Dröhnen, Krachen. Es gibt weitaus mehr Klänge, also Worte, um sie zu beschreiben. Sie variieren von Ort zu Ort und Zeit zu Zeit, aber wirkliche Stille existiert nirgendwo. Sogar beim Betreten eines „schalltoten“ Raumes<sup>1</sup> stellt man fest, dass man selbst pulsiert und dröhnt, als man seinen eigenen Körper dabei hört **lebendig zu sein**. Mit zunehmendem technologischen Fortschritt scheint sich das Verhältnis zwischen natürlichen und künstlichen Geräuschen mehr und mehr zugunsten Letzterer zu verschieben. Insbesondere die zunehmende Dominanz von Geräuschen mit gleichförmigem Pegel, also ohne erkennbarem Anfang und Ende, führte mit der Gründung des „*World Soundscape Project*“ durch Murray Schafer et. al. in den späten 1960-Jahren zum Aufstieg der Akustischen Ökologie (engl. *acoustic ecology*) und in der Folge zu einem neuen Bewusstsein für akustische Umweltverschmutzung sowie einer breiten Akzeptanz von Lärmschutzbestimmungen. Diese frühen Klangforscher waren getrieben von der romantischen Sehnsucht nach bukolischen und prä-industriellen Klangumgebungen (engl. *soundscape*s). Die Grundprämisse des World Soundscape Project war, von der Verdrängung bedrohte Klänge und Soundscapes zu sammeln und zu dokumentieren und strikte Schallschutzmaßnahmen durchzusetzen. Während die zugrundeliegenden Motive zwar als hehr aufgefasst werden können, bergen sie doch eine tiefsitzende Problematik: die Unterscheidung zwischen gewünschten und ungewünschten Klängen basiert sehr stark auf persönlichen Präferenzen. Schafer selbst erklärt konstante Dauertöne wie etwa Maschinenlärm zu seinem Feindbild, und nimmt somit eine tendentiell technologiefeindliche Haltung

---

1 Ein stark gedämmter Raum, der Umgebungsreflexionen nahezu vollständig eliminiert.

ein. Wo ist die Trennlinie zwischen natürlicher und artifizieller Klangquelle zu ziehen? Beim Netzbrummen? Beim Verbrennungsmotor? Bei der Dampfmaschine? Beim Nebelhorn? Beim Mühlrad? Oder gar beim Faustkeil?

Man könnte argumentieren, dass viele vermeintlich künstliche Geräusche, die uns umgeben, doch natürlich sind in dem Sinn als sie als akustische Nebenprodukte eines Mechanismus entstehen, der eine spezifische Funktionalität ermöglicht. So ist das Ticken der mechanischen Uhr ein Nebenprodukt der Mechanik der Uhrwerks, das Brummen des Autos ist ein Resultat der Explosionen in den Zylindern des Motors und das Summen des Server-Racks ist das Resultat der sich drehenden Kühlventilatoren. Die Wahrheit aber ist, dass viele dieser mechanischen Klänge das gezielte Werk von Product Sound Designern ist sind: Getränkeflaschen werden entworfen um beim Öffnen einen befriedigenden Knall zu erzeugen (Der Standard, 2005), Computerspieler zahlen hohe Aufpreise für Keyboards, die mit einem Klicken antworten, wenn man ihre Tasten drückt (Crecente, 2015), Fahrzeughersteller investieren in die Entwicklung eines Türschluss-Geräuschs das Vertrauen und Solidität vermitteln soll (Adweek, 2012), ja sogar der Klang von Filzstift-Spitzen, die über ein Blatt Papier gleiten, wird bewusst entworfen (Byron, 2012) und um mögliche Gefühle der Peinlichkeit zu vermeiden werden Tampon-Packungen extra abgestimmt um möglichst leise zu sein (Byron, 2012).

Eine der ersten Industrien, die Product Sound Design zur Anwendung gebracht hatten, war die Fahrzeugindustrie, in der Bestrebung, den Klang des Motors im Fahrzeuginneren angenehmer zu gestalten. Heute gibt es kaum ein akustisches Detail an einem Fahrzeug, das nicht bewusst gestaltet ist. Audio-Branding nimmt mittlerweile 5% des Entwicklungsbudgets eines neuen Fahrzeug in Anspruch. (Schleufe, 2010) Dabei führte uns der technologische Fortschritt in eine paradoxe Situation: Viele der Nebengeräusche, die uns umgeben, wurden durch den technologischen Fortschritt eliminiert. Und doch sind sie immer noch für uns hörbar.

Zum Beispiel bestand jahrzehntelang ein direkter Zusammenhang zwischen der physischen Bauweise seines Motors und dessen Lärmentwicklung. Motoren könnten inzwischen so leise gebaut werden, dass sie kaum hörbar wären und doch wird unser Stadtbild noch heute von zum Teil sehr lauten Fahrzeugen beherrscht. Für manche Fahrzeugklassen ist Lautstärke untrennbar mit dem Produkt verwoben. Ein leise schnurrender Ferrari würde beispielsweise sein Zielpublikum enttäuschen. Röhren und Lautheit symbolisieren Kraft – das Hauptmerkmale eines Sportwagens. Und so wurde auch in der Formel 1 Mitte 2014 öffentlich diskutiert, ob Eingriffe getätigt werden sollten, um die Wagen wieder lauter werden zu lassen, um diese Kategorie des Motorsports wieder attraktiver werden zu lassen. (Noble, 2014) Unter den am Markt erhältlichen Straßenfahrzeugen gehören Harley-Davidson Motorräder zu den lautesten. Ihre akustische Identität beruht auf deren

Blubbern und Krachen im Leerlauf und deren donnerndem Röhren bei Vollgas – einem Klangbild, das Rohheit symbolisiert und Bilder einer archaischen Bestie beschwört, die nur darauf wartet (von einem zahlenden Kunden) gezähmt zu werden. Und so betreiben Harley-Davidson auch eine eigene Forschungsabteilung, die den vielsagenden Namen Harshness, Vibration & Noise trägt. Sie ist gleichermaßen verantwortlich für Lärmschutz als auch Lärmdesign. Harley-Davidson ist die Bedeutung des Klangs ihrer Motoren so wichtig, dass sie im Jahre 1994 sogar versuchten diesen patentieren zu lassen, was schlussendlich vor Gericht abgelehnt wurde. (Lyon, 2003, S19) Es wurde argumentiert, dass der Klang ein Nebenprodukt der in den 45° V2 Motoren zur Anwendung kommenden Kurbelwelle mit nur einem Hubzapfen ist. Da diese Technik bereits lange verfügbar ist und von zahlreichen Motorradbauern verwendet worden war kann der Klang als Nebenprodukt dieses Designs kein schützenswertes Marken-Charakteristikum darstellen. Das ist zwar richtig, wirft aber gleichzeitig ein interessantes Licht auf die Bedeutung, die Harley-Davidson dem Klang ihrer Motoren beimessen. Das Grunddesign ihrer Motoren geht nämlich auf das Jahr 1909 zurück und wurde von den meisten Herstellern bereits in den 1940er Jahren aufgrund technologischer Mängel<sup>2</sup> aufgegeben. (Moser, 1996) Harley-Davidson setzen also bewusst Techniken ein, die ihre Produkte stark erscheinen lassen, aber das Gegenteil bewirken. Das Image welches sie bewerben ist ihnen wichtiger als dessen Versprechungen einzulösen.

Ähnliche Muster können auch beim VW Golf GTI DSG (Santer, 2009) und dem Mini Cooper S (Kurier, 2013) beobachtet werden: bei diesen kommen künstliche (elektronisch gesteuerte) Fehlzündungen zum Einsatz, welche die Wagen kraftvoller erscheinen lassen, aber auf Dauer die Motoren schädigen können. Bei Sportwagenherstellern wie Jaguar (Viamichelin.de, 2013) und Porsche hingegen setzt man auf ausgeklügelte Resonatoren und Ventilsysteme, um den Klang des Motors künstlich zu verstärken. Beim Studium von Autoforen im Internet kann man beobachten, dass viele Kunden sich dieser Entwicklungen bewusst sind und diese sogar begrüßen.

Der Grat zwischen der Steigerung von Kundenzufriedenheit und Täuschung ist schmal. Was mit mechanischen Manipulationen begonnen hatte, führt im Zeitalter fortschreitender elektronische „De-Mechanisierung“ zu regelrechtem Schwindel. Autohersteller wie Ford, VW und BMW statten einige Modelle inzwischen mit gänzlich synthetische Motorengeräuschen aus, die über Lautsprecher in der Motorhaube wiedergegeben werden. (Harwell, 2015; Bimmertoday.de, 2011) Beim Elektroauto gehen die Hersteller allerdings in Ermangelung einer historischen Klangvorlage gänzlich neue Wege und lassen die Antriebsgeräusche von Grund auf neu entwerfen. So kann etwa der „Komponist“ des Sportwagens „R8 e-tron“ in einem Werbevideo von Audi bei der Arbeit beobachtet werden. (Auto-News.de, 2012) Der Akt des Sounddesigns wird hier offen und stolz als Innova-

---

<sup>2</sup> Durch den unrunder Lauf, aus dem sich das synkopierte Knattern ergibt, geht viel Energie verloren, und die ungleichmäßige Belastung der Kurbelwelle führt zu Spannungen im Material und in der Folge zu erhöhtem Verschleiß.

tion präsentiert. Während aber der Klang des letztendlich vorgestellten Wagens noch Ähnlichkeiten mit einem fossil angetriebenen aufweist (Ausfahrt.tv, 2013), ist der Kompaktwagen Nissan „Leaf“ mit einem Sounddesign ausgestattet, das an ein Düsentriebwerk gemahnt. (Ambler, 2013; Nissanleafaustralia, 2012) Wie so oft in der Geschichte übernehmen hier Wissenschaft und Technik Anleihen an populär gewordenen Werken der Fiktion. „Knight Rider“ lässt grüßen: Der Triebwerksklang von „K.I.T.T.“ ist nicht allzu weit entfernt von dieser Klangstilistik, die mit der Elektro-Motorrad-Studie Harley-Davidson „Live-Wire“ auf die Spitze getrieben wurde. (Tsay, 2014). In Anbetracht der ikonischen Bedeutung des klassischen Harley-Davidson Motorenklanges birgt dieses Experiment für die Firma das große Risiko die eigene Marke zu verwässern und wird von vielen Anhängern kritisch diskutiert. Für manche ist allein der Gedanke einer elektrischen Harley ein Affront. Dieses Beispiel illustriert jedoch sehr gut, welche Emotionen Product Sound Design hervorrufen kann, wenn es eng abgestimmt mit anderen Parametern des Produktes ist. Der Klang eines Produktes ist sowohl für Hersteller als auch Kunden so bedeutend, dass viel Zeit, Geld und Energie investiert wird, um diesen zu gestalten, wobei Hersteller so weit gehen, Unzulänglichkeiten in Kauf zu nehmen, um den Klang zu erzielen, der sich am besten verkauft. Kunden schließlich nehmen in vollem Bewusstsein dieser Unzulänglichkeiten selbige in Kauf, um diesen Klang für sich in Anspruch nehmen zu können. Aber was sind die Gründe für diesen Wert, den Sound Design mehr oder weniger unbemerkt in den letzten Jahrzehnten erlangt hat?

Schall als Informationsträger erzählt uns Geschichten über die Welt in der wir leben. So können etwa durch die Kombination von akustischen Symbolen komplexe Bedeutungen in eine Klangsequenz kodiert werden. Menschliche Sprache zum Beispiel hat seinen Ursprung in Gurren und Schreien, erlaubt uns aber abstrakte Gedankenmodelle zu vermitteln. Allerdings repräsentiert menschliche Sprache eine Art von Inhalten, die in erster Linie absichtlich in Klänge kodiert werden und dazu gedacht sind, bewusst dekodiert zu werden. Zusätzlich zu dieser semiotischen Ebene, deren Inhalte intellektuell verarbeitet werden gibt es aber noch eine weitere, emotionale Ebene, deren Bedeutung sich in der Modulation erschließt. So kann die Sprachmelodie des Gesagten dessen Inhalt sowohl verstärken als auch konterkarieren und es kann dadurch auf die Intention des Sprechers geschlossen werden. Und schlussendlich gibt es eine die physi(kali)sche Bedeutungsebene, in der vom Klang der Stimme auf die physische Konstitution der Sprechenden geschlossen werden kann. Sie ist das Resultat der Konfiguration von Stimmbändern, Zunge, Mund und Resonanzräumen.

Als eines der drei Resultate von Energiefreisetzung – die anderen beiden sind Licht und Hitze – ist Schall das direkte Ergebnis einer physischen Aktion. Wenn wir ein bestimmtes Geräusch hören, schließen wir von diesem auf die ursprüngliche Aktion. Im Umkehrschluss erwarten wir aber auch eine passende akustische Antwort wenn wir Zeuge eines physischen Vorgangs werden. In einer Zeit

in der mechanische Prozesse zunehmend durch immaterielle elektrische Lösungen ersetzt werden, löst sich die Verbindung zwischen Aktion und akustischem Feedback langsam auf. Schall aber ist uns ein essentielles Mittel, um mit unserer Umwelt in Verbindung zu treten: er gibt uns „Feedback“ über die physische Beschaffenheit unserer Umgebung und die Vorgänge, die sie bewegen. Sein Fehlen hinterlässt eine Lücke, die einer sensorischen Behinderung gleichkommt. Diese Lücke mit artifiziellen Schallereignissen zu füllen ist eine Möglichkeit, um dieses Feedback wiederherzustellen. Wir sind also gewohnt, dass Aktionen Vibrationen zur Reaktion haben und dass diese Vibrationen uns von den Konsequenzen unser Handlungen künden.

Zum einen können wir aus dem Frequenzbild eines Klangs viele Informationen darüber herauslesen wie er erzeugt wurde. Dies funktioniert zum einen mit aktiven Klangquellen: dass wir tiefes Knattern als kraftvoll empfinden ist kein Zufall, sondern resultiert daraus, dass größere Volumina tiefere Resonanzfrequenzen aufweisen und es mehr Energie benötigt, um diese in Schwingung zu versetzen als um etwa ein hohes Surren zu erzielen. Des weiteren kann derselbe Klang uns aber auch etwas über die Materialität der involvierten Körper mitteilen. Unser Gedächtnis und unsere Erfahrung spielen dabei eine große Rolle. So lernen wir etwa von klein auf, wie verschiedene Materialien wie Glas, Holz, Stein, Kunststoffe oder Textilien klingen. Und so wird der empirische Wert einer multisensorischen audio-taktilen Erinnerung auch wiedererkannt, wenn nur die klangliche Komponente vernommen wird. Durch Hören können wir Materialität erfahren. Als Beispiel sei das Medium Film erwähnt: hier wird die taktile Materialität zumeist über die Tonspur transportiert. Szenen die einer hohen körperlichen Immersion bedürfen werden in hohem Masse vom Klang getragen. Das gilt genauso für gewalttätige Ereignisse wie Explosionen oder Schlägereien wie für intime und zärtliche Darstellungen. Das Fehlen einer Tonspur wiederum kann Szenen geisterhaft und entkörperpert wirken lassen – eine Technik, die häufig bei der Darstellung von Rückblicken oder Traumsequenzen Anwendung findet und die jeweilige Szene als unreal kennzeichnet und in der Gedankenwelt eines Protagonisten verortet.

Unser Hirn verarbeitet den Großteil aller auraler Informationen unbewusst. Der Thalamus, der als „sensorischer Türsteher“ für unser Bewusstsein fungiert, trifft eine Vorauswahl aus den eintreffenden Sinnesreizen. Während ein Großteil der sensorischen Informationen ausgefiltert werden, werden Klänge, die eine starke emotionale Reaktion hervorrufen, bevorzugt behandelt. Aufgrund der Vorrangstellung, die die auditive Verarbeitung beim Menschen einnimmt, kann Klang auch ein Machtfaktor sein. Zum Beispiel kann der Soundtrack eines Filmes uns zum Weinen bringen oder vor Schreck hochspringen lassen. Er kann in uns also physische Reaktionen auslösen und das gegen unseren Willen. Seit die Erkenntnisse der Psychoanalyse mithilfe von Edward Bernays Eingang in

die Werbung gefunden haben, um so das moderne Marketing zu begründen<sup>3</sup>, hat die Industrie den Wert erkannt, welchen Wert das „Triggern“ des Unterbewusstseins darstellt. Klang ist hierfür ein hochgradig effektives Mittel. Dies ist der Grund warum die akustischen Eigenschaften von Produkten oft nicht dem Zufall überlassen sondern manipuliert werden. Es ist interessant zu beobachten, dass das Wissen um die Möglichkeiten von Sound Design bis jetzt nur bei Objekten zur Anwendung kommt, deren Maßstab sie als Produkt erkennbar und greifbar macht. Während Autos aural „gestimmt“ werden, bleibt unsere architektonische Umgebung größtenteils vernachlässigt.

Murray Schafer bezeichnete in seinem Werk „The Soundscape or Our Sonic Environment And The Tuning Of The World“ hören als „Berühren aus der Ferne“: *„Touch is the most personal of the senses. Hearing and touch meet where the lower frequencies of audible sound pass over to tactile vibrations (at about 20 hertz). Hearing is a way of touching at a distance and the intimacy of the first sense is fused with sociability whenever people gather together to hear something special.“* (Schafer, 1977, S. 11) Während Schafer diese Zeilen als Metapher für die intimen, emotionalen Reaktionen, die Hören auslösen kann intendierte, gibt es noch eine andere mögliche – wörtliche – Lesart: Schall ist die mechanische Verformung von Luft und anderer elastischer Materie, also die physische Verformung unserer Umgebung, und in tieferen Frequenzbereichen können wir ihn als Erschütterungen in unserem Körper spüren. Er kann als Echo einer vergangenen Konfiguration unserer Welt in der Gegenwart verstanden werden. Unser ganzer Kosmos kann so gesehen als gigantische Symphonie verstanden werden. Jeder Körper, der sich bewegt und jede Oberfläche die vibriert, trägt dazu bei, so wie der kleinste Fisch, der die Oberfläche eines Teichs durchbricht seine Wellen schlagen wird. Jedem Klang, der gehört werden kann, geht eine „Umordnung“ der physischen Welt voraus. Obwohl nicht jeder Klang mit dem menschlichen Ohr gehört werden kann, so ließe sich doch argumentieren, dass jede Bewegung Klang und jeder Klang Bewegung ist. Einen Klang zu erzeugen bedeutet demnach unsere Umwelt zu berühren und sie zum Schwingen zu bringen. Hören bedeutet berührt zu werden und zum Schwingen gebracht zu werden.

Aber es ist nicht nur die Körperlichkeit von Schallerzeugern, die wir hören: wir sind auch in der Lage eine akustisch passive Umgebung zu hören, ähnlich Fledermäusen, die Ultraschall-Echolocation zur Navigation benutzen. Im Gegensatz zu diesen Säugetieren haben wir Menschen jedoch keine Organe, die dezidiert Signale erzeugen, um unsere Umgebung abzutasten. Wir sind jedoch bis zu einem gewissen Grad fähig, die Reflexionen zu interpretieren, die verschiedene Räume und Hindernisse von Klangereignissen wie Händeklatschen, Fußtritte oder dem Klang einer Stimme zurückwerfen.

---

<sup>3</sup> Bernays, E. (1928), Propaganda, Routledge, Online-Version: <http://www.historyisaweapon.com/defcon1/bernprop.html>. (Retr. June 2nd, 2015)



Die Körperlichkeit von Schall bedingt seine Räumlichkeit. Er hüllt uns ein und pulsiert um uns herum und durch uns hindurch. Im Gegensatz zu Licht ist er sehr langsam, wird aber nicht so schnell von Oberflächen absorbiert. Stattdessen breitet er sich aus und springt hin und her zwischen den Begrenzungen eines abgeschlossenen Raumes bis seine Energie durch die Reibung an Luft und Oberflächen verloren geht. Jede Oberfläche an der er reflektiert wird und jeder Körper, den er durchdringt, verändert seinen spektralen Gehalt, während die Geometrie unserer Umgebung diese Reflexionen lenkt. Wenn die Reflexionen schließlich unsere Ohren erreichen, tragen sie einen einzigartigen akustischen „Abdruck“ in sich, der uns eine Geschichte über die Reise des Schalls durch den Raum erzählen kann. Man kann diese Reflexionen in zwei Gruppen unterteilen: Erstreflexionen und diffuser Nachhall. Während erstere unserer Fähigkeit in geschlossenen Räumen zu navigieren zuträglich ist, indem sie uns erlauben, Distanzen zu reflektierenden Oberflächen in unserer Nähe zu „messen“, verwischen Zweitere die Reflexionen in ein diffuses Rauschen, das zwar unseren navigatorischen Fähigkeiten und der Verständlichkeit abträglich ist, uns aber erlaubt, das Volumen und die Materialität eines Raumes zu begreifen. Wir reagieren instinktiv auf diese Reflexionen und können mit etwas Training lernen, sie bewusst zu lesen und Räume aural zu visualisieren.

Die meisten Menschen sind sich dieser Effekte nicht bewusst oder bemerken sie erst, wenn sie in extremer Form auftreten (z.B. bei einem starken Echo oder dem typischen stark resonierenden Badezimmer-Nachhall). Der Grund hierfür ist darin zu suchen, dass unser Gehirn darauf ausgelegt ist, diese Effekte vor uns zu verbergen, denn sie können die Verständlichkeit von Signalen beeinträchtigen. Viel Energie wird also darauf verwendet, aus den verhallten und verfremdeten Signalen, die unsere Ohren erreichen, das Originalsignal so gut wie möglich zu rekonstruieren. Sich auf diese Verfremdungen des originalen Signals durch den Raum zu konzentrieren erfordert sich ihrer bewusst zu werden. Die Fähigkeit den Raum bewusst hörend wahrzunehmen kann aber trainiert werden. Experten dieser Fähigkeit finden sich etwa unter blinden Menschen, deren sensorisches Defizit nahezu immer durch gesteigerte auditive Fähigkeiten kompensiert wird.

Da unser auditiver Horizont annähernd kugelförmig ist, können wir mit unseren beiden Ohren jeweils die Reflexionen, die aus allen Richtungen kommen, hören. Unsere Hörbahn vergleicht die Signale, die unsere beiden Ohren erreichen, was es unserem Gehirn erlaubt, die Umwelt buchstäblich räumlich rekonstruieren. *„In the superior colliculus, a midbrain nucleus involved in the control of orienting behaviour, visual, auditory, and tactile inputs converge to form superimposed maps of sensory space. Maps of visual space and of the body surface arise from spatially ordered projections from the retina and skin, respectively. In contrast, the map of auditory space is computed within the brain by tuning the neurons to different localization.“* (King, 1999, S.900) Man vermutet, dass die colluculi inferiores innerhalb der Neuronenbahnen des Gehirns ein maßstäbliches Miniaturmodell der akustisch gehör-

ten Umgebung generieren. Dies passiert noch bevor die gehörten Informationen zur Vorauswahl in den Thalamus und in der Folge zum primären Auditiven Cortex zur bewussten Verarbeitung weitergeleitet werden. Als allererstes – und noch vor jedem Warnsignal, emotionalen „Trigger“ oder sinnbehaftetem Inhalt – hören wir also Raum. (Vgl. Kapitel 1)

Die menschlichen Fähigkeiten Raum zu er-hören dient und beeinflusst uns gleichermaßen und auf vielfältige Art und Weise. Neben der bewusst wie unbewusst zum Einsatz kommenden navigatorischen Aspekte hat sie auch Einfluss auf unsere Emotionen und unser soziales Miteinander. So wie jeder Klang sanft oder kräftig, düster oder fröhlich und entspannt oder energetisch auf uns wirken kann, hat auch der Klang eines Raum(bereich)es eine akustische Färbung, die auf uns einwirkt und so unsere Stimmung beeinflusst. Beispielsweise kann eine Struktur, die Hintergrundlärm verstärkt, auf der einen Seite Stress erhöhen, während auf der anderen Seite gefühlte Privatheit reduziert wird. Eine große, lange nachhallende Halle (!) erschwert die Kommunikation und kann ein Gefühl der Bedeutungslosigkeit hervorrufen, wenn man beim Sprechen gezwungen ist, die Stimme zu senken, um nicht im Nachhall unterzugehen. Ein Raum, der mit Teppichen gedämpft ist, wird kaum Nachhall aufweisen und so das Gefühl der Intimität steigern, wenn sogar weiter voneinander entfernte Personen einander gut verstehen können ohne die Stimme erheben zu müssen. Barry Blesser und Linda Salter haben in ihrem Buch „Spaces Speak, are you listening?“ den Terminus „*Auditory Spatial Awareness*“ (auditives räumliches Bewußtsein) eingeführt, um die vielfältigen Phänomene zu beschreiben, die die akustische Gestaltung einer räumlichen Struktur auf Individuen und Sozialgefüge haben kann. Unter ihren Gesichtspunkten wird aus der Perspektive der „Hörer“ aus akustischer Architektur *aurale Architektur* – ein Begriff der sich der soziokulturellen Aspekte einer gestalteten Klangwelt annimmt. (Vgl. Kapitel 3)

Während sich traditionellerweise die Akustik in der Architektur vor allem mit leicht messbaren und begreiflichen Zielen wie Schallschutz und Verständlichkeit beschäftigt<sup>4</sup>, versucht die Aurale Architektur eine adäquate Begrifflichkeit für Aspekte zu finden, welche die eigentlich architektonischen Aspekte des Klangs von Raum darstellen, aber noch kaum Eingang in das kollektive Bewusstsein gefunden haben: die Gestaltung von Klangraum als Umgebung, die unser tägliches Leben strukturiert. So wie Schafer & Co es geschafft haben, den Begriff Soundscape populär und die akustische Umweltverschmutzung bewusst zu machen, ist für die Entwicklung eines Bewusstseins für die klangliche Gestaltung unserer passiven Umgebung noch Arbeit nötig. Elektronische Medien bieten uns vielfältige und günstige Möglichkeiten, um mit diesem Feld zu experimentieren und dieses neue Bewusstsein zu schaffen.

---

<sup>4</sup> Obwohl selbst daran Zweifel bestehen, wenn man die große Zahl an aural dysfunktionalen Klassenzimmern, Lesesälen, Büroräumen, Konzertsälen, Opernhäusern etc in Betracht zieht



Wie ich versucht habe hier zu erklären, ist Schall immer körperlich und als Konsequenz immer räumlich. Als solcher stellt er eine körperliche Verbindung zum Raum dar. Weiters ist er eine Informationsquelle, die uns dazu dient, die sich ständig verändernde Konfiguration der Welt um uns zu erfassen. Hörend können wir unsere Umwelt ertasten. Außerdem kann Schall auf einer persönlichen Ebene unsere Emotionen und auf einer zwischenmenschlichen unser soziales Gefüge beeinflussen. Verschiedene Industrien haben sich die Macht des Schalls schon lange zunutze gemacht, aber jedoch fehlt bis heute ein breites gesellschaftliches Bewusstsein für die weitreichenden Konsequenzen dieser Macht. Die passiven auralen Eigenschaften unserer Umgebung sind nicht so einfach erfassbar wie ein spezifischer aktiver Klang (beispielsweise ein Motor oder ein „Sound-Logo“). Diesem mangelnden Bewusstsein ist auch die gegenwärtig geringe Aufmerksamkeit bei der Gestaltung auraler Aspekte von Architektur geschuldet. Vermehrte Investitionen in diesem Bereich würden einen breiten gesellschaftlichen Konsens über ihre Bedeutung und die resultierenden Konsequenzen erfordern. Neue Technologien würden uns zum einen erlauben, diesen Konsens mit Hilfe kostengünstiger Eingriffe und Simulationen herzustellen, zum anderen würden sie Experimente ermöglichen die zeigen könnten, dass nicht nur Raum Klang determiniert, sondern Vice Versa Klang selbst Raum schafft.

Während Schall in der materiellen Welt von räumlichen Strukturen geformt wird, bieten uns digitale Technologien Werkzeuge, um räumliche akustische Umgebungen zu entwerfen, die keinen direkten physischen Ursprung haben abseits der Schaltkreise des Rechners und der vibrierenden Membranen des Wiedergabesystems. Schallereignisse können von Reisen durch Räume und Welten erzählen, die sie unserem Verständnis nach gar nie zurückgelegt haben. Nun ist die Erzeugung von künstlichen Klangwelten kein gänzlich neues Phänomen, denn bereits im analogen elektronischen Zeitalter waren teils plausible Ergebnisse machbar. Auch Methoden um künstliche Schallfelder aufzubauen gehen der Durchdringung unserer Gesellschaft mit Computern bereits um Jahrzehnte voraus. Aber aurale Immersion beruht nicht nur auf der Glaubwürdigkeit und Genauigkeit der erzeugten Klangumgebung, denn unsere akustische Umgebung kann nicht von uns „Zuhörern“ getrennt werden. Mit jeder Aktion tragen wir zu ihrem Zustand bei – ja mit unserer bloßen Anwesenheit verändern wir sie bereits. Wahre Immersion kann also nur erzielt werden durch die Interaktion mit dieser künstlichen Umgebung. Es reicht nicht, die künstliche Umgebung zu simulieren. Unsere *eigene* Präsenz muss in ihr simuliert werden, um an dieser künstlichen Umgebung tatsächlich Anteil nehmen zu können. Erst durch den Aufbau eines auralen Regelkreises aus Aktion und Reaktion wird es möglich, die künstliche Umgebung zu betreten und durchwandern. Digitale Technologien erlauben uns einerseits, diese Feedback-Schleife zwischen uns und der künstlichen Welt herzustellen und andererseits, die materielle Welt aural mit künstlichen Räumen zu verschränken und zu erweitern.

Die resultierenden Möglichkeiten sind nahezu endlos. Existierende Räume könnten aural modifiziert werden; architektonische Strukturen gehört (und aural gespürt) werden bevor sie gebaut sind; ferne Orte besucht werden; ein Fenster in einen fensterlosen Raum gesetzt werden; eine Kathedrale inmitten einer Küche erstehen; entfernte Räume in einen einzigen Raum verschmolzen werden; ein Zimmer unter Wasser gesetzt werden; zwei Räume zur gleichen Zeit betreten werden. Neue Welten könnten erstehen. Unmögliche – aurale – Geometrien realisiert werden.

Auf der visuellen Seite können immersive künstliche Umgebungen bereits sehr effektiv mit neuen Produkten<sup>5</sup> erzeugt werden. Authentische aurale Immersion ist dagegen ungleich schwieriger zu erreichen, denn um Raum zu hören müssen die eigenen akustischen „Lebenszeichen“ – seien es Fußtritte, Rascheln der Kleidung oder Stimmen – eine Antwort in der künstlichen Umgebung erhalten.<sup>6</sup> Um das zu bewerkstelligen müssten diese „Äußerungen der eigenen Anwesenheit“ erfasst, isoliert und integriert *oder* vermessen, klassifiziert und (re)synthetisiert werden. In beiden Fällen handelt es sich um hoch komplexe Vorgänge. Sollten wir sie meistern, tun sich aber gänzlich neue Möglichkeiten auf. Völlig virtuelle Welten werden uns zur Erkundung offen stehen, die wir physisch erfahren werden können. Diese Welten werden im Mikrokosmos unserer Neuronen erstehen, ganz wie es solche tun würden, die aus Stein, Glas Holz oder Erde sind.

---

5 Wie etwa Oculus Rift, SteamVR / HTC Vive, Microsoft Hololens oder Sony PlaystationVR (Vgl. Kapitel 4)

6 Eine visuelle Analogie wäre die Begegnung mit dem eigenen Spiegelbild.

## Räumlich Hören

### 1.1 Licht & Schall

Licht und Schall verhalten sich auf vielerlei Art ähnlich: Beides sind Schwingungen, die sich, einmal von einer Quelle abgegeben, in Wellen fortbewegen. Treffen sie auf ein Hindernis, so werden sie von diesem reflektiert. Damit enden aber die Gemeinsamkeiten, denn Schall und Licht sind doch grundlegend verschiedene Phänomene. Licht besteht aus Transversalwellen die orthogonal zur Ausbreitungsrichtung oszillieren, während Schall aus eine Longitudinalwellen besteht, die in Ausbreitungsrichtung schwingen. Der markanteste Unterschied zwischen den beiden Phänomenen liegt aber darin, woraus sie sich zusammensetzen, denn Licht ist elektromagnetische Strahlung und besteht aus Teilchen, den Photonen, die zur Ausbreitung kein Medium erfordern und sich deshalb auch im Vakuum des Weltraums fortbewegen können. Zur Beschreibung einer elektromagnetischen Welle benötigt man in jedem Punkt eine elektrische Feldstärke (Vektor) und eine magnetische Feldstärke (Vektor), die aufeinander und zur Ausbreitungsrichtung senkrecht stehen. Verknüpft sind beide durch die sogenannten Maxwell-Gleichungen. (Westphall, 1958, S.447) Im Gegensatz dazu besteht Schall nicht aus Teilchen, sondern ist vielmehr die physische Verformung von Materie selbst, benötigt daher zur Ausbreitung also auch einen materieerfüllten Raum. Bei Schall führen also die Partikel des Trägermediums die Schwingungen aus. Zur Beschreibung benötigt man in jedem Punkt die Schallschnelle (Vektor, der die Geschwindigkeit der Materie-Teilchen darstellt) und den Schalldruck (ein Skalar). Diese sind durch die Bewegungsgleichung der Mechanik (Kraft ist Masse mal Beschleunigung) verknüpft. (Westphall, 1958, S.25)

Schall ist also Veränderung im Zustand der materiellen Umgebung. Spürbar wird dies bei sehr niederfrequentem Schall, den wir als Erschütterung wahrnehmen können. Schall ist somit ein grundlegend räumliches Phänomen. Es handelt sich um periodische Druckveränderung, die nicht nur

die uns umgebende Luft betreffen, sondern jede Form von Materie. Je nach Steifigkeit des Trägermediums breitet sich Schall verschieden schnell aus, wobei sich Schwingungen mit verschiedenen Wellenlängen (also unterschiedlichen Frequenzen) je nach Materialeigenschaften verschieden gut fortbewegen können. Dabei bewegt sich Schall nicht im sprichwörtlichen Sinne weiter. Viel eher muss man sich sein Verhalten vorstellen wie die Energieweitergabe bei kinetischem Spielzeug wie etwa einem Kugelstosspendel. Ein Molekül übt Druck auf das Nachbar-Molekül aus, welches diesen wiederum an den nächsten Nachbarn weitergibt, etc. In der Luft breitet sich Schall beispielsweise kugelförmig aus. (Kuttruff, 2009, S.35) Bei dieser Weitergabe des Drucks entstehen Reibungsverluste, wodurch der Schall mit zunehmender Distanz leiser wird. (Sabine, 1923, S.9) Diese Reibungsverluste sind in der Luft weitaus höher als in Wasser, weshalb sich Schall unter Wasser mehr als 4 mal so schnell ausbreitet.

Obwohl unsere Welt durchdrungen von komplexen Schallwellen ist, lässt sich jede dieser Wellen als Summe verschiedener Sinuswellen beschreiben. Die Gelegenheit reine Sinusschwingungen wahrnehmen zu können, haben wir in der Natur allerdings praktisch nicht. Manche Musikinstrumente erzielen annähernd reine Sinustöne, etwa das Klavier, allerdings unter der Bedingung, dass jeweils nur ein Ton erzeugt wird. Erklängen etwa mehrere Klavierseiten gleichzeitig, überlagern sich die Sinusschwingungen und resultieren in komplexen Wellen. Tatsächlich kann man jede komplexe Welle als Summe von Sinus-Partialtönen sehen.<sup>1</sup> Dies ist insofern von Bedeutung, als in der Umkehrung dieses Prinzips die Funktionalität des Ohres zu finden ist, denn während die Klänge der Welt sich aus Sinuswellen zu komplexen Wellen zusammensetzen, zerlegt unser Ohr diese wieder.

## 1.2 Hören & Sehen

Hören und Sehen gehören zu den externen Sinnen des Menschen, deren Aufgabe es ist, die Umwelt zu überwachen, um Aktionen zu veranlassen, die das Überleben gewährleisten sollen. Aufgrund der Ausrichtung nach außen werden viele sensorische Zeugnisse unseres eigenen Körpers von unseren Sinnen ausgeblendet. So ist etwa unsere Nasenspitze nicht ständiger prominenter Protagonist unseres Gesichtssinns und wir nehmen auch unsere körpereigenen Geräusche die meiste Zeit über nicht bewusst wahr. Stattdessen ist es die Aufgabe dieser Sinne unsere Umgebung zu erfassen. In dieser Funktion sind diese beiden Sinne inhärent räumliche Sinne. Die unterschiedlichen Eigenschaften von Schall und Licht haben aber zur Folge, dass die räumliche Abbildung von Gesichtssinn und Hörsinn grundlegend anders zustande kommen.

---

<sup>1</sup> Das Kunstprojekt „Speaking Piano“ von Peter Ablinger macht sich dieses Prinzip zunutze. Aufgenommene Sprache wurde hier in Teilschwingungen zerlegt und mit dem chromatischen Tonumfang des Klaviers abgestimmt. Servomotoren betätigen die Klaviatur entsprechend der aus Sprache bestehenden Partitur, welche durch die Töne des Klaviers hörbar wird. Die Tonauflösung des Klaviers ist durch seine 12 Töne pro Oktave sehr beschränkt jedoch ausreichend um annähernde Verständlichkeit zu erzielen. (Ablinger, 2009)

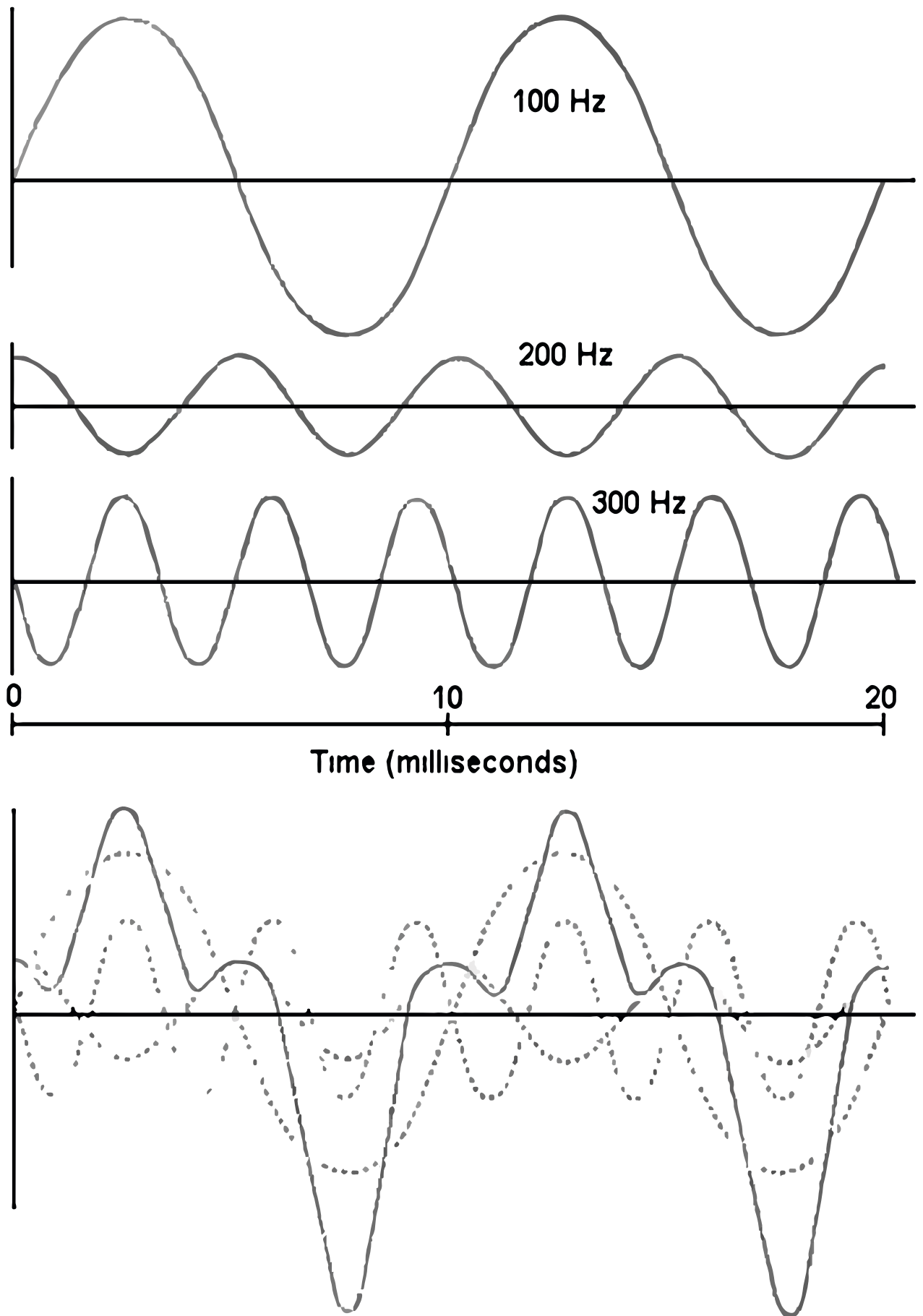


Abbildung 1: Addition von Sinuswellen ( Mannell, 2015)

Sehen und Hören funktionieren jeweils durch die Übersetzung von Wellenfronten in neuronale elektrische Impulse. Auge und Ohr benötigen jeweils zwei Stück ihrer Art um räumliche Informationen wahrnehmen zu können, allerdings unterscheidet sich ihre Funktionsweise grundsätzlich. Als Lichtwellen bezeichnet man den sichtbaren Bereich des elektromagnetischen Spektrums zwischen  $3,8 \cdot 10^{14}$  und  $7,9 \cdot 10^{14}$  Hz. Schallwellen sind für den Menschen im Bereich von 20 Hz – 20.000 Hz hörbar, wobei ältere Menschen graduell unempfindlicher für hohe Frequenzen werden. Eine Welle mit einer Frequenz von 1 Hertz vollendet einen ganzen Schwingungszyklus in einer Sekunde. Eine Welle mit 20.000 Hz muss in dieser Sekunde 20.000 mal schwingen. Anhand dieser Zahlen wird klar, dass Licht um viele Potenzen schneller schwingt und schon allein aus diesem Grund kaum mechanisch erfassbar sein kann wie dies etwa beim Hören der Fall ist.

Stattdessen ist Sehen ein chemischer Prozess. Das Auge erfasst eine Vielzahl von Lichtstrahlen, die die Hornhaut und Linse in geraden Linien passieren, um von den Nervenzellen auf der Retina simultan erfasst zu werden. Das Resultat ist ein Abbild der Welt, das bereits räumlich auf einer zweidimensionalen Fläche ausgerichtet ist. Die weitere Interpretation dieses 2D-Bildes ist Aufgabe des visuellen Cortex. Das Ohr andererseits muss mit einer Mischung aller Schallwellen arbeiten, die es erreichen. Das Innenohr weist eine komplexe Mechanik auf, um dieses Klanggemisch in spektrale Komponenten zu zerlegen, allerdings keine Mittel, um unterschiedliche Klangquellen räumlich voneinander zu trennen. Diese Arbeit wird zur Gänze von der Hörbahn bewerkstelligt, die die gewonnenen Daten trennen und sortieren muss, um die ursprünglichen Klangquellen aus der Summe von überlagerten Geräuschen zu rekonstruieren. Während das Auge also bereits räumlich „geordnete“ Informationen zur Interpretation an den visuellen Cortex leiten kann, muss die Hörbahn ungleich mehr leisten, um zu einer vergleichbaren Ausgangslage in der „Datenverarbeitung“ zu kommen. Um eine Vorstellung von der Leistung des Gehörs zu bekommen, sei hier eine Analogie erlaubt: Würde das Auge nach den Prinzipien des Ohres funktionieren, könnte es nur einen einzigen Farbpunkt erfassen. Aus diesem müsste das Gehirn alle Lichtquellen sowie die Oberflächen, an denen diese am Weg zum Auge reflektiert wurden, rekonstruieren. Es müsste aus einem einzigen Bildpunkt ein zweidimensionales Bild errechnen können und um eine dreidimensionale Umgebung zu rekonstruieren, stünden ihm nur zwei Bildpunkte zur Verfügung.

Aber auch in der Abbildung und Verarbeitung unterscheiden sie sich grundlegend. So ist der Bereich, den wir optisch erfassen können, ein kegelförmiger, während unser Gehör eine annähernd sphärische Richtcharakteristik aufweist. Das bedeutet, dass wir in unserem auditiven Horizont zentriert sind. Während unsere Augen also auf ein Ziel gerichtet werden müssen, um es erfassen zu können, befinden sich unsere Ohren in einer permanenten „Alarmbereitschaft“ und können unsere Umgebung relativ ungerichtet und gleichmäßig erfassen. Obwohl es nicht nur unser bewusster

Wille ist, der den Fokus unseres Auges steuert, sondern auch neurologische Automatismen hierbei eine Rolle spielen, wird das Auge doch durch Muskelkraft gesteuert, während der auditive Fokus ausschließlich ein Produkt neuronaler Aktivität ist. Die Aktivität unseres Gehirnes zu kontrollieren ist ungleich schwieriger als die Kontrolle unserer Muskel zu trainieren.<sup>2</sup> „Schärfe“ im Fokus der Vision ist ein physikalischer Zustand für den es keine Entsprechung im Hören gibt. Visuelle Elemente in „idealer“ Sichtdistanz können fokussiert oder unscharf sein, je nach physischem Zustand unserer Augenmuskeln und unserer Linsen. Akustische Elemente in „idealer“ Hörweite sind alle gleichermaßen „scharf“, vorausgesetzt, dass keine Störquellen den auditiven Fokus erschweren. Das Fokussieren wird ausschließlich durch die Hörbahn in unserem Gehirn bewerkstelligt.

Allerdings bedeutet physische Schärfe des Gesehenen keineswegs, dass das fokussierte Element gleichzeitig das Zentrum der visuellen Aufmerksamkeit ist. Genauso wie beim Hören kann der Fokus im Gesehenen durch Hirnaktivität verschoben werden. Obwohl das Gehirn in der Lage ist im Ohr motorisch Frequenzbereiche zu verstärken, entfällt aber gegenüber dem Sehen die physische Komponente des Fokussierens, und somit eine Ebene mechanischer Komplexität, die hier durch neurologische Aktivität ersetzt wird. Der menschliche Blick kann einen ungefähren Bereich von 180° horizontal und 135° vertikal erfassen. Beim Fokussieren auf einzelne Elemente in diesem Sichtfeld können andere leicht übersehen werden, selbst wenn sie unweit des fokussierten Elements sind. Jeder Schüler fernöstlicher Kampfkunst lernt, dass ein Zustand visueller Wachsamkeit bewusstes De-fokussieren erfordert, also das Entspannen der Augenmuskeln und das Vermeiden des optischen Fokus auf ein bestimmtes Element, um dynamische Änderungen im ganzen Blickfeld erfassen zu können. Aber selbst mit dem Erlernen dieses Blicks bleiben Objekte, die hinter dem Kopf oder von anderen verdeckt sind unsichtbar. Licht das welches mit festen Objekten kollidiert, kann sich je nach den Eigenschaften des Materials unterschiedlich verhalten. Es kann reflektiert, absorbiert, gebrochen oder gestreut werden, ähnlich wie Klang. Aber während Licht von den meisten opaken Festkörpern blockiert wird, verhält sich Schall anders. Je nach Frequenz kann er seinen Weg durch vielerlei Materialien hindurch finden, die für Licht vollkommen undurchlässig sind. Ein Traktor, der auf der Rückseite einer Scheune „dahintuckert“ könnte von ihrer Vorderseite aus gehört werden, ohne gesehen zu werden. Das Rauschen eines Autos, das die Straße hinunter fährt, könnte im hinteren Ende einer Wohnung im zweiten Stock gehört werden, ohne dass eine Sichtverbindung besteht. Laute Musik oder Fußstritte können den Fußboden in einer Wohnung über der eigenen zum Schwingen bringen. Schall wird also nicht linear übertragen, wie dies bei Licht der Fall ist, das sich in Strahlen ausbreitet. Er breitet sich kugelförmig aus und kann an Ecken gebogen werden, verschiedene Materialien in Form von Vibrationen durchdringen oder den Hörer über vielerlei Weisen reflektiert erreichen. Visuell verdeckte Elemente können also nicht gesehen, aber

---

2      Sehkraft kann bei Kurzsichtigkeit mittels Augen-Übungen verbessert werden.

unter Umständen gehört werden, wenn sie schwingen. Überdies können die Augen geschlossen werden, was etwa in Schlafphasen generell der Fall ist, während der Auditive Cortex auch in dieser Phase aktiv bleibt. Der Grad der Reaktivität gegenüber akustischen Signalen im Schlaf variiert von Individuum zu Individuum und Schlafphase zu Schlafphase, trotzdem ist Hören auch im Schlaf ein wirksamer Sicherheitsmechanismus.

Somit ist Hören auf vielerlei Weise unmittelbarer als Sehen, weshalb man spekulieren könnte, dass der Hörsinn primär als Warnsinn von der Evolution hervorgebracht wurde. (Blesser, Salter, 2007, S.333) Gegenüber dem Sehsinn arbeitet er auch mit weitaus höherer Geschwindigkeit, worauf ich in diesem Kapitel noch zurückkommen werden. Wenn sich unsere prähistorischen Vorfahren in der Wildnis bewegten, war es von überaus großer Bedeutung für sie, ihre Umgebung präzise erfassen zu können. Ein Geräusch im Unterholz konnte jederzeit Beute oder Raubtier sein und eine falsche Reaktion konnte den unmittelbaren Tod bedeuten. Vielleicht ist dies der Grund dafür, dass wir auf akustische Signale sehr schnell reagieren können. So hat etwa unsere Hörbahn die Fähigkeit unsere optische Aufmerksamkeit zu lenken. Es existiert eine direkte Nervenverbindung zwischen dem Bereich der Hörbahn, der für die Lokalisierung zuständig ist und den Augenmuskel-Kernen, die den Blick lenken. (Berghaus et.al., 1996, S.37) Hierbei handelt es sich um Vorgänge, die weitgehend reflektorisch ablaufen. Das bedeutet, dass Reaktionen erfolgen können, noch bevor die Inhalte im Bewusstsein verarbeitet werden.

Wie bereits in der Einleitung erwähnt deuten jüngste Forschungen darauf hin, dass eine Vorauswahl von Schallinformationen im Thalamus, dem sensorischen „Türsteher“ unseres Gehirns – erfolgt. (Koelsch, 2012, S.256f.) Der Thalamus liegt im limbischen System, von dem vermutet wird, dass es für die Verarbeitungen unserer Emotionen zuständig ist. Emotionen könnten also die treibende Kraft hinter der grundlegenden Regulierung von Aufmerksamkeit und Wachsamkeit sein. (Yokosawa et.al., 2013) Der „Corpus Geniculatum Laterale“ (ein Teil des Thalamus) entscheidet, wie viele und welche Schallinformationen zur weiteren Verarbeitung an das primäre auditive System weitergeleitet werden. Dies deutet darauf hin, dass vom Thalamus auch eine Überprüfung emotionaler Resonanz auf das Gehörte durchgeführt wird. Der Thalamus entscheidet, was einerseits gefiltert oder durchgelassen wird und kann andererseits sogar selbständig Aktionen provozieren mittels Überbrückung des bewussten Entscheidungsapparats. Es handelt sich sozusagen um eine emotionale Vorauswahl, die uns dazu bringen kann, vor Geräuschen welche Furcht erzeugen zu fliehen, aber auch genauso Tränen zu vergießen zu rührseligen akustischen Klangkulissen in Blockbuster-Filmen.

So zeigen sich die vielfältigen – vor allem emotionalen und raumbildenden – Effekte von Schall auch im Medium Film. Interessanterweise wird selbiges in erster Linie als visuelles Medium wahrgenom-



men. Die Sprache verrät dies zu allererst, denn ein Film wird gemeinhin *gesehen*. Ein gebräuchlicher Ausdruck in Englischer Sprache ist auch „motion picture“. Dies ist deshalb bemerkenswert, da die Verbindung von Bild und Ton den Film eigentlich erst zu dem gemacht hat, was er ist. Auch der frühe Stummfilm wurde nicht in Stille vorgeführt, sondern mit Musik begleitet, wobei die Musiker ihre Darbietung an Dynamik, Stimmung und Intensität der Ereignisse im Film anpassten. Der Schlüssel zum Wahrnehmungsphänomen, welches diese Verbindung von Bild und Ton zu solch einem immersiven Ereignis werden lässt, wird Synchrese (engl. *synchresis*) genannt – ein Terminus, der vom französischen Komponisten und Filmtheoretiker Michel Chion geprägt wurde. (Chion, 1990) Er beschreibt das Phänomen, daß Bild und Ton einander gegenseitig ergänzen um ein funktional neues Paket zu schaffen welches andere Informationen transportieren kann als jede der Teilkomponenten für sich. Nach den Prinzipien der Gestaltpsychologie kann das Prinzip der Synchrese mit folgendem berühmten Satz auf den Punkt gebracht werden: „*Das Ganze ist etwas anderes als die Summe seiner Teile.*“ (Metzger, 1975, S.6) Chion selbst beschreibt Synchresis so: „*The spontaneous and irresistible mental fusion, completely free of any logic, that happens between a sound and a visual when these occur at exactly the same time.*“ (Chion M., 1990, S.XIII) Um sein Konzept weiter zu illustrieren führt Chion eine Vielzahl von Beispiele an die nach dem folgenden Prinzip funktionieren: Trennt man bei einem Film Ton- und Bildspur führt das Resultat in beiden Fällen zu einem Bedeutungswandel. Ton und Bild erfüllen im Film jeweils einen anderen Zweck, und sehr oft wird Körperlichkeit und Materialität mithilfe der Tonspur vermittelt: die Sinnlichkeit der Berührung eines Fingers mit einer Oberfläche wird durch die Geräusche am besten dargestellt, genauso wie die Kraft einer Explosion oder das Blasen eines starken Windes. Die enge Verbindung zwischen den auditiven und sensomotorischen Cortices in unserem Gehirn machen Klang und Berührung zu natürlichen Komplementären. Überdies dient Schall dazu eine Szene mit emotionaler Konnotation zu versehen.

Im Rahmen dieser Arbeit hat sich der Autor entschieden, Michel Chions Experimente mithilfe der Eröffnungssequenz von Stanley Kubricks Film „The Shining“<sup>3</sup> nachzustellen, in welcher eine Kamera einem Auto durch eine bewaldete Berglandschaft folgt. Die Szene wird in Weitwinkel-Luftaufnahmen dargestellt und von einem „Dies Irae“-Thema begleitet, das von Wendy Carlos am Moog Synthesizer komponiert wurde. Die Weitwinkelkamera und die karge Landschaft der Sequenz schaffen bereits eine entrückte Atmosphäre, aber die Unheil verkündende Musik machen diese erst wirklich bedrohlich. In einem Versuch im Rahmen der TU Kinderuni 2007<sup>4</sup> tauschte der Autor die Originalmusik und spielte sie einer Gruppe von Kindern zwischen 7 und 12 vor. Alpine Zitter-Musik verwandelte die Sequenz in den Augen der Kinder in ein Fragment einer anthropo-

---

3 Kubrick, Stanley (1980): The Shining. Warner Bros.

4 Hausch, Clemens (2008): Wie beeinflussen Räume den Sound?, <http://oe1.orf.at/artikel/213203>.

logischen Dokumentation über den Alpenraum. US-amerikanischen Surf-Punk wiederum ließ aus der Eröffnungssequenz eines Horrorfilms das Intro zu einer Teenager-Komödie über eine Gruppe von Freunden am Weg zum Hormon-geladenen Wochenendausflug werden.

Film in seiner jetzigen Form benötigt die Tonspur, um wie erwartet zu funktionieren und die gewünschte Immersion zu erzeugen. Die Tonspur hat gleichen Anteil an der Rezeption des Filmes wie die Bildspur, in der öffentlichen Wahrnehmung ist und bleibt Film aber primär ein visuelles Medium und Filmton ein Element der Unterstützung. Die Ursachen hierfür könnte man einerseits in der Dominanz visueller Medien suchen, andererseits in der flüchtigen Natur von Schall, welchen man nicht einfach Einfrieren kann um eine Momentaufnahme zu reproduzieren, wie dies bei Bildsequenzen möglich ist. Schall kann jedenfalls unsere Wahrnehmung beträchtlich verändern und Gesehenes in völlig anderem Kontext erscheinen lassen.

Aber Schall kann auch auf körperlichem Wege unser Wohlbefinden und somit Wahrnehmung beeinflussen. Gerüchte berichten etwa von sogenannten „Demutspfeifen“ in Kirchenorgeln, deren Infraschall den Gläubigen Ehrfurcht einflößen sollten. (Pade, 1993) Obwohl die Bezeichnung höchstwahrscheinlich aus der Gegenwart stammt und die bewusste Anwendung von Infraschall in der Kirchenmusik nicht bewiesen ist, existieren zumindest zwei Orgeln weltweit, die echte vollständige 64 Fuss Register aufweisen, bei deren grösste Pfeife eine Länge von beinahe 20 Metern beträgt. Diese kann das Subsubcontra-C erzeugen, dessen Frequenz 8,2 Hertz beträgt und damit deutlich unter der Hörschwelle liegt. Die betroffenen Orgeln stammen aber rund um das Jahr 1900 und fallen so weder in die Hochphase religiöser Orgelmusik, die im Barock lag, noch in eine Zeit gesellschaftlich tief verwurzelter Religiosität. Infraschall kann jedenfalls sowohl Unwohlsein als auch Erregung verursachen. Ein Effekt solcher Pfeifen kann nachgewiesen werden, eine Absicht dahinter allerdings nicht. (Döring, 2015)

Die Funktion des menschlichen Gehörs ist jedenfalls erst seit wesentlich kürzerer Zeit hinreichend aufgeklärt worden als dies beim Sehsinn der Fall ist. Auch heute sind noch einige Details ungeklärt, insbesondere die Hörbahn und ihre Fähigkeit, Karten der Umgebung anzufertigen sind noch Gegenstand andauernder Forschung. Aber auch die Mechanik des menschlichen Ohres wurde erst vor relativ kurzer Zeit aufgeschlüsselt. Der ungarische Biophysiker Georg von Békésy entschlüsselte Mitte des 20. Jhdts. mittels komplizierter Experimente die Funktionsweise des Innenohrs, wofür er 1961 den Nobelpreis für Medizin erhielt. (Nobelprize.org, 2015). Auch die Fähigkeit der Hörbahn aus einem Strom von Klangdaten einzelne Elemente herauszufiltern – bekannt als „*auditory stream segregation*“ und benannt von Albert Bregman – wird erst seit Ende der 1960er Jahre erforscht. Insgesamt ist das Hören also noch ein weithin unerforschter Sinn und wurde über weite Strecken

mystisch verklärt.<sup>5</sup>

### 1.3 Das Ohr

Die Aufgabe unserer Ohren ist es, Druckschwankungen in der Umgebung in elektrische Impulse zu übersetzen. Allerdings wird, anders als beispielsweise beim Digitalisieren von Audio-Material mittels eines AD-Wandlers, nicht die Amplitude der Summe des vom Ohr registrierten Schalls über die Zeit weg aufgezeichnet. Vergleichen lässt sich die Leistung des Ohrs eher mit den Ergebnissen, welche die Fourier-Analyse liefert, denn bereits im Innenohr wird der Schall in Frequenzbänder getrennt. Die Schwingungen, die unser Ohr erreichen, dringen überlagert in unseren Gehörgang ein und treffen im Mittelohr auf eine Membran, die Trommelfell genannt wird. Ähnlich wie ein Lautsprecher mittels einer Membran elektrische Potentiale durch Bewegung in Schall wandelt wird am Trommelfell Schall in elektrische Potentiale gewandelt. Mit dem Trommelfell verwachsen ist ein Knöchelchen, welches als Hammer (lat. *malleus*) bezeichnet wird. Der Hammer ist über den sogenannten Amboss (lat. *incus*) gelenkig mit dem Steigbügel (lat. *stapes*) verbunden. Erzittert das Trommelfell, wird der Hammer mitbewegt. Die Vibration überträgt sich auf den Steigbügel, der diese an das Innenohr weitergibt.<sup>6</sup> (Mathews, 1999, S.4ff.) Das Innenohr ist ein komplexes Organ, das sowohl für den Gleichgewichtssinn, als auch für das Gehör zuständig ist. Essentielle Komponente ist die Schnecke (lat. *cochlea*): Hier erfolgt die eigentliche Umwandlung des Schalls in neuronale Impulse.

Die Schnecke ist ein sich verjüngender Kanal im Schädelknochen und trägt ihren Namen daher, dass sie sich in der Form eines Schneckenhauses windet. Sie ist entlang ihrer Achse in 3 Kammern unterteilt. Abbildung 2.4 erleichtert das Verständnis durch eine schematische Darstellung. Die beiden äußeren Kammern – die Vorhoftrappe (lat. *scala vestibuli*) und die Paukentrappe (lat. *scala tympani*) – sind am hinteren (*apikalen*) Ende der Schnecke über eine Helicotrema genannte Passage miteinander verbunden und enthalten eine lymph-ähnliche Flüssigkeit, die Perilymphe. In der Mitte der Schnecke verläuft entlang ihrer Achse eine dritte, von den ersten beiden isolierte Kammer – der endolymphatische Gang oder Schneckengang (lat. *scala media*) – die mit einer weiteren lymph-artigen Flüssigkeit gefüllt ist, der Endolymphe, welche gegenüber der Perilymphe (und den meisten andern Körperzellen) mit 80 Millivolt positiv geladen ist. Sie enthält das Corti-Organ,

---

5 Diese Verklärung hält zum Teil bis in die Gegenwart an, weshalb es nicht verwunderlich ist, dass die Menge an spekulativer Literatur, die sich der Bedeutung des Hörens widmet, gewaltig ist. Als Beispiel sei hier das in den 1980er Jahren sehr erfolgreiche Werk „Nada Brahma“ des deutschen Musikkritikers Joachim-Ernst Berendt erwähnt, in dem das Hören zum eigentlichen Ur-Sinn erklärt wird, welcher Vorrang vor allen anderen Sinnen hat oder haben sollte. Auch Werke der sogenannten Toronto-Schule wie etwa Marshall McLuhans „Understanding Media“ oder R. Murray Schafers „The Soundscape“ tendieren in diese Richtung.

6 Diese „Entkoppelung“ der Übertragungsleitung besteht zum Schutz unseres Gehörs vor zu lauten Geräuschen: so kann der Steigbügel-Muskel (*musculus stapedius*) im Bedarfsfall entspannt werden, wodurch die Verbindung lockerer und unempfindlicher für Erschütterungen wird.

welches die eigentliche „Digital-Wandlung“ bewerkstelligt. (Van Bergeijk et.al., 1960, S.117ff.) Das Corti-Organ ist zur Vorhoftrappe durch die Reissnersche Membran und zur Paukentreppe durch die Basilarmembran abgegrenzt. Es enthält die inneren und äußeren Haarzellen. Letztere fungieren als „Motorzellen“ und können vom Gehirn angesteuert und zum Schwingen gebracht werden, um bei schlechten akustischen Verhältnissen das Gehörte zu verstärken.<sup>7</sup> (Mathews, 1999, S.9) Erstere dienen dazu, bei mechanischer Erschütterung die neuronalen Impulse zu erzeugen, die wir als Hören wahrnehmen.

Sowohl Vorhoftrappe als auch Paukentreppe verfügen am vorderen (*basalen*) Ende der Schnecke über eine Öffnung, die mit einer elastischen Membran verschlossen ist. Während das ovale Fenster in der Vorhoftrappe direkt mit dem Steigbügel verwachsen ist, dient das runde Fenster am basalen Ende der Paukentreppe dem Druckausgleich, da die Perilymphe nicht kompressibel ist. Dringt nun eine Vibration über den Steigbügel an das ovale Fenster, wird diese an die Flüssigkeit weitergegeben. Die (Schall)Welle wandert durch die Flüssigkeit. Das runde Fenster am entgegengesetzten Ende wird nach außen gedrückt. Je nach Wellenlänge dringt Schall weiter oder weniger weit in die Schnecke ein, da für jeden Abschnitt der Schnecke eine Art Grenzfrequenz schlagend wird. Die Basilarmembran verbreitert sich mit zunehmender Länge der Schnecke vom basalen zum apikalen Ende hin, während der Durchmesser der Schnecke abnimmt. Diese besondere Konstruktion hat zur Folge, dass sich im Verlauf der Schnecke die Schwingungseigenschaften ändern. Das Verhalten des Schalls in der Schnecke wird als Wanderwelle bezeichnet. Die Welle dringt in die Perilymphe ein und durchwandert sie bis zu der Stelle, an der die korrespondierende Resonanzfrequenz beheimatet ist.

Über diesen Punkt hinaus verliert die Welle schnell an Energie. Entspricht die Schallfrequenz der Resonanzfrequenz verlangsamt sich die Welle, wobei sie an Amplitude gewinnt und die inneren Haarzellen beginnen Impulse abzufeuern. (Mathews, 1999b, S.19) Hohe Frequenzen liegen nahe dem basalen Ende der Schnecke, tiefe Frequenzen beim apikalen Ende. Wellen verschiedener Frequenzen haben also ihre jeweils größte Amplitude an verschiedenen, aber fix zugewiesenen Orten in der Schnecke, die von der Resonanz an dieser genauen Position abhängen. Beginnt nun die Basilarmembran in einem Abschnitt zu schwingen, feuern die inneren Haarzellen Impulse ab und werden über den Hörnerv (lat. Nervus cochlearis) an die Hörbahn weitergegeben. Das Ohr hat somit mit mechanischen Mitteln den Schall in Frequenzbänder aufgeteilt. Da die Haarzellen direkt in den Hörnerv münden, ist beim Empfang eines Impulses über eine bestimmte Nervenleitung schon festgelegt, welche Frequenz gehört wurde.

---

<sup>7</sup> Störungen im Nervensystem können die äußeren Haarzellen dazu bringen permanent zu schwingen, ein Phänomen, das wir als Tinnitus bezeichnen.

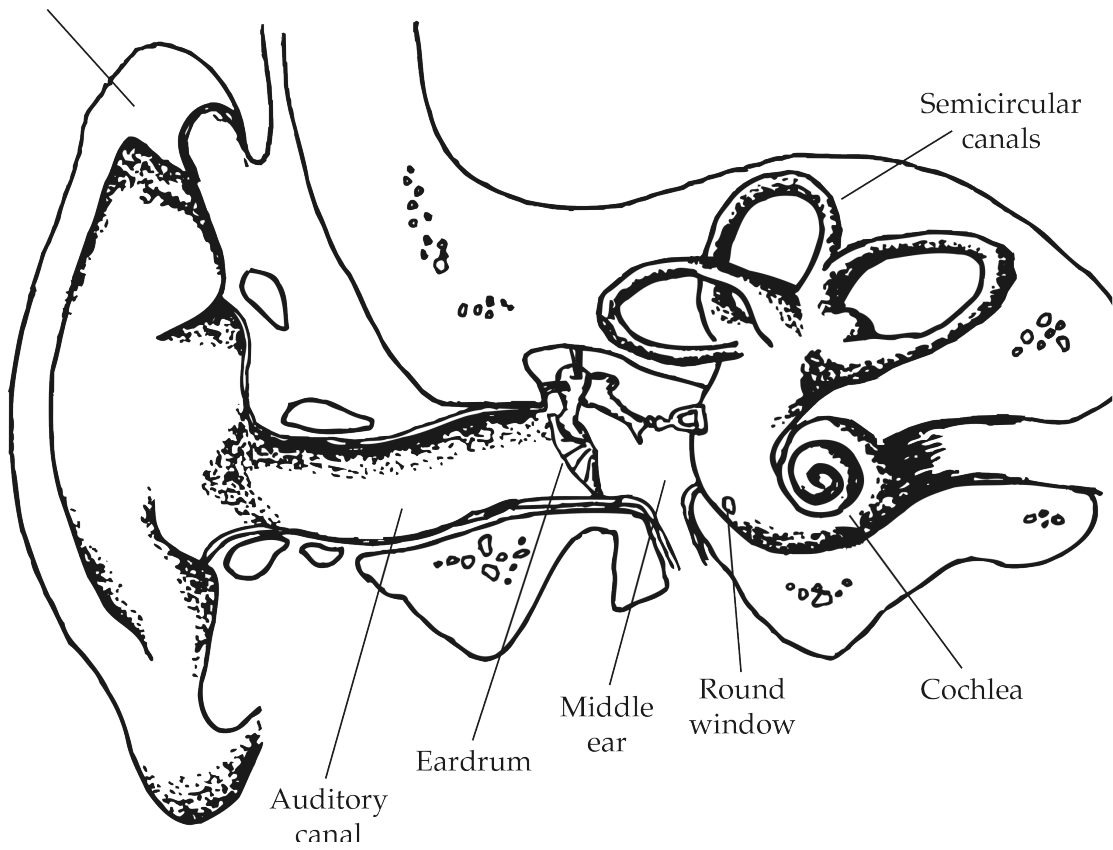


Abbildung 2.2: Schalleitungs & -empfindungssystem (Everest, Pohlmann, 2009, S.40)

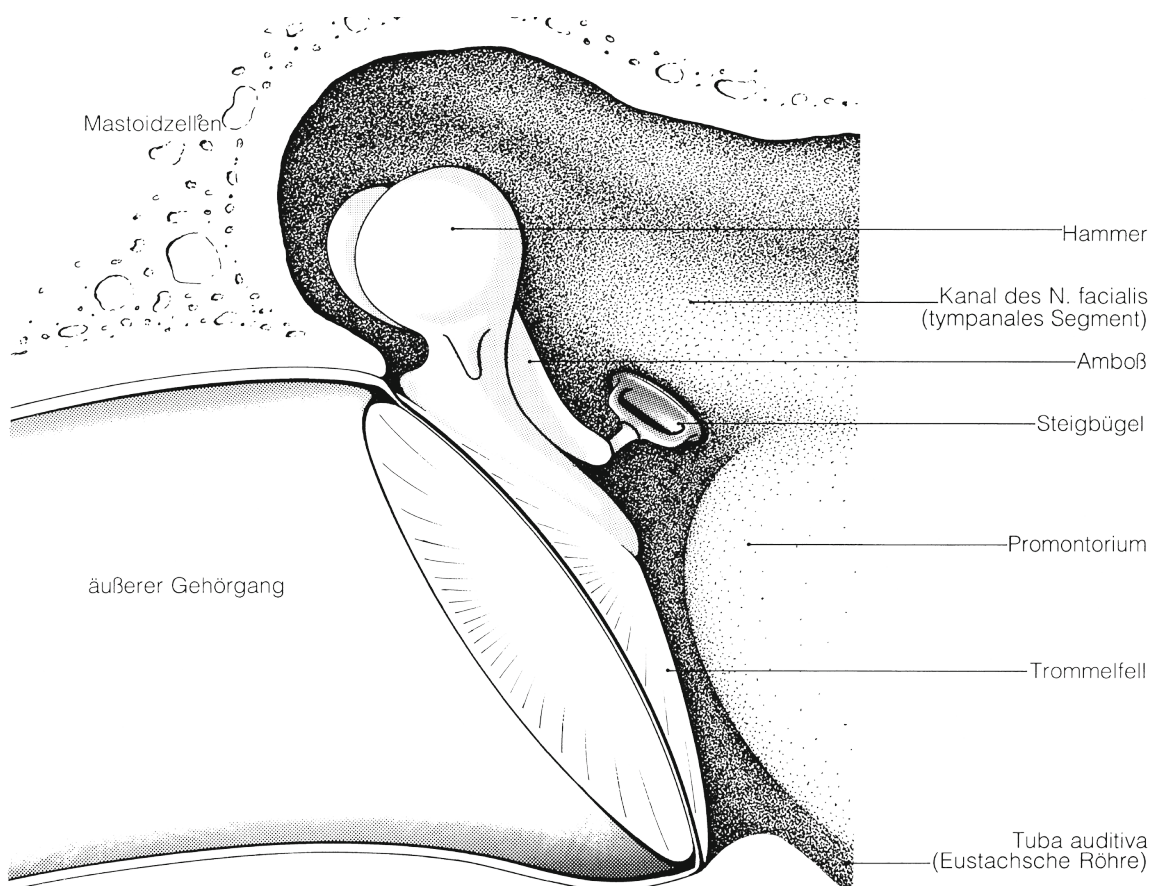


Abbildung 2.3: Trommelfell, Hammer, Amboß und Steigbügel (Berghaus, 1996, S.28)

## 1.4 Die Hörbahn

Es ist erstaunlich, wie gut räumliches Hören beim Menschen funktioniert, wenn man bedenkt, dass unsere beiden Ohren mit „Mono“-Signalen arbeiten müssen. Während das Auge theoretisch den gleichen Farbton an vielen verschiedenen Stellen im Gesichtsfeld erkennen kann, bietet das Ohr dem Gehirn zur Rekonstruktion der Umgebung nur eine Art Spektrogramm (Vgl. Abbildung 2.7) des Gehörten an. Erkennbare Faktoren sind also nur Intensität, Zeitlichkeit und Frequenz. Mehrere identische Frequenzen von verschiedenen Quellen werden im gleichen Abschnitt der Schnecke abgebildet und können also theoretisch nicht unterschieden werden. Wie wird aus diesen Nerveimpulsen eine räumliche Rekonstruktion unserer Umgebung?

Zuallererst sei erklärt, aus welchem Grund es überhaupt möglich ist, aus zwei „Kanälen“ mit Klanginformationen ein dreidimensionales Modell von Raum abzuleiten. Der Schlüssel hierzu liegt in der Form unseres Kopfes und der getrennten räumlichen Lage unserer Ohren. So sind es drei Faktoren, die ausschlaggebend sind: Spektrum, Intensität und Zeitlichkeit. Schall, der von links auf einen Hörer einströmt, wird auf dessen rechtem Ohr vom Kopf abgeschattet, was Änderungen im Frequenzspektrum des Gehörten mit sich bringt. Die Besondere Form von Ohrmuschel und Gehörgang verändern den Klang weiter. Diese Veränderungen werden mittels der sogenannten „*head related transfer function*“ (HRTF) beschrieben. Auch wird ein solcher Klang am linken Ohr lauter vernommen werden als am Rechten.<sup>8</sup> Zu guter Letzt kommen aufgrund des räumlichen Abstands unserer Ohren zueinander Schallereignisse, die die selbe Ursache haben je nach Lage unseres Kopfes zeitlich verzögert in beiden Ohren an. (Blauert, 1997, S.138) Die spektrale Analyse des Klangs erlaubt bis zu einem gewissen Grad sogar die Lokalisierung von Klängen mit nur einem Ohr. In Kombination mit Zeit- und Intensitätsunterschieden der Signale in linkem und rechtem Ohr erlaubt es dies unserer Hörbahn, die räumliche Umgebung zu erfassen.

Die Hörbahn ist symmetrisch aufgebaut und besteht aus einem komplexen Netzwerk aus Nervenleitungen in der weißen Substanz des Gehirns, die die Impulse an als Verarbeitungsknoten fungierende Zellkerne, bestehend aus grauer Substanz, weiterverteilen. Dabei werden die Leitungen, die vom rechten und linken Ohr ausgehen, im Laufe der Verarbeitung immer wieder verschränkt, um mit dem linken und rechten Ohr empfangene Reize vergleichen zu können und damit das eigentlich räumliche Hören zu ermöglichen. Hierbei sei angemerkt, dass die „Rekonstruktion“ unserer auditiven Umgebung nicht zentral in einem Zellkern geschieht, sondern im Laufe der verschiedenen Stationen der Hörbahn sukzessive erweitert wird. Hierzu entstehen „Karten“ der Umgebung in unterschiedlichen Repräsentationsformen.

---

<sup>8</sup> Bei sehr hohen Laustärken setzt diese Funktionalität aus, da ab einem bestimmten Schwellenwert keine graduellen Intensitätsunterschiede mehr registriert werden können, was zum vorübergehenden Verlust des Orientierungssinns führt.



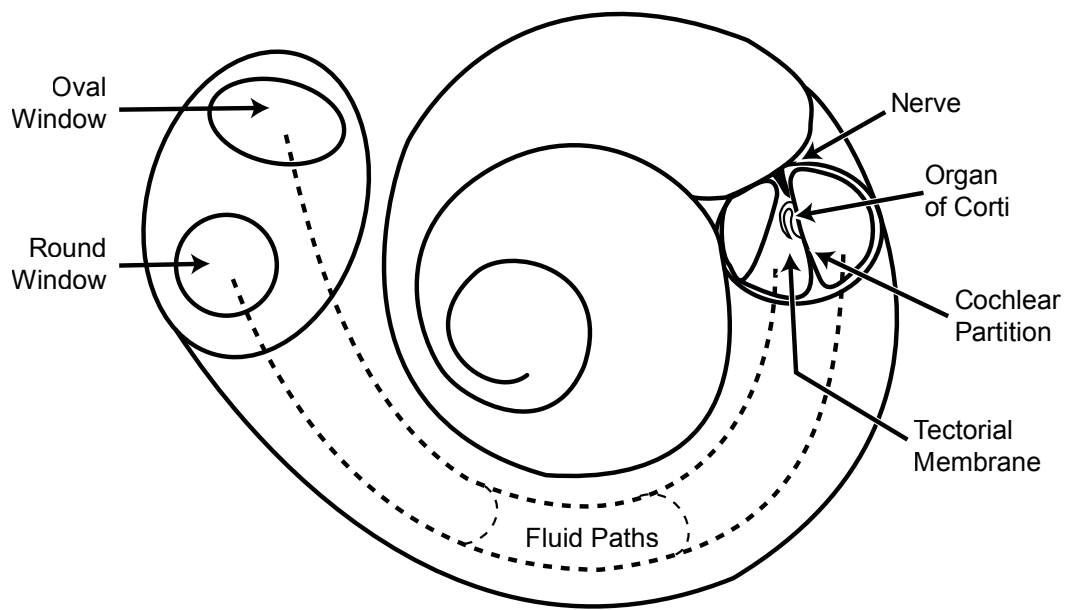


Abbildung 2.4: Die Hörschnecke (Cochlea). (Liberman et. al. 2002)

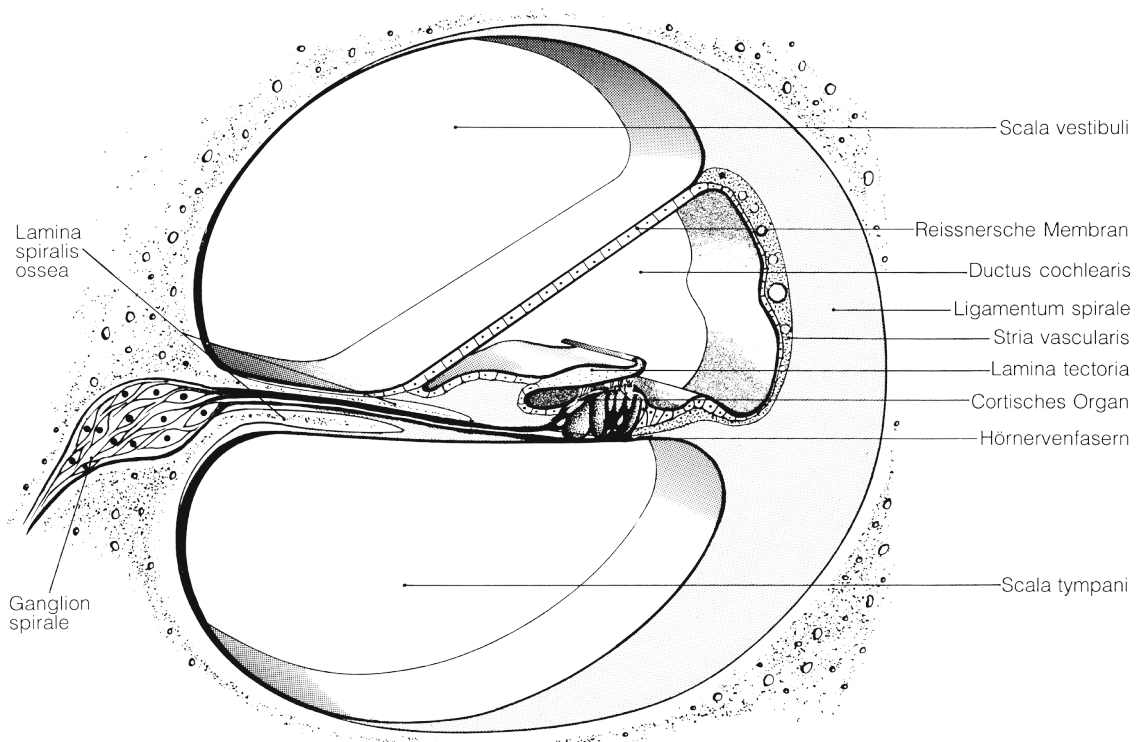


Abbildung 2.5: Schnitt durch die Schnecke (Berghaus, 1996, S.33)

Der Hörnerv überträgt die von der Schnecke spektral getrennten Frequenzen als Impulse an den Nucleus Cochlearis (Schneckenkern), der im Hirnstamm sitzt. Im Schneckenkern erfolgen erste Analysen des Gehörten. So werden hier beispielsweise monaurale (d.h. einkanalige) Informationen verarbeitet und es erfolgt eine Modulationsfrequenz-Analyse. Vom Schneckenkern aus verlaufen ein Teil der Nervenfasern zum Colliculus Inferior, der als wichtigste Schaltzentrale der Hörbahn verstanden werden kann und ein Teil zum Oberen Olivenkomplex (lat. *nucleus olivaris superior*), der die Hauptarbeit bei der Rekonstruktion der auditiven Umgebung leistet. Hier geschehen vor allem zweierlei Dinge: Einerseits werden Impulse, die auf äquivalenten Nervenleitungen vom linken und rechten Ohr empfangen werden – also Schallinformationen der gleichen Frequenz enthalten – auf ihre Intensität hin überprüft, andererseits werden asynchron eintreffende Impulse auf diesen Leitungen auf ihren zeitlichen Abstand hin verglichen. Die jeweiligen resultierenden Differenzen werden ITD (engl. Inter-aural Time Difference) und ILD (engl. Inter-aural Level Difference) genannt. ITD und ILD Messungen funktionieren auf verschiedenen Frequenzbändern unterschiedlich gut. So nutzt das menschliche Gehör zur Lokalisierung von Schallquellen bei Frequenzen unter 800 Hz vor allem ITD-Messungen, während Frequenzen über 1600 Hz in erster Linie mittels ILDs lokalisiert werden. Vom Oberen Olivenkomplex werden die stark aufbereiteten Hörimpulse nun an den Colliculus Inferior übertragen, der mithilfe der nun vorliegenden binauralen (also zweikanaligen) Daten eine *topologische* Karte der Umgebung anfertigt. (Cohen, Knudsen, 1999, S.130)

Anhand der Differenzen zwischen den Impulsen des linken und rechten Ohrs kann die Richtung des Gehörten in Relation zum Kopf relativ genau bestimmt werden. Der Schlüssel zu dieser Leistung liegt in der zeitlichen Leistungsfähigkeit unserer Hörbahn. Während der Sehnerv ungefähr 770.000 – 1.000.000 Nervenfasern enthält, weist der Hörnerv nur ca. 30.000 Nervenfasern auf. (Jonas et. al. 1992; Spöndlin, Schrott, 1989) Unsere Augen erfassen also von der Menge der Daten her gesehen ein Vielfaches dessen, was unsere Ohren aufnehmen können, dafür benötigt unser Gehör eine wesentlich höhere Geschwindigkeit, um aus dem Gehörten vergleichbare Informationen gewinnen zu können. Analog zur Datenverarbeitung von Audio und Video in der Informatik könnte man sagen, dass das Auge riesige Datenblöcke langsam zum Gehirn bewegt, während unser Ohr einen kontinuierlichen Strom an kleinen Datenpaketen sehr schnell verarbeitet.

Unsere Hörbahn ist also wesentlich zeitkritischer als unser visuelles System. Der Olivenkomplex muss Laufzeitdifferenzen zwischen den Signalen, die am linken und am rechten Ohr ankommen auswerten, die so kurz wie 10 Mikrosekunden sein können, um Positionen von Klangquellen zu extrapolieren. Dies erlaubt es, Klangquellen auf ein Grad genau im Hörbereich zu lokalisieren. (Pierce, 1999, S.98f.) Unser visuelles System erkennt ab 16 Bildern pro Sekunde flüssige Bewegungen, während es maximal 60 Bilder pro Sekunde verarbeiten kann, was einer Taktung von ca.



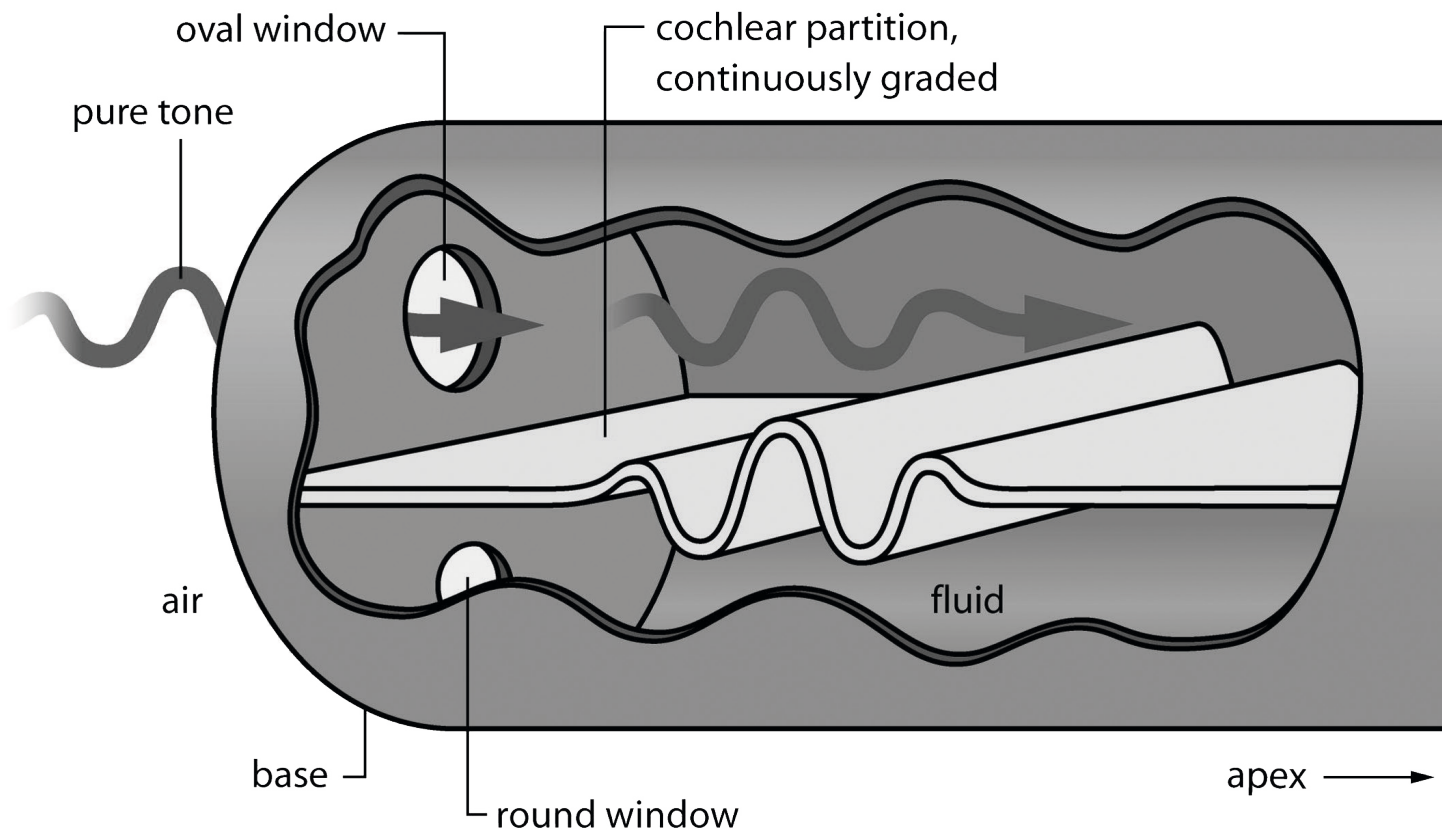


Abbildung 2.6: Bewegung der Wanderwelle an der Basilarmembran (Bell, 2004, S.1522)

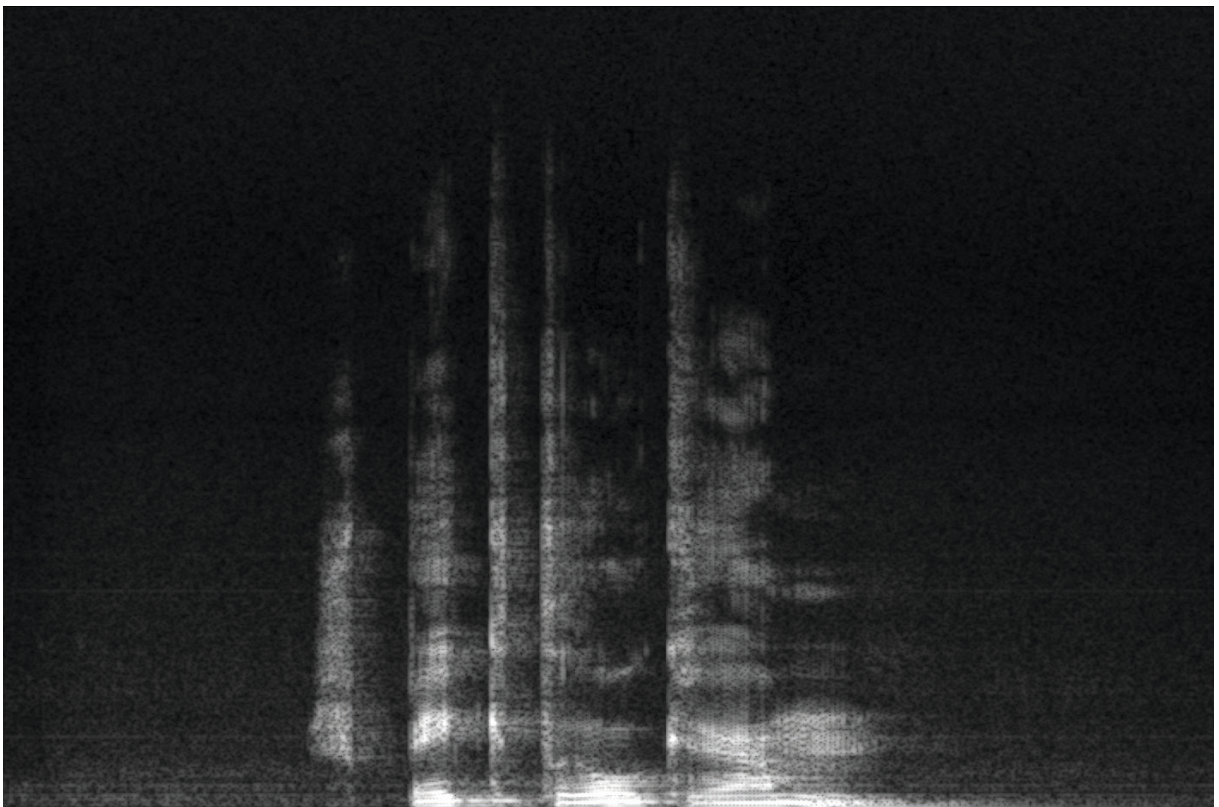


Abbildung 2.7: Spektrogramm einer Aufnahme des Wortes "Spektrogramm", aufgenommen und bearbeitet vom Autor.

16.6 Millisekunden gleichkommt. Bilder die für eine kürzere Dauer angezeigt werden, können vom Gehirn nicht mehr richtig verarbeitet werden und können sogar unbemerkt bleiben, während unser Gehör beispielsweise Unterbrechungen von bis zu 8 Millisekunden in einem kontinuierlichen Strom von weißem Rauschen identifizieren kann, was einer Rate von 125 Bildern pro Sekunde entspricht. Impuls-Klicks wiederum können bereits bei einem zeitlichen Abstand von 2 Millisekunden als unabhängige Geräusche identifiziert werden.<sup>9</sup> (Pierce, 1999, S.102)

Im Colliculus Inferior liegen die gehörten Geräusche nun also als *tonotope* Karte (bestehend aus den von der Schnecke getrennten Frequenzbänder), *topologische* Karte („errechnet“ vom Colliculus Inferior) und Modulationskarte (bereitgestellt vom Nucleus Cochlearis) vor. Von hier werden die Informationen einerseits an den Corpus geniculatum mediale weitergeleitet, der im Thalamus liegt, und andererseits an den Colliculus Superior, der dafür zuständig ist, die sensorischen Karten verschiedener Sinne zusammenzuführen und zu synchronisieren. (King, 1999) Der Thalamus, der als sensorischer Filter unseres Bewusstseins fungiert, leitet das Gehörte und Verarbeitete an den primären auditiven Cortex weiter, wo die bewusste Verarbeitung geschieht. Ein interessanter Aspekt des Colliculus Superior ist, dass er ermöglicht, die vorliegenden auditiven Karten nachträglich von den optischen und sensomotorischen Karten zu „überschreiben“. (Rees, 1996; King, 2002) Über die Gründe hierfür kann man nur spekulieren. Höchstwahrscheinlich ist aber diese Fähigkeit unseres Gehirns dafür verantwortlich, dass wir beim Ansehen eines Filmes auch dann den Eindruck haben die Dialoge direkt aus den Mündern der Schauspieler zu vernehmen, wenn in Wirklichkeit ein Mono-Lautsprecher links von der Leinwand den Ton bereitstellt. (Vgl. Teil 2, Kapitel 6.2) Was auffällt, wenn man diesen „Signalfluss“ der Hörbahn studiert ist, dass räumliches Hören einen vorrangigen Stellenwert in der Klangverarbeitung hat. So ist die räumliche Rekonstruktion der Umgebung der erste „Arbeitsschritt“, der von der Hörbahn bewerkstelligt wird noch bevor das Gehörte bewussten kognitiven Prozessen unterworfen wird. Wir hören also als allererstes Raum.

---

<sup>9</sup> Die zeitliche Einordnung, d.h. die Bestimmung der Reihenfolge der Klicks, ist erst bei größeren zeitlichen Abständen möglich.

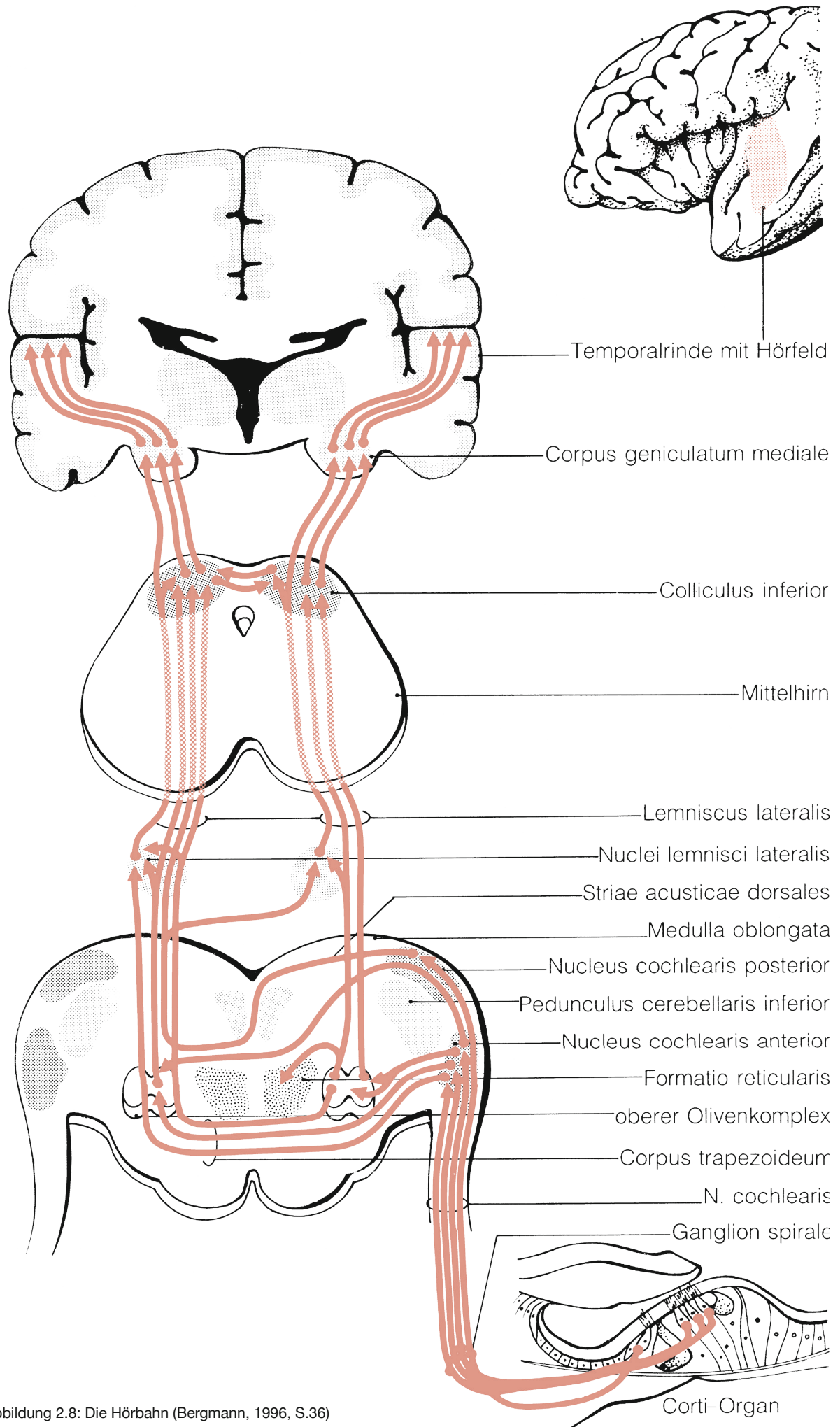


Abbildung 2.8: Die Hörbahn (Bergmann, 1996, S.36)



## Sound Studies

### 2.1 Soundscapes oder: Die Audiovisuelle Litanei

Die intensive Auseinandersetzung mit unserer klanglichen Umgebung ist in den Sozial- und Geisteswissenschaften ein relativ neues Phänomen. Während Werke zu Musiktheorie und Akustik bereits aus der Antike überliefert sind und die Beschäftigung mit diesen Arbeitsgebieten über die Jahrhunderte konstant aktiv betrieben worden ist, lässt sich der Ursprung dessen, was wir heute als „Sound Studies“ oder „Soundscape Studies“ bezeichnen, auf die Fünfziger und Sechziger Jahre des 20. Jahrhunderts zurückführen. Als Schlüsselwerk dieser neuen Disziplin, die eigentlich eine Anti-Disziplin – weil per definitionem interdisziplinär – ist, wird im Allgemeinen das Werk „The Soundscape: Our Sonic Environment or The Tuning Of The World“ des kanadischen Komponisten und Klangforschers Murray Schafer aus dem Jahr 1977 betrachtet. Schafer legte erstmals eine umfassende Theorie zur Untersuchung unserer klanglichen Umgebung vor, die weder rein musikalisch, physikalisch, psychologisch oder soziologisch ist, sich aber mit all diesen Disziplinen auseinandersetzt. Im Zentrum seiner Schriften steht der „*soundscape*“ – ein Kunstwort, das er als Pendant zum englischen Landscape bildete, um die Fülle an akustischen Ereignissen und Charakteristika eines Ortes beschreiben zu können. Auf dem Rücken der Technik der Field Recordings (Feldaufnahmen) schaffte es dieser Begriff im Laufe der Jahre in den sprachlichen Mainstream. Schafer definiert den Soundscape folgendermaßen: „*Soundscape: The sonic environment. Technically, any portion of the sonic environment regarded as a field for study. The term may refer to actual environments, or to abstract constructions such as musical compositions and tape montages, particularly when considered as an environment.*“ (Schafer, 1977, S.274) Schafer entwickelte seinen Theoriekanon unter maßgeblichem Einfluss dreier Kräfte: der Philosophie der sogenannten Toronto-Schule, der Musiktheorie der Musique Concrete und der Politik der erstarkenden Ökologie-Bewegung. (Schafer, 2010, 0:11, 4:18)

Prominenteste Vertreter der „Toronto Schule“ sind der Medientheoretiker Marshall McLuhan und dessen Schüler Walter Ong. Letzterer entwickelte die Konzepte McLuhans über Jahrzehnte zu ei-



nem umfassenden philosophischen Werk weiter. Einer der zentralen Punkte in den Theorien der Toronto-Schule ist die Sichtweise, dass die Medien, die einer Gesellschaft zur Verfügung stehen, diese formen. Demnach entwickelt eine Gesellschaft, die vermehrt unter dem Einfluss akustischer Kommunikationsformen steht, eine Kultur, deren Auffassungen und Weltverständnis sich von den physikalischen Aspekten von Schall sowie der menschlichen Fähigkeit zu hören ableiten. So beschreibt Ong die Gegensätze von Hören und Sehen folgendermassen: *„Vision comes to a human being from one direction at a time: to look at a room or a landscape, I must move my eyes around from one part to another. When I hear, however, I gather sound simultaneously from every direction at once: I am at the center of my auditory world, which envelopes me, establishing me at a kind of core of sensation and existence. This centering effect of sound is what high-fidelity sound reproduction exploits with intense sophistication. You can immerse yourself in hearing, in sound. There is no way to immerse yourself similarly in sight.“* (Ong, 1982, S.70)

Von Beschreibungen der Sinne wie der obigen leiten McLuhan und Ong ab, dass sich das Selbstverständnis einer Kultur, die vor allem mündliche oder akustische Medien nutzt, anders entwickeln muss, als dies bei einer schriftbasierten (engl. *literate*) Kultur der Fall wäre. So wird etwa erklärt, dass schriftbasierte Kulturen wie die „westliche“, die vor allem visuell geprägt sind, durch die Direktivität des menschlichen Blicks dazu neigen, die Welt in ihrer Wahrnehmung in separate Elemente zu „zerlegen“, während etwa orale Kulturen die Welt verstärkt als Kontinuum und „Gehörtes“ als Einheit wahrnehmen. Aufbauend auf solchen grundlegenden Annahmen über die physikalischen Ursachen und psychologischen Effekte von Hören und Sehen bauen McLuhan und Ong ein Theoriesystem auf, das Kulturen unter dem Gesichtspunkt von jeweils akustischer oder visueller Dominanz klassifiziert. Diese Einteilung ganzer Kulturen aufgrund ihrer Fertigkeit der (Re)Produktion und Tendenz zur Konsumation bestimmter Medien wird etwa von Veit Erlmann als „*great divide*“ bezeichnet, denn sie positioniert Gesellschaften (und in der Folge Individuen) auf einer Seite eines „Grabens“, während eine eindeutige Tendenz der Entwicklung prophezeit wird. Je nachdem, auf welcher der beiden Seiten man sich als Leser wiederfindet, werden also Siegeshoffnungen oder Verlustängste geschürt.

Tatsächlich sind viele der Grundannahmen der Toronto-Schule stark vereinfachend. Sie umfassen Dualitäten kontrastierender Begriffspaare, die man gar als Klischees bezeichnen könnte, wie etwa *„hearing immerses its subject, vision offers a perspective; hearing is concerned with interiors, vision is concerned with surfaces; sounds come to us, but vision travels to its object; hearing is about affect, vision is about intellect; hearing is a sense that immerses us in the world, while vision removes us from it.“* Jonathan Sterne bezeichnet diese Argumente als „audiovisuelle Litanei“ und kritisiert sie folgendermassen: *„The problem with the litany is that it elevates a set of cultural prenotions about the senses (prej-*

*udices, really) to the level of theory. To figure sound in these terms is to misattribute causes and effects.*“ (Sterne et.al., 2012 , S.9) Erlmann teilt Sternes Sicht in seinem Buch „Reason and Resonance“ und schreibt *„the problem with the audiovisual „litany“ lies in the sense of totality and inevitably that it imputes to social processes.“*; sowie, Sterne zitierend *„The audiovisual litany renders the history of the senses as a zero-sum game, where the dominance of one sense by necessity leads to the decline of another sense.“* (Erlmann, 2010, S.14)

Übersetzt bedeutet dies, dass die technische und kulturelle Dominanz eines Sinnes notwendigerweise zum Verkümmern der anderen führen muss – eine Sichtweise, die einige Probleme in sich trägt. Zum einen negiert diese die multisensorische Natur des Menschen, zum anderen sieht sie die Menschheit bzw. ihre kulturellen und technischen Errungenschaften als ausserhalb der Natur stehend an, was insofern interessant ist, da etwa McLuhan die vom Menschen geschaffenen Medien als Erweiterungen (engl. extensions) des Menschen versteht. Das Paradoxon ist also, dass den Menschen als Geschöpfen der Natur das Unnatürliche bereits innewohnen soll, was sie in gewisser Weise selbst als ausserhalb der Natur stehend positioniert. Diese Sichtweise hat bis heute zahlreiche Anhänger, obwohl sich einige Prophezeiungen etwa McLuhans leicht falsifizieren lassen. So ist etwa das „Globale Dorf“ durch Internet und globalen Handel Realität geworden, aber dieses ist kein oral dominiertes. Vielmehr entstand beispielsweise das Internet auf Basis eines visuellen und schriftbasierten Interface, und hat diesen Charakter trotz des Vormarsches von multimedialen Inhalten bis heute nicht abgestreift. Obwohl akustische Interaktions-Methoden zunehmend Verbreitung finden (z.B. Apple Siri, Google Now, Microsoft Cortana), so ist doch die visuelle Kommunikation durch Wort und Bild jetzt die dominante Interaktionsmethode. Wenn sich eine Tendenz der digitalen Medien erahnen lässt, dann viel eher hin zur multisensorischen Erfahrung.

Die weitere Auseinandersetzung mit diesem „Disput“ ist nicht von besonderem Interesse für diese Arbeit, allerdings beleuchtet dessen Beschreibung das Gärbecken, in dem die Sound(scape) Studies entstanden sind, sowie deren Gründungsgeist. So ist auch aus Schafers Texten ein nahezu kulturkämpferischer Tenor herauszuhören, der klar eine Präeminenz des Hörens fordert. So schreibt er etwa: *„We are coming to believe that dependence on the eye as the gatherer and orderer of environmental information is directly related to literacy and is therefore a habit that has been learned by Westerners as far back as late Greek civilization, but that as the West begins to enter its postliterate phase, the ear will return as a primary sensing instrument, just as it still is in many parts of the world.“* (Schafer, 1976, S.4) Zeitgenössische Autoren stehen dieser exklusiven Dualität kritisch gegenüber – als Stufe in einem Evolutionsprozess, hin zu einer sensorisch balancierten Sicht auf die Menschen und ihre Kultur gesehen allerdings könnte man Verständnis aufbringen. So kann man die harten Positionen der „Toronto-Schule“ und ihrer Abkömmlinge als Radikalismus einer Avantgarde verstehen, die den

Weg geebnet hat für eine ausgewogenere Beschäftigung mit den Aspekten der Kulturgeschichte der Sinne und somit der akustischen Umwelt.

Tatsache ist, dass die Beschäftigung der geisteswissenschaftlichen Disziplinen mit unserer akustischen Umgebung vor dem Aufstieg der Sound Studies gering war. So sollte man Schafers Werk bei aller berechtigter Kritik dafür würdigen, unserer klanglichen Umgebung mehr Aufmerksamkeit gespendet zu haben, denn bis heute wird akustischen Inhalten oft weniger Aufmerksamkeit geschenkt, als ihren visuellen Pendanten. Dies lässt sich insbesondere bei multimedialen Inhalten, wie etwa Filmen oder Videospielen, feststellen. Von dieser Beobachtung sollte man allerdings nicht den vorschnellen Schluss ableiten, dass akustische Inhalte weniger Einfluss auf uns nehmen. Oft sind es nur die Rezipienten, die sich dieses Einflusses nicht bewusst sind.

Eine weitere „Schule“ in deren Tradition Schafer steht, ist die Musiktheorie und das Wirken des französischen Komponisten und Tontechnikers Pierre Schaeffer, der im Jahre 1948 die *Musique Concrete* begründete. Schaeffer, damals Toningenieur beim Pariser Radio Television Francais, war fasziniert von den neuen technischen Möglichkeiten der Elektroakustik und gleichzeitig abgestossen von der immer noch dominanten 12-Ton Musik. In einem Interview aus dem Jahr 1988 sagt er „[A]fter the war, in the '45 to '48 period, we had driven back the German invasion but we hadn't driven back the invasion of Austrian music, 12-tone music. We had liberated ourselves politically, but music was still under an occupying foreign power, the music of the Vienna school.“ (Hodgkinson, 1987) Er begann mit Tonaufnahmen zu experimentieren – erst auf Schellack, bald darauf auf Magnetophon –, in der Absicht, einen neuen Weg für die Musik in der Elektronik zu finden. Douglas Khan beschreibt in „Noise, Water, Meat“, wie Schaeffer beim Versuch, aus aufgenommenen Geräuschen Klangcollagen zu assemblieren, mehr aus Verzweigung als aus Plan zur Manipulation griff, denn die verwendeten Geräusche blieben selbst in unkonventionellen Arrangements doch immer Symbole für die ihnen zugrundeliegenden Ereignisse, etwa ein Knarzen für eine Tür oder ein Bellen für einen Hund. Diese Symbolhaftigkeit der Klangcollagen jedoch war in seinen Augen ein Hindernis für das Entstehen von Musik im Kopf des Publikums. (Kahn, 1999, S.110) Trotz seiner Manipulationen und obwohl sie musikhistorisch großen Einfluss auf die weitere Entwicklung im 20. Jahrhundert hatte, betrachtete Schaeffer selbst die *Musique Concrete* am Ende seiner Karriere als gescheitert und wandte sich gänzlich von Musik ab.

R. Murray Schafer beginnt zwar als Student der *Musique Concrete*, aber wo Pierre Schaeffer aufgibt, beginnt sein Interesse und Wirken erst. Denn dieser zugrunde liegend war der Gedanke, Musik im „Unmusikalischen“ zu finden – im Speziellen in den Geräuschen der Natur und der Welt. Wo aber die *Musique Concrete* Ton-Aufnahmen in erster Linie als Rohmaterial für die Manipulationen und in letzter Instanz Komposition sah, beginnt Schafer sich für die Naturgeräusche selbst zu interessie-



ren. Er unternimmt intensive Untersuchungen von selbst angefertigten Field Recordings und beginnt das Gehörte in graphischen Partituren festzuhalten. Das Einsetzen von Vogelgesang zu frühen Morgenstunden, die Wachphasen von Kröten und Eulen in der Nacht und der Chor der Zikaden am Abend erscheinen ihm wie fertige Kompositionen. In der Folge ist es nicht verwunderlich, dass er beginnt, die Elemente eines Soundscapes auch mit musikalischen Begriffen zu klassifizieren.

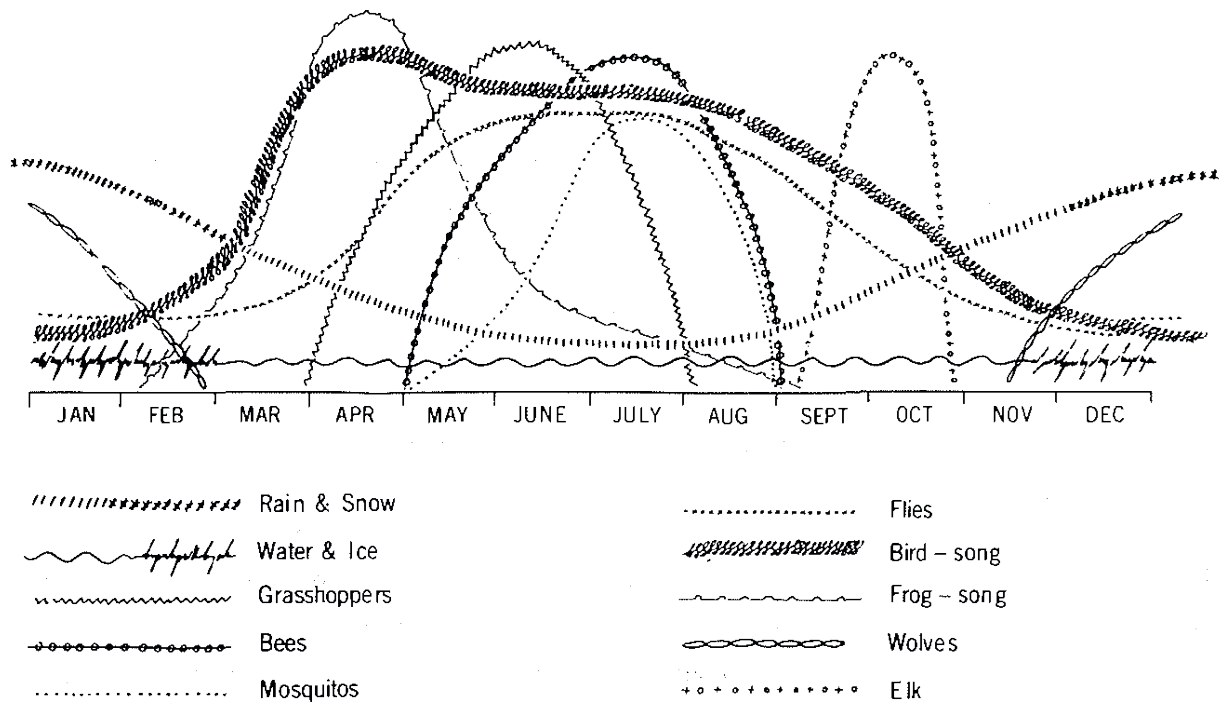


Abbildung 2.1: Monatliche Zyklen im Soundscape der Westküste von British Columbia (Schafer, 1977, S.229)

Schafer zerlegt die Soundscapes, die er studiert, in drei Gruppen von Klängen: *Keynote Sounds*, *Sound Signals* und *Soundmarks*. Der Begriff *Keynote Sounds* leitet sich vom englischen Wort für Tonart ab. Er beschreibt also den Grundton oder die Stimmung eines Soundscapes. *Keynote Sounds* sind also Hintergrundgeräusche, die die Szene „einfassen“, indem sie ein Bezugssystem herstellen. Dabei kann es sich beispielsweise um Regen oder Verkehrslärm handeln. Aber auch Wellen, die an einen Strand gespült werden, sind *Keynote Sounds*. *Sound Signals* stehen im Kontrast zu *Keynote Sounds*. Hierbei handelt es sich um Vordergrundgeräusche, bei denen es sich oftmals um bewusst kodierte Informationen handelt. Ein beliebtes Beispiel hierfür bei Schafer ist etwa das Nebelhorn – ein Klang, dessen langsames Verschwinden aus dem Vancouver der 1960er Jahre ihn betrübt. Aber auch eine Stimme kann ein *Sound Signal* sein, sowie das Zwitschern eines Vogels, oder ein einzelnes Auto in einer ländlichen Umgebung. Wesentlich anzumerken ist hierbei, dass *Keynote Sound* und *Sound Signals* ihre jeweilige Kategorie allein durch die Beurteilung der Hörenden erhalten. Diese

bezeichnet Schafer als „Ohrenzeugen“ (engl. *ear-witnesses*). Die Einteilung folgt dabei analog zur visuellen Figur-Grund-Wahrnehmung. Dabei kann der Grund in den Ohren der Ohrenzeugen auch zur Figur werden, und vice versa. Keinesfalls aber nehmen Klänge laut Schafer beide Rollen gleichzeitig ein. (Schafer, 1977, S.152) Soundmarks schließlich stellen einen Sonderfall unter den drei Gruppen dar. Wieder handelt es sich hierbei um ein Kunstwort das in Analogie zum englischen Landmark gebildet wurde. Übersetzen könnte man es also als „Hörenswürdigkeit“ (von Sehenswürdigkeit). Entscheidend für die Klassifizierung als Soundmark ist die Einzigartigkeit des Klangs, der mit symbolischer Bedeutung aufgeladen sein kann und den Charakter eines Ortes unverwechselbar macht. (Schafer, 1977, S.10)

Für die Qualitative Bewertung eines Soundscapes schließlich bedient sich Schafer eines Begriffs der Audiotechnologie: Fidelity – zu deutsch etwa Wiedergabetreue. Fidelity wurde besonders in der Blütezeit der analogen Langspielplatte als Gütesiegel benutzt um den Rauschabstand einer Aufnahme zu kategorisieren, d.h. das Verhältnis von Nebengeräuschen wie etwa Rauschen der Aufnahme oder des Tonträgers zum Maximalpegel des erwünschten Anteils der Wiedergabe. Analog zu diesem Begriff bezeichnet Schafer einen qualitativ hochwertigen Soundscape als Hi-Fi (also *high fidelity*). Er zeichnet sich dadurch aus, dass Hintergrundlärm gering ist und man die Elemente des Soundscapes klar heraushören kann. Als Beispiele nennt er hier vor allem rurale Szenerien. (Schafer, 1977, S.43f.) Ein Lo-Fi (engl. *low fidelity*) Soundscape auf der anderen Seite ist bestimmt von störenden Nebengeräuschen, die das Verstehen des eigentlich fokussierten Elements der Aufmerksamkeit im Soundscape wesentlich erschweren. In einer Lo-Fi Umgebung etwa steigert sich die Lärmbelastung immer weiter, da hinzukommende Geräusche und Signale dazu neigen, den Grundgeräuschpegel zu übertönen, um noch wahrgenommen zu werden. Schafer bezieht hier klare Position: *„The transition from rural to urban life can be characterized generally as a passage from the hi-fi to the lo-fi soundscape. A lo-fi soundscape is one in which trivial or conflicting acoustic information masks the sounds we want or need to hear. For a sound to catch one’s attention at all it must be monstrously loud or insistent. Radios, the birdsong of modern life, do not go south in winter; bulldozers do not hibernate and traffic does not sleep at night. Everything operates simultaneously with much wasted acoustic energy and attendant destruction of nerves and eardrums.“* (Schafer, 1976, S.6)

Hiermit bin ich beim letzten großen Einflussfaktor in Schafers Theoriekanon angekommen, nämlich der aufkommenden Ökologie-Bewegung der 1960er Jahre. Schafer, damals Dozent an der Simon Fraser University in Vancouver, Kanada, beginnt sich für die zunehmende Veränderung der klanglichen Umgebung Vancouvers durch Industrie, Motorisierung, Flugverkehr und elektrisch verstärkte Musik zu interessieren. Er bemerkt die zunehmende Verdrängung interessanter und markanter Geräusche aus dem Soundscape durch – aus seiner Sicht austauschbaren und generischen

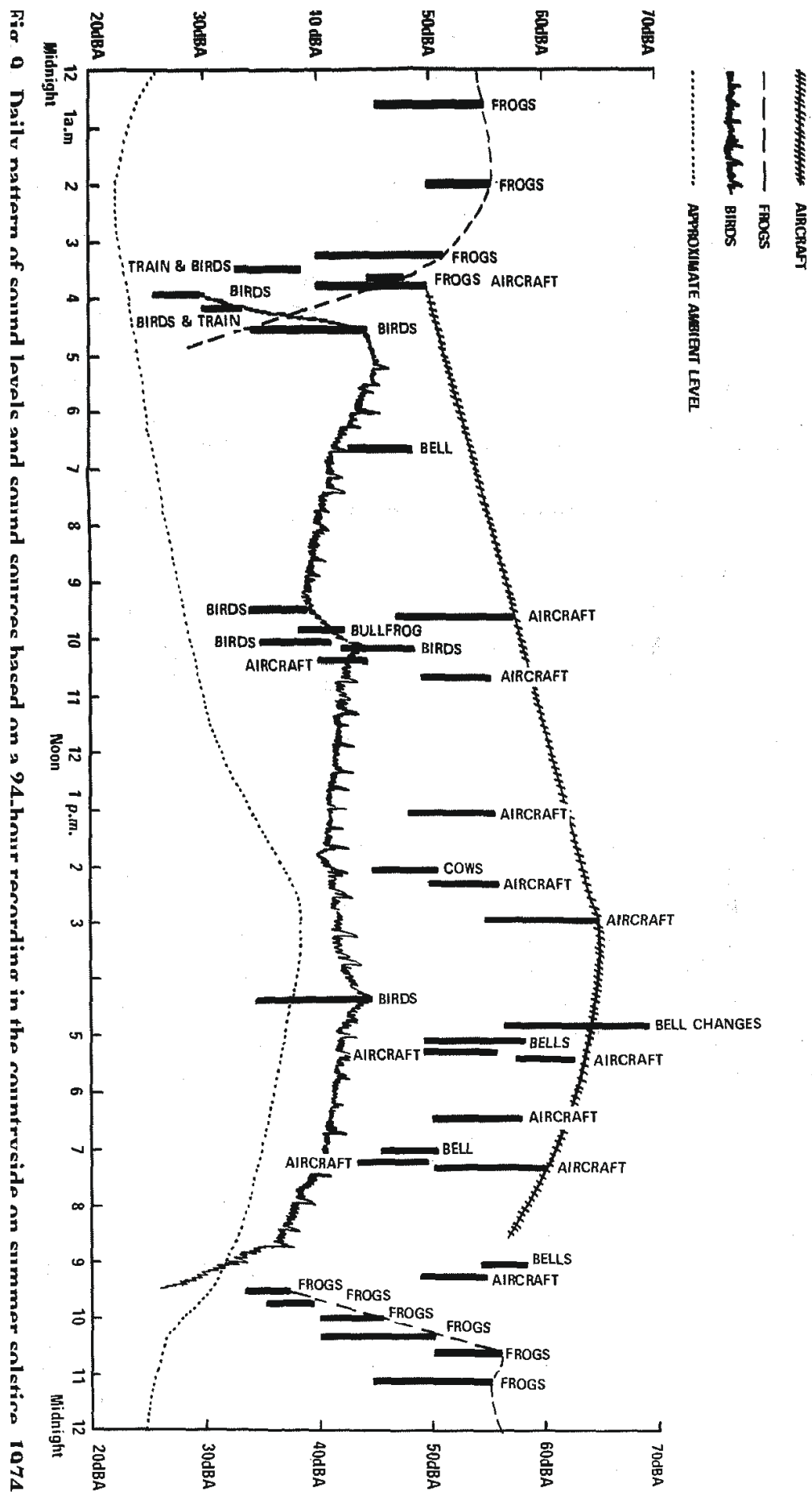


Abbildung 2.2: Graphische Dokumentation von 24 Stunden in einem ländlichen Soundscape in Kanada (Truax, 1984, S.68)

– Maschinenlärm und bringt eine Reihe von Artikeln und Schriften heraus, in denen er den Kampf gegen akustische Umweltverschmutzung (engl. *noise pollution*) zu seinem Ziel erklärt. Aus dieser Position heraus gründete er Ende der 1960er Jahre das World Soundscape Project (WSP). Eine Gruppe junger engagierter Mitarbeiter schloss sich ihm an, unter den prominentesten sind Barry Truax und Hildegard Westerkamp. (Truax, *The World Soundscape Project*) Ziel des WSP war es, ein Bewusstsein für den damals gegenwärtigen Zustand der akustischen Umgebung, deren Veränderung durch Technologie und deren Auswirkungen auf die Menschen zu schaffen, und die Öffentlichkeit über die Problematik der akustischen Umweltverschmutzung aufzuklären. Durch die Ausbildung von Studierenden und die Erarbeitung von Klangkonzepten sollte die Qualität der gegebenen Soundscapes letztlich verbessert werden. (Westerkamp, 1991)

Das WSP hat eine Vielzahl an Soundscapes weltweit in einem Katalog von mehr als 300 Tonbändern und CDs dokumentiert. Zwar löste sich die Forschungsgruppe des WSP mit dem Verlassen Schafers der Simon Fraser University ab 1975 langsam auf, jedoch ging 1993 initiiert durch Hildegard Westerkamp aus dem WSP das „World Forum For Acoustic Ecology“ (WFAE) hervor, das bis heute Bestand hat. (WFAE, 2015) Ein Ziel des WSP kann man heute als erreicht betrachten: Ein Bewusstsein für Lärmbelastung hat weithin Einzug gehalten in der öffentlichen Wahrnehmung. Zwar sind wir auch heute noch in so gut wie allen Städten der Welt mit Kaufhausmusik und Verkehrslärm konfrontiert, Schafer selbst gibt aber in einem Interview an, dass der Lautstärkepegel einer typischen Stadtumgebung in den späten 1960er Jahren weitaus höher war als in der Gegenwart und bezeichnet gar die Zeit um 1965 rückblickend als „*about the noisiest time in the western world*“. (Schafer, 2010, 2:39) Die Einführung gesetzlicher Lärmschutzmaßnahmen kann man also als positive Entwicklung betrachten. Gleichzeitig stellt sie aber ein zweischneidiges Schwert dar, denn bevor Regeln und Gesetze erlassen werden können, muss zuerst eine zentrale Frage geklärt werden: Was ist Lärm?

## 2.2 Noise: Lärm und Geräusch

Das zentrale Element der von Schafer bekämpften akustischen Umweltverschmutzung ist nicht unbedingt einfach Lärm – es ist *noise*. Das englische Wort *noise* bedeutet zwar Lärm, es kann aber genauso als Geräusch übersetzt werden. Da sowohl Schafers Texte als auch der Großteil der Sound Studies Literatur in englischer Sprache verfasst sind, heißt es hier Vorsicht walten zu lassen, denn bei der Übersetzung ins Deutsche kann eine Vielzahl von Deutungsmöglichkeiten verloren gehen. Das englische *noise* ist als Begriff zwar nicht wertfrei, repräsentiert aber besser den subjektiven Charakter des Bezeichneten: des einen Geräusch kann des anderen Lärm sein. Diese Ambivalenz des englischen Begriffs erhält bei der Übersetzung ins Deutsche automatisch eine Wertung und wird

so subjektiviert. Schafer selbst bietet vier Definitionen für den Begriff noise an: Unerwünschter Klang, unmusikalischer Klang, (jeglicher) lauter Klang und Störung einer Signalkette. Er erklärt in der Folge, dass er die Definition von noise als „unerwünschtem Klang“ als am befriedigendsten erachtet, räumt aber selbst ein, dass diese Bedeutung problematisch – da hochgradig subjektiv – ist. Diesen Zwiespalt zwischen seiner eigenen Präferenz und dem Wissen um deren Subjektivität versucht er durch das Berufen auf einen gesellschaftlichen Konsens des Erwünschten aus dem Wege zu räumen. (Schafer, 1977, S.182f.) Durch die fortschreitende (mediale) Globalisierung und gleichzeitige Fragmentierung lokaler gesellschaftlicher Normen wird solch ein Konsens aber zunehmend schwieriger zu erzielen. Spätestens hier wird klar dass „noise control“ als Kampf gegen „akustische Umweltverschmutzung“ ein hochgradig politisches Thema ist.

Die Wegbereiter der musikalischen Schule, in deren Tradition Schafer steht, waren selbst oft dem Geräusch und dem Lärm gegenüber aufgeschlossen und auch haben mitunter auch gefordert, überkommene Normen des „Wohlklangs“ zu bekämpfen. Die Beschäftigung mit Naturgeräuschen als musikalische Elemente wäre möglicherweise ohne die zunehmende Dominanz des Lärms der Maschinen nie zustande gekommen. Einer der ersten, der für noise offen Partei ergreift, ist der italienische Futurist Luigi Russolo. Sein 1913 verfasstes Manifest „L'Arte dei Rumori“ fordert vehement die Überwindung eines „beschränkten Kreises von reinen Tönen“ und stattdessen die Einbeziehung von Geräuschen in die Musik. (Russolo, 1916, S.11) Sein 1913 erbautes Geräuschorchester – genannt „intonarumori“ – setzt sich aus einer Vielzahl von Klangmaschinen zusammen, die entworfen wurden, um die komplexen Klänge einer industrialisierten Welt zu imitieren. (Russolo, 1913) Das italienische rumore kann ähnlich offen verstanden werden, wie das englische noise. Rumore wurde laut Russolo erst mit der „Erfindung der Maschinen“ im 19. Jhd. als Begriff geboren. (Russolo, 1916, S.19) Er selbst war ganz auf Linie des Futurismus von der Gewaltigkeit von Maschinenlärm begeistert. So schreibt er etwa „*Wir Futuristen haben die Harmonien der grossen Meister alle tief geliebt und genossen. Beethoven und Wagner haben während vieler Jahre unsere Nerven erschüttert und Herzen bewegt. Heute sind wir ihrer überdrüssig und geniessen es viel mehr, die Geräusche der Tram, der Explosionsmotoren, Wagen und schreienden Menschenmengen in unserer Vorstellung zu kombinieren, als beispielsweise die «Eroica» oder die «Pastorale» wiederzuhören.*“ (Russolo, 1916, S.11) Aus dieser Beschreibung kann man anhand des Wortes „kombinieren“ bereits die spätere Musique Concrete anklängen hören.<sup>1</sup> Gleichzeitig war Russolo aber – wie die Bewegung des Futurismus insgesamt, die anfänglich große Sympathien für den Faschismus aufbrachte – begeistert von der kriegerischen Dimension des Maschinenlärms. Der industrialisierte Krieg und seine ohrenbetäubenden Geräusche wurden als reinigende Kraft gesehen, aus der eine neue Welt hervorgehen sollte. So widmete auch

---

1 Auch kann er als geistiger Vater der Industrial Music der spätem 1970er Jahre gesehen werden. Allerdings muss hier angemerkt werden, dass diese die brachiale Ästhetik der Industrie zumeist als Mittel der Kritik an der zunehmend unmenschlichen Gesellschaft eingesetzt und folglich im Gegensatz zu den Futuristen als ironisches Stilmittel genutzt hat.

Russolo in der 1916 erschienen erweiterten Fassung der *L'Arte dei Rumori* ein ganzes Kapitel den Geräuschen des Krieges. Die Gewaltphantasien des Futurismus müssen freilich im Kontext der Zeit gesehen werden. All dies ereignete sich am Vorabend des ersten Weltkrieges, in dessen Verlauf viele Futuristen ihre Einstellung zu Krieg und Faschismus änderten. So wandte sich auch Russolo gegen den Faschismus, wurde allerdings nie zum Pazifisten. (Kahn, 1999, S.66)

Nach der *Musique Concrete*, in welcher durch Manipulationen an aufgenommenem Material Musik entstehen sollte, war es vor allem der amerikanische Komponist John Cage, der sich des *noise* annahm. Während Ersterer aber vor allem an den Klangmöglichkeiten interessiert war, die sich durch neue Technologien ergaben, war John Cage viel eher darauf erpicht, Musik in jedwedem Klang selbst zu finden: „*to hear a sound in itself*“. (Kahn, 1999, S.164f.) Während *noise* für Schaffer unerwünschten Klang darstellte, bezeichnete Cage *noise* schlicht als Klang (engl. *sound*), der uns (noch) fremd ist und forderte, jegliche Geräusche vorurteilsfrei als potentiell musikalisch zu betrachten. (Cage, 1961, S.87). Ähnlich wie bei Russolo lässt sich sein Begriff von *noise* eher als „Geräusch“ übersetzen. Cage war sich der politischen Dimension von *noise* als erklärter Anarchist genau bewusst, aber mit fortschreitendem Schaffen wurde sein bevorzugtes Werkzeug die Stille. In seinem wohl berühmtesten Stück *4'33"*, einem Solostück für Klavier, sitzt der Pianist still vor dem Konzertflügel. Eine klare Provokation, denn das Publikum erwartet eine Darbietung, stattdessen werden seinen eigenen Geräusche wie Räuspern, Rascheln oder Sesselknirschen durch die unerwartete Stille in den Mittelpunkt gerückt. Die Stille bei John Cage ist aber eine andere als die Murray Schafers. Der französische Ökonom (und ehemalige Finanzminister Françoise Mitterrands) Jacques Attali widmete der politischen Dimension von Musik 1977 sein Buch „*Bruits. Essai sur l'économie politique de la musique*“ und deutet in diesem Cages Komposition als politischen Akt: „*When Cage opens the door to the concert hall to let the noise of street in, he is regenerating all of music: he is taking it to its culmination. He is blaspheming, criticizing the code and the network. When he sits motionless at the piano for four minutes and thirty-three seconds, letting the audience grow impatient and make noises, he is giving back the right to speak to people who do not want to have it.*“ (Attali, 1985, S.136f.) Der gesellschaftliche Code, der verlangt, dass bei einer Aufführung Stille zu herrschen hat, ist also auch eine Kulturtechnik der Repression, deren Aufhebung dem Individuum eine Verantwortung zurückgibt, die es unter Umständen gar nicht haben will. So betont Attali, dass das Recht „Lärm“ zu erzeugen noch vor der industriellen Revolution ein Recht jedes Menschen war – ein Zeugnis der persönlichen Autonomie. Der *noise*-Begriff bei Attali hat jedoch gewalttätige Züge. Ähnlich dem italienischen *rumore* ist das französische *bruit* weiter gefasst in seiner Bedeutung und gleicht damit eher dem englischen *noise*. Es weckt aber sowohl für deutsch- als auch englischsprachige Ohren eine Assoziation – *bruit* entspringt dem gleichen Wortstamm wie *brutal*, das sich vom lateinischen *brutalis* („*unvernünftig, tierisch*“) ableitet. (Kluge, 2002, S.155) In ihm steckt die Gewalttätigkeit.



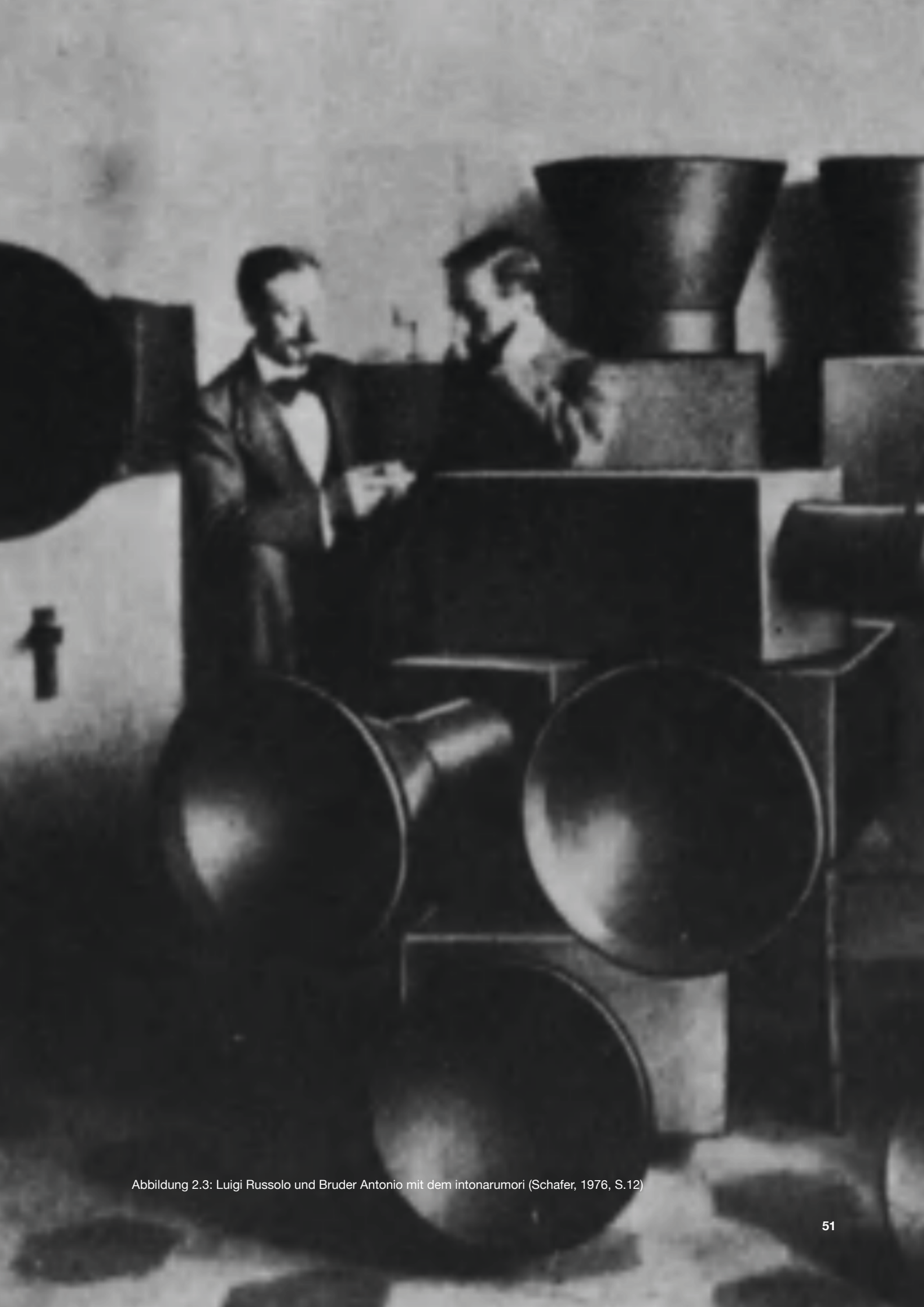


Abbildung 2.3: Luigi Russolo und Bruder Antonio mit dem intonarumori (Schafer, 1976, S.12)

Und so betont Attali nicht nur die potentiell schädigende Wirkung von noise – er erklärt diesen zur tödlichen Waffe. „*Since it is a threat of death, noise is a concern of power; when power founds its legitimacy on the fear it inspires, on its capacity to create social order, on its univocal monopoly of violence, it monopolizes noise. Thus in most cultures, the theme of noise, its audition and endowment with form, lies at the origin of the religious idea.*“ (Attali, 1985, S.27) Musik stellt einen Versuch der Zähmung und Ritualisierung dieser Macht dar – eine sublimierte rituelle Schlachtung (im Original „*simulacrum of ritual murder*“).

Auch bei Schafer finden der gewalttätige und der kultische Aspekt von noise in der Beschreibung des „*Sacred Noise*“ Niederschlag. Sacred Noises haben Schafer zufolge ihren Ursprung in den gewaltigen Klangereignissen der Natur wie etwa Donner, Vulkanausbrüchen und Stürmen, denen einst religiöse Bedeutung beigemessen wurde, umfassen aber genauso soziale Klänge wie Kirchenglocken, Polizeisirenen oder den Lärm der Industrie. Es handelt sich um Geräusche, die eine Macht repräsentieren. In Schafers Augen hat mit zunehmender Säkularisierung der Welt die Industrie die Sacred Noises von den religiösen und (später) politischen Institutionen übernommen. Noise equals power titelt Schafer ein Unterkapitel in dem er schreibt: „*The association of Noise and power has never really been broken in the human imagination. It descends from God, to the priest, to the industrialist, and more recently to the broadcaster and the aviator. The important thing to realize is this: to have the Sacred Noise is not merely to make the biggest noise; rather it is a matter of having the authority to make it without censure.*“ (Schafer, 1977, S.76) So drückt sich etwa Macht auch durch im Hintergrund spielende Kaufhausmusik aus, die sich nicht unbedingt durch Lautstärke auszeichnet, sondern durch unausweichliche Präsenz: Muzak – von Schafer verballhornt als Moozak – ist der zum Genrebegriff gewordene Markenname der amerikanischen Muzak Corporation, die seit den 1930er Jahren ein zentralisiertes elektronisches Distributionsnetzwerk bereit stellt, über das Funktionsmusik verbreitet wird, die Kunden beruhigen oder zum Kauf anregen soll. Es ist dies laut Schafer der Klang des Imperialismus, dessen sich die Bürger nicht erwehren können. Auch Attali nimmt auf Muzak Bezug und beschreibt, dass die Muzak Corporation sich selbst in den 1970ern als größter privater Sicherheitskonzern beworben hat, da sie in Kontrolle der akustischen Sphäre von Einkaufszentren auch die größtmögliche Reichweite zur Übermittlung von Befehlen erzielt. (Attali, 1985, S.8)

Macht manifestiert sich nicht nur in der Erzeugung von Schall, sondern auch in dessen Kontrolle. Schafer setzt jedwede Produktion von lauten Schallereignissen durch eine Person oder Institution mit Imperialismus gleich (Schafer, 1977, S.77) und fordert umfangreiche gesetzliche Vorgaben zur Begrenzung von erlaubten Lautstärkepegeln. Da jede Produktion von noise also Imperialismus bedeutet, muss auch das laut(!)starke Aufbegehren gegen diesen schon inhärent imperialistisch sein.



Unsere Welt werde im Wahn des Kapitalismus lauter und lauter, und Mensch und Natur seien die Opfer. Schafers Antwort sind rigorose Massnahmen zum Schutz vor noise pollution. Möge die Menschheit schweigen und die Natur klingen lassen– Schafers Utopie ist die Ruhe. Als einzige Alternative zur Durchsetzung von Massnahmen zur „noise control“ nennt er den Zusammenbruch der technologischen Infrastruktur – vielleicht insgeheim eine wilde Hoffnung seinerseits? (Schafer, 1977, S.181)<sup>2</sup>

Jedenfalls sind Maßnahmen zum Lärmschutz heute vor allem durch Lautstärkebegrenzungen umgesetzt – also eine quantitative Beschränkung der tolerierten Maximalpegel, die zumeist mit Zeitplänen gekoppelt sind. Was aber klassifiziert noise? Wie Schafer einräumt, handelt es sich dabei nicht unbedingt um das lauteste Geräusch. Vielmehr ist noise subjektiv. Jede/r, der/die selbst schon jemanden friedlich auf dem vibrierenden Subwoofer eines Klubs schlafen gesehen hat, wird dies bestätigen. Jede/r, der/die aufgrund eines tropfenden Wasserhahnes schlaflos bleiben musste, ebenso. Hillel Schwartz formuliert diese Problematik in seinem Buch „Making Noise“ folgendermaßen: *„Across the last two thousand years, humankind on occasion has matched the volume of the rest of the natural world. Across the last two hundred we have regularly surpassed it at every point along the spectrum: sirens and steam whistles on the high end, more piercing than screech of parrots or shriek of sandstorm; klaxons and loudspeakers at midpoint, howling above wolves and high winds; supersonic jets and nuclear explosions on the low end, booming above calving glaciers. Yet the cry of a peacock in an abandoned courtyard, the whimper of a baby in an empty stadium, or the groan of oaks under the winds of dark woods may feel louder, noisier, than any siren; [...] True, sentiments of noise are often bound up with sensations of loudness, and above 130 db all sound, alas, is fatal to the cilia of our inner ears, but what one anywhere encounters as noise are sounds that must be listened through, regardless of decibels: hard candies slowly unwrapped at harpsichord concerts, a chainsaw upriver of a wilderness camp.“* (Schwartz, 2011, S.37)

Im Bewusstsein der Vielschichtigkeit des noise-Begriffs fordert Schafer vor allem qualitativen Lärmschutz. (Schafer, 1977, S.183) Schafer nennt die Öffentlichkeit als möglichen Entscheidungsträger (engl. *public opinion*). Aber wer ist die Öffentlichkeit? Ist es die demokratische Mehrheit? Die Jurisdiktion eines Bezirks, einer Stadt oder eines Staats? Und wer überprüft, ob die Entscheidungsträger auch die Betroffenen sind? Eine homogene Meinung wird in den wenigsten Fällen zu erzielen sein. Hier beißt sich die Katze in den Schwanz: Sind nicht Beschluss und Ausübung solch qualitativen Massnahmen ebenfalls subjektiv, und folglich nach Schafers Logik imperialistisch? Und wer darf sie beschließen? Folglich ist das Problem mit der Umsetzung solch qualitativer Massnahmen immer

---

2 Erwähnenswert ist hier, dass Schafer 1975 seine Lehrtätigkeit beendete, um auf einen Bauernhof 200 Meilen von Toronto zu ziehen. (de Oliveira, Duarte, Lima, 2014, 6:54) Ein Zusammenhang zwischen den erwähnten Textpassagen und diesem Faktum drängt sich somit auf.

ein politisches – vielleicht mehr noch als dies bei quantitativen Massnahmen der Fall wäre. Der Mann mit einem Lautsprecher in der Hand ist imperialistischer als einer ohne, der Arbeiter mit dem Vorschlaghammer imperialistischer als der mit der Schaufel in der Hand, schreibt Schafer. (Schafer, 1977, S.77) Aber kann man etwa die mit Lautsprechern „bewaffneten“ Pussy Riot, die in der Christ-Erlöser-Kathedrale in Moskau gegen die dominante politische Kaste protestieren, tatsächlich als imperialistisch bezeichnen? Und birgt nicht das von Schafer indirekt geforderte Regime des Wohlklangs schon Züge von Totalitarismus?

Da eine objektive Beschreibung von Lärm unmöglich ist, könnte man versucht sein, Schafers gesamtes Werk aufgrund seines wertenden Blicks auf diese Thematik abzulehnen. Zur Ehrenrettung der Sound Studies Schaferscher Prägung trägt Barry Truax bei, der ab 1973 Mitstreiter Schafers am World Soundscape Project ist. In „Acoustic Communication“ (1984) erweitert er den von Schafer beschriebenen Konzeptkanon um Elemente der Kommunikationstheorie und schafft damit ein flexibles Modell, das auch erlaubt, gegensätzliche Positionen zu klanglichen Umgebungen zu vereinen. Er beschreibt diesen Perspektivenwechsel folgendermassen: „*[A] communicational approach includes the notion of context. The exchange of information is highly dependent on context, whereas the transfer of energy is not. [...] In a communicational approach, context is essential for understanding the meaning of any message, including sound.*“ (Truax, 1984, S.10) und weiter „*[...] [A] communicational model will not deal with linear chains of energy or signal transfers, but with systems of related elements operating at different hierarchic levels. Instead of thinking of sound as coming from the environment to the listener and perhaps being generated back again, we will think of it as mediating, or creating relationships, between listener and environment.*“ (Truax, 1984, S.11) Diese Re-Fokussierung von technischen Parametern, wie etwa Schallintensität, hin zur Aspekten der Kommunikation, läutet auch einen bedeutenden Paradigmenwechsel in den Sound Studies ein. Der Mensch und seine Wahrnehmung treten mit diesem Zugang in die Mitte des Soundscapes. So ersetzt Truax auch Schafers Konzepte von Hi-Fi und Lo-Fi durch den Begriff der *Balance*. Was hier als semantische Spitzfindigkeit aufgefasst werden könnte, birgt eine tiefere Bedeutung: Hi-Fi setzt als Lehnwort aus der Tontechnik durch seine technische Konnotation auch eine technische, messbare Bewertung des Soundscapes voraus und öffnet damit eben jener problematischen Interpretation von noise Tür und Tor, die wir in den vorangegangenen Seiten behandelt haben. Der allgemeinere Begriff der Balance erlaubt uns, Soundscapes aus einer relativen Perspektive zu analysieren, die am jeweiligen Individuum orientiert ist. Entsprechend kritisch nähert sich Truax auch der Thematik der acoustic ecology an.<sup>3</sup> So beschreibt er etwa eine Gemeinde, in der bestimmte Kräfte versuchten, die Signalpfeife einer lokalen Fabrik zu verbieten, ohne deren Bedeutung in Betracht zu ziehen und ersetzt die dogmatische

---

<sup>3</sup> Truax spricht sich nicht gänzlich gegen acoustic ecology aus, jedoch schreibt Jonthan Sterne in seinem Blog, dass Truax ihm offenbart habe, sich selbst nicht als „Acoustic Ecologist“ zu sehen. (Sterne, 2010)

Gleichsetzung von Lärm und Lautstärke durch die Forderung nach einer kritischen Beurteilung der umstrittenen Geräusche hinsichtlich ihrer Bedeutung für die Gemeinde. „*The equating of loudness to pollution in this case ignores the way in which the sound functions in the community, probably as a useful community signal that reflects an important institution and contributes to community identity. Another way to consider the problem would be to ask what the result would be if all “loud” sounds were removed from the community and the result happened to be a dull, homogeneous environment of low level hums and drones! The lack of acoustic information and interesting sonic relationships would hardly produce a balanced soundscape.*“ (Truax, 1984, S.24)

## 2.3 Raum und Maske

Diese Standpunktverschiebung durch Truax bedingt, dass wir uns eingehender mit der Art und Weise beschäftigen müssen, wie wir Menschen den Soundscape wahrnehmen. Denn völlig unabhängig davon, welche Geräusche man nun als Lärm bezeichnet, ist der Effekt „konkurrierender“ Klangereignisse – die Basis für die Beurteilung der Balance eines Soundscape – von hoher Relevanz für das menschliche Hören, im Speziellen für dessen räumliche Aspekte: Auditive Maskierungseffekte. Maskierungseffekte sind Phänomene, die auf physischen Eigenheiten des menschlichen Gehörs beruhen, und bewirken, dass Geräusche von anderen Geräuschen „verdeckt“ werden. Obwohl ein maskiertes Geräusch weiterhin präsent ist, wird es nicht mehr (oder nur mehr teilweise) wahrgenommen – die Maskierung findet also im menschlichen Gehör statt. Es gibt zwei Arten der auditiven Maskierung: Frequenz-Maskierung und temporale Maskierung. Die Frequenz-Maskierung, die eigentlich Simultanmaskierung genannt werden sollte, da die temporale Maskierung genauso frequenzabhängig ist, tritt auf, wenn das maskierende und das maskierte Geräusch gleichzeitig ertönen. Hierbei hebt das Erklingen des Ersteren die Hörschwelle an, wodurch Zweiteres, wenn es unter dieser Intensität liegt, nicht mehr hörbar ist. Der Grund hierfür ist im Aufbau des Ohres zu suchen. Einerseits lockert sich die Spannung des Steigbügel-Muskels bei lauten Tönen, um das Trommelfell zu schützen, wodurch die Schallübertragung abgeschwächt wird. (Vgl. Kapitel 1) Andererseits können tiefe Töne höhere verdecken, da erstere in der Hörschnecke am hinteren Ende registriert werden. Das hat zur Folge, dass tiefe Töne beim Durchlauf der Schnecke als Wanderwelle die hohen Töne überlagern. Die Maskierung ist am erfolgreichsten, wenn das maskierende Geräusch auf der selben Frequenz liegt wie das maskierte. Mit zunehmendem Frequenzabstand wird das maskierte Geräusch wieder wahrnehmbar. Die temporale Maskierung wiederum erfolgt, wenn das maskierende und das maskierte Geräusch zeitversetzt erklingen, und kann auf zweierlei Art erfolgen: Bei der Rückwärts-Maskierung erklingt das Maskierungsgeräusch nach dem Zielgeräusch, von einer Vorwärts-Maskierung spricht man, wenn das Maskierungsgeräusch vor dem Zielgeräusch erklingt. Bei beiden Phänomenen ist die Dauer des Maskierungsgeräusches von Bedeutung. Man

kann vereinfachend sagen, dass die Vorwärts-Maskierung bis etwa 200 ms nach dem Zielgeräusch erfolgreich ist und bei der Rückwärts-Maskierung das Maskierungsgeräusch maximal 50 ms vor dem Zielgeräusch erklingen darf. (Necciari, 2010, S.49-66)

Ein Klangereignis, das also in einen bestehenden Soundscape eingebracht wird, verändert die Wahrnehmbarkeit der Klangquellen. Maskierung ist somit ein wesentliches Phänomen für die räumliche Gliederung unserer akustischen Umwelt. Durch unterschiedliche Lautstärke und Frequenzgehalt von Klangquellen kann der Raum in Zonen eingeteilt werden, in denen die Elemente verschieden gut hörbar sind, bzw. im Vordergrund oder Hintergrund stehen. Diese Gliederung erfolgt aus der Position der menschlichen Wahrnehmung heraus, muss sich aber nicht unbedingt mit der physikalischen Realität des Schalls decken. Beim Versuch, Naturgeräusche mit einem Mikrophon aufzunehmen, kann es beispielsweise passieren, dass sich die vermeintlich geglückte Aufnahme bei der Kontrolle an der heimischen Stereoanlage als missglückt erweist, wenn plötzlich ein dumpfes Rauschen im Hintergrund hörbar wird, aus dem sich immer wieder erkennbare Verkehrsgeräusche erheben, die bei der ersten Begegnung mit dem Klangereignis von unserem Gehirn ausgeblendet wurden. Durch den Umweg der Tonaufnahme wird deutlich, dass das Mikrophon anders „hört“ als wir Menschen.

Was hier geschehen ist, lässt sich in der Sprache der Soundscapes als Verringerung des akustischen Horizonts (engl. *acoustic horizon*) beschreiben. Ein vages Grundgeräusch, das unter Umständen nicht wahrgenommen wird, so man nicht konzentriert hinhört, maskiert durch seine bloße Präsenz andere Geräusche. Der akustische Horizont ist, äquivalent zum visuellen Horizont, definiert durch die maximale Distanz bei der akustische Ereignisse wahrgenommen werden können. In gewisser Weise könnte man sagen, dass durch die Präsenz von Klängen der uns umgebende akustische Raum verkleinert wird. Barry truax notiert hierzu: „*The presence of a steady level of sound reduces what we may call the “acoustic horizon” of an environment, that is, the farthest distance from which sound may be heard. The steady sound masks low level sounds, thereby producing a reduced sense of space. In the most extreme case, each individual is surrounded in a cocoon of sound with no aural contact with others.*“ (Truax, 1984, S.23) Ein Beispiel für den von Truax angesprochenen Fall wäre das Tragen von Kopfhörern oder auch der Grundgeräuschpegel einer Party. Durch das Einbringen oder Reduzieren von Klängen wird die Kommunikationsdistanz zwischen Individuen verändert. Gleichzeitig ändert sich etwa auch der private akustische Raumbereich. Das Gegenstück zum akustischen Horizont ist die akustische Arena (engl. *acoustic arena*). Sie bezeichnet die maximale Distanz, bei der ein Geräuschereignis noch hörbar ist.

In einem Videovortrag am Massachusetts Institute of Technology beschreibt die Kulturhistorikerin Emily Thompson einen bedeutenden Moment, in dem das Bewusstsein für die raumbildenden Ef-

fekte der Maskierung einsetzte. Die Erforschung der Raumakustik um 1900 (Vgl. Kapitel 3) hatte die Entwicklung von schalldämmenden Materialien zur Folge. Gleichzeitig wurden Stadtgebiete durch zunehmende Motorisierung lauter und lauter und Stille zu einem kommerziell verwertbaren Gut. Industriepsychologen in den USA begannen den Effekt von Lärm auf den Menschen zu erforschen und fanden heraus, dass Lärmbelastung die Effizienz von Arbeitenden um ein Drittel reduzieren kann. Gesteigerte Bemühungen, ruhige Bürogebäude zu realisieren, waren die Folge. Das 1932 realisierte Philadelphia Savings Fund Society Building wurde innen grossflächig mit Dämmmaterial ausgekleidet, was zur Folge hatte, dass 95% der im Gebäude emittierten Geräusche absorbiert wurden. Auf einer Werbeanzeige prangte in Blockbuchstaben ein riesiges „QUIET!“. Die Dämmeigenschaften der fertigen Räumlichkeiten waren jedoch so gut, dass die resultierenden Arbeitsumgebungen zu still waren. Plötzlich wurde jedes Räuspern und Rascheln hörbar. Dies resultierte darin, dass die Planer im ganzen Haus ein Leitungsnetz installierten, das es erlaubte in den Büros Radios anzuschliessen. (Thompson, 2002, 25:50-33:00) Das Phänomen der Maskierung begann bewusst eingesetzt zu werden, um den akustischen Raum zu formen. Nicht zufällig fällt in die selbe Dekade auch die Gründung der bereits erwähnten Firma Muzak, denn die Ökonomisierung der akustischen Umgebung ist die logische Konsequenz der Erforschung der Akustik und des menschlichen Hörvermögens.

Diese bewusste Urbarmachung von Maskierungseffekten behandelt auch Jonathan Sterne in seinem Buch „MP3 – The Meaning Of A Format“.<sup>4</sup> Er berichtet, wie die aktive Nutzung von Schall ab den 1960er Jahren im großen Stil in der architektonischen Praxis Eingang findet. Anstatt Schall zu eliminieren, werden seine maskierenden Effekte nun bewusst eingesetzt, um störende Geräusche zu überdecken. Sterne zitiert den Psychoakustiker Leslie Doelle, dessen Buch „Environmental Acoustics“ (1972) zum Teil als Handbuch der Maskierung werden kann. Neben dem Klang von Klimaanlage und Hintergrundmusik beschreibt er auch die Benutzung von elektronisch generiertem Rauschen als akzeptables Maskierungsgeräusch, das etwa in Grossraumbüros eingesetzt werden kann, um die Mitarbeiter voneinander akustisch zu isolieren. Er bezeichnet die Einbringung derartiger Klangkulissen als akustisches Parfum und betont, dass zu ihrer effektiven Nutzung dieses bedeutungslos (engl. *meaningless*) – also frei von ablenkenden Informationen – sein muss. (Sterne, 2012, S.120ff.) *„Doelle’s architectural acoustics used masking to create an acoustic space that was smaller than the physical space inhabited by corporate workers and machines. One form of distraction enables or facilitates a second form of productive distraction — for instance, when the hum of a HVAC system smoothes out the noise of conversations in the next cubicle. Like an animal previously thought to be wild, noise was domesticated and put to use in Doelle’s architectural acoustics. In the process, noise took up a*

---

4 Die Effizienz von Kompressionsformaten wie MP3 oder AAC beruht auf der Eliminierung von Schallinformationen, die auf Grund von Maskierungseffekten nicht hörbar sind.

*position as part of the communicative order, rather than as a form of entropy that threatened it.*“ (Sterne, 2012, S.141) Wahrscheinlich ist es kein Zufall, dass Schafer seine Soundscape-Lehre ausgerechnet in dem Jahrzehnt entwickelt, in dem die kommerzielle Nutzung von Schall als Maskierungselement für die Architektur erstmals in Handbüchern erläutert wird. Maskierung erlaubt es also, unsichtbare akustische Barrieren zu errichten und folglich, den Soundscape in Zonen zu gliedern. Der Raum wird akustisch geformt und die entstehenden Grenzen müssen sich nicht mit den baulichen und visuellen Grenzen decken. Auch die „unberührte“ Natur bietet reichlich Klangmasken. Ein Wasserfall oder ein Brunnen haben eine ähnliche Wirkung wie die besagte Klimaanlage und der Wind in den Blättern eines Waldes kann so manches Tier akustisch verbergen. Jede Landschaft, die durchwandert wird, wird an bestimmten Stellen bestimmte Geräusche preisgeben und andere verbergen.

## 2.4 Karten

Eine vom World Soundscape aufgebrachte und als besonders wichtig erachtete Praktik beim Studium von Soundscapes ist das Anlegen von Landkarten – eine Praktik die Schafer „*aerial sonography*“ nennt. (Schafer, 1977, S.131f.) Schafer beschreibt zwei Techniken der Kartendarstellung: Die Isobel Konturlinien-Karte (engl. *isobel contour map*, von der Einheit *bel* abgeleitet) und die Ereignis-Karte (engl. *events map*). Erstere ist abgeleitet von den Höhenschichtlinien (Isohypsen) der topographischen Darstellung in der Geographie, in der Punkte gleicher Höhe verbunden werden, um Höhenunterschiede in einem zweidimensionalen Grundriss darstellen zu können. Statt Höhenmetern bezeichnen die dargestellten Konturlinien bei der Isobel-Karte aber Intensitätspegel, gemessen in (Dezi)bel. Diese Darstellung bezieht die Verteilung unterschiedlicher Klangquellen oder deren Effekt auf die Wahrnehmung nicht mit ein. Stattdessen beruht sie auf Messungen des Schalldruckpegels, die mit einem Messgerät vorgenommen wurden. Zweitere ist wesentlich interessanter für uns: Die Ereignis-Karte verzeichnet die Position von Klangereignissen, beziehungsweise den Hörbereich, in dem diese wahrnehmbar sind. Obwohl Schafer selbst einräumt, dass graphische Darstellungen von Soundscapes Behelfsmittel darstellen, die zwangsweise die gegebenen Verhältnisse nur unzureichend wiedergeben können, stellen diese doch wichtige Werkzeuge dar, um die räumlichen Relationen eines Soundscapes erfassen und die Verbreitung von Geräuschen dokumentieren zu können. Die Ereignis-Karte bringt ein wichtiges Element ins Spiel, das in den „klassischen“ Sound Studies vernachlässigt wird: Die Räumlichkeit eines Soundscapes.

Um eine derartige Karte eines Gebiets anzulegen, muss dieses gezielt abgegangen und konzentriert „belauscht“ werden. Das Resultat kann dann zu Papier gebracht werden. Schafer nennt diese Art des bewegten Hörens „*listening walk*“. (Schafer, 1977, S.212f.) Gemeinhin werden unter Soundscap-





Abbildung 2.4: Isobel-Karte von Dollar, Schottland (Copyright: Murray Schafer) (Truax, 1984, S.63)

pes Klangbilder, wie etwa atmosphärische Tonaufnahmen, verstanden, die meistens aus einem statischen Punkt heraus aufgenommen werden. Schafers Begriff des Soundscape leitet sich aber aus der Landschaft ab, und diese kann durchwandert werden. So steht dem Listening Walk steht der wesentlich bekanntere „*soundwalk*“ gegenüber. Im Gegensatz zum Listening Walk ist das Ziel des Soundwalk nicht die Analyse einer Umgebung und das Ergründen der absoluten Lage von Klangquellen. Vielmehr steht die persönliche Erfahrung einer klanglichen Umgebung im Vordergrund. Der Pfad des Spaziergangs ist vorgegeben und dient so als Partitur für eine Symphonie der „gefundenen“ Klänge. Der Spaziergang wird so zur inversen Choreographie der Klangquellen im Raum. Die „Komposition“ des Soundscape wird durch die Bewegung räumlich erfahrbar.

Die Erkundung eines Soundscapes mittels Listening Walk oder Sound Walk ist zweifelsohne eine dreidimensionale akustische Erfahrung. Dennoch fällt auf, dass die Ausrichtung auf die Landschaft gleichzeitig eine Orientierung an der Fläche darstellt. Der Raum der klassischen Sound Studies ist bis zu einem gewissen Grad ein zweidimensionaler. Auch wenn die Darstellung in Karten nur ein Behelfsmittel zur Analyse eines Soundscapes darstellen soll, so ist bereits durch die Wahl dieses Repräsentationsmittels schon eine Wertung vollzogen und ein inhärentes Raumverständnis erkennbar. Soundscapes sind Klanglandschaften, und Landschaften werden aufrecht erkundet, zumeist jedenfalls in konstanter Höhe. Variationen in der Höhe sind wenn dann durch das Terrain gegeben. Der architektonische Raum allerdings ist ein dreidimensionaler. Er windet sich in Wendeltreppen und Garageneinfahrten und schichtet sich in Stockwerken und Schubladen. Er erschließt sich im Gehen, Sitzen, Hocken, Kriechen und Liegen. Die Wege des Architektonischen Raumes kreuzen sich hundertfach in der Ebene wie in der Höhe, während die Wege der Landschaft zumeist Pfade mit Ziel und Ende sind. Selbstverständlich schließt die Landschaft Architektur mit ein, und auch der Soundscape tut dies – obwohl die Architektur aber den Soundscape lenkt und „färbt“, bleibt sie in diesem zumeist Staffage, denn der Raum der Sound Studies ist ein Raum der Relationen – ein Raum zwischen Klangobjekten und deren Einflusszonen: Ein Zwischenraum, der erst erst durch das Klangereignis entsteht.

Obwohl die Sprache der Soundscape-Schule unsere gebaute Umwelt mit einschließt, beschäftigt sie sich nur in Sonderfällen mit den passiven akustischen Aspekten unserer Umgebung, etwa beim Studium von Repräsentationsräumen oder Kirchen. Die vom World Soundscape Project beeinflussten Denker richten ihre Aufmerksamkeit in erster Linie auf aktive Geräuschquellen. Wir Menschen hören aber viel mehr als nur Lebewesen, Motoren, Lautsprecher und Naturereignisse. Wir hören Gras, Bäume, Wände, Fenster, Säulen, Gruben, Hallen etc. Insofern eignet sich das Konzept der Soundscapes nur begrenzt, um die klanglichen Eigenschaften von Raum zu beschreiben. Viel eher beschreiben Soundscapes Orte im Raum und deren Relationalität. Ein solcher Raumbegriff deckt sich aber nur begrenzt mit dem der Architekten. Selbstverständlich sind Zonen und Relationen von großer Bedeutung. Genauso aber ist die architektonische Hülle essentielle Komponente des architektonischen Raumverständnisses: Sie ist unsere dritte Haut. Insbesondere die auf Kommunikationstheorie basierende Deutung der Sound Studies eines Barry Truax bereitet aber einen Wahrnehmungswandel vor, um diese erfassen zu können. Zwar ist Truax' Fokus auf die Kommunikation gleichzeitig eine Konzentration auf die Botschaft, und somit auf deren Übermittler – also aktive Klangquellen. Durch das Verständnis von Soundscapes als Regelkreise der Kommunikation wird aber auch die Rückkoppelung des aktiven Klangs in der passiven Umwelt angedeutet. Die Verheiratung der Sound Studies mit Raum- und Psychoakustik erlaubt uns zu studieren, wie die passive akustische Umgebung den Schall formt und so auf uns wirkt.



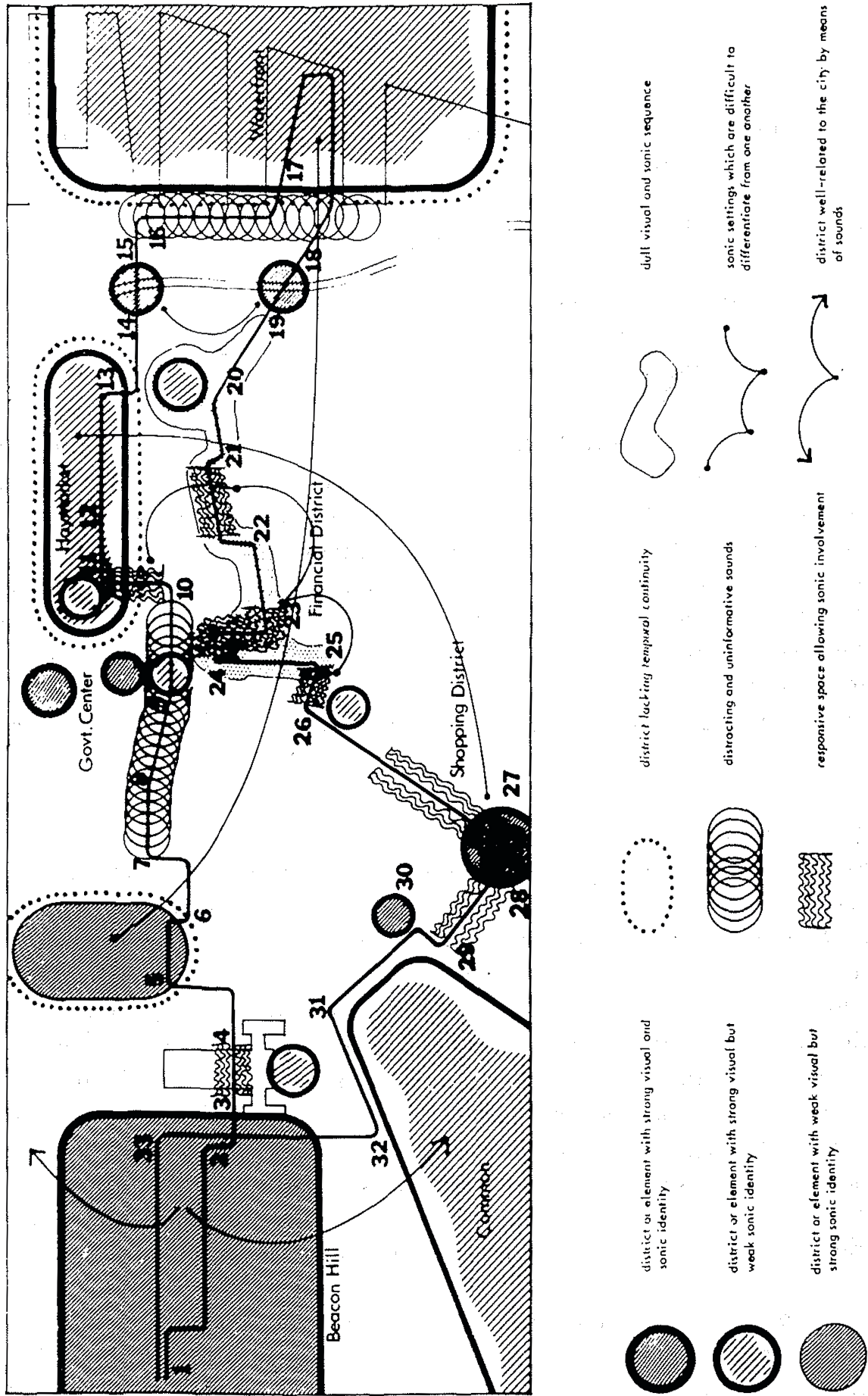


Abbildung 2.5: Ereigniskarte eines Teiles des Soundscape von Boston (Copyright: M. Southworth) (Truax, 1984, S.67)



## Aurale Architektur

### 3.1 Übersicht

Der Soundscape Schafers ist ein Konstrukt aus *aktiven* Schallquellen. Als Musiker interessierten Schafer und seine Schüler neben diesen Klängen und Geräuschen selbst zumeist nur deren Häufungen und Relationen. Dem Weg des Schalls von der schwingenden Saite, Luftsäule oder Membran hin zu seinem Rezeptor, unserem Ohr, bringen sie wenig Interesse entgegen. Aber gerade die räumliche Ausformung dieses Weges beeinflusst den Schall auf vielfältige Weise. Vom kleinen Resonanzraum im Inneren eines Instruments, über die Räume, in denen wir leben, bis in unser Innenohr durchläuft der Schall eine Kette von Transformationen, die ihn erst zu dem werden lassen, was wir als Soundscape verstehen können. Was ist die vom Bogen gestrichene schwingende Seite ohne den Resonanzraum der Geige, der ihr erst die eigentliche Klangfarbe gibt? Was der Klang eines Orchesters ohne den Nachhall des Konzertsaals, der die einzelnen Stimmen akustisch zu einer Einheit werden lässt? Der musikalische Hintergrund der Protagonisten der Sound Studies führt dazu, dass diese Aspekte vernachlässigt werden. Ihr Interesse gilt den Klängen selbst, deren Zusammenspiel und Wirkung. In einzelnen Beispielen mit herausragender Akustik findet selbige zwar Eingang in die Beschreibung des Soundscapes (Vgl. Schafer, 1977, S.214-225), aber in dem meisten Fällen liegt das Zentrum der Aufmerksamkeit wo anders: bei den aktiven Klangquellen und nicht bei unserer passiven akustischen Umgebung. Die Wissenschaft der Akustik auf der anderen Seite befasst sich zwar mit diesen Klangeigenschaften eines Raumes, nicht aber mit den sozialen und kulturellen Effekten des Klanges einer Architektur.

Die Modifikationen, die Raum am Klang vornimmt, können mit den Modifikationen durch das Effektgerät, das in der Tontechnik benutzt wird, verglichen werden. Es kann verzögern, verzerren, das Frequenzbild verändern, Hall erzeugen und noch viel mehr. Es existiert eine Vielzahl von Effektgeräten am Markt, die alle nur denkbaren Modifikationen vornehmen. Genauso gibt es Millionen von verschiedenen Räumen, die alle möglichen Modifikationen mit Schall anstellen, der

in diesen erzeugt wurde. Insofern ist es passend, dass ein Hallgeräte-Designer einer der Ersten ist, der sich den schallformenden Aspekten von Architektur aus der Perspektive der Sound Studies her annähert: Barry Blesser, in den 1970er Jahren Entwickler des berühmten EMT 250 Hall-Algorithmus, bringt 2006 zusammen mit der Umweltpsychologin Linda-Ruth Salter „Spaces Speak – Are You listening?“ heraus. „Spaces Speak“ stellt den Versuch dar, die akustischen Eigenschaften von Architektur und deren Wirkung auf uns Menschen begrifflich zu formulieren. Der überwiegende Teil der Beispiele, die sie untersuchen, findet sich zwar schon bei Schafer, Truax und vielen anderen Autoren, allerdings wechseln Blesser und Salter den Standpunkt. Was ihr Buch heraushebt, ist die Leistung, etwas herauszuarbeiten, was gemeinhin auf wenig Aufmerksamkeit stösst, und diesem einen Namen zu geben: aurale Architektur (engl. *aural architecture*). „Aural“ bezeichnet als Pendant zum soziokulturell geprägten „visuell“ die Effekte der Wahrnehmung des Akustischen und dessen kulturelle Bedeutung. Was ist nun das Besondere an der Konzeption der auralen Architektur? Sie wendet sich vom semiotischen Fokus der Sound Studies ab und bezieht naturwissenschaftliche Aspekte der Akustik und Psychoakustik mit ein, um aufzuzeigen, dass die akustischen Eigenschaften von Raum weit mehr Aufmerksamkeit verdienen, als sie gemeinhin etwa durch Massnahmen des Schallschutzes erfahren. Obwohl akustische Beobachtungen mit eingeschlossen sind, ist der Kern der Betrachtung die Wahrnehmung und Wirkung der klangformenden Eigenschaften von Architektur auf den Menschen.

Wie Blesser und Salter anmerken, besteht ein grundlegender Unterschied in der Konzeption des Soundscape-Begriffs und der Bedeutung des Wortes Landschaft. Während Landschaften leblos sein können, muss ein Soundscape belebt sein, um zu klingen. (Blesser, Salter, 2006, S.15) Und doch hat jede Landschaft, genau wie jede bauliche Struktur, ganz spezifische Eigenschaften, die zwar nicht unbedingt als aktive Geräusche in den Vordergrund treten müssen, aber etwaigen aktiven Geräuschquellen einen spezifischen Charakter verleihen, der ganz klar einen „Abdruck“ des Ortes im Klang erkennbar macht. Aber während die Soundscape Studies den resultierenden Klang beschreiben, versucht die aurale Architektur die Quelle des „Abdrucks“ aufzuspüren, zu beschreiben und aktiv zu gestalten. Michael Fowler beschreibt diese gegensätzliche Position in einem Artikel von 2008 folgendermassen: Im Soundscape ist der Raum im Klang enthalten, ja er produziert den Raum erst (*space-in-sound*), während bei der auralen Architektur der Klang im Raum erzeugt wird. (*sound-in-space*). (Fowler, 2008, S.3)

Blesser und Salter selbst beschreiben die aurale Architektur so: „*Aural architecture refers to the properties of a space that can be experienced by hearing [...] Hearing, together with its active complement, listening, is a means by which we sense the events of life, aurally visualize spatial geometry, propagate cultural symbols, stimulate emotions, communicate aural information, experience the movement of time,*

*build social relationships, and retain a memory of experiences. To a significant but underappreciated degree, aural architecture influences all of these functions.*“ (Blesser, Salter, 2006, S.4f.) Als aurale Architektur bezeichnen Blesser und Salter also vor allem die passiven akustischen Eigenschaften einer räumlichen Struktur. Sie stellt kein konkurrierendes Konzept zum Soundscape dar. Vielmehr beinhaltet jeder Soundscape eine aurale Architektur, die sich „zwischen“ dessen Klangquellen befindet und diese akustisch färbt und formt, und somit erst räumlich einfasst. Um noch einmal die Metapher aus der Tontechnik zu benutzen, ist also der Soundscape das musikalische Endprodukt eines Mischvorgangs, während die aurale Architektur von den zahlreichen Effekten, Equalizern und Kanalzügen der Mischkonsole des Tontechnikers repräsentiert wird. Wie aber kann die aurale Architektur aus dem Soundscape „herausgehört“ werden, wenn sie durch passive Elemente charakterisiert ist? Um dies herauszuarbeiten haben Blesser und Salter hierfür eine visuelle Metapher gewählt: Die „akustische Beleuchtung“ (engl. *sonic illumination*). Ähnlich wie Licht einen Raum erhellt, und diesen somit sichtbar werden lässt, macht erst ein Schallereignis einen Raum hörbar. (Blesser, Salter, 2006, S.15) Genauso wie dies bei visueller Beleuchtung der Fall ist, sind der Schlüssel zur akustischen Beleuchtung der auralen Architektur Reflexionen, und somit auch deren sensorischer Wahrnehmung.

### **3.2 Auditory Spatial Awareness**

Den kognitiven Mechanismus, der uns aurale Architektur wahrnehmen lässt, bezeichnen Blesser und Salter als „auditives räumliches Bewusstsein“ (engl. *auditory spatial awareness*). Sie beschreiben dieses als „[...] *a complex amalgam of spatial attributes, auditory perception, personal history, and cultural values* [...]“ (Blesser, Salter, 2006, S.11) Es repräsentiert jenen Teil unseres Bewusstseins, der auf die räumlichen Aspekte von Klang reagiert und manifestiert sich auf vier Arten: Erstens prägt es unser Sozialverhalten im Raum, zweitens ist es für unsere räumliche Orientierung zuständig, drittens prägt es unser ästhetisches Empfinden und viertens erweitert es unsere musikalische Erfahrung. Die vier Kategorien des auditiven räumlichen Bewusstseins korrespondieren laut Blesser und Salter mit vier Aspekten der auralen Architektur: sozial, navigatorisch, ästhetisch und musikalisch. Alle vier dieser Aspekte der auralen Architektur erschliessen sich durch die Reflexionen von Schall im Raum. Bei manchen ist dies offensichtlicher, bei anderen weniger. Am leichtesten nachvollziehbar ist dies bei den musikalischen Aspekten. So kann ein Raum als erweiterter Resonanzkörper eines Instruments betrachtet werden, denn seine akustischen Eigenschaften können den Klang von Musik färben. Dies kann sowohl zum Vorteil als auch zum Nachteil einer Aufführung ausfallen. Dass die Akustik bei der Gestaltung von Konzerthäusern von grosser Bedeutung ist, ist den meisten Menschen bewusst, und bedarf deswegen am wenigsten Erläuterung. Die ästhetischen Aspekte auraler Architektur können auf ähnliche Weise verstanden werden wie die musikalischen. Sie decken ab,

inwieweit wir einen Raum akustisch als angenehm oder ausgeglichen empfinden. Sie sind in erster Linie soziokulturell geprägt, denn was in einem Kulturkreis als schön oder hässlich betrachtet wird, kann in einem anderen völlig unterschiedlich verstanden werden. Die navigatorischen Aspekte der auralen Architektur erlauben durch die Interpretation der Reflexionen in einem Raum, Distanzen zu Objekten und sogar die Materialität von Oberflächen herauszuhören. Sie decken also sowohl die Orientierung im Raum ab, als auch die Geometrie desselben. Die sozialen Aspekte der auralen Architektur schließlich betreffen zu einem grossen Teil den Raumklang, und wie sehr er der Kommunikation dienlich ist, bzw. wie er die akustischen Arenen beeinflusst. So kann die Akustik eines Raumes öffentlichen Charakter haben, oder die aurale Privatsphäre stärken. Je nachdem wird sich die Akustik des Raumes auch auf seine Nutzung auswirken.

### **3.3 Raum Formt Schall**

In Kapitel 1 habe ich beschrieben, wie die Lage von Klangquellen im Raum dadurch erfasst werden kann, dass unser Gehirn mit Hilfe unsere Ohren Karten unserer Umgebung anlegt, die in der Folge mit sensomotorischen und optischen Karten synchronisiert werden. Allerdings bin ich bei der Beschreibung der Lokalisierungsfähigkeiten des Menschen von einem Idealzustand ausgegangen, nämlich dass eine Klangquelle störungsfrei wahrgenommen werden kann. Wenn wir uns jedoch durch unsere Umwelt bewegen, ist es in den seltensten Fällen der Direktschall von isolierten Klangquellen, die wir hören. Was unser Ohr tatsächlich erreicht, ist ein komplexes Klanggemisch, das seine Ursachen in einem System hat, das man als verschachtelte räumliche Anordnung von Klangregnern, Resonatoren und Filtern begreifen kann. Direktschall ist nur eine Komponente in dem Gemisch an Geräuschen und Klängen, welches tagtäglich an unsere Ohren dringt, denn Schall wird an Hindernissen (z.B. Wänden) reflektiert, ähnlich wie dies mit Licht an einem Spiegel geschieht. Dabei gleicht gemäß dem Reflexionsgesetz der Eingangswinkel dem Ausgangswinkel, sofern die Oberfläche glatt und hart ist. (Kuttruff, 2009, S.102) Mit zunehmender Rauheit der Oberfläche werden die Reflexionen unregelmäßiger gestreut, mit zunehmender Weichheit selbiger werden sie gedämpft. Kanten, Öffnungen oder Unterbrechungen in der reflektierenden Oberfläche streuen die Reflexionen. (Kuttruff, 2009, S.35)

Beispielsweise wird eine Klangquelle, die ein rhythmisches Klicken aussendet und im Freien positioniert ist (und in unserem idealisierten Beispiel von keinen reflektierenden Oberflächen umgeben ist) mittels ITDs und ILDs (und HRTF) klar lokalisierbar sein. (Vgl. Kapitel 1) Stellen wir uns jedoch vor, dass in einigen Metern Abstand zur Klangquelle eine glatte ebene Wand positioniert ist, ändert sich die Lage drastisch, denn was wir hören, ist nicht nur der direkte Schall der Klangquelle. Hinzu gesellt sich eine Reflexion, die man als „Spiegelquelle“ bezeichnet. Sie liegt von der ursprüng-



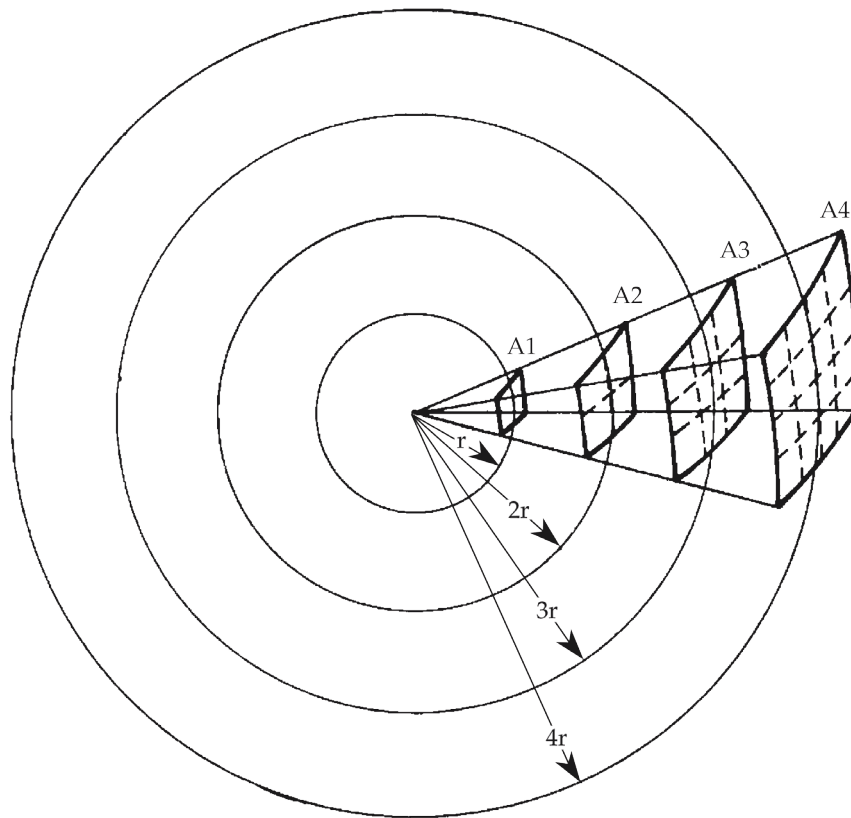


Abbildung 3.1: Zusammenhang zwischen Schallausbreitung, Intensitäts- und Druckabnahme. (Everest, Pohlmann, 2009, S.34)

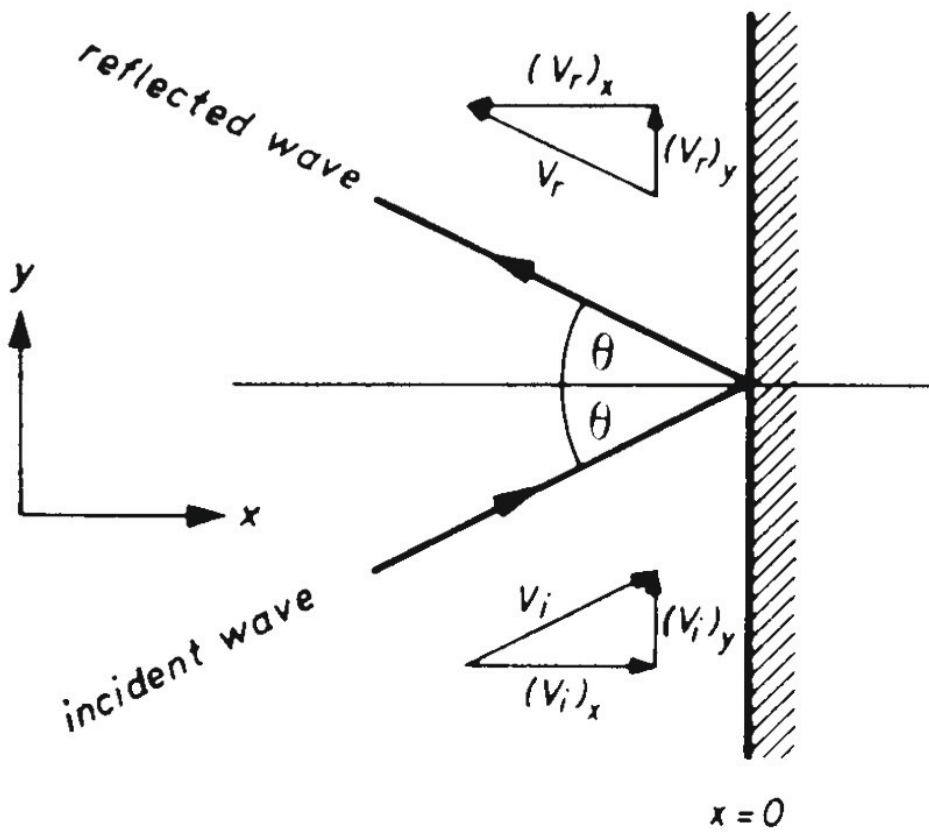


Abbildung 3.2: Reflexion gemäß dem Reflexionsgesetz. (Kuttruff, 2009, S.42)

lichen Klangquelle im Lot zur reflektierenden Fläche, und hat, wie ein Spiegelbild, den gleichen Abstand zur Wand wie das Original. Es handelt sich also um eine Kopie des Original-Signals, das uns jedoch aus einer anderer Richtung erreicht. Diese Spiegelquelle ist aber auch in anderer Hinsicht gegenüber dem Original leicht verändert. So erreicht sie uns Erstens leicht verzögert, weil der Schall über den Umweg der reflektierenden Oberfläche einen längeren Weg zurücklegen musste. Zweitens ist sie leiser als das Original, weil sie auf dem Weg an Energie verloren hat. Drittens hat sie womöglich spektrale Veränderungen durchgemacht, weil durch den Reibungswiderstand der Luft und die Beschaffenheit der Oberfläche manche Frequenzen stärker abgeschwächt werden als andere. So „reisen“ langwellige (tiefe) Frequenzen weiter als hohe, kurzwellige Frequenzen, welche wiederum schneller abgedämpft werden.

Aufgrund des sogenannten „Gesetzes der ersten Wellenfront“ (engl. *precedence effect*) werden Direktschall und Spiegelquelle aber nicht als getrennte auditive Ereignisse wahrgenommen. (Blauert, 1999, S.279) Stattdessen wird die Reflexion aufgrund der späteren Ankunftszeit dem Direktschallereignis zugeordnet. Dies funktioniert in einem gewissen Zeitfenster, das von Frequenzgehalt und Ereignisdauer des Original-Signals abhängt. Direktschall und Reflexionen die in einem Bereich von ungefähr 2 – 40 Millisekunden nach ersterem eintreffen, werden vom Gehirn zu einem Auditiven Ereignis zusammengefasst. (Pierce, 1999, S.94)<sup>1</sup> Ab einer Verzögerung von etwa 40 Millisekunden spricht man von einem Echo, das als eigenständiges Klangereignis registriert wird.

Diese Reflektionen bezeichnet man als *Erstreflektionen*, da sie der erste Teil der sogenannten Hallfahne<sup>2</sup> sind, der den Hörer erreicht. Man könnte annehmen, daß diese Erstreflektionen die Lokalisierungsfähigkeiten des Menschen beeinträchtigen, da zusätzlicher Schall ja das deutliche Hören des Direktschalls erschwert, aber tatsächlich können diese Reflexionen für die Ortung sogar nützlich sein. Unsere Hörbahn kann die Erstreflektionen dekodieren, und aus Verzögerung und Veränderung gegenüber dem Direktschall die reflektierenden Hindernisse erkennen, ähnlich wie dies in der Schifffahrt mittels Sonar und bei Fledermäusen oder Zahnwalen mittels Ultraschallortung bewerkstelligt wird. Bei Lebewesen spricht man bei dieser Fähigkeit von Echolokation. Zwar ist sie bei den meisten Menschen nicht sehr stark ausgeprägt und den wenigsten auch bewusst, aber im Grunde besitzt jeder die Anlage dazu. Beispielsweise haben viele Menschen kein Problem dabei, im Dunkeln durch ihre Wohnung zu navigieren wenn sie in der Nacht die Toilette aufsuchen müssen. Dabei spielt zwar auch die Gewöhnung an den Wohnraum und das „sensomotorische Gedächtnis“ an dessen Raumstruktur eine wichtige Rolle, aber man kann aufgrund des Vertrauens in die ge-

---

<sup>1</sup> Bei sehr vielen ähnlichen Ereignissen können diese in unserer Wahrnehmung verschmelzen, wodurch die Lokalisierung einzelner Ereignisse unmöglich wird. Ein Beispiel hierfür ist etwa ein Schwarm Grillen oder Zikaden, ein anderes die einzelnen Regentropfen bei starkem Regen.

<sup>2</sup> Hallfahne ist ein umgangssprachlicher Ausdruck, der sich vom optischen Erscheinungsbild eines Echogramms ableitet.



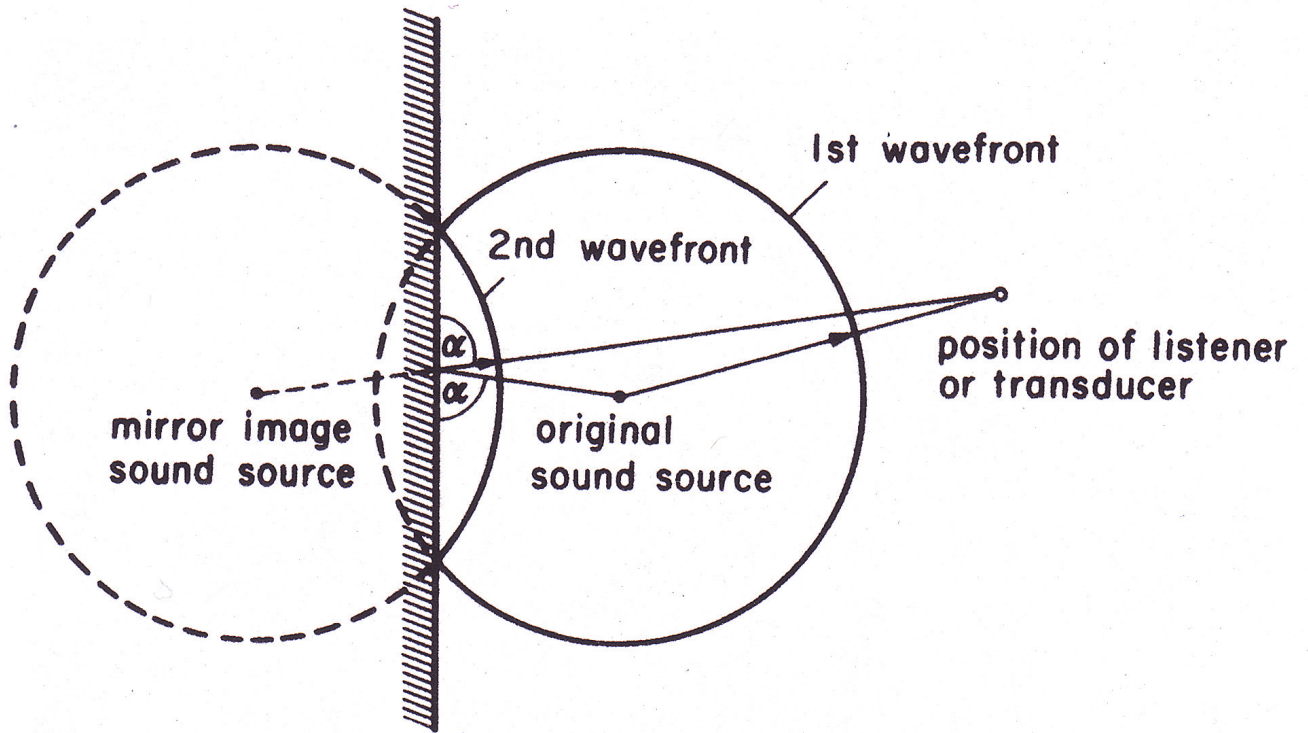


Abbildung 3.3: Veranschaulichung der ersten Wellenfront. (Blauert, 1997, S.31)

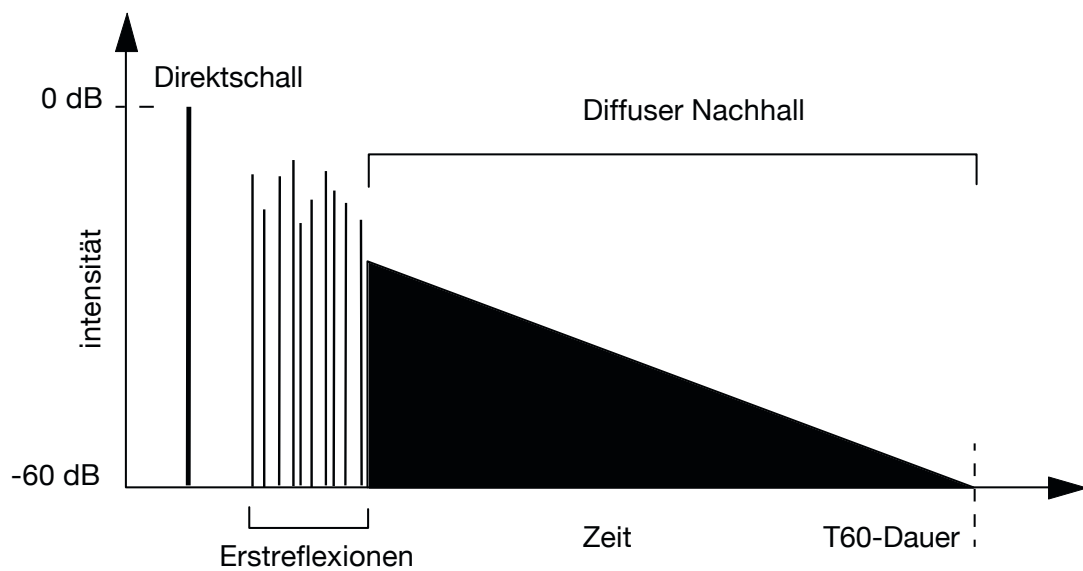


Abbildung 3.4: Hallfahne (als Impulsantwort, bzw. vereinfachtes Echogramm) (Dahl, Jot, 2000, S.5)

wohnte Umgebung ein Gefühl der Sicherheit erlangen, das es erst erlaubt, auf die Hinweise unserer Ohren zu vertrauen. In einer solchen Situation wird bewusst, dass es möglich ist, Wände, Türen und Fenster zu erhören. Man könnte also sagen, dass den Menschen generell nur etwas Training (und Anlass bzw. Gelegenheit hierzu) in diesem Bereich fehlt. Blinde Menschen zeigen vor, wie stark ausbaubar diese Fähigkeit eigentlich ist, denn für viele von ihnen bedeutet der Verlust des Augenlichts ein erhöhtes Bewusstsein für unsere klangliche Umgebung, und in der Folge, gesteigerte Fähigkeiten mittels Echolokation navigieren zu können. (Blessner, Salter 2007, S.37-41) Barry Truax berichtet in „Acoustic Communication“ von der Technik von kanadischen Fischern, die Pfiffe benutzten, um ihre Boote bei schlechten Sichtverhältnissen in Ufernähe gefahrlos navigieren zu können, um Kollisionen zu vermeiden. (Truax, 1984, S.18) Überdies können Erstreflexionen uns auch die Materialität bzw. Härte der Oberfläche eines Objektes erkennen lassen. Materialien reflektieren verschiedene Frequenzbereiche je nach Oberflächenbeschaffenheit unterschiedlich gut. Eine Stahlplatte wird beispielsweise sehr helle und klare Erstreflexionen zurückwerfen, während eine Holzwand dumpfer klingt. (Vgl. Teil 2, Kapitel 7.2) Die Interpretation dieser Reflexionen lässt sich also bis zu einem gewissen Grade lernen.

Erstreflexionen gibt es wie in unserem obigen Beispiel auch im Freien, so reflektierende Oberflächen vorhanden sind. In geschlossenen Volumina aber (sowie in Situationen mit mehreren gegenüberliegenden und reflektierenden Oberflächen) gibt es noch ein weiteres Phänomen, das mit etwas zeitlichem Abstand nach den Erstreflexionen hörbar wird. Graphik 3.4 zeigt die Struktur der Reflexionen in einem Punkt eines Raumes. Die Ankunftszeiten werden hier auf der x-Achse und die Intensitäten auf y-Achse aufgezeichnet. Man bezeichnet eine derartige Darstellung einer Hallfahne auch als Echogramm. (Kuttruff, 2009, S.109) Der Schall wird nach dem Eintreffen der Erstreflexionen immer weiter hin und her reflektiert und gestreut, bis mit zunehmender Zeit die Dichte der Reflexionen immer mehr zunimmt bis sie einen Grad erreicht, bei dem das ursprüngliche Schalleignis nicht mehr identifiziert werden kann. Man nennt diesen Bestandteil des Nachhalls „diffuser Nachhall“ (engl. *reverberation*). Der US-amerikanische Physiker Wallace Clement Sabine definiert diesen so: „*The prolongation of sound in a room after the cessation of its source may be regarded either as a case of stored energy which is gradually suffering loss by transmission through and absorption by the walls and contained material, or it may be regarded as a process of rapid reflection from wall to wall with loss at each reflection.*“ (Sabine, 1923, S.74) Sabine gilt als Pionier der Erforschung von Raumakustik und entwickelte um das Jahr 1900 als erster Kennwerte für Schallübertragung, Schallabsorption und Nachhallzeit in Räumen.<sup>3</sup> Seine Formel für die Errechnung der Nachhallzeit von Räumen (die heute als T60 bzw. im englischen Sprachraum als RT60 bezeichnet wird) beschreibt den Zeitraum

---

<sup>3</sup> Sabine konnte seine Erkenntnisse anhand des Baus der Boston Symphony Hall austesten, die heute neben dem Wiener Musikverein und dem Leipziger Gewandhaus als einer der besten Konzertsäle der Welt gilt. (Schwartz, 2011, 465ff.)

vom ersten Erklängen des Direktschalls, bis zum Zeitpunkt, in dem der diffuse Nachhall in einem Raum um 60dB gegenüber dem Ursprungspegel abgeklungen ist, was einem Tausendstel an Intensität entspricht. Diese Nachhallzeit ist proportional zum Raumvolumen und umgekehrt proportional zur Absorptionsfläche des Raumes. (Sabine, 1923, S.25) (Vgl. Teil 2, Kapitel 7.3)

Während Erstreflexionen der menschlichen Lokalisierungsfähigkeit zuträglich sind, hat diffuser Nachhall den gegenteiligen Effekt, denn er verwischt das Schallfeld je nach Größe und Rauheit der das Volumen umschließenden Oberflächen. Diffuser Nachhall ist per definitionem ungerichtet, denn er entsteht überhaupt erst, wenn der ursprüngliche Direktschall alle möglichen Winkel des Raumes durchlaufen hat. Es handelt sich bei ihm um Reflexionen, die bereits so breit gestreut und zeitlich verzögert sind, dass aufgrund von reinem diffusen Nachhall eine Lokalisierung unmöglich wird. (Pierce, 1999, S.98) Je größer die Distanz der Klangquelle zum Hörer ist, desto lauter ist der diffuse Nachhall im Verhältnis zum Direktschall und zu den Erstreflexionen und desto schwieriger wird es auch, eine Klangquelle zu orten. Unser Gehirn nimmt Klangereignisse, die einen hohen Anteil an diffusem Hall aufweisen, automatisch als weit entfernt wahr, und zwar unabhängig von der Gesamtlautstärke. (Blauert, 1999, S.279; Shepard, 1999, S.26) Nachhall erschwert aber nicht nur das Lokalisieren von Klangquellen aufgrund seiner Diffusität, sondern erschwert auch die Verständlichkeit von Klanginformationen, die er auch zeitlich „verwischt“. Und schließlich ist das Verhältnis zwischen Direktschall, Erstreflexionen und diffusem Nachhall von zentraler Bedeutung dafür, als wie angenehm wir den Klang eines Raumes empfinden. All dies hat auch soziale Konsequenzen. So vermittelt etwa ein Raum, der einen starken Nachhall-Anteil aufweist, ein Gefühl der Öffentlichkeit, das sich auf die gesellschaftlichen Interaktionen auswirkt.

Insgesamt wird starker diffuser Nachhall zumeist als zu vermeidendes Übel gesehen, aber obwohl er die Lokalisierung und das Verständnis von Informationen erschwert, kann er trotzdem wertvolle Informationen über die Struktur eines Raumgefüges vermitteln. Im Speziellen die Ausdehnung und Materialität eines Raumes lässt sich auf Basis des Nachhalls und dessen Klangfarbe ermitteln, sowie des Grades der „Glattheit“ beziehungsweise „Gefülltheit“ des Volumens. So mag uns Nachhall in einem mittleren Raum in Wohnzimmergröße kaum bewusst sein, aber in einer großen Halle bemerkt werden. Was wir dann wahrnehmen wird unter Umständen nicht der Nachhall selbst sein, sondern die Größe der Halle, die uns plötzlich bewusst wird. (Shepard, 1999, S.29) Blesser und Salter schreiben hierzu: *„[...] [V]ision both decodes size as length, width, and height, and organizes distance by the way objects obscure one another or change their relative size. In contrast, hearing decodes size as the global metric of volume because sound permeates air as a fluid, flowing around objects and into crevices. We cannot see volume, but we can hear it. Aurally, we sense the volume of a large space by its long reverberation time and the volume of a small space by its sharp frequency resonances. Visually, we can*

*sense volume only by mentally multiplying the three dimensions of a space.*“ (Blessner, Salter, 2006, S.21)

Eine Sonderstellung unter den Schallreflexionen nehmen die Resonanzen – sogenannte „stehenden Wellen“ – ein. Stehende Wellen entstehen aus der Überlagerung zweier gegenläufig fortschreitender Wellen gleicher Frequenz und gleicher Amplitude. Es handelt sich also um Wellen, deren Auslenkung an einem bestimmten Punkt immer Null ist – sie bleiben also an einer Position im Raum „stehen“. Sie machen sich bemerkbar, wenn beispielsweise in einem Raum ein Klangereignis ertönt, welches eine Frequenz enthält, deren Wellenlänge gleich der Länge einer Raumachse ist. Voraussetzung dafür ist, dass die begrenzenden Oberflächen parallel zueinander stehen. In solch einem Fall „passt“ die betreffende Welle exakt in diesen Bereich hinein. Wird sie nun hin und her reflektiert, hat sie an den Endpunkten (dh. an den Wänden) immer ihren Amplituden-Nullpunkt, während sie in der Mitte zwischen den reflektierenden Oberflächen einen „Bauch“ aufweist. D.h. an den Endpunkten herrscht ein Schalldruck-Maximum, aber ein Schallschnelle-Minimum (die Welle „stösst“ an die Oberfläche, die Schnelle sinkt auf Null, der Druck steigt, worauf sie sich in die Gegenrichtung entlädt), in der Mitte aber ein Schalldruck-Minimum und ein Schallschnelle-Maximum (die Welle hat ihre maximale Auslenkung erreicht). Dies gilt auch für ganzzahlige Vielfache der Grundfrequenz. Das dreifache der Grundresonanz wird dementsprechend zwei Nulldurchgangs-Punkte und drei Auslenkungmaxima erreichen. (Westphal, 1958, 194-198)

Hörbar wird dies dadurch, dass die betreffende Frequenz gegenüber anderen Frequenzen im Raum weitaus weniger abgeschwächt wird. Zwei parallele Wände etwa, die in einer Distanz von 3.5 Metern zueinander stehen, weisen eine Resonanzfrequenz von 47 Hertz auf. Ertönt zwischen diesen beiden Wänden ein Schallereignis, welches Anteile dieser Frequenz enthält werden eben diese – sowie ganzzahlige Vielfache dieser Frequenz – länger und stärker reflektiert werden als andere. Die bekannten Verzerrungen im Frequenzbild beim Singen im Badezimmer ergeben sich aus diesem Phänomen: Da die meisten Bäder verflieste Wände und somit besonders harte und glatte Oberflächen aufweisen, haben sie zumeist auch besonders starke Reflektionseigenschaften. Aus diesem Grund sind Resonanzen hauptsächlich dafür verantwortlich, als wie schön oder hässlich wir die Akustik eines Raumes empfinden. Da sie je nach Akustik des betreffenden Raumes stark in den Vordergrund treten können, werden sie unter manchen Bedingungen als eigenständige Klangergebnisse wahrgenommen. Stehen die Teilfrequenzen, genannt Raummoden, etwa in harmonischen Intervallen zueinander, können Resonanzen als musikalisch empfunden werden. Unter bestimmten Bedingungen können sie aber auch kalt oder bedrohlich wirken und sogar klaustrophobische Gefühle auslösen.

Eine effektvolle Demonstration von Raumresonanz stellt das berühmte Stück „I am sitting in a

room“ des US-amerikanischen Komponisten Alvin Lucier dar.<sup>4</sup> In einem Hotelzimmer sitzend nahm Lucier einen einzigen Satz mit einem Mikrofon auf. *„I am sitting in a room different from the one you are in now. I am recording the sound of my speaking voice and I am going to play it back into the room again and again until the resonant frequencies of the room reinforce themselves so that any semblance of my speech, with perhaps the exception of rhythm, is destroyed. What you will hear, then, are the natural resonant frequencies of the room articulated by speech. I regard this activity not so much as a demonstration of a physical fact, but more as a way to smooth out any irregularities my speech might have.“* Die Aufnahme spielte er also wieder über eine Audioanlage im gleichen Raum ab und nahm währenddessen den Klang im Raum erneut auf. Diesen Vorgang wiederholte er so lange, bis letzten Endes die Stimme vollkommen in den Resonanzen des Raumes unterging. Resonanz wird hier zum kompositorischen Element. Gleichzeitig vermag die Aufnahme auf etwas aufmerksam machen, das manchem Hörer vielleicht nicht bewusst gewesen war, denn sie hebt akustische Eigenschaften hervor, die in vielen Räumen kaum merkbar sind. Salome Voegelin drückt es so aus: *„The repetitions erase its architectural certainty rather than stabilizing it. In the end he is not sitting in a room at all anymore. Instead he is sitting in pure sound; the reverb and repetition having performed an acousmatic reduction to the core of sonic timespace: that of his enunciation and that of my listening“* (Voegelin, 2010, S.127f.)

Reflexionen sind also der Schlüssel zur Raumakustik und somit auch zur menschlichen Fähigkeit, architektonische Strukturen zu *er*-hören. Wir hören also nicht nur aktive Klangquellen, sondern auch unsere passive Umgebung. Hörend rekonstruiert unser Gehirn ein räumliches Abbild unserer Umwelt; zwar wird dieses erst im Zusammenspiel mit den anderen Sinnen komplettiert, aber trotzdem ist es erstaunlich, wie viel von dem unser Gehör leisten kann, was wir sonst etwa eher bei Seh- und Tastsinn verortet hätten. Roger Shepard bemerkt hierzu: *„It is curious that the addition of walls and boundaries, essentially limiting space, gives the sense of spaciousness in audition. In a purely anechoic room (a specially constructed space that minimizes reflections from the walls, floor, and ceiling) we get no reverberation, and thus no sense of space. In vision, too, if an observer were in space with no objects around, there would be no sense of the space. Gibson pointed out that we do not perceive space but, rather, objects in space. In audition, we need surfaces to give us the sense of the space they define.“* (Shepard, 1999, S.27f.)

### 3.4 Auditive Aufmerksamkeit

Das Er-hören von Raum ist ein auditiver Prozess, der zum Teil unbewusst abläuft. Allerdings erlaubt erst das bewusste *Zuhören*, Meisterschaft in dieser Technik zu erlangen. Die meisten Menschen

---

4 Lucier, Alvin (1969): I Am Sitting In A Room. New York, Lovely Music Ltd.

werden problemlos in der Lage sein, aufmerksam einer Stimme zu lauschen oder in auditiver Bereitschaft zu verbleiben, um auf bestimmte Klangsignale zu reagieren zu können. Dies ist zum Teil möglich, weil Stimmen und Signale Elemente sind, die als Symbole wiedererkannt werden können. Diesem Wiedererkennungsprozess können aber Jahre des Lernens vorausgegangen sein. An dieser Stelle ist es vonnöten, über die Art und Weise zu sprechen, wie wir Menschen zuhören. Dies ist keine exklusive Domäne der auralen Architektur. Auch ein Soundscape kann auf verschiedene Arten wahrgenommen werden. Darum haben zahlreiche Theoretiker versucht, die unterschiedlichen Arten und Weisen wie wir dies bewerkstelligen zu beschreiben. Blesser und Salter beschreiben diese Wahrnehmungsmodi als „auditives Bewusstsein“ (engl. *auditory awareness*). Dieses liegt dem auditiven räumlichen Bewusstsein zugrunde. Sie gruppieren diese in drei Ebenen namens *raw sensation*, *perception* und *attentive* (bzw. *emotionally engaged*) *listening*. (Blesser, Salter, 2007, S.14) Der „rohe Sinneseindruck“ (engl. *raw sensation*) umfasst die biomechanische- und neurologische Wandlung von Schalldruck in neuronale Impulse, umfasst somit also auch die in Kapitel 1 behandelte erste Rekonstruktion der räumlichen Umgebung in den Olivkernen. Aufbauend auf dieser Ebene entsteht die bewusste Wahrnehmung (engl. *perception*) durch die persönliche Geschichte sowie Vorkenntnisse des Individuums, für welches das Gehörte Sinn ergibt. Diese Ebene der Wahrnehmung ist zu einem grossen Teil kulturell geprägt. Es handelt sich also um Schallinformationen, die verstanden werden und deshalb verarbeitet werden. Das aufmerksame Zuhören (engl. *attentive listening*) schließlich umfasst die Aspekte des Gehörten, die hohe emotionale Resonanz und Erregung erzeugen. Dies umfasst sowohl positive wie negative Reaktionen. Diese Begriffe sind nicht als naturwissenschaftliche Fakten zu verstehen. Vielmehr ist das Gebiet des auditiven Bewusstseins Gegenstand andauernder Forschung und Diskussion.

Tatsächlich finden sich erste Hinweise für verschiedene Hörmodi bereits in der (englischen) Sprache. *Hearing* wird gemeinhin für das passive Hören verwendet, während *listening* zumeist für ein aktives Zuhören steht und im Deutschen etwa auch mit Horchen oder Lauschen übersetzt werden kann. Auch Murray Schafer unterscheidet in „The Soundscape“ unter aktivem, aufmerksamem „*concentrated listening*“ und „*peripheral hearing*“ (Schafer, 1977, S.117). Ein probates Mittel um das Gehör zu schärfen ist laut Schafer „*ear cleaning*“, womit nicht die wörtliche Reinigung des Ohres von physischen Verunreinigungen gemeint ist. Stattdessen ist der Begriff im Kontext von Schafers akustischer Ökologie zu verstehen und beschreibt eine Reinigung des von Lärm beanspruchten Gehörs und auditiven Bewusstseins. So beschreibt Schafer im gleichnamigen Band „Ear Cleaning“ verschiedene Übungen und Experimente, um das aktive Zuhören (insbesondere bei Musikern) zu trainieren. (Vgl. Schafer, 1967) Wesentlich komplexer ist aber das dreistufige Hörmodell, das Barry Truax in „Acoustic Communication“ definiert, welches auch als Basis für Blessers und Salters Modell gedient haben dürfte. Truax beschreibt „*hearing*“ als biologisch-neurologischen Akt der



Schallwandlung. (Truax, 1984, S.13) Die Hörmodi, allesamt als „*listening*“ klassifiziert, unterteilt er in *listening-in-search*, *listening-in-readiness* und *background listening*. *Listening-in-search* stellt das Äquivalent zu Bessers und Salters *attentive listening* dar. Es zeichnet sich durch einen hohen Grad von fokussierter Aufmerksamkeit aus, in welchem ein Individuum die Umgebung aktiv nach Hinweisen durchsucht. Die Fähigkeit, bestimmte Klangereignisse auszublenden, um zu fokussieren – der sogenannte „Cocktail Party Effekt“ – ist essentieller Bestandteil dieser Art, zuzuhören. (Truax, 1984, S.19) *Listening-in-readiness*, also Hören-in-Bereitschaft, ist stark abhängig von erlernten Assoziationen und erlaubt, auch im Hintergrund vertraute Signale zu identifizieren, um jederzeit ins *Listening-in-search* zu wechseln. Truax nennt als Beispiel für diesen Hörmodus eine Mutter, die im Schlaf sofort auf das Schreien ihres Kindes reagieren kann, aber von etwaigem Strassenlärm nicht geweckt wird. *Listening-in-readiness* ist stark abhängig von dem vorherrschenden „*signal-to-noise*“-Pegel. (Vgl. Kapitel 2.3) (Truax, 1984, S.19f.) Beim *Background listening* schließlich verbleibt das Gehörte im Hintergrund und nimmt so keine Aufmerksamkeit ein, obwohl man sich, darauf angesprochen, sehr wohl an die Geräusche erinnern würde. *Background listening* korrespondiert laut Truax mit den *Keynote sounds* der *Soundscapes*. „[...] [*T*]his type of listening differs from “subliminal” perception which is defined as the total lack of conscious awareness of a perception, but with later behavioral evidence that something has been experienced at a subconscious level (Dixon, 1971).“ (Truax, 1984, S.21) Eine umfassende Übersicht über die grosse Zahl an unterschiedlichen Begriffen und Konzepten des Zuhörens in der Literatur der *Sound Studies* bietet Florian Hollerweger in seiner Dissertation „The Revolution is Hear!“. (Hollerweger, 2011, S.54-63) Gemein ist all diesen, dass sie versuchen, die verschiedenen Grade an Aufmerksamkeit und Bewusstsein des Hörens zu kategorisieren.

Das Zuhören ist also Grundvoraussetzung, um die aurale Architektur wahrnehmen zu können. Je nach persönlichem Hintergrund nehmen verschiedene Individuen in bestimmten Situationen automatisch bestimmte Hörmodi ein. Je nach Person unterscheidet sich auch der Grad der Aufmerksamkeit, der bestimmten Aspekten der akustischen Umgebung entgegengebracht wird, denn die Bereiche des auditiven räumlichen Bewusstseins sind üblicherweise nicht gleichmässig ausgebildet. Die Effekte der Maskierung beispielsweise haben direkte Auswirkung auf die Schallwandlung, und treten deswegen bei trainierten wie untrainierten Personen gleichermassen in Erscheinung. Emotionale Effekte der auralen Architektur könnten wiederum unbemerkt bleiben und trotzdem ihren Einfluss auf Menschen ausüben. Ein Individuum könnte sich also in einem Raum unwohl oder wohl fühlen, ohne die Ursache dieses Gefühls – in Ermangelung von auraler Erfahrung – ausmachen zu können. Die bewusste Interpretation navigatorischer Hinweise der Umgebung durch Reflexionen wiederum erfordert ein hohes Mass an Konzentration und Training. Zwar können die meisten Menschen etwa ihre Hand „hören“, wenn sie diese direkt vor das Ohr halten, wie Besser

und Salter beschreiben (Blessner, Salter, 2006, S.36), der sichere Gang aber durch eine unbekanntere Umgebung im Dunklen stellt aber eine wesentlich komplexere räumliche Herausforderung dar. Die musikalischen Qualitäten einer auralen Architektur wiederum erschliessen sich erst im jeweiligen präferierten Stil und den Hörgewohnheiten sowie der Sensibilität der Hörer. Historische Musikstile sind in enger Verbindung mit dem designierten Aufführungsraum entstanden. Die Nachhallzeit der Konzerthäuser des 19. Jhdts. sind genau darauf ausgelegt, die Orchester diese Zeit akustisch zu einer Einheit zusammenzuschweißen. Der Nachhall einer gotischen Kathedrale wiederum simuliert durch seine Dauer eine Polyphonie, die in den monophonen Chorälen eigentlich nicht vorhanden ist. Umgekehrt wiederum kollidiert die elektrische Verstärkung mit den akustischen Eigenschaften einer solchen Kathedrale. Erst mit der „Entkörperung“ des Klanges durch die Aufnahme- und Wiedergabe-Techniken des 20. Jhdts. wird ein „trockener“ (also hallarmer) Aufführungsraum zum Ideal. Soziale Effekte wiederum werden vor allem durch Maskierung verursacht, ergeben sich aber in gewissem Masse auch durch die drei anderen Kategorien.

Einer der Gründe für die Notwendigkeit, den Begriff der auralen Architektur zu definieren, liegt in den vielfältigen Problemen die sich ergeben können, wenn die auralen Eigenschaften einer räumlichen Struktur nicht in Einklang mit deren intendierter Nutzung sind. Blessner und Salter stützen die sozialen Aspekte der auditiven räumlichen Wahrnehmung auf Barry Truax' Konzepte der akustischen Arena und des akustischen Horizonts. (Blessner, Salter, 2007, S.21-37) Diese werden zwar wie in Kapitel 2 erwähnt durch Maskierung bestimmt, aber auch durch die passiven akustischen Eigenschaften eines Raumes geformt. „*Although echoes and reverberation are the space's response to the target sound, we can think of the space as creating its own sonic noise by accumulating old and obsolete target sounds.*“ (Blessner, Salter, 2007, S.23) Das bedeutet, dass nicht nur Klangereignisse selbst, die innerhalb eines Raumes erzeugt werden, verantwortlich für Maskierungseffekte sind, sondern dass die Struktur einer Architektur selbst die Bildung selbiger in sich trägt. Ein Extrembeispiel hierfür wäre etwa die „whispering gallery“ in der St. Paul's Cathedral in London, die aufgrund ihrer Form Reflexionen bündelt, und so Punkte in der Kuppel, die eigentlich ausserhalb der Hörweite liegen, zu einer gemeinsamen akustischen Arena vereint. Besonders gekrümmte Oberflächen, aber auch gerade Zwischenwände, Säulen und Stützen, aber auch Möbel und andere Gegenstände in einem Raum können den Schall lenken und streuen oder bündeln und so akustische Arenen und Horizonte umformen. Schall kann also durch gezielte Eingriffe an der Geometrie gelenkt werden. Dies kann der Funktionalität einer baulichen Struktur zugute kommen. Es gibt aber zahlreiche Fälle, wo das Gegenteil der Fall ist. So ziehen Blessner und Salter das Konzept der sozialen Sphären von Edward T. Hall heran.<sup>5</sup> Hall unterschied vier soziale Sphären, die Individuen umgeben: Die intime

---

<sup>5</sup> Hierbei muss angemerkt werden, dass diese sozialen Sphären und ihre Ausdehnung kulturell bedingt sein können, weshalb die jeweilige Distanzmetrik mit Vorsicht zu geniessen ist.



Sphäre (engl. *intimate sphere*) reicht ungefähr einen halben Meter und ist reserviert für Familie und enge Freunde. Die persönliche Sphäre (engl. *personal sphere*) endet bei etwa einem Meter Distanz und ist reserviert für Bekannte. Die Konversations-Sphäre (engl. *conversational sphere*) endet bei etwa vier Metern und ist für den Austausch mit Fremden reserviert. Die öffentliche Sphäre (engl. *public sphere*) schließlich wird vom akustischen Horizont begrenzt. Diese Distanzsphären, von Hall „*Proxemics*“ betitelt, können von Kultur zu Kultur unterschiedlich interpretiert werden, und sollten deshalb nicht als feste Werte betrachtet werden. Die Existenz solcher Sphären erschliesst sich in deren „Verletzung“, die als Akt von Aggression oder Zurückweisung aufgefasst werden kann. Und so versagt auch die aurale Architektur, wenn ein Konflikt zwischen den sozialen Sphären in einem Raum und den gegebenen akustischen Arenen auftritt. (Blessner, Salter, 2007, S.34f.)

### 3.5 Kritik

Wiewohl ich die Prägung des Begriffs der auralen Architektur als längst ausständigen Fortschritt erachte, der in der konventionellen Architektur bei ausreichender Propagierung zu einem Bewusstseinswandel bezüglich der gestalterischen Bedeutung von Raumakustik führen könnte, erscheint mir die Subkategorisierung des auditiven räumlichen Bewusstseins durch Blessner und Salter willkürlich. Insbesondere die musikalische und ästhetische Erfahrung auf eine kategorische Ebene zu stellen, wie soziale und navigatorische Effekte, erscheint mir didaktisch unsauber. Die Navigation im Sinne Blessners und Salters definiert die räumliche Orientierung des Individuums im Raum und kann als biologisch-neurologische Grundfunktion des menschlichen Organismus betrachtet werden. Das ästhetische als persönliche Dimension ist eine soziokulturelle Kategorie. Das Soziale wiederum ist Konsequenz der ersten beiden Kategorien. Als wissenschaftliche Kategorie definiert es das Verhalten von Menschen untereinander. Diese drei Komponenten decken also von biologischen Grundfunktionen bis zur gesellschaftlichen Organisation alles ab. Musik aber ist eine komplexe Kulturtechnik, die auf diesen und vielen anderen Parametern aufbaut. Sie zu einem Grundparameter des räumlichen Bewusstseins zu machen trägt essentialistische Züge in sich, die die Musik zu einem Grundbaustein der menschlichen Psyche machen. Nun ist es möglich, dass Musik tatsächlich ein ähnliches Grundwerkzeug für die Menschheit darstellt wie etwa Sprache. Der Umstand, dass, soweit dies bekannt ist, alle Kulturen der Erde Musik auch unabhängig voneinander entwickelt haben, legt dies nahe. Die Entstehung von Musik in unterschiedlichen Kulturen könnte aber genauso Konsequenz völlig verschiedener Ursachen sein. Auch Funktion und Bedeutung von Musik variieren unter einzelnen Kulturen stark. Nicht nur das, ist sogar die Einschätzung dessen, was Musik von sonstigen Schallereignissen unterscheidet subjektiv. (Vgl. Kapitel 2.2) Da also Musik so schwer fassbar und definierbar ist, sollte sie meines Erachtens nach aus der Kategorisierung der auralen Architektur ausgenommen werden.

Stattdessen plädiere ich dafür, diese nach rein psychoakustischen Gesichtspunkten zu überarbeiten. Nachhall, Resonanzen und Maskierungseffekte sind messbar, und somit auch vermittelbar, und zwar sowohl in der architektonischen Praxis durch Architekten oder Akustiker an die Auftraggeber, als auch in der Lehre. Die Kategorisierung nach Musikalität aber fügt dieser theoretisch wichtigen neuen Disziplin einen universalistischen Beigeschmack hinzu, der dezidierten Räumen der Musik – und somit indirekt auch spezifischen Formen und Stilen von Musik – eine Sonderstellung einräumt. Ein weiterer Problempunkt in der gegenwärtigen Konzeption ist, dass sie voraussetzt, dass die Architekten gleichzeitig perfekt ausgebildete Zuhörer sind. Dies zu erwarten ist viel verlangt. Aus diesen Gründen bin ich der Meinung, dass die gegenwärtige Kategorisierung der auralen Architektur so keinen Eingang in die tatsächliche architektonische Praxis finden wird. Lässt sich das Tätigkeitsfeld eines auralen Architekten aber eingrenzen, könnte dies positive Effekte auf die Architektur haben, wenn etwa eine Profession entsteht, die in die Planungsphase eines Bauwerks eng eingebunden ist.<sup>6</sup> Jedenfalls finden sich in der gegenwärtigen Literatur zwar zahlreiche Erwähnungen von „Spaces Speak“, die Auseinandersetzung mit ihrem Modell der auralen Architektur bleibt aber zumeist aus, was ich auf die konzeptionellen Schwächen der Ausformulierung des Begriffs zurückführe. Möglicherweise ist es vonnöten, dass sich erst jemand aus architektonischer Perspektive dieser Thematik annähert, damit der auralen Architektur mehr Beachtung geschenkt wird.

### 3.6 Schall Formt Raum

In Kapitel 1 habe ich die physischen Aspekte von Schall beschrieben und festgestellt, dass Schall als Verformung unserer Umwelt über die Zeit begriffen werden kann. Er wird durch unsere materielle Umgebung erzeugt und geleitet, aber auch von dieser absorbiert. Er selbst verformt somit wiederum unsere Umwelt. Am Eindringlichsten kann man dies an unserem Gehör erkennen: Es ist ein mechanischer Schallwandler. Das Trommelfell erzittert, wenn Druckwellen darauf treffen. Schall zu erzeugen ist also ein physischer Akt – ein Akt des Berührens. Schall zu empfangen wiederum einer des Berührt-werdens. Die Geschichte, die uns erzählt wird und das Lied, dem wir lauschen, berühren uns – im wörtlichen, nicht nur im metaphorischen Sinne. Der Knall der Pistole wiederum verletzt uns. Laute abrupte Geräusche können zum Riss des Trommelfells führen, laute akustische Dauerbelastung zum Absterben der Haarzellen im Innenohr. Schall kann sogar zum Tod führen. Unter diesem Gesichtspunkt möchte ich die Metapher von der „akustische Beleuchtung“ von Architektur, die Blesser und Salter vorgeschlagen haben, gegen das „akustische Ertasten“ tauschen, denn der Hörsinn ist weitaus verwandter mit dem Tast- als dem Gesichtssinn.

Schall formt also die materielle Umgebung, genauso wie er von dieser geformt wird. Als der deut-

---

6 Leidenschaftliche Sänger würden bestimmt ein vorteilhaft „gestimmtes“ Badezimmer begrüßen..

sche Physiker Ernst Florens Friedrich Chladni 1808 in Paris vor Publikum Sand auf Metallplatten streute und begann, selbige mit einem Geigenbogen zu streichen, eröffnete sich erstmals der körperliche und raumformende Charakter von Schall vor den Augen der Welt. (Schwartz, 2011, S.211) Der Sand hatte begonnen, sich nach den Resonanzfrequenzen der Platten in räumliche Muster zu gruppieren. Chladnis Entdeckung war so faszinierend für die Augen der Menschen des frühen 19. Jhdts., dass dieser den Rest seines Lebens von Demonstrationen der Muster vor Publikum leben konnte. Mit der Nutzbarmachung der Elektrizität fand schliesslich auch die Industrie vielfältige Anwendung für Schall. Neben der Tonaufzeichnung und -übertragung, deren Nutzen für Unterhaltung, Kommunikation und Dokumentation erschlossen wurde, konnten erstmals gezielt Frequenzen jenseits des menschlichen Hörbereichs erzeugt werden, die zahlreiche Effekte auf Materie haben. So wird etwa Margarine mittels Ultraschall emulgiert. (Hielscher.com, 2015) Bis heute findet Ultraschall vielfältige Anwendung in der Lebensmitteltechnologie. Aber auch Reinigung kann durch Ultraschall erzielt werden. Medizinische Geräte etwa werden mit Ultraschall gereinigt, genauso wie Kontaktlinsen, Zahnspangen, etc. Bereits in den 1930er Jahren kamen als Vorläufer der Trommelwaschmaschinen Ultraschallgeräte auf den Markt. Der umgangssprachlich „Waschbär“ genannte „Schallwäscher“, ein Gerät, das die Firma Bosch in Stuttgart auf den Markt brachte, versprach, die Wäsche mittels Ultraschall zu reinigen noch bevor konventionelle Waschmaschinen am Markt waren. Der Apparat wurde in einen mit Schmutzwäsche gefüllten Waschzuber gehängt, und löste aufgrund der starken Vibrationen Schmutz aus der Wäsche. In der DDR wurde es bald wieder vom Markt genommen, da es ein heimliches zweites Anwendungsgebiet fand: Es wurde verwendet, um Schwangerschaften abzubringen. Auf den Bauch einer Schwangeren angebracht sollen die Vibrationen Fehlgeburten eingeleitet haben. (De.muvs.org, 2015)

Nun könnte man dem entgegenhalten, dass solche Phänomene zwar rein wissenschaftlich gesehen räumlicher Natur sind, aber den architektonischen Raum eigentlich kaum tangieren. Beispielsweise sind Wände mittels Schall nur sehr schwer zu bewegen. Hier möchte ich die verheerenden Wirkungen einer Bombenexplosion in Erinnerung rufen. Was etwa Nagasaki im August 1945 am grossflächigsten zerstört hat, war weder die Radioaktivität, noch die Flammenentwicklung der Explosion, sondern die Druckwelle, die bis zu einem Radius von 16.000 Fuss (etwa 4,8 Kilometer) von der Bomben-Abwurfstelle Häuser zum Einsturz gebracht hatte. (Abomb1.org, 2003; Avalon Project, 2008)<sup>7</sup> Wenn man die Verwüstungen des Tsunami von 2005 und des Erdbebens 2015 in Nepal als Resultate von Druckwellen begreift, wird deutlich, welche Dimension die „raumformende“ Wirkung von Schall annehmen kann. Freilich können auch kleine Schallereignisse konkrete Wirkung auf den behausten Raum haben. Laute Erschütterungen in einem darüberliegenden Stockwerk

---

<sup>7</sup> Wiewohl angemerkt werden muss, dass „nachhaltigeren“ Schäden wahrscheinlich durch die radioaktive Verstrahlung verursacht wurden.

können den Putz in einer Wohnung zum bröckeln bringen. Derartige gewalttätige Folgen haben ihre Ursache zumeist in Infraschall, also Schall, der sich unter der Hörschwelle bewegt und sehr langwellig ist. In der Natur tritt dieser Neben Erdbeben und Flutwellen ausserdem bei Gewitter auf. Aber auch Ultraschall, also Schall, der jenseits der maximalen menschlichen Hörschwelle von 20.000 Hertz schwingt, kann desaströse Auswirkungen auf Organismen haben.

In Anbetracht der mächtigen Effekte, die Druckwellen haben können, begannen schon im zweiten Weltkrieg sowohl die USA als auch Nazi-Deutschland, die potentielle Nutzbarkeit von Schall für militärische Zwecke zu erforschen. Zwar erzielte man hier einige Erfolge, aber Anwendung fanden diese in der Kriegsführung nicht. Ihren Höhepunkt dürften die Bestrebungen, Schall als Waffe zu nutzen in den 1960er Jahren gefunden haben. So zitiert etwa Murray Schafer in seinem Band „The New Soundscape“ einen Artikel von Max Gunther aus dem Playboy namens „The Sonics Boom“. (Schafer, 1969, S.19-23) Gunther berichtet, wie der Wissenschaftler Isadore Rudnick am Pennsylvania State College mit einer speziellen Sirene experimentierte, die Frequenzen von 34 kHz bei 175 dB erzeugen konnte. Mit seinem Team konnte er damit nicht nur Murmeln und Münzen zum schweben bringen. Auch Glas und elektronische Geräte wurden von dem hochkonzentrierten Strahl förmlich zerlegt. Nicht nur das, stellten sie auch fest, dass hoch dosierter Ultraschall Materie erhitzen kann. So gelang es etwa, die Pfeife eines Kollegen anzuzünden. Im Tierversuch schliesslich demonstriert die Schallkanone sein ganzes grausiges Potential. *„The biologist has brought a white rat into the room in a small cage. The rat is running around the cage, looking unhappy about all the noise. But his worries don't last long. The biologist lifts the cage into the sound field. The rat stiffens, rises up to the full stretch of his legs, arches his back, opens his mouth wide and falls over. He is dead. An autopsy will reveal that he had died of instant overheating and a massive case of the bends. There are bubbles in his veins and internal organs.“* (Gunther, 1967, S.113)

In Frankreich wiederum wurde durch Vladimir Gavreau an Infraschallwaffen geforscht. Gavreau soll in den 60er Jahren eine Infraschall-Sirene gebaut haben, die dem Gebäude der Forschungseinrichtung Risse zugefügt hatte. Aber auch auf die Menschen wirkt Infraschall. Verschiedene Frequenzen können unterschiedliche Hohlräume des Körpers in Schwingung versetzen, was *„Seerkrankheit, Erschrecken, Übel und Panik“*, aber auch eine Trübung der Sicht auslösen kann. *„Bei einem der französischen Forscher liess eine der seltsamen Pfeifen die Nüstern so stark erbeben, daß der Mann plötzlich seinen Geruchssinn wiedergewann, den er mehrere Jahre zuvor verloren hatte.“* (Dietrich, 1968) Eine 10-minütige Beschallung mit Infraschall bei 170dB schliesslich führt zum Platzen der Lungenbläschen, und in der Folge zum Tod. Aber auch in niedrigeren Dosen kann Infraschall zu körperlichen und physischen Beschwerden führen. *„Atem- und Kopfschmerzen, Abnahme des Leistungs- und Konzentrationsvermögens, allgemeine Stressreaktionen, Ohrenklingen und -rau-*

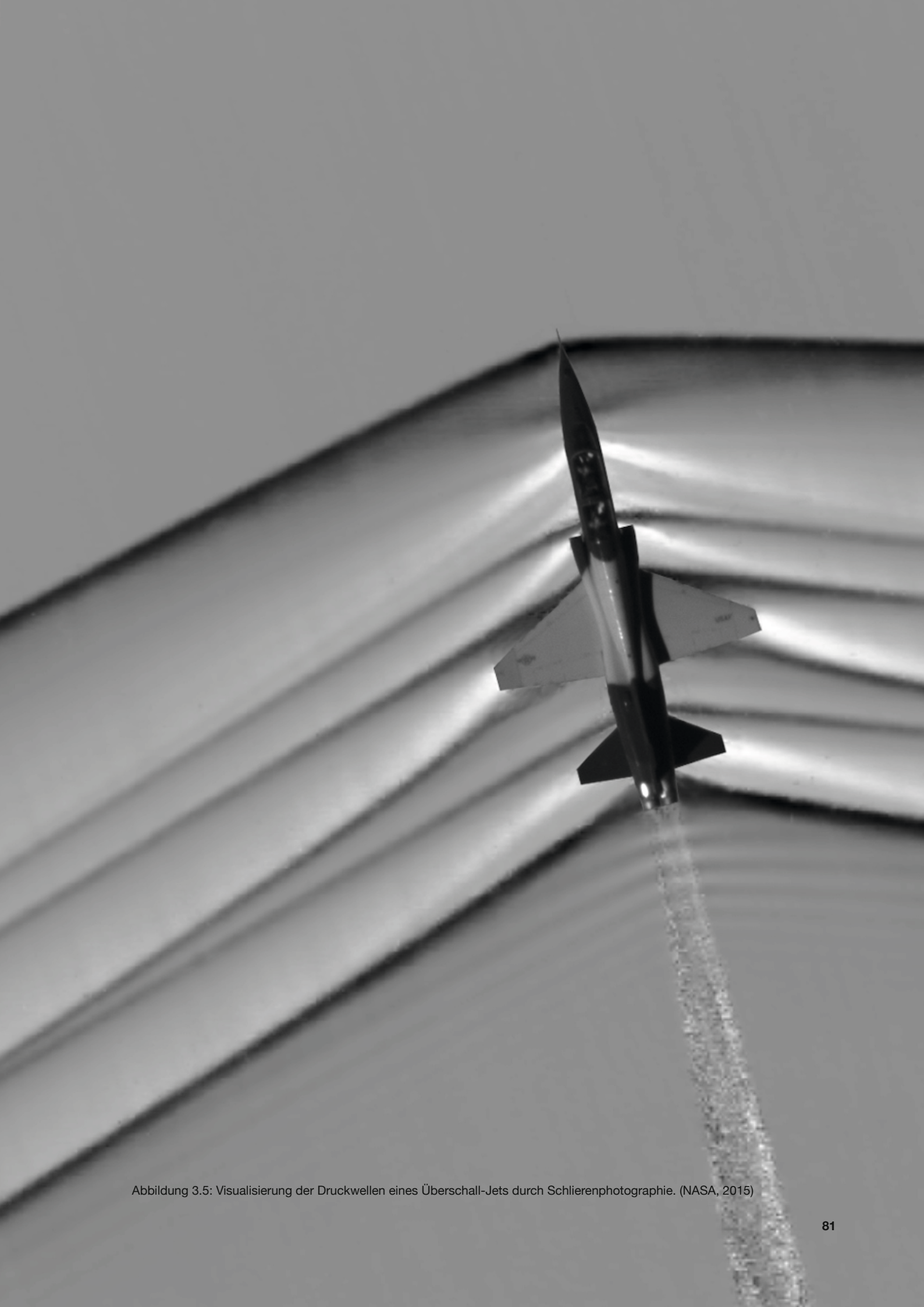


Abbildung 3.5: Visualisierung der Druckwellen eines Überschall-Jets durch Schlierenphotographie. (NASA, 2015)

schen, Benommenheit“ können die Folge sein. (Pade, 1993) Aus diesem Grunde gibt es Bestrebungen, Infraschall-Waffen als nicht tödliches Mittel einzusetzen, um etwa Aufruhre zu zerschlagen. Bis heute haben sie aber nicht die notwendige technologische Reife erreicht, um eingesetzt zu werden, was zum Teil mit dem enormen Energiebedarf und der damit einhergehenden Grösse der Geräte zu tun hat, aber auch mit der Schwierigkeit, den Schall zu lenken und gezielt einzusetzen. Ultraschall-Waffen allerdings können bereits jetzt gebaut werden.

Schall ist also sowohl unter und über der Hörschwelle, als auch im hörbaren Bereich eine Kraft, die grossen Schaden anrichten kann. Ich habe diesen Aspekt weit ausgeführt, um vor Augen zu führen, wie sprichwörtlich die Körperlichkeit von Schall zu verstehen ist. Es ist diese Körperlichkeit, die man sich vor Augen halten sollte, wenn man Murray Schafers Metapher vom Hören als „*touching from a distance*“ liest. Schall formt den Raum. Dies ist nicht nur metaphorisch, sondern auch wörtlich zu verstehen. Schall zu erzeugen bedeutet zu berühren. Hören bedeutet berührt zu werden. Und hörend können wir auch die Umwelt erspüren und den Raum ertasten.







## Aurale Multiplizitäten

### 4.1 Virtualität

Mit der fortschreitenden Entwicklung digitaler Technologie ist in den späten 1980er Jahren etwas möglich geworden, von dem die Menschheit lange geträumt hatte: künstliche Welten zu erschaffen, die betreten werden können. Der virtuelle Raum konnte erstmals sensorisch erschlossen werden. Mit dieser Technologie und im Speziellen durch den Terminus „Virtual Reality“ begann sich jedoch ein Bedeutungswandel zu vollziehen. Virtualität wurde nunmehr einseitig mit dem Computer als Mittel einer neuen Wirklichkeitskonstruktion verbunden. Im alltäglichen Sprachgebrauch haben sich Wirklichkeit und Virtualität zu einem Gegensatz-Paar entwickelt. Allerdings stellt dieses Verständnis eine Begriffsverengung dar. Der Duden definiert „Virtualität“ als „*innewohnende Kraft oder Möglichkeit*“ (Vgl. Duden, 2015), und „virtuell“ als „entsprechend seiner Anlage als Möglichkeit vorhanden“ und „*die Möglichkeit zu etwas in sich begreifend*“ bzw. „*nicht echt, nicht in Wirklichkeit vorhanden, aber echt erscheinend*“ (Vgl. Duden, 2015b). Virtualität meint also die Eigenschaft einer Sache, nicht in der Form zu existieren, in der sie in unserer Vorstellung oder Wahrnehmung nach zu existieren scheint, aber in ihrem Wesen oder ihrer Wirkung nach doch vorliegt. Diese Bedeutung erschließt sich auch über den französischen Begriff „*virtuel*“ (*fähig zu wirken, möglich*), der aus dem lateinischen „*Virtus*“ (*Kraft, Tüchtigkeit, Mannhaftigkeit*) hervorgegangen ist. (Kluge, 2002, S.961)

Clara Völker lokalisiert in ihrer Ideengeschichte der Virtualität den Ursprung des Virtualitätsbegriffs bereits in der Philosophie der griechischen Antike. So begann als erster Parmenides damit, das Wesen des „Seins“ anhand des logischen Ausschlusses des nicht-Seienden zu ergründen und andererseits das „Wahre“ durch die „Elimination der nicht-möglichen Möglichkeiten“ zu finden. (Völker, 2010, S.44f.) Platon schliesslich legte in seiner Ideenlehre eine Zwiespaltene Realität vor, die zum einen sinnlich erfahren werden, zum anderen durch Verstand begriffen werden kann. So muss die sinnliche Wahrnehmung durch den Verstand überwunden werden, um an den Kern des Wirklichen zu gelangen, der in den der äusseren Realität innewohnenden Ideen zu finden ist.

Aristoteles baut als Schüler von Platon auf dessen Elementen auf, kehrte die Konzeption jedoch um und sollte so bis in das Mittelalter die dominante philosophische Ontologie beeinflussen. In seiner „Metaphysik“ begreift er die Wirklichkeit als Dualität aus Form (griechisch εἶδος) und Stoff (griechisch ὕλη, im Gegensatz zu Platons ἰδέα). (Aristoteles, 1995, S.180-198) Form und Stoff bedingen einander gegenseitig. Das Wesen einer Sache ist in dessen Form enthalten, aber damit etwas Form werden kann, bedarf es des Stoffes, der allerdings nach Aristoteles ontologisch minderwertiger ist. (Völker, 2010, S50) Die Wirkungsmechanismen des Seins sind ebenfalls als Dualität aufgebaut. Sie bestehen aus dem Vermögen (griechisch δύναμις) und der Wirklichkeit, bzw. „wirklichen Tätigkeit“ (griechisch ἐνέργεια). Die Dynamis kann sowohl als „Möglichkeit“ als auch als „Vermögen“ verstanden werden, wobei Aristoteles erstere Bedeutung als unerheblich für die Frage nach dem Sein erachtet, da diese bloss Aussagen über das Sein trifft, dieses aber nicht bestimmt. Laut Aristoteles muss das Mögliche notwendigerweise eintreffen, weswegen er es in der „Metaphysik“ nicht weiter behandelt. Das „Vermögen“ aber kann sich verwirklichen, muss dies aber nicht. So verstanden kann man diese Dualität nach moderner Lesart auch als Trinität interpretieren. (Vermögend – Wirklich – Möglich). Das Vermögen hat seine Entsprechung im Stoff, das Wirkliche wiederum in der Form. *„Die Wirklichkeit eliminiert die Optionen des Vermögens, Verschiedenes werden zu können, indem sie Eines geworden ist. Bedingt durch das Vermögen als Prinzip der Bewegung in einem anderen bewegt sie sich als wirkliche Tätigkeit auf eine vollendete Wirklichkeit hin.“* (Völker, 2010, S.67) Das Vermögen stellt somit eine erste Inkarnation dessen dar, was später als Virtualität bezeichnet werden sollte.

Wie Clara Völker anmerkt, ist zu beobachten, dass die aristotelische Trennung von „wirklichem Sein“ und „sinnlich wahrnehmbarer Erscheinung“, die auch die christliche Philosophie durch Thomas von Aquin (Akt und Potenz) prägte und vorübergehen durch ein naturwissenschaftlich determiniertes Weltbild in den Hintergrund gedrängt wurde, angesichts der Durchdringung der Gesellschaft mit digitalen Medien wieder in leicht abgewandelter Form die Oberhoheit gewinnt. (Völker, 2010, S68) Obwohl nicht in Auseinandersetzung mit neuen Technologien gedacht, enthält etwa die Ontologie des französischen Philosophen Henri Bergson in Grundzügen eine ähnliche Konzeption von Virtualität, wie sie bei vielen zeitgenössischen Autoren nach wie vor Aktualität besitzt. Bergson begreift die Wirklichkeit als etwas durch ein subjektives Bewusstsein zu Empfindendes, das nicht durch den Verstand erfasst werden kann, sondern in erster Linie durch eine wahrgenommene Dauer gegliedert wird. Kern der Wirklichkeit ist nach Bergson die Bewegung, die nicht als räumlicher Prozess der Ortsveränderung gedacht wird, sondern als Prozess der ständigen Neudifferenzierung und Neuordnung der Wirklichkeit. Diese kann nur durch Intuition erfahren werden. Bergson grenzt das Virtuelle gegenüber dem Möglichen durch eine zeitliche Achse ab, wobei das Mögliche zeitlich nach dem Wirklichen entsteht, während das Virtuelle bereits zeitlich vor dem Wirklichen existiert. (Völker, 2010, S.210f.) Das Virtuelle ist geistiger Natur, und gleichsam ontologischer Art,

während das Mögliche Materieller Natur ist, und logischer Art. Unser Körper wiederum ist Schnittpunkt zwischen dem Virtuellen und Materiellen, und die Wahrnehmung ist der Menschen virtuelle Tätigkeit, die nicht die reale Wirklichkeit erfassen kann, sondern nur deren virtuelle Wirkung auf uns Menschen. *„Nicht jede Virtualität wird verwirklicht, es wird also angenommen, dass es mehrere auseinanderfaltende und in diverse Richtungen strebende Virtualitäten gibt, die jedoch, da sie sich nicht aktualisiert haben, nicht erkennbar sind. Die Dauer ist insofern eine virtuelle und kontinuierliche Vielheit.“* (Völker, 2010, S.213)

Als geistige Kategorie verstanden ist der Begriff der Virtualität eng verbunden mit dem der Fiktion. Fiktionen sind Darstellungen ohne notwendigen Wirklichkeitsbezug. Nach Aristoteles ist es nicht Aufgabe der Fiktion, mitzuteilen, was geschehen ist, sondern vielmehr nachahmend darzustellen, was nach den Regeln der Wahrscheinlichkeit oder Notwendigkeit möglich wäre. (Aristoteles, 2005, S.28f) Die Fiktion kann in diesem Sinne also nicht von der Realität losgelöst betrachtet werden, da sie sich meist nachahmend auf diese bezieht, was die Fiktion in die Nähe des Begriffs des Virtuellen rückt. Auch die Mimesis, ein wesentliches Merkmal der Fiktion, ist nicht die bloße Nachahmung der Wirklichkeit, sondern eine Darstellung oder Simulation einer möglichen Realität. Diese Realität muss aber nicht den selben Gesetzmässigkeiten wie der realen Welt gehorchen. So ist es auch nicht verwunderlich, dass der Begriff „virtuelle Realität“ von einem Dramatiker, dem französischen Surrealisten Antonin Artaud, geprägt wurde. Artaud forderte in seinem 1938 veröffentlichten Text „Das Theater und sein Double“ eine Transformation des klassischen Theaterbegriffs in eine Kunst, in deren Mittelpunkt kein „Meisterwerk“ stehen sollte, sondern das Publikum, in dessen Denken die Stücke direkt eingreifen sollten. Dieser Theaterbegriff ist vergleichbar mit den Inszenierungen der Happenings oder Aktionskunst. (Günzel, 1998, S.14ff.) Wie Erik Davis anmerkt betrachtete Artaud das Theater viel eher als alchemistisches Werkzeug, anstatt als Mittel, um die Realität abzubilden. Die virtuelle Realität Artauds ist somit eine rituell heaufbeschworene Manifestation einer Vision in einem kollektiven Zustand der Trance. (Davis, 2004, S.226f.)

Der Begriff der Vielheit – oder auch Multiplizität – spielt auch im Denken des französischen Philosophen Gilles Deleuze eine grosse Rolle. Deleuze übernimmt die Konzeption des Virtualitätsbegriffs von Henri Bergson, stellt diese in seinem Kommentar „Bergsonism“ (1966) aber nicht dem „Realen“ gegenüber, sondern dem „Aktuellen“, während aber das „Reale“ in Opposition zum „Möglichen“ steht. (Günzel, 1998, S.70) Hermann Doetsch erklärt das Verhältnis von Realem zum Möglichen als Konkurrenz zwischen zwei Wirklichkeitsmodellen, von denen das eine realisiert ist, während das andere noch auf die Realisierung wartet, während das Verhältnis zwischen Aktuellem und Virtuellem den „dynamischen Prozess der Genese von Form in einem Kräftefeld verschiedener Tendenzen“ beschreibt. (Doetsch, 2006, S.208) *„Dem Realisierungsprozeß der Dialektik, der*

*nur statische Vielheiten hervorbringt, steht der Aktualisierungsprozeß des Differenzdenkens gegenüber. Virtuelles, d.h. zu anderer Zeit Unbedeutendes oder Unwirksames, aktualisiert sich in seiner Zeit als 'originäre Produktion' zur Erzeugung von dynamischer Vielheit - also von Heterogenität(en). Differenz und Kreation sind deren Funktionen, Differenz als Zeit aus sich und Kreation als Aktualisierung. Realisierung erzeugt Kopien einer fertigen Welt. Aktualisierung schafft Karten einer werdenden Welt.*“ (Günzel, 1998, S.71f.) Manuel DeLanda beschreibt in seinem Kommentar „Intensive science and virtual philosophy“ das Deleuzsche Konzept von Multiplizität als „*concrete universals*“, die aus einer bestimmten Gruppe von Attraktoren bestehen, die bestimmte Tendenzen zur Zustandsänderung in einem physischen System beschreiben und durch Bifurkationen verbunden sind, welche abrupte Übergänge in den Tendenzen des physischen Systems definieren. (DeLanda, 2002, S.22). Multiplizitäten bestimmen aber nicht die endgültige Form eines Systems, sondern definieren Prozesse, die in unterschiedlichen Situationen zu gänzlich anderen physischen Ergebnissen führen können. Als komplementäres Konzept führt DeLanda Mannigfaltigkeiten (engl. manifolds) ein, die eine Art mehrdimensionalen Möglichkeitsraum repräsentieren.

Die materielle Realität kann, als Mannigfaltigkeit begriffen, als Zustandsraum (engl. *state space*) verstanden werden, der Reales mit Virtuellem, Möglichem und Aktuellem in einem System vereint, in dem der momentane Zustand etwa eines Objekts durch einen Punkt repräsentiert wird, der innerhalb dieser Dimensionen beweglich bleibt, und so die mögliche andere Zustände abbildet, die zwar möglich, aber nicht aktualisiert sind. DeLanda versucht diese Konzeption anhand des physischen Beispiels eines Fahrrades zu erklären. „(...) [*A*]n object's instantaneous state, no matter how complex, becomes a single point, a great simplification, but the space in which the object's state is embedded becomes more complex (e.g. the three-dimensional space of the bicycle becomes a ten-dimensional state space).“ (DeLanda, 2002, S.14) Diese zehndimensionale Mannigfaltigkeit des Fahrrades erklärt sich also durch die Grade der Bewegungsfreiheit bestimmter Elemente: Lenker, Vorderrad, Hinterrad, Kette und beide Pedale haben jeweils zwei Richtungen, in die sie sich verändern können, von denen jede eine Dimension in dieser Mannigfaltigkeit darstellt. Freilich ist dies eine Vereinfachung, denn natürlich könnte auch der Fahrradrahmen brechen, und so ließe sich die Anzahl der Dimensionen also in die Unendlichkeit erhöhen. Die Konzepte von Multiplizität und Mannigfaltigkeit stellen also hochgradig komplexe Versuche dar, philosophische Überlegungen zur Wirklichkeit in einem materialistischen Modell abzubilden, das mit Entwicklungen der Mathematik und Naturwissenschaften im Einklang steht. Hermann Doetsch schreibt hierzu in „Raumtheorie“: „*Diese Theorie hat weit reichende Konsequenzen: Das, was wir mit dem Wahrnehmungsapparat unseres Körpers als Raum erfahren, stellt also lediglich die Aktualisierung von bestimmten Tendenzen einer Multiplizität dar, die jederzeit durch die Aktualisierung von anderen Tendenzen mit Hilfe anderer Apparate komplexer gestaltet werden kann. Raum wird somit zur Variablen verschiedener Aktualisierungsprozesse, abhängig*

*von Verfahren und nicht zuletzt von technischen Medien. Die Theorie des virtuellen Raums verhindert somit eigentümliche Kurzschlüsse, welche die Körpererfahrung von der medial induzierten Erfahrung trennen und welche den körperlich erfahrbaren Raum als realen Raum in ontologischer Differenz etwa zu mittels optischer Medien erzeugten Räumen — das, was man dann Illusionen nennt — begreifen. Jeglicher Raum bildet die Aktualisierung einer virtuellen Räumlichkeit. Jeder aktualisierte Raum enthält virtuelle Tendenzen einer Entgrenzung, die beispielsweise in der Telekommunikation erfahrbar wird.“* (Doetsch, 2006, S.208f.)

So betrachtet stellt jede Form des künstlerischen Ausdrucks, aber auch jede Technologie, immer schon eine Realisierung des Virtuellen dar, durch die der gegebene Realraum entgrenzt wird. Das beginnt mit der Höhlenmalerei und der ersten Verwendung des Faustkeils und führt über Literatur und Plastik bis in die von elektronischer Technologie geprägten Medien der Gegenwart. Als Multiplizitäten und Mannigfaltigkeiten verstanden gelingt es, das Verständnis von Virtualität als rein geistige Kategorie mit dem Virtualitäts-Begriff der „Virtual Reality“ zu vereinen. Mit der Entkopplung des Klangs von seinem Ursprung etwa durch Methoden der Aufzeichnung und in der Folge mit der elektrischen Verstärkung kann akustisch ein Zustand geschaffen werden, der die Quelle des Klangs wahrnehmbar werden, aber doch nicht für unser Verständnis in dem Sinne „real“ werden lässt. Das Radio beispielsweise erzeugt eine sonderbare Meta-Realität, in der eine Stimme klar vernommen werden kann und nach sensorischem Massstäben eine Präsenz im Raum einnimmt, die dennoch nicht dauerhaft, ja nicht einmal wahrscheinlich ist.<sup>1</sup> Und doch vermittelt uns der Klang der Stimme, der, als physisches „Abbild“ einer Person die von ihr erzeugten Druckwellen an unser Ohr und unsere Haut trägt, deren Präsenz. Eine Mensch wird also in unseren Raum hereingelassen, ohne hier zu sein. Mehr noch, mit der Person wird auch ein Stück weit der sie umgebende Raum in den unseren hereingelassen, und unser gewohnter Raum „verformt“. Dies ist allerdings ein einseitiges Phänomen, denn weder können wir in diesen Raum hineingreifen, noch können wir uns selbst in diesem widerhallen hören.

Was beim Radio scheitert, funktioniert allerdings beim Telephon. Die Leitung ist bidirektional. Neben meinem Gesprächspartner lasse ich auch dessen Umraum in meine aurale Umgebung herein, während ich am anderen Ende der Leitung in einen anderen Raum dringe. *„Die abwesende Wirklichkeit wird partiell und in modifizierter, simulativer Form anwesend, und umgekehrt wird meine anwesende Wirklichkeit abwesend. Daraus resultiert die Simulationstheorie der Medien.“* (Weibel, 1989, S.104) Während Peter Weibel im Telephonat in gewisser Weise die Gesprächspartner ihre Räume wechseln, also eine Reise antreten sieht, stellt Michel Serres eine andere Frage: Wo findet

---

<sup>1</sup> Murray Schafer bezeichnete diesen „entkörpernten“ Klang als „Schizophon“ (engl. Schizophonic)(Schafer, 1977, SX). Die negative Konnotation der Silbe Schizo-, die an einen Krankheitszustand gemahnt, ist bewusst gewählt. Für Schafer stellt die Verbreitung elektrisch verstärkter Klänge und Klangaufzeichnungen eine potentielle Belastung des natürlichen „Ökosystems“ eines Soundscape dar.

das Gespräch statt? Findet es hier statt, oder dort? Oder ist es ein anderer, dritter, Raum, der hier entsteht? Einer, der weder hier, noch dort ist – ein Raum dazwischen also? *„Stellen wir die Frage nach dem Ort [...] , so zeigt sich, dass wir dort mit neuen Mitteln unsere alten virtuellen Habitate erschliessen, die wir einst mit der Technologie der Schrift erschlossen. Und dort zeichnen wir Wege auf paradoxe Karten, die unsere Teilhabe am Lokalen ins Universelle verlängern.“* (Serres, 1994, S.176f.)  
Denn nach Serres ist dieser andere, dritte Raum Teil von uns selbst. *„Seit Homo Sapiens Sapiens vor Millionen Jahren Afrika verließ, irrt er auf der Erde umher und lebt in seinem Kopf [...] Das ist der Mensch, der sich so deutlich von den Tieren und Pflanzen unterscheidet [...] Im Gegensatz dazu sind wir alle Wesen, die da draußen sind.“* (Serres, 1994, S.179)

## 4.2 Virtual Reality

Die Popularisierung des Begriffs „Virtual Reality“ – oder kurz VR – als Synonym für eine im Computer generierte Simulation einer künstlichen Realität wiederum geht laut Davis auf Jaron Lanier zurück. Mitte der 80er Jahre brachte der ehemalige Atari-Mitarbeiter mit seiner Firma VPL Research eine der ersten kommerziell erhältlichen Datenbrillen namens „EyePhone“ (!) sowie den Datenhandschuh „DataGlove“ auf den Markt.<sup>2</sup> Michael Heim beschreibt in seinem am Höhepunkt des VR-Booms entstandenen Buches „The Metaphysics of Virtual Reality“ den Begriff folgendermassen: *„[...]Virtual reality means sensory immersion in a virtual environment. Such systems, known primarily by their head-mounted displays (HMD) and gloves [...]. The HMD cuts off visual and audio sensations from the surrounding world and replaces them with computer-generated sensations. The body moves through artificial space using feedback gloves, foot treadmills, bicycle grips, or joysticks.“* (Heim, 1993, S.113) Im Kern der Erfahrung einer „virtuellen Realität“ steht also die Immersion. *„Immersion creates the sense of being present in a virtual world, a sense that goes beyond physical input and output.“* (Heim, 1993, S.154f.)

Das Gefühl, in einer künstlichen Umgebung physisch präsent zu sein, erfordert, den Sinnen des Benutzers möglichst umfassend Eindrücke von dieser zu vermitteln. Umso bemerkenswerter ist es, dass die Entwicklung von VR in erster Linie im Bereich visueller Technologie vorangetrieben worden ist. So hat am nachhaltigsten die allgemeine Vorstellung dessen, was Virtual Reality zu sein hat, der amerikanische Autor William Gibson geprägt. Mit seinem berühmtesten Roman, „Neuromancer“ schuf er nicht nur einen neuen Begriff, der zum Modewort der 1980er und 90er Jahre

---

<sup>2</sup> Interessant ist hierbei, dass sowohl bei Davis, als auch in vielen anderen Quellen, etwa dem englischsprachigen Wikipedia-Eintrag über VPL Research, sowohl EyePhone als auch DataGlove Erwähnung finden, aber erst die Firmen-Webseite berichtet von einem ebenfalls entwickelten 3D-Audio-System, genannt „AudioSphere“. (vrs.org.uk, 2015) (Wikipedia, 2015) VPL ging 1990 in Insolvenz und wurde 1999 von Sun Microsystems aufgekauft. Die Autorenschaft der erwähnten Web-Seite, die wie ein Internet-Denkmal einer der beteiligten Personen wirkt, ist ungeklärt. Der erwähnte Wikipedia-Eintrag findet sich unter [https://en.wikipedia.org/wiki/VPL\\_Research](https://en.wikipedia.org/wiki/VPL_Research), abgerufen am 12. 10. 2015.



Aug. 28, 1962

M. L. HEILIG

3,050,870

SENSORAMA SIMULATOR

Filed Jan. 10, 1961

8 Sheets-Sheet 3

Fig. 5.

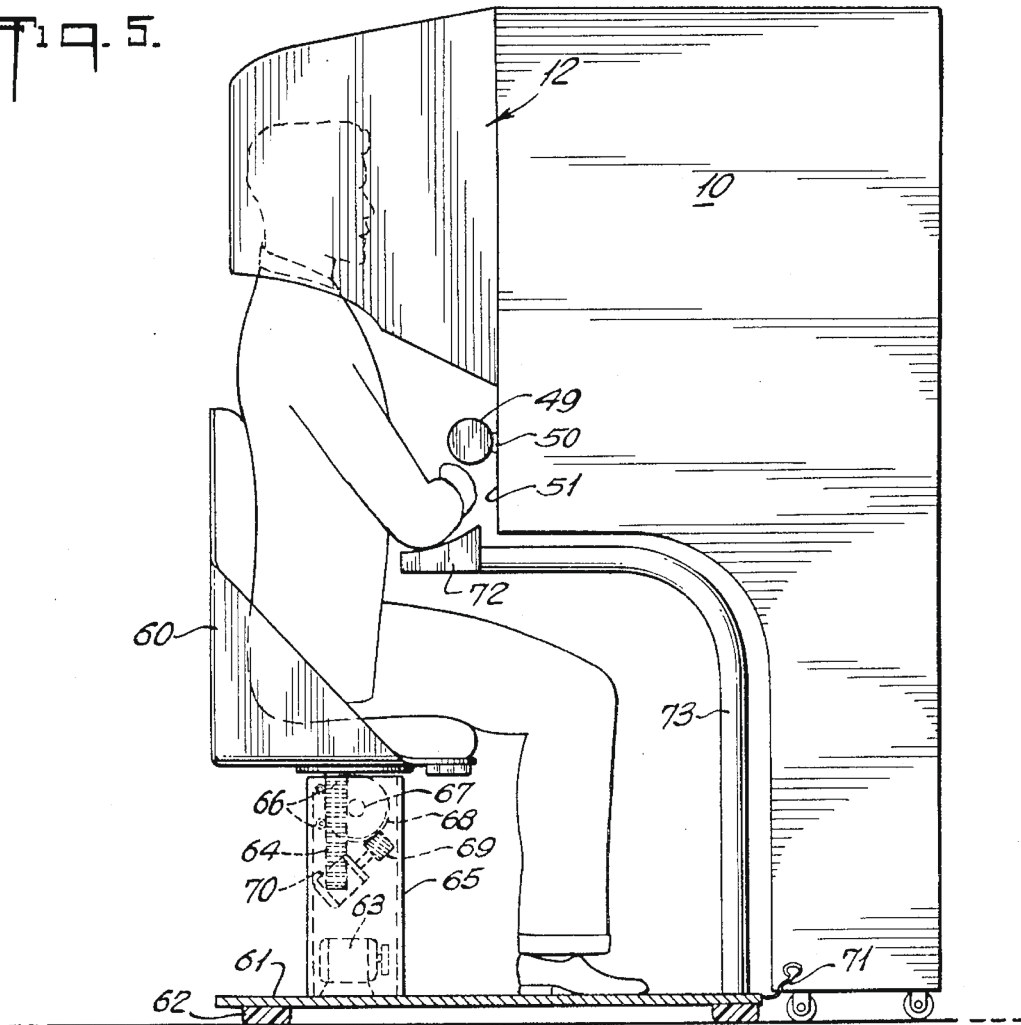
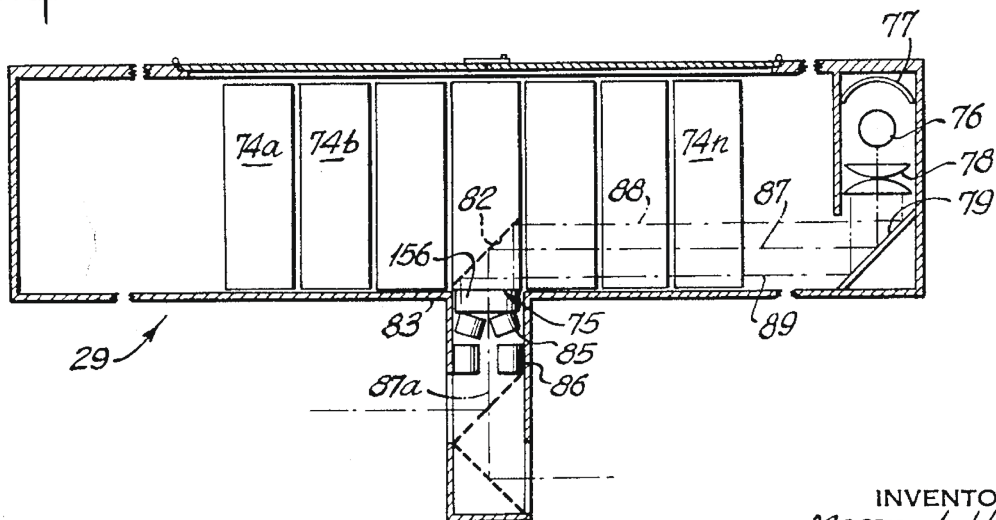


Fig. 6.



INVENTOR  
MORTON L. HEILIG

BY

*Douglas M. Clarkson*  
ATTORNEY

Abbildung 4.1: Auszug aus der Patentschrift zu Heiligs „Sensorama“ (Heilig, 1961, S.3)

werden sollte – der „Cyberspace“ –, sondern gibt diesem auch ein visuelles Interface: *„Cyberspace. A consensual hallucination. [...] A graphic [sic!] representation of data abstracted from the banks of every computer in the human system. [...] Lines of light ranged in the non space of the mind, clusters and constellations of data. Like city lights, receding...“* (Gibson, 1984, S.51) Was hier in Gibsons Beschreibung anklingt, ist symptomatisch für die gesamte Entwicklungsgeschichte von VR-Technologie: Die relative Indifferenz dem Auralen gegenüber, oder mehr noch, das Verständnis des virtuellen Raumes als primär visuellem.

Das Konzept der Virtual Reality ist aber weitaus älter. Der visionäre Autor, brillante Phantast und Weltraum-Satiriker Stanislaw Lem beschreibt etwa in seinem theoretischen Werk *„Summa Technologiae“* aus dem Jahre 1964 auf äußerst akkurate Weise das Prinzip einer virtuellen Realität, die er „Phantomatik“ tauft. (Lem, 1964, S.319-392) Lems Ansatz resultiert weniger aus gängigen Motiven fiktionaler Unterhaltung, sondern aus naturwissenschaftlichen Überlegungen der Neurologie und Kybernetik. Obwohl er selbst Protagonist einer fiktiven Disziplin ist, ist sein biographischer Hintergrund ein wissenschaftlicher. So überrascht es auch nicht, dass seine Konzeption grundlegend multisensorischer Natur ist. Die Erkenntnis, dass zur Simulation einer fiktiven Realität zu allererst die Sinne, bzw. die Nervenleitung zu den reizverarbeitenden Hirnbereichen mit korrespondierenden Eindrücken gespeist werden müssen, führt ihn zu der Untersuchung, wie viele und welche Sinne eigentlich bedient werden müssen, damit das Individuum eine immersive Erfahrung erlebt, die nicht mehr von der Realität unterschieden werden kann. Folgerichtig unterscheidet er zwischen peripherer und zentraler Phantomatik. Erstere stimuliert externe Sinne (Hören, Sehen, Schmecken, etc.), während letztere die internen Sinne (Gleichgewicht, Lage des Körpers, Zustand der Organe, etc.) betrifft, die laut Lem etwa mit chemischen Hilfsmitteln stimuliert werden könnten. Dieser Multisensorische Fokus unterscheidet Lems Konzeption grundlegend von vielen anderen Autoren, welche etwa primär die visuelle Erfahrung betonen.

Die erste kommerzielle Anwendung dessen, was man im Rückblick als geistigen Vorläufer einer virtuellen Realität bezeichnen kann, war höchstwahrscheinlich Morton Heiligs *„Sensorama“*, das dieser ab dem Jahre 1954 entwickelte. Untergebracht in einer Kabine ähnlich einem Photo-Automaten, versuchte es eine multisensorische Erfahrung zu ermöglichen. (Heilig, 1961) *„The Sensorama user sat in a wooden booth, peered into a large viewer, and gripped a set of handlebars. Sensorama took its “rider” on a motorcycle trip through Brooklyn, a California helicopter ride, a drive in a convertible with a young blonde companion, and to a personal bellydancer performance. Besides creating a visual journey, Sensorama provided appropriate smells, stereoscopic sounds, a lurching seat, vibrating handlebars, and “wind” blowing at its rider.“* (Cryptologic Quarterly, 1993) Während das Sensorama aber noch auf aufgezeichnetes Medienmaterial zurückgreifen musste, wurde bald schon daran geforscht,

die neuen Möglichkeiten der Computertechnologie zur Simulation einer künstlichen Realität heranzuziehen. Das „Ultimate Display“ von Ivan E. Sutherland nutzte bereits einen stereoskopischen Datenhelm, um dreidimensionale Drahtgitter-Figuren in den Raum zu „projizieren“. Da die Projektion transparent über ein Linsensystem in das Blickfeld des Benutzers geblendet wurde, man also die Gitterfiguren im Realraum „schwebend“ wahrnahm, kann man das „Ultimate Display“ also auch als Vorläufer der „Augmented Reality“ bezeichnen. (Sutherland, 1965) (Sterling, 2009)

Einen grossen Beitrag zu den Fortschritten im VR-Bereich steuerte auch das kalifornische Kollektiv und Forschungsinstitut EVL (Electronic Visualization Laboratory) bei. Auf ihr Konto geht etwa die Entwicklung des ersten Datenhandschuhs – des Sayre Glove – im Jahre 1976 (Sturman, 1994, S.32)<sup>3</sup>, und in späteren Jahren des CAVE-Systems (Cruz-Neira et. al., 1992), das beispielsweise jahrelang im Ars Electronica Center in Linz zum Einsatz kam. (Evl.uic.edu, 1996) Gegen Ende der 1980er und Anfang der 1990er Jahre war Virtual Reality als Begriff in aller Munde. Zahlreiche Hersteller brachten Interfaces auf den Markt, die die Erschliessung dieses verheissungsvollen künstlichen Raumes für die Massen vorantreiben sollten. Die Firma „Virtuality“ setzte auf VR-Systeme, die auf handelsüblichen „Amiga 3000“- Rechnern basierten und zur Anwendung in Spielhallen konzipiert wurden. Die Bewegungsfreiheit der Benutzer im Realraum wurde hier durch einen Sitzgurt und eine ringförmige Halterung begrenzt, wodurch wiederum Bewegungsfreiheit im virtuellen Raum erlangt werden sollte, ohne dass es zu Kollisionen kam. (South, 2015) Mattel brachte den PowerGlove für das Nintendo Entertainment System heraus (Sturman, 1994, S.33). Sega plante die Veröffentlichung eines HMD für das „MegaDrive“, Atari ebenso für die aufkommende Konsole „Jaguar“. Auch der japanische Spielkonsolenhersteller Nintendo versuchte mithilfe der Firma „Reflections“, die eine Technik zur LED-Retina-Projektion entwickelt hatte, auf den VR-Zug aufzuspringen und brachte den „Virtual Boy“ auf den Markt. (Edwards, 2015) Publikum, Medien wie Industrie waren rund um das Jahr 1990 geradezu besessen von den Prospekten der neuen virtuellen Welten.

### 4.3 Unsichtbare Architekturen

Darstellungen des Cyberspace oder der Virtuellen Realität waren jahrelang exklusive Domäne der Science Fiction. Im Film erreichten insbesondere grosse Hollywood-Blockbuster aufgrund ihrer Prominenz und ihres Budgets ein breites Publikum und formten somit nachhaltig das kollektive Bewusstsein. Angefangen von Tron (1982), über Lawnmower Man (1992) bis hin zu Matrix (1999)

---

<sup>3</sup> Interessant ist, dass frühe Datenhandschuhe Ultraschall zur Lokalisierung von Position und Orientierung der Hand und ihrer Finger benutzten.(Sturman, 1994, S32)

waren und sind diese Darstellungen überwiegend dystopischer Natur.<sup>4</sup> Aber neben der Verarbeitung in fiktionalen Medien wie Literatur und Film (und selbstverständlich, auf selbstreferentielle Weise Computerspielen), begann auch die Architektur sich mit der Virtuellen Realität auseinanderzusetzen. CAD (engl. *computer aided design*) war zwar als Hilfsmittel, um Gebäude zu planen, ein schnell Verbreitung findendes Werkzeug, aber CAD befasste sich mit der Gestaltung von Form und Raum unter Berücksichtigung der Beschränkungen, die diesem durch Statik, Gravitation und Gesetzgebung auferlegt wurden. Befreit von der „Last“ der physischen Welt allerdings vermochte es der Computer, Räume zu synthetisieren, die weit entfernt waren von den symbolischen Imitationen einer gegenständlichen Realität durch den kinematographischen Mainstream.

So begann Mitte der 1980er Jahre Marcos Novak, venezuelanischer Architekt griechischer Herkunft, die Möglichkeiten und Konsequenzen der Ausbreitung der digitalen Informationstechnologie auf die Architektur zu erforschen. In einem Interview aus dem Jahre 1996 konstatiert er, dass, nachdem die Architektur mithilfe von dreidimensionalen Darstellungsformen in den Cyberspace „abgewandert“ ist, der nächste Schritt der ist, dass der Realraum vom Cyberspace erobert werden wird. „[...] *[D]er körperliche Raum [wurde] durch technologische und soziale Arrangements abstrakt gemacht, die es uns erlauben, fast jeden Raum für fast jeden Zweck zu nutzen. Durch die Loslösung der Form von der Funktion und der Realität von der Aktualität haben wir Bedeutung in eine Reihe von unabhängigen Dimensionen vektorisiert.*“ (Novak, 1996) Durch das Internet, das sich wie eine unsichtbare Parallelwelt über die materielle Architektur legt, erschliessen sich Möglichkeiten, die erlauben, dass die vom Utilitarismus ihrer „Rolle als gebaute Offenbarung des Denkhorizonts“ beraubte Architektur, deren Entwicklung aus seiner Sicht nicht Schritt halten konnte mit den Fortschritten der Wissenschaft, (wieder) beseelt wird von einer Art platonischem Ideenraum. Architektur selbst definiert Novak als das, was über die blossen Notwendigkeiten hinaus geht. Eine Tür beispielsweise ist nicht nur Passage, sondern Idee. Es reicht nicht, die mechanische Tür zu elektrifizieren, um eine Entsprechung zu finden auf den Raum der Informationen, der sich „zwischen“ unserer materiellen Welt entfaltet hat. Zu Architektur wird eine Tür erst, wenn sie den Akt des Durchschreitens auf irgend eine Weise perfektioniert.

Zentrales Motiv in den Projekten Novaks ist die Übersetzung der künstlichen Architektur in Sicht-, Hör- und Tasteindrücke. Die algorithmisch generierten Formen stehen in Opposition zu den gängigen Nachahmungen des vertrauten kartesischen Raumes unserer Erde durch Computerspiele und Filme. Oben und Unten, Links und Rechts, sind relative Begriffe, die bedeutungslos sind, wenn sie einen Moment später schon in ihr Gegenteil verkehrt oder gar gleichbedeutend sein können. An-

---

4 Eine prominente Ausnahme mag hier das humanistisch motivierte Star-Trek-Universum darstellen, in dem das Holo-Deck ein radikales Gegenstück zu diesen Interpretationen des „Cyberspace“ bot, da sich hier das Virtuelle im Realraum manifestiert.

fänglich konzentriert er sich noch ausschliesslich auf den virtuellen Raum. In seinem Essay „Liquid Architectures in Cyberspace“ beschreibt er eine neue digitale Architektur – eine flüssige Architektur – die befreit von den Beschränkungen der materiellen Welt Formen und Zustände ändern und auf den Benutzer reagieren kann. Für Novak, der eine musikalische Ausbildung in klassischer Gitarre genossen hatte, stellte diese flüssige Architektur eine Art von visueller Musik dar. So erachtet er die Zeit als wesentlichen Transformationsfaktor für die neuen künstlichen Räume. *„If we described liquid architecture as a symphony in space, this description would still fall short of the promise. A symphony, though it varies within its duration, is still a fixed object and can be repeated. At its fullest expression a liquid architecture is more than that. It is a symphony in space, but a symphony that never repeats and continues to develop. If architecture is an extension of our bodies, shelter and actor for the fragile self, a liquid architecture is that self in the act of becoming its own changing shelter. Like us, it has an identity; but this identity is only revealed fully during the course of its lifetime.“* (Novak, 1991, S.284) In diese Zeit fällt etwa seine Kollaboration „Dancing with the Virtual Dervish“ mit Diane Gromala und Yacov Sharir am Banff Center in Alberta, Kanada. (Gromala, Sharir, 1994) (The World Technology Network, 2005)

Ab Mitte der 1990er Jahre ergänzt er aber seine Konzeption einer flüssigen Architektur um die einer unsichtbaren Architektur. Er erkennt, dass die eigentliche architektonische Herausforderung darin liegt, die künstlichen Räume des Computers zurück in die reale Welt zu holen. Diesen Vorgang tauft er als Gegenstück zur Immersion „*Eversion*“. (Novak, 2002a, S.120) In der Fusion von virtuellem mit Realraum entsteht eine sogenannte „Trans-Architektur“ (engl. TransArchitecture). *„transArchitektur, die Architektur jenseits der Architektur, ist eine Architektur aus ehemals unsichtbaren Gerüsten. Sie hat ein zweifaltiges Wesen: im Cyberspace liegt sie als liquide Architektur vor, die über die globalen Informationsnetze übertragen wird; im materiellen Raum existiert sie als unsichtbares elektronisches Double, das über unsere materielle Welt gelegt wird.“* (Novak, 1996) Das Projekt „Invisible Architectures“, das Novak für den Griechenland-Pavillon auf der Biennale 2000 konzipierte, stellt den Versuch dar, diese verschiedenen Dimensionen in der Realität fassbar zu machen. *„Investigating the invisible in a broader sense, the components of the installation were all derived from inherently invisible higher dimensional geometries.“* (Novak, 2002b, S.69) Durch Transformation dieser unsichtbaren komplexen Form in vier Dimensionen – Raum und Zeit – wird die unsichtbare Architektur auf mehreren Ebenen wahrnehmbar. Einmal als flüssige Architektur – durch eine Videoprojektion; weiters als unsichtbare Architektur – durch eine Vertonung der sich wandelnden Form und Repräsentationen dieser Form als Infrarot-Sensornetz, das eine Interaktion der Besucher ermöglicht; schliesslich als reale, greifbare Form – durch vier Momentzustände dieses komplexen Objektes, die durch Rapid Prototyping hergestellt wurden. (Novak, 2002a) Mit dieser Rückführung der flüssigen Architektur in den Realraum nimmt Novak eine Entwicklung vorweg, die sich wenige Jahre nach

der Präsentation dieser Installation vollziehen sollte: die Durchdringung unserer physischen Welt durch die virtuelle. Und so schreibt Novak schon 1996: *„Der Cyberspace wird in seinen vielen Formen nicht den wirklichen Raum ersetzen. Beide werden vielmehr in einem Amalgam verschmelzen, das sich bereits in Begriffen wie erweiterte Wirklichkeit und intelligente Umgebungen herauskristallisiert.“* (Novak, 1996)

#### 4.4 Eversion

Novaks Re-Fokussierung von der liquiden Architektur zur unsichtbaren und Trans-Architektur ist symptomatisch für einen Bruch, der sich Mitte der 1990er Jahre vollzog, denn die ursprüngliche Begeisterung für Virtual Reality, die ab Mitte der 1980er Jahre die Medienwelt erfasste hatte, ließ zu dieser Zeit stark nach. Was könnten die Ursachen hierfür gewesen sein? Ein nicht zu unterschätzender Faktor ist gewiss die krude Diskrepanz zwischen den vorherrschenden Darstellungen eines „Cyberspace“ in Film und Fernsehen und dem tatsächlichen Stand der Technik. Während bei Filmproduktionen auf „vorgerenderte“ Sequenzen zurückgegriffen werden konnte, deren Herstellung eine wochenlange Berechnung auf Supercomputern erfordert haben konnte, mussten die Ein- und Ausgabe-Daten bei der tatsächlichen Nutzung eines VR-System in Echtzeit berechnet werden. Die Verbrauchergeräte dieser Zeit waren nicht annähernd in der Lage, eine ähnliche Qualität der Darstellung zu erreichen, wie sie von den Medien suggeriert wurde. Auch die sensorische Unausgewogenheit der virtuellen Erfahrung war beträchtlich. Während das Eintauchen in den Cyberspace im Film eine komplette sensorische Immersion zur Folge hatte, mussten sich die Benutzer dieser frühen Systeme mit dreidimensionalem Bild und Datenhandschuh zufrieden geben. Je vollständiger die sensorische Erfahrung sein sollte, desto teurer wurde auch die benötigte technische Ausstattung, was dazu führte, dass sich diese zumeist nur in akademischen und militärischen Einrichtungen, sowie in manchen Spielhallen befanden. Die Umständlichkeit in der Bedienung der Gerätschaften schliesslich war beträchtlich: Datenhelme waren zumeist schwer und durch dicke Kabelstränge mit dem Computersystem verbunden. All dies führte zu einem Interessensverlust an Virtual Reality, der auch in einer erlahmenden Forschung resultierte.

Es gibt aber noch einen wesentlich naheliegenderen Grund für das Verschwinden von VR: Just selbiges Medium, durch welches VR eigentlich versinnbildlicht wurde, machte sich selbstständig und eroberte erst Universitäten, dann Büros und schließlich Wohnzimmer: das Internet. Die fallenden Preise der webfähigen Endgeräte und Internettarife und die Miniaturisierung der Halbleitertechnologie hatten zur Folge, dass es weitaus bequemer und praktikabler wurde, den Cyberspace in den Realraum „hereinzulassen“, als mühsam mittels komplizierter „Prothesen“ in diesen „hineinzuklettern“. *„[D]er virtuelle Raum und der Weltraum beginnen, in den Lebensraum einzubrechen, ihn*



*teilweise zu überdecken (overlap), und einander zu überdecken. Dadurch werden wir vom Boden (aus dem Hier und Jetzt) gerissen und gezwungen, vogelfrei zu werden.*” (Flusser, 1991, S.280) Was Vilém Flusser bereits 1991 antizipierte, kulminierte Ende des letzten Jahrzehnts mit der Durchdringung der Gesellschaft mit günstigen Smartphones. „Wearables“ versprechen, diesen Trend fortzusetzen. Der Cyberspace steckt nun in der Hosentasche, erlaubt dass wir ihn jederzeit in unseren Realraum hereinlassen und meldet sich in unterschiedlichen Intervallen sogar von selbst, wenn etwa die Benachrichtigung eines sozialen Netzwerkes uns zur Interaktion auffordert. Was Novak prophezeit hatte wurde Realität, allerdings auf andere Art und Weise, als er sich dies erträumt hatte: der von ihm kritisierte enzyklopädische Raum des Internet, der unsere Aufmerksamkeit auf sich saugt, indem er uns von Link zu Link jagt, hat sich verselbständigt. (Novak, 1996) Anstatt, dass wir ihn zu unseren Gunsten nutzen, werden wir selbst von diesem benutzt. Wir selbst wurden zu Einträgen in dieser Enzyklopädie. Das getriebene Surfen wiederum wurde entkoppelt von der zugrundeliegenden Idee des Wissenserwerbs. Stattdessen werden Informationen extra erzeugt, um unsere Konsumationsmanie aufrechtzuerhalten.

Aber nicht nur das: durch das Einbrechen in den Realraum hat der Cyberspace auch seinen figurativen Charakter abgestreift. Seine Räumlichkeit ist keine, die direkt mit unseren raumerfassenden Sinnen korrespondiert. Es ist nicht die optische Perspektive, der dreidimensionale Klang oder das Gleichgewicht, das seine Räumlichkeit bestimmt. Seine Räumlichkeit ist abstrakter Natur. In einem Interview mit dem italienischen Magazin Neural aus dem Jahre 2001 prophezeite Novak, dass in baldiger Zukunft jeder Mensch eine virtuelle Immobilie sein eigen nennen wird, die er oder sie ausgestalten und pflegen wird, und die zum Treffpunkt mit anderen werden kann. (Ludovico, 2013, S.75) Was er sich erträumte wurde nur für kurze Zeit Realität. In dem frühen Online-Rollenspiel „Ultima Online“ etwa kauften die Spieler Grundstücke, bauten Häuser, und handelten sogar mit Gegenständen auf Auktionsplattformen wie ebay. Seine Blüte erreichte diese Vision rund um das Jahr 2007 mit der Popularitätswelle, die „Second Life“ erfuhr. Aber diese Popularität hielt nicht an. Novak hat die Geschwindigkeit der modernen Kommunikation unterschätzt, und ihre Konsequenz: das menschliche „Multitasking“, das zu einem grundsätzlichen Aktivitätsschema unserer digitalen Gesellschaft geworden ist. Die Form der Interaktionen im virtuellen Raum macht eine vollständige Immersion unpraktikabel und schliesst derartige Raumrepräsentationen bis zu einem gewissen Grad aus. Auch die Video-Spieler von heute sind virtuell nicht sesshaft. Zwar gibt es Aufbauspiele zu Hauf, aber das Erbaute wird mit dem Spiel gewechselt. Was ausserhalb des Spiels steht sind Benutzerkonten (engl. *accounts*). Diese beherbergen Highscores, Freundeslisten, Nachrichten, Einkäufe, etc. Soziale Netzwerke sind die virtuelle Immobilie geworden, von der Novak sprach, wahrscheinlich aber nicht die, von der er träumte.



Und doch erlebt die Virtual Reality seit wenigen Jahren wieder ein Revival. Als das Projekt „Oculus Rift“ am 1. September 2012 bei einer Kickstarter Kampagne mit über 2,4 Millionen Dollar finanziert wurde, entfachte es eine neue Begeisterungswelle für VR. Mehrere grosse Hersteller haben eigene VR-Systeme angekündigt, unter diesen Sony<sup>5</sup>, Samsung<sup>6</sup>, HTC in Kooperation mit Valve<sup>7</sup> und Microsoft<sup>8</sup>. Im Schatten dieser technologisch aufwändigen Projekte, die gleichzeitig teure Hardwareanschaffungen erfordern, werden auch schlaue Lösungen entwickelt wie etwa Google Cardboard, die nur eine Karton-„Brille“ und ein Smartphone voraussetzen, und folglich um wenig Geld Zugang zur virtuellen Realität bieten.<sup>9</sup> In den rund 20 Jahren seit dem Abebben des ersten Virtual Reality Booms haben grosse technologische Fortschritte stattgefunden. Mit der Rechenkraft moderner Computer ist „Raytracing“ in Echtzeit möglich geworden. Einige Probleme der ersten VR Versuche konnten beseitigt werden, andere verbleiben.

## 4.5 Immersion

Ein grosses Problem ist etwa die Latenz, also die Dauer der Verzögerung zwischen Aktion des Benutzers und Verarbeitung durch den Computer. Sie korrespondiert direkt mit der verfügbaren Rechenkapazität und der Übertragungs-Bandbreite der Repräsentationssysteme. Dreht etwa ein Benutzer seinen Kopf im Realraum, muss diese Bewegung in der virtuellen Umgebung auch vollzogen werden. Erfolgt die Aktualisierung der Repräsentation zu langsam, kann dies das Gefühl der Immersion zerstören. In Spielen der späten 80er und frühen 90er Jahren waren Frame-Raten von 10 Bilder pro Sekunde keine Seltenheit. (Cryptologic Quarterly, 1993, S.36) Das bedeutet, dass zwischen einer Bewegung und der Aktualisierung der visuellen Darstellung eine Verzögerung von bis zu 100 Millisekunden liegen konnte. Der Entwickler Michael Abrash vom Spielehersteller „Valve Software“ beschreibt 20 Millisekunden als seines Erachtens nach maximal tolerable Verzögerung für VR-Anwendungen, und räumt ein, dass Forschungen darauf hinweisen, dass die Toleranzgrenze noch niedriger liegen könnte. Eines der Probleme, das sich ergeben kann, wenn die Latenz zu hoch ist, ist, dass die virtuelle Umgebung an „Glaubwürdigkeit“ einbüßt. Hiermit ist kein etwaiger wahrgenommener Realismus der Simulation gemeint, der leiden könnte, sondern interne Logik und Gesetzmässigkeiten der Umgebung, die in so einem Fall nicht mit den vom eigenen Körper registrierten Zuständen übereinstimmen, und dem Gehirn somit zu verstehen geben, dass an der wahrgenommenen Umgebung etwas nicht stimmt. „[...]Virtual objects have to stay in very nearly the same perceived real-world locations as you move; that is, they have to register as being in almost ex-

5 PlayStation VR, 2015, abgerufen am 20. 10. 2015, <https://www.playstation.com/de-at/explore/ps4/features/playstation-vr/>.

6 Samsung GALAXY Note Gear VR, 2015, abgerufen am 20. 10. 2015, <http://www.samsung.com/at/promotions/galaxy-note4/feature/gearvr/>.

7 HTC Vive, 2015, abgerufen am 20. 10. 2015, <http://www.htcvr.com/>.

8 Microsoft HoloLens, 2015, abgerufen am 20. 10. 2015, <https://www.microsoft.com/microsoft-hololens/en-us>.

9 Google Cardboard, 2015, abgerufen am 20. 10. 2015, <https://www.google.com/get/cardboard/>.

*actly the right position all the time. Being right 99 percent of the time is no good, because the occasional mis-registration is precisely the sort of thing your visual system is designed to detect, and will stick out like a sore thumb.*“ (Abrash, 2012)

Ein anderes Problem, das durch hohe Latenzzeiten ebenfalls noch verschlimmert wird, ist durch die Qualität der Simulation nicht gänzlich zu lösen. Viel eher erfordert es die Eröffnung gänzlich neuer sensorischer Aspekte der Simulation. Michael Heim bezeichnet es als „*Alternate World Syndrome*“. (Heim, 1993, S.148f.) Man kann es aber auch als See- oder Reise- oder Simulatorkrankheit bezeichnen. Stanislaw Lem beschreibt es als Konflikt der „äußeren“ mit den „inneren“ Sinnen: Wenn die an den Organismus gemeldete Orientierung des Kopfes nicht identisch mit der durch die Bogengänge im Mittelohr registrierten ist, können Schwindel, Übelkeit und Kopfschmerzen die Folge sein. (Lem, 1998) Die Anfälligkeit für derartige Symptome variiert stark von Mensch zu Mensch. Bei vielen Individuen können sie sich nach einer anfänglichen Gewöhnungszeit legen, allerdings nach der Rückkehr in die Realität – ähnlich der Landkrankheit bei Seeleuten – wieder einsetzen. Bei intensiver Nutzung von VR-Anwendungen kann das Krankheitsbild chronisch werden. Heim spricht in diesem Fall von der „*Alternate World Disease*“. Während die Verstärkung des „*Alternate World Syndrome*“ durch Vermeidung hoher Latenzen vermieden werden kann, ist dem Grund-Problem momentan nicht gänzlich beizukommen. Um es gänzlich zu beseitigen, müsste man neben den äußeren Sinnen auch die inneren Sinne mit künstlichen Reizeindrücken speisen, was nach Lems Begriffskanon gleichbedeutend wäre mit dem Konzept der „zentralen Phantomatik“. In einer solchen würde sich die Simulation durch nichts mehr als solche verraten, vorausgesetzt, dass die künstliche Umgebung neben sinnlichem „Output“ auch ihrer internen Logik entsprechend genügend „Input“ des Benutzers akzeptiert.

Bevor aber an so etwas gedacht werden kann, sollte zuerst die „periphere Phantomatik“ vervollständigt werden. Das heißt also, es sollten alle, oder zumindest die dominanten Sinne angesprochen werden, um die Immersion in der künstlichen Umgebung zu steigern. Wie bereits mehrfach angedeutet, wurde der klangliche Komponente der VR bisher weitaus weniger Aufmerksamkeit gespendet als ihrem visuellen Gegenpart. In einem Essay von 1993 weist Steven Jones hin auf das überwiegend visuelle Verständnis von „Cyberspace“ durch dessen Keimzellen, wie etwa „VPL Research“ oder das „MIT Media Lab“, und deutet an, dass sogar die Sprache des frühen Cyberspace das Aurale ausschließt. Er konstatiert, dass dem Desinteresse an auralen Techniken der virtuellen Realität, die er „virtual audio“ nennt, eine fortschreitende „De-Auralisierung“ (engl. deauralization) zugrunde liegen könnte. *„The goal of virtual reality technology is creation of space [...] This is a feat already accomplished by a variety of audio technologies, [...] Yet virtual reality is considered a primarily visual space. As such it's creation can be understood as a part of the ongoing technological visulisation and*

*deauralization of space.*“ (Jones, 1993, S.246)

Eric Davis wiederum berichtet in einem Artikel auf seinem Blog aus dem Jahre 2000, wie stark die immersive Erfahrung einer virtuellen Umgebung für ihn durch künstlichen dreidimensionalen Klang geworden war, und stellt die Frage, was der Grund hierfür sein mag. Seine Antwort findet er in den emotionalen Aspekten auraler Wahrnehmung: *„Affect is a tremendously important dimension of experience, and one of the most difficult to achieve in a visual environment. “Atmosphere” might be a good way to describe this aspect: sound produces atmosphere, almost in the way that incense [...] can do. Sound and smell carry vectors of mood and affect which change the qualitative organization of space, unfolding a different logic with a space’s range of potentials. Ambient music, or an ambient soundscape, can change the quality of a space in subtle or dramatic ways.*“ (Davis, 2000) Jones dagegen bezieht eine Position, die abgeleitet von den Theorien Walter Ongs ist (Vgl. Kap. 2), wenn er die Ursache in der immersiven Wirkung von *„virtual audio“* in dessen einhüllendem Wesen sieht: *„Virtual audio re-centers the listener by providing a sense that audio is no longer a “surface” projected from a speaker (or stereo speakers). It is once again “all around,” and that all around moves with “me.”“* (Jones, 1993, S248) Zweifellos ist die emotionale Wirkung von Klang auf den Raumeindruck ein nicht zu vernachlässigender Faktor bei der Untersuchung von Immersion. Auch der Umstand, dass Schall uns einhüllt, und wir so im Zentrum unserer klingenden Umgebung zu liegen kommen, ist sicher nicht zu unterschätzen. Ich denke jedoch, dass beide Autoren hier Effekt mit Ursache verwechseln. An dieser Stelle wage ich mit Verweis auf Kapitel 1 und 3 zu behaupten, dass der gewichtigste Grund für die immersiven Effekte einer räumlichen Klangsimulation im Cyberspace in den körperlichen Eigenschaften von Schall liegen könnte. Durch ihn können wir die virtuelle Umgebung erspüren. Durch ihn wird die virtuelle Erfahrung physisch, und somit „real“.

## **4.6 Das Aurale in virtuellen Umgebungen**

Wie wohl sich die Audiotechnik über die Jahre ebenfalls verbessert hat, hinkt diese der Graphik in seiner Entwicklung stark hinterher. Dies mag auf den ersten Blick nicht offensichtlich sein, darum möchte ich diesen Punkt genauer ausführen. Obwohl den Gehversuchen der 1980er Jahre in Virtual Reality kein Erfolg beschieden war, hat sich am Rücken sogenannter „Ego-Shooter“ – initiiert durch das hochgradig erfolgreiche „Doom“ – in der Videospiele-Industrie die dreidimensionale Darstellung in Form einer perspektivischen Projektion weitgehend durchgesetzt. Anstatt wie in der Anfangszeit der Computer & Videospiele auf Bitmap-Graphik zu setzen (d.h. Pixelgraphik, bei der einzelne Rasterpunkte in fixer Farbe vorliegen) basiert die Darstellung der meisten modernen Spiele auf Vektordaten. Bitmaps werden zumeist nur zur Texturierung der dreidimensionalen Modelle herangezogen. Ein Effekt dieser Entwicklung war, dass die Hersteller von Graphikhardware begannen,

sogenannte 3D-Beschleuniger zu entwickeln. Es handelt sich hierbei um spezielle Recheneinheiten, die die Berechnung von Fließkomma-Operationen und Vektor-Transformationen optimieren. Inzwischen sind diese so hoch entwickelt, dass sie sogar Raytracing in Echtzeit bewerkstelligen können. Was bedeutet das also? Nichts weniger, als dass das physikalische Verhalten von Licht im Raum simuliert wird. Die perspektivische Verzerrung einer geometrischen Struktur in der Optik einer „virtuellen“ Kamera wird errechnet. Künstliche Lichtquellen, die im virtuellen Raum positioniert sind, emittieren Strahlen, die von den Oberflächen dieser Geometrie reflektiert, gebrochen und absorbiert werden. Das Resultat ist eine (je nach Einstellung der Render-Engine) realistische Simulation des physikalischen Verhaltens von Licht. Es existieren zahlreiche Softwarepakete, sogenannte 3D-Engines, die diese Simulation je nach Funktionsumfang qualitativ hochwertig bewerkstelligen. Die Entwicklung hin zu dieser Technik war eine graduelle. Schon frühe Vektordarstellungen (wie etwa „Maze War“ von 1974 (Olivetti, 2012)) errechneten mittels Drahtgitter-Modellen Perspektiven. In späteren Spielen kamen Verdeckungen hinzu. Noch später wurden diese um künstliche Lichtquellen und Oberflächensimulationen ergänzt. Parallel steigerten sich Auflösung des resultierenden Bildes und der geometrischen Objekte in den Szenen immer mehr. Das bedeutet also, dass es in der Computergraphik eine grundsätzliche Fokusverschiebung vom „Aufgezeichneten“ (das punktgenaue Bild) zum „Generierten“ (die Repräsentation einer Menge von Daten durch einen Standpunkt) gegeben hat.

In der Audiotechnik wiederum gibt es nichts vergleichbares. Zwar hat spätestens mit der Popularisierung des Mediums DVD ein allgemeiner Trend hin zu Surround-Raumklangsystemen eingesetzt, der auch in Videospiele Niederschlag gefunden hat, aber das zugrundeliegende Prinzip ist ein gänzlich anderes. Bei 3D-Audio, wie es zur Zeit in Verwendung ist, werden Klangquellen (aufgezeichneter oder synthetisierter Art) in einem Koordinatensystem positioniert und deren Lautstärke, und womöglich Frequenzspektrum, abhängig von der Position des Spielers verändert. Im Gegensatz zu den dominanten Graphik-Engines sprechen wir hier aber nicht von einer naturgetreuen Simulation eines physikalischen Phänomens: die üblichen Methoden, Raumklang zu erzeugen, sind Näherungen. Sie erzielen dies nach einer Methodik, die vergleichbar ist mit dem Prinzip der frühen 3D-Engines, wie sie etwa in „Doom“ Verwendung fanden. In diesen wurde nur das Terrain (zumeist ein Keller, bzw. „Dungeon“) perspektivisch berechnet, während die Spielfiguren aus Pixelgraphiken bestanden, die im Raum platziert und in Größe und Helligkeit an die eingenommene Position in Relation zum Spieler angepasst wurden.

Und so ist es auch kein Wunder, dass sich die Audio-Hardware von heute nur marginal von der der späten 1990er Jahre unterscheidet. Mögen Aus- und Eingänge hinzugekommen und Abtastrate und Dynamikumfang gestiegen sein – substantielle technische Fortschritte in der Schallsimulation

sind noch ausständig. Um das Verhalten von Schall im virtuellen Raum adäquat zum Verhalten von Licht berechnen zu können, wäre eine „Wave-Tracing“-Engine vonnöten.<sup>10</sup> Diese würde die Ausbreitung von Druckwellen im Raum, sowie deren Reflexion, Brechung und Absorption berechnen können. Obwohl es immer wieder Bestrebungen gab, Wavetracing umzusetzen, so etwa durch die Firma Aural, die Ende der 1990er Jahre sogar spezielle Software und Hardware für den Zweck auf den Markt gebracht hatte, um physikalisch korrekt simulierten Klang zu ermöglichen (Schneider, 2009; Abel, Foster, 1994; Abel, Foster, 1997), ist mir zum gegenwärtigen Zeitpunkt kein einziges Projekt bekannt, das sich in aktiver Entwicklung befindet, um diese Bestrebungen marktreif zu machen. Nun könnte man fragen, was denn der Vorteil einer solchen realistischen Berechnung von Schall wäre. Mit Verweis auf Kapitel 3 möchte ich hierauf antworten: die Simulation der *passiven* akustischen Aspekte einer virtuellen Umgebung: die **aurale Architektur**. Diese fehlt nämlich in den gegenwärtigen virtuellen Welten gänzlich. Der Zugang zum Raumklang der Videospiele- und VR-Industrie ist eher ein Soundscape-Zugang. Der Versuch, aurale Architektur im Computer zu synthetisieren ist im Moment noch ein Unterfangen, bei dem der jeweilige Autor bei Null anfangen muss. Es existieren kaum Audio-Engines, die das Verhalten von Schall simulieren<sup>11</sup>; es existiert keine Audio-Hardware, die der CPU bei der Berechnung des Schallfeldes assistiert. Bei der physikalischen Simulation von Schall im Raum handelt es sich um ein gänzlich unbestelltes Feld. Die Frage nach dem „Warum?“ muss ich im Raum stehen lassen. Höchstwahrscheinlich mangelt es bei Konsumenten wie Produzenten an Bewusstsein für den Nutzen einer solchen realistischen Berechnung von Schall. Ohne Nachfrage wird sich auch das Angebot nicht entwickeln. Andererseits wird sich aber auch die Nachfrage ohne Bewusstsein schwerlich vergrößern. Möglicherweise sind die gängigen Näherungen schlicht und einfach „good enough“. Jedenfalls kann eine gewisse Gleichgültigkeit gegenüber dem Auralen nicht von der Hand gewiesen werden.

## 4.7 Feedback

Immersion ist ein relativ schwammiger Begriff, der in etwa gleichbedeutend mit „Versinken“ ist. Wie Mark Grimshaw anmerkt, existieren mehrere Arten der Immersion. „*Building upon Pine and Gilmore's (1999) work on experience, Ermi and Mayra (2005) claim that "immersion [...] means becoming physically or virtually a part of the experience itself."* Furthermore, they distinguish between three forms or states of immersion: sensory; challenge based; and imaginative immersion.“ (Grimshaw, 2012, S.358) In einem Computerspiel kann man als Spieler vollständig versinken, und Zeit und Raum rundherum vergessen, selbst wenn der Klang aus einem 1-Bit IBM-PC-Speaker „quäkt“ und das Bild sich aus vier „schroffen“ CGA-Farben zusammensetzt, die in einen Raster von 160

<sup>10</sup> Näherungsweise könnte auch mit Raytracing-Engines gearbeitet werden. (Funkhouser et. al., 1998)

<sup>11</sup> Eine Ausnahme stellen spezifische Anwendungen, dar, die zur Berechnung der Raumakustik im Bauingenieurwesen herangezogen werden, sowie einzelne Hallgeräte, die für die Erzeugung von Nachhall in der Musik verwendet werden.

mal 120 Pixeln gepresst wurden. Genauso kann ein Musiker sich im Spiel seines Instruments verlieren, ein Hörer in der dargebotenen Musik, ein Leser in einem Roman, in welchen wiederum ein Schriftsteller beim Verfassen komplett versunken war. Immersion ist also keine sensorische Exklusivdomäne. Allerdings ist körperliches Raumempfinden eine rein sensorische Domäne. Zwar kann ich mich in Gedankenräumen verlieren, in diesem Fall verlasse ich aber auch in gewisser Weise meinen Körper. Das Ziel von VR-Umgebungen ist aber nicht, die Benutzer dazu zu bringen, ihre Körper zu verlassen, sondern ebendiese Körper in eine künstliche Welt zu transportieren. Körperliches Raumbewusstsein kann uns allerdings von jenem imaginären Raum entfernen, den wir durch andere Formen der Immersion betreten können. Das Jucken auf der Fußsohle inmitten der Lektüre eines spannenden Buches oder der Wadenkrampf im Fußballspiel sind Extrembeispiele dafür, wie der Zustand unseres Körpers uns aus einer Tätigkeit reißen kann. Aus diesem Grunde ist die sensorische Immersion Vorbedingung für ein Funktionieren einer virtuellen Raumerfahrung digitalen Verständnisses. Die gegenwärtigen Praktiken dreidimensionalen Raumklangs in virtuellen Umgebungen sind durchaus Effektiv: Sie erlauben es, glaubwürdig darzustellen, dass sich etwa ein Klangereignis von der Seite oder von hinten nähert. Sie vermitteln Distanzen, und schaffen so einen belebten auralen Relationalraum, der dem Benutzer das Gefühl vermittelt, einen Platz einzunehmen in einem dynamischen Beziehungsnetz aus Figuren und Ereignissen. Warum also die Mühen auf sich nehmen, die passive Umgebung auch noch darzustellen?

Unsere externe sensorische Wahrnehmung hat zwar primär die Aufgabe, die Umwelt zu erfassen, sie nimmt darüber hinaus aber noch etwas anderes wahr: uns selbst in dieser Umwelt. Wenn wir vor einem Spiegel stehen, erwarten wir ein Abbild unserer Selbst zu sehen. Aber auch unsere Hände sind ständig im Blickfeld, wenn wir physische Tätigkeiten verrichten. Und während ich diese Zeilen hier auf einem Keyboard an einem Computer sitzend verfasse, schenke ich meinen Händen zwar kaum Beachtung, denn mein Blick ist auf die schwarzen Buchstaben auf weißem Grund der Textverarbeitung fokussiert, aber dennoch befinden sie sich in meinem Blickfeld. Ist dieses Beispiel noch relativ offensichtlich, so erfordert das nächste einen kurzen Moment der Konzentration (und womöglich eine „Verrenkung“ des Blicks): unsere Nase. Sie ist immer in unserem Blickfeld. Sie liegt zu nahe an unseren Augen, als dass wir sie scharf wahrnehmen können, und höchstwahrscheinlich spenden sehr wenige Menschen ihr Aufmerksamkeit, aber wir können sie sehen. Sie ist externes Zeugnis davon, dass wir uns in unserem Körper befinden. Wenn diese eigenen Zeugnisse der eigenen Präsenz nun fehlen, fehlen sie auch in unserem Sensorium, welches zwar gewohnt ist sie auszublenden, aber diese doch heranzieht, um uns selbst im Raum wiederzufinden. Es kommt nicht von ungefähr, dass die sogenannten „First Person“-Computerspiele zumeist ein Paar Hände am unteren Bildschirmrand, oder verlängerte Extremitäten in Form einer Schrotflinte oder eines Langschwertes darstellen. Nach einiger Spielzeit nehmen wir auch diese nicht mehr wahr. Trotzdem



sind sie wichtige sensorische Hinweise, um uns in der dargestellten Welt zu verorten.

Wir sollten nicht annehmen, dass dieses Phänomen rein optischer Natur ist. Selbstverständlich nehmen wir auch unbewusst unseren eigenen Geruch wahr, und mit unserer Haut erspüren wir nicht nur unsere physische Umgebung, sondern auch unsere Gliedmassen selbst. In Kapitel 3 habe ich beschrieben, wie die Reflexionen eines Raumes uns von ihrer auralen Architektur berichten. Verschieben wir den Blick auf dieses unbestreitbare Faktum etwas, kann dieses auch anders verstanden werden: nicht nur hören wir den Raum – wir hören uns selbst in diesem Raum. Beim Schritt durch das Stiegenhaus des Altbaus, in dessen Souterrain dieser Text entstanden ist, wird mir an der Färbung des Klangs meiner eigenen Fusstritte nicht nur bewusst, dass ich gehe; es wird mir nicht nur Volumen und Struktur des Eingangsbereichs mit seinem hohen Tonnengewölbe bewusst: im Klang meines Schrittes und im Klang des Raumes wird mir vielmehr bewusst, dass *ich* hier bin. Es besteht eine aktive aurale Rückkoppelungsschleife zwischen mir und dem Raum, der mich umgibt. Erst in diesem „Feedback“ erkenne ich mich selbst im Raum. Schall als physisches Medium verbindet mich mit der räumlichen Struktur der Architektur, in der ich mich bewege. Dies ist der Grund, warum es von überaus grosser Bedeutung ist, die aurale Architektur künstlicher Räume darzustellen.

## 4.8 Aurale Multiplizitäten

In Kapitel 1 habe ich beschrieben, mit welchen Mitteln unser Gehör den uns umgebenden Raum rekonstruiert. In Kapitel 3 habe ich beschrieben, wie Raum Schall, und wiederum Schall Raum formt, und zwar sowohl physisch als auch durch unsere Wahrnehmung. Der uns umgebende Raum wird strukturiert durch Bezugspunkte, die durch Klangereignisse in diesen eingebracht werden, sowie durch die aurale Wahrnehmung der passiven Welt. Im Laufe dieses Kapitels habe ich schließlich beschrieben, welche Bedeutung in einer virtuellen Welt neben der Simulation einer aktiven Klangkulisse die Darstellung der passiven auralen Umgebung hat. In gewisser Weise wird die künstliche Umgebung durch deren „Vertonung“ physisch erfahrbar. Wir können also einen künstlichen Raum körperlich spüren. Die allgemeine Tendenz der letzten 20 Jahre war bestimmt durch das, wie Vilém Flusser es formulierte, „Herüberschwappen“ der Virtualität in den Realraum.<sup>12</sup> Marcos Novak schöpfte für diesen Prozess den Begriff „Eversion“. *„Eversion is [...] a motion complementary to the familiar notion of immersion. Whereas immersion describes a vector moving from ordinary to virtual space, eversion describes the counter-vector of the virtual leaking out into the actual. Eversion predicts that the content of augmented reality and ubiquitous computing will be the population of the physical world with phenomena and entities first encountered in virtual space.“* (Novak, 2002a, S.10)

---

<sup>12</sup> Tatsächlich hat das „Herüberschwappen“ des virtuellen in die Realität mit der Verfügbarkeit billiger 3d-Drucker eine neue Dimension erreicht.



Vollziehen wir nun als Gedankenexperiment einen letzten Schritt: holen wir die aurale Architektur eines virtuellen Raumes in den Realraum zurück. Welche Möglichkeiten und Konsequenzen ergeben sich daraus? Schall könnte die Komponente sein, die das Virtuelle erst vollständig in die materielle Umgebung zurückholt. Spannend wird diese Überlegung, wenn man einen Moment zurücktritt und die Aspekte des Bewusstseins und der Kommunikation beiseite lässt: Die Vorstellung, dass die Druckwelle, die durch eine Zustandsveränderung in der Raumkonfiguration an einem Ort der Welt entstanden ist, sich an einem ganz anderen Ort der Welt manifestiert, illustriert, dass hier unser metrischer Raum physisch überwunden wird. Versetzen wir uns in eine hypothetische Betrachtungsposition, in der diese Druckwelle verlustfrei transportiert werden könnte: dann könnte das Erdbeben, das sich 2015 in Nepal ereignete per digitaler Leitung in Wien weitergegeben werden. Dieses Gedankenspiel könnte man als zynisch erachten, aber es dient auf markante Weise zur Illustration, wie die Körperlichkeit von Schall in Verbindung mit modernen Technologien unseren physischen Raum aufbrechen könnte.

Mit digitalen Mitteln wird es plötzlich möglich, den Raum nicht nur zu verkleinern, wie dies durch Maskierung machbar ist, sondern ihn zu verändern oder zu erweitern. Synthetische aurale Architektur in den Realraum zu integrieren würde erfordern, die akustischen Artefakte seiner Bewohner erfassen zu können, um diese, vom virtuellen Raum reflektiert, im realen Raum zum Klingen zu bringen. Wohlgermerkt ist dies ist bei weitem kein triviales Unterfangen, und mit gegenwärtigem Stand der Technik nur sehr schwer zu erzielen. Um dies zu erreichen, müssten die Klangereignisse in einem Raum von dessen auraler Architektur getrennt werden können, um den Direktschall im künstlichen Raum erklingen zu lassen. Weiters müsste mit elektronischen Mitteln der synthetisierte Klang akustische Eigenheiten des bestehenden Realraums überdecken bzw. kaschieren. Ersteres könnte entweder durch Mikrophonierung geschehen, gepaart mit einer intelligenten Filterung die den „Raumabdruck“ entfernt. Alternativ könnten erkannte Bewegungen und Laute neu synthetisiert werden. Zweiteres ist noch komplizierter und erfordert ein hohes Mass an psychoakustischen Massnahmen. Beides ist zur Zeit noch schwer zu erreichen, aber nicht unmöglich. Gelingt es, wird ein Dialog zwischen dem realen und dem künstlichen Raum möglich, und der Klang einer Architektur kann verändert werden. Man könnte nun fragen, was es bringen soll, ein kleines Zimmer gross, oder ein großes klein klingen zu lassen. Dies kann rein ästhetische Gründe haben, kann aber auch nachhaltig die Nutzung verändern. Wie ich beschrieben habe hat der Klang einer Architektur vielfältige Effekte auf uns Menschen. Er kann unser Wohlbefinden, aber genauso auch unsere alltäglichen Handlungen und unser Sozialgefüge beeinflussen. Akustisches Feedback im Raum zeigt uns auch unsere Bewegungsgrenzen auf. Die akustischen Karten im Gehirn sind ja synchronisiert mit den sensomotorischen und optischen. Das bedeutet auch, dass durch unsere akustische Raumwahrnehmung unser Handlungshorizont mitdefiniert wird. Durch akustische Simulation könnten

aber nicht nur künstliche Räume entstehen, sondern auch künstliche Objekte. Leitsysteme würden etwa eine praktische Anwendung für solch eine Technologie bieten.

Auch könnten Raumbereiche künstlich zu privaten oder öffentlichen Zonen gemacht werden. All dies wäre in dynamischer Art und Weise variabel. Durch digitale Interventionen würde der uns umgebende architektonische Raum tatsächlich flüssig, wie Novak dies erträumte. Gänzlich neue Räume würden möglich. Einige davon könnten nützlich sein – andere einfach nur spannend.

# **Teil 2**

# **Sonic**

# **Manifolds**



## Einleitung

„*Sonic Manifolds*“ ist eine explorative Installation in deren Mittelpunkt das Erkunden von Raum steht. Auf Basis der Überlegungen zur Auralen Architektur und zur Virtualität soll aus Klang Raum erschaffen werden, den die Besucher eingeladen sind zu betreten. Der Fokus ruht hierbei auf der Interaktion mit diesem künstlichen akustischen Raum. Dadurch, dass die Bewegung des Benutzers im Realraum eine Antwort in der künstlichen Umgebung erhält, entsteht „aurale Immersion“. Dynamik und Kontrast sind die bestimmenden Elemente dieser auralen Umgebung. Bewegung und Innehalten sind die Methoden, mit denen sie erforscht werden kann. Das Ohr ist das Organ, das uns leitet.

Zur Umsetzung dieser Installation mussten vielfältige Gebiete untersucht werden. Dabei befasste ich mich unter anderem mit Techniken wie Videotracking, Statistik, Klangerzeugung, Raumakustik und der Simulation von virtuellen Räumen. So ist auch die Vermittlerrolle zwischen verschiedenen Disziplinen wie Architektur, Kunst, Tontechnik, Musik, Informatik, Mathematik und Human Interface Design essentieller Charakter des Projektes.

### 1.1 Überblick der Funktionsweise

*Sonic Manifolds* besteht aus einer von mir programmierten Software, die Bildmaterial von einer Infrarot-Kamera verarbeitet, aus diesen Klangdaten errechnet und ein Array von acht Lautsprecher mit diesen beschickt. Diese Lautsprecher befinden sich an den Eckpunkten eines quaderförmigen Raumbereichs von  $3,6 \times 3,6 \times 2,2$  Metern, der von besagter Videokamera erfasst wird. Tritt ein Benutzer in diese Zone ein, werden dessen Position, Körperhaltung sowie Art und Geschwindigkeit von dessen Bewegungen registriert. Anhand dieser Daten werden Schallereignisse in einem virtuellen Raum berechnet, die dann wiederum über die Lautsprecher im Realraum wiedergegeben

werden. Dabei werden die künstlichen Signale „maßstäblich“ auf den Realraum übertragen - eine Bewegung von 50 cm im Realraum hat auch eine Bewegung von 50 cm im simulierten Raum zur Folge. Die simulierten Räume selbst jedoch können beliebige Dimensionen haben und können weiters in beliebiger Reihenfolge und Anordnung im realen Raum positioniert werden. So ist es möglich, dass der Benutzer mit einem Schritt aus der Mitte einer Kathedrale in ein Badezimmer treten kann. Dies wird dadurch möglich, dass die Ausdehnungen der Raumbereiche im Realraum unabhängig sind von der Ausdehnung des simulierten Raumes. Um dies zu erreichen, wird es notwendig, Zonen zu definieren, in denen entweder der eine oder der andere Raum betreten und gehört wird. Weiters können Zwischenstufen zwischen verschiedenen Räumen interpoliert, sowie die physikalischen Parameter der simulierten Räume variiert werden, um geometrisch „surreale“ Szenarien zu erzeugen, die in einem realen Raum aufgrund der Gesetzmäßigkeiten der Physik unmöglich wären.

Der Kern von *Sonic Manifolds* ist die Exploration durch das „Abtasten“ von Raum durch den Prozess des Hörens. In Realräumen geschieht dies durch die Interpretation der Reflexionen eigener akustischer Äußerungen, wie z.B. Fußstritte, Rascheln oder vokale Laute. Da diese in eine Simulation einzubinden höchst komplexe Mikrophonierungs-Techniken erfordern würde, der Fokus der Installation aber auf der Simulation neuer Räume liegt, habe ich mich entschlossen, die Bewegungen der Benutzer durch künstliche Geräusche zu abstrahieren, die durch ihren akustischen Minimalismus als Klangereignisse in den Hintergrund treten, während die transformativen Effekte von Raum auf Schall in den Vordergrund treten sollen. Um dies zu erzielen, wurde als grundlegendes Schallereignis der Impuls gewählt, wie er auch verwendet wird, um den Nachhall von architektonischen Strukturen zu messen.

## 1.2 Software

Das Herz von *Sonic Manifolds* ist eine in der graphischen Programmiersprache „Max“ von mir programmierte Software. Max – benannt nach dem US-amerikanischen Pionier der Computermusik Max Mathews – wurde ursprünglich Mitte der 1980er-Jahre unter dem Namen „Patcher“ von Miller Puckette am IRCAM („Institut de Recherche et Coordination Acoustique/Musique“) als Sprache zur Steuerung elektroakustischer Kompositionen entwickelt. Ursprünglich für die Verarbeitung von MIDI-Daten ausgelegt, erhielt es Mitte der Neunziger mit der Erweiterung „MSP“ die Fähigkeit, DSP (engl. *digital signal processing*) zu bewältigen. 2003 wurde mit Jitter eine Erweiterung vorgestellt, um Videomaterial in Echtzeit zu verarbeiten. Max zeichnet sich dadurch aus, dass mit relativ bescheidenen Programmierkenntnissen sehr professionelle Ergebnisse erzielt werden können, so die logische Lösung zur Erzielung des jeweiligen Ergebnisses bekannt ist. Im

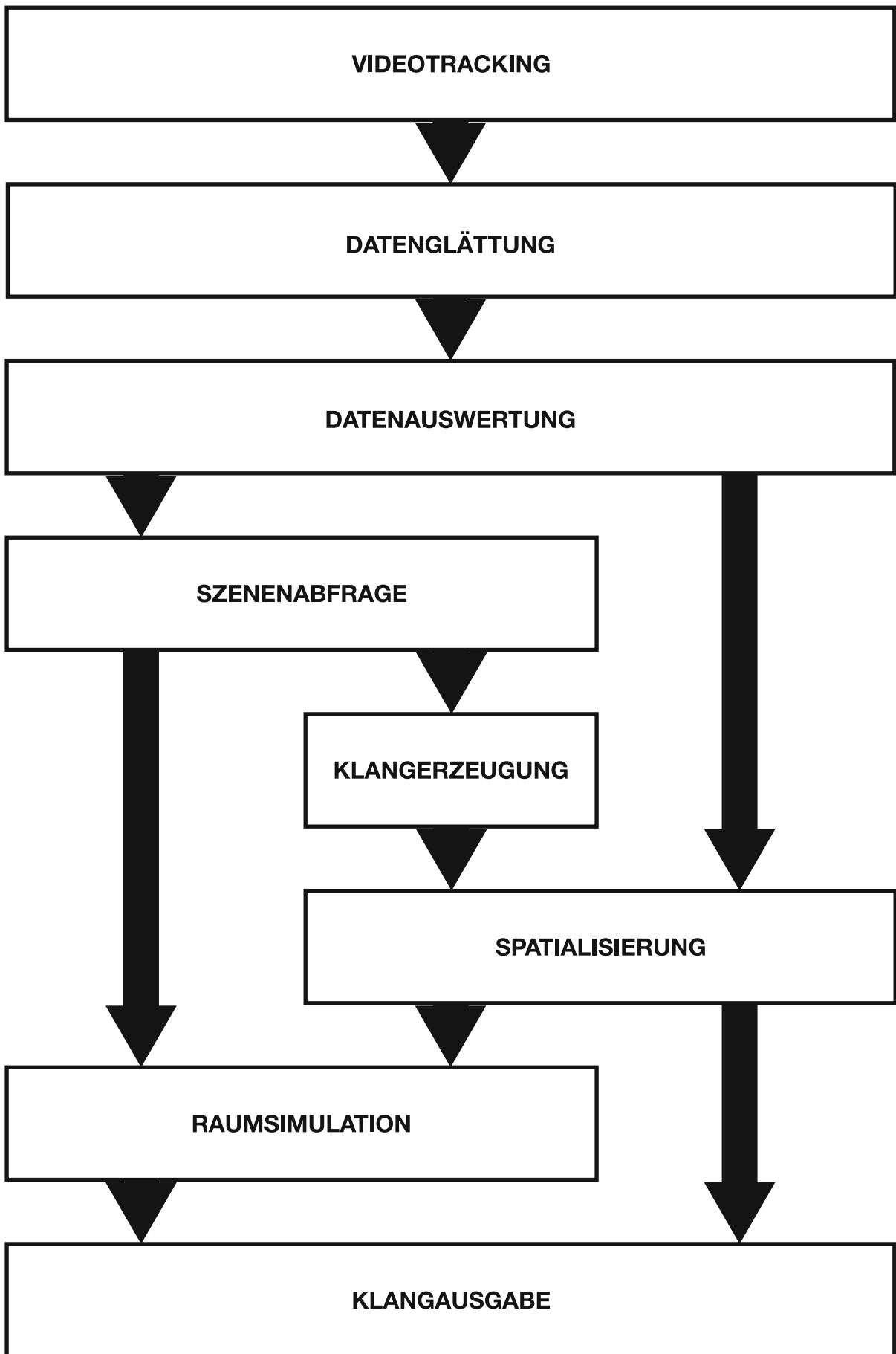


Abbildung 1.1: Struktureller Aufbau der Datenverarbeitung (Hausch, 2015)



Gegensatz zu anderen Programmiersprachen muss sich der Benutzer also kaum mit den Interna des Computers bzw. Betriebssystems beschäftigen. Anstatt also Variablen zu deklarieren oder Speicherbereiche zuzuweisen, kann direkt vom Programmstart weg mit für Medienkünstler und Musiker optimierten Befehlen angefangen werden, Bilder zu bearbeiten, Klang zu synthetisieren oder Interaktionen zu entwerfen. Befehle – sogenannte „Objekte“ – werden graphisch aufbereitet und präsentieren sich dem Benutzer in Form von Bausteinen mit Ein- und Ausgängen. Durch das „Verkabeln“ der Objekte mittels Verbindungslinien wird der Datenfluss festgelegt. Die Objekte selbst sind für relativ spezialisierte Aufgaben ausgelegt wie z.B. Farbfilter (für Bilder & Video), Oszillatoren (zur Klangerzeugung) oder User-Interface-Elemente (für die menschliche Interaktion). Gleichzeitig existieren aber auch eine Reihe von „Low-Level“-Befehlen, um komplexere Objekte selbst zu entwerfen. Der modulare Charakter von Max erlaubt es, die bestehende Bibliothek durch selbst geschriebene, oder von Dritten programmierte Objekte, zu erweitern. Darüber hinaus liegen auch Schnittstellen zu anderen Programmiersprachen wie C, Java, Java-Script, Ruby, php, GLSL (OpenGL), u.a. vor, womit in diesen jeweiligen Sprachen Programmiertes leicht in ein Max-Projekt integriert werden kann. Zu guter Letzt sind Schnittstellen zur Integration von Netzwerk, Sensoren, Motoren, Kameras etc. im Überfluss vorhanden.

## **Struktur Sonic Manifolds**

Funktional lässt sich die Struktur von *Sonic Manifolds* in drei Bereiche einteilen: Videoerkennung (Input), Datenverarbeitung und Klangerzeugung (Output). Die Videoerkennung, genannt Videotracking, liefert Informationen über die Benutzer an die Datenverarbeitung, wo anhand der Positionskoordinaten errechnet wird, was für Klänge berechnet und ausgegeben werden sollen. Im Folgenden will ich jeden dieser Bereiche behandeln und sowohl seine Funktionsweise, wie die dahinter liegende Theorie erläutern, sowie auch auf die geschichtliche Entwicklung der zur Anwendung kommenden Komponenten eingehen.

## Videotracking

Um Audioereignisse zu simulieren, die vom Benutzer ausgelöst und in den bestehenden Raum übertragen werden sollen, muss eine Methode gefunden werden, den Benutzer elektronisch zu erfassen, zu vermessen und seine Handlungen zu interpretieren. Dabei erfasst eine digitale Videokamera ein Bild, das anschließend mittels Bilderkennungsverfahren vom Computer ausgewertet wird. Die Auswertung der erfassten Bilder kann dabei mittels verschiedener Methoden erfolgen. Dabei ergeben sich verschiedene Probleme. Einerseits muss der Inhalt des Bildes erkannt, Personen oder Objekte von der Umgebung getrennt und eindeutig identifiziert werden. Andererseits muss die Identifikation über einen längeren Zeitraum hinweg gewährleistet werden können, um sinnvolle Daten zu liefern. Wird dies nicht gewährleistet, können einzelne Personen oder Objekte die Identifikationsnummer tauschen, was zu „Sprüngen“ der zu verwendenden Koordinatenpunkte im Raum führen kann. Im Falle von *Sonic Manifolds* hätte dies etwa Störgeräusche zur Folge. Dabei können verschiedene Algorithmen zum Einsatz kommen.

Eine früher sehr verbreitete Methode war das „Blob-Tracking“. Hierbei wurden zusammenhängende Bereiche erfasst – sogenannten „Blobs“ – um anschließend auf ihre Größe, Position etc. untersucht zu werden. Um die Blobs isolieren zu können, mussten sie aber zuerst vom Hintergrund getrennt werden. Dies kann entweder dadurch geschehen, dass ein Standbild des Hintergrunds vom aktuellen Kamerabild subtrahiert wird, oder durch das Analysieren von Bewegung durch Subtraktion aufeinanderfolgender Bilder. Ein gängiges Problem bei all diesen Methoden war, dass z.B. erkannte Personen optisch verschmelzen, wenn sie sich berühren, eng beieinander stehen oder sich verdecken. Eine andere gebräuchliche Methode ist das „Konturen-Tracking“, bei dem ebenfalls mittels Kontrast und Farbe zusammenhängende Bereiche gesucht werden, die dann aber geometrisch analysiert werden, um mit einer Datenbank verglichen zu werden in der z.B. Körperposen oder typische Gesichtsproportionen gespeichert sind. Zur kontinuierlichen Identifikation über mehrere Bild-Frames hinweg wiederum können die erkannten Objekte entweder nach Größe, Umriss oder Nähe zu zuvor erfassten Koordinaten sortiert werden. All diese Methoden

leiden darunter, dass nur ein zweidimensionales Bild zur Verfügung steht, was die Trennung von zu erkennenden Elementen von den Sichtbedingungen abhängig macht. Seit der Markteinführung der Microsoft Kinect Kamera im Frühjahr 2010 haben die Möglichkeiten der Datenverarbeitung mittels Videotracking jedoch einen Quantensprung erlebt. Erstmals war eine kostengünstige und standardisierte Methode verfügbar, mittels derer man direkt dreidimensionale Daten der Umgebung erhalten konnte. Mittels passender Treiber konnte außerdem direkt auf sogenannte "Skeleton Data" der Benutzer zugegriffen werden. Obwohl Kinect eigentlich als Interface für die X-Box 360 auf den Markt gebracht worden war, fand sie schnell Verbreitung in der DIY-Community. Eine Vielzahl von Hackern machte sich daran, offene Treiber für die Kamera herzustellen. Die Folge war eine explosionsartige Verbreitung der Kamera unter Nicht-Spielern – im Speziellen unter Kunstschaffenden. Aufgrund der vielfältigen Möglichkeiten zur Einbindung in Max/MSP wählte ich die Kinect als Steuerinterface für *Sonic Manifolds*.

Die Kinect zugrundeliegende Sensortechnik wurde ursprünglich von der israelischen Firma „PrimeSense“ entwickelt und 2010 von Microsoft für die Verwendung in der X-Box lizenziert. Bei dem Gerät handelt es sich allerdings nicht um eine stereoskopische Videokamera, sondern um einen Hybrid aus regulärer Kamera und Infrarot-Sensor. Die VGA Kamera liefert ein Farbbild in den Dimensionen  $640 \times 480$ , während die Tiefen-Information vom integrierten Infrarotsensor geliefert wird. Ein Projektor im Gehäuse erzeugt ein gitterförmiges Infrarot-Signal, das auf die zu erfassende Szene projiziert wird. Dieses Infrarot-Muster bleibt für die VGA Kamera unsichtbar, wird aber vom Infrarotsensor erkannt und aufgezeichnet. Dieser berechnet anhand der Verzerrung des Gitters auf der Szene die Tiefen-Information. Der Treiber kombiniert VGA und Tiefenbild und liefert diese an den Computer weiter.

Das erwähnte „Skeleton Tracking“ wird von der Treibersoftware „OpenNI“ besorgt. Diese kann anhand der Tiefen- und VGA-Bilder Umrisse mit menschlichen Proportionen erfassen. Dabei werden bis zu 20 „Joints“ (Gelenke) ausgewertet und zur Verfügung gestellt. (Vgl. Graphik 2.1) Weiters ist die Software in der Lage, mehrere Personen gleichzeitig zu erfassen, und sogar eindeutig zu identifizieren. Wenn z.B. 2 Personen im Bild sind, erhalten diese die Identifikationsnummern #1 und #2. Verlässt eine der beiden daraufhin das Bild, um von einer Dritten ersetzt zu werden, wird die Dritte die ID #3 erhalten. Kehrt wiederum die 2. Person zurück, wird ihr wieder ihre ursprüngliche ID #2 zugewiesen. Die Skeletterkennung ist erstaunlich zuverlässig und funktioniert auch einigermaßen gut, wenn Teile eines Skeletts verdeckt sind. Die Software interpoliert in diesem Fall die verdeckten „Joints“ (Gelenke). Wenn z.B. eine Person hinter einer anderen vorbeigeht ist die Interpolation so gut, dass die Verdeckung gänzlich unbemerkt bleibt. Dauert die Verdeckung längere Zeit an, können die interpolierten Joints unnatürliche Posen einnehmen.

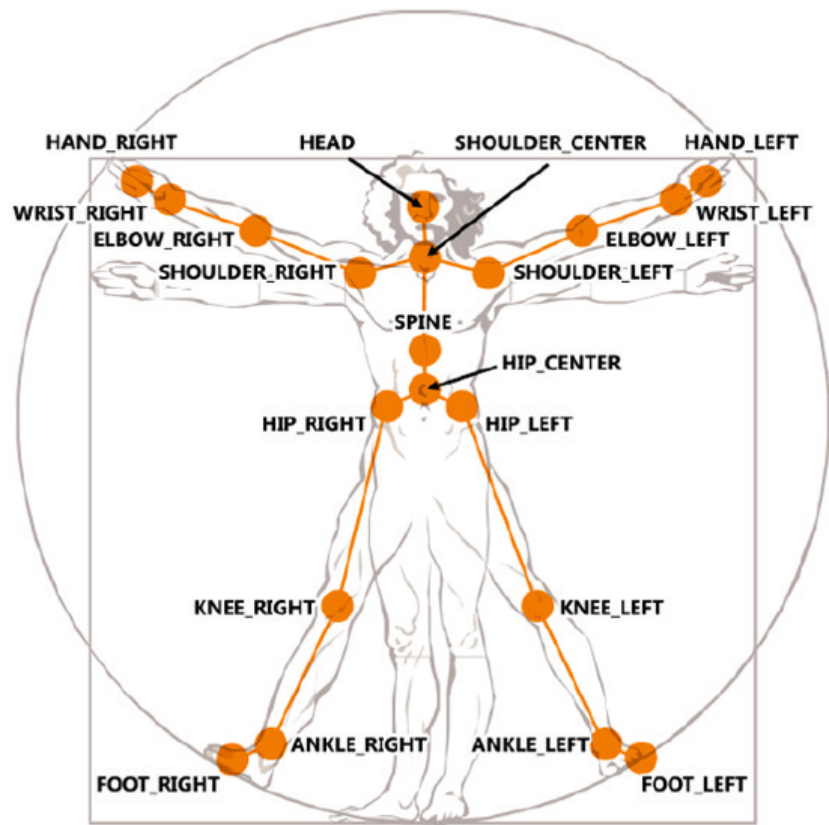


Abbildung 2.1: Vom Skeleton-Tracking der Kinect erfasste "Joints". (Microsoft, 2012)

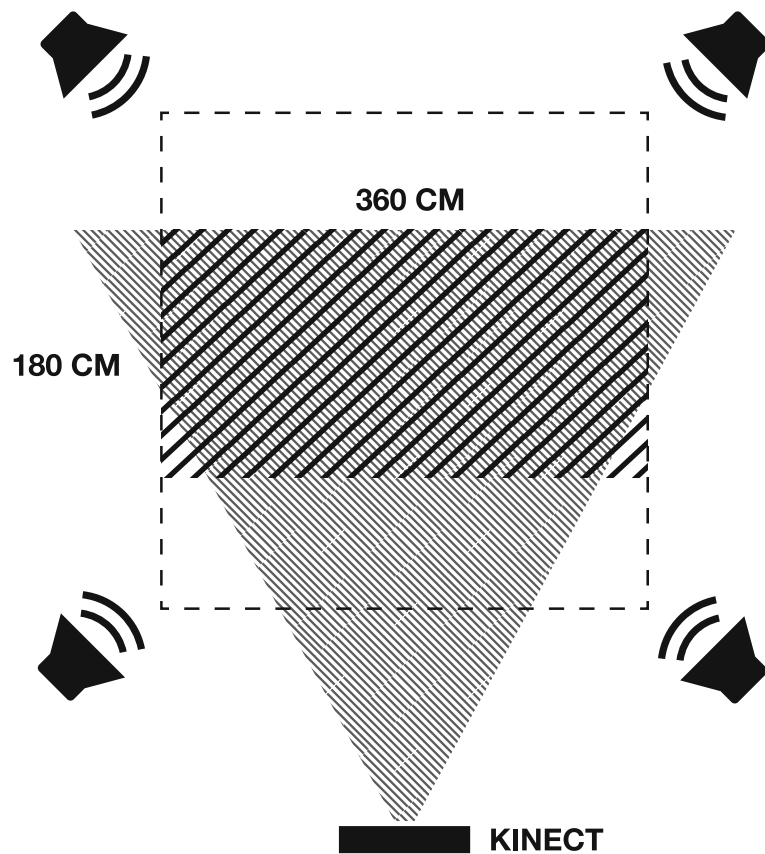


Abbildung 2.2: Versuchsaufbau des Projektes

Der Sichtbereich, den Kinect erfassen kann, liegt innerhalb eines Winkels von ca. 57° Horizontal und 43° vertikal. Die minimale Nähe zu zu trackenden Personen beträgt in etwa 80 cm, während die maximale Entfernung ca. 6 m beträgt. Aufgrund des begrenzten Sichtwinkels ist die Kinect bei *Sonic Manifolds* mit einer Entfernung von 2,6 Meter vor dem zu überwachenden Raumbereich positioniert, um trotzdem einen quaderförmigen Bereich ohne abgeschnittene Ecken erfassen zu können. Allerdings wird die Auflösung der Tiefen-Information mit großer Distanz immer gröber, weshalb ich in *Sonic Manifolds* als maximale Entfernung 4,5 m angenommen habe. Aufgrund der Konfiguration des mir zur Verfügung stehenden Lautsprecher-Setups, der Notwendigkeit, den Tracking-Bereich in der Mitte dieses Setups anzulegen, sowie des Sichtwinkels der Kinect, musste *Sonic Manifolds* für die Demonstration in den mir zu Verfügung stehenden Räumlichkeiten auf eine zu trackende Grundfläche von 1,8 × 3,6 Meter reduziert werden. (Vgl. Grapik 2.2)

## Datenauswertung

Obwohl die „Open NI“-library bereits eine integrierte Datenglättungsmethode mitbringt, ergaben meine Experimente, dass diese für meine Zwecke nicht ausreichend ist. Das Tiefenbild, das die Microsoft Kinect Kamera liefert, neigt dazu stark zu rauschen, was zum Resultat hat, dass ermittelte Koordinaten springen. Mit zunehmender Entfernung von der Kamera werden die Daten auch immer ungenauer. So springen die Koordinaten bei einer Entfernung von 5 Metern von der Kamera bereits in 10cm-Schritten in der Tiefen-Achse. In der Folge beschäftigte ich mich mit verschiedenen Methoden, gleitende Mittelwerte zu berechnen, mit dem Ziel, einerseits Kamerarauschen und Datenausfälle auszufiltern, jedoch gewollte Bewegungen sehr wohl zu registrieren. Im Weiteren entwickelte ich ein Analysewerkzeug, um die Koordinatendaten sowie die verschiedenen Glättungsmethoden graphisch vergleichen und die geeignetste Methode feststellen zu können. Im Folgenden finden Sie die Beschreibung der implementierten Varianten sowie eine Vergleichsgraphik.

### 3.1 Datenglättung

Beim (linearen) gleitenden Mittelwert, im folgenden SMA für „*Simple Moving Average*“ genannt, wird eine Liste der letzten  $n$  Datensätze gespeichert. Die Werte der letzten  $n$  Datensätze werden addiert und durch die Anzahl  $n$  dividiert. Kommt ein neuer Wert hinzu, wird der älteste Wert (Index  $n$ ) aus dem Speicher gelöscht, und die bestehenden Datensätze rücken um jeweils eine Position nach hinten. Der SMA bietet eine sehr gute Glättung, reagiert allerdings äußerst träge auf Auslenkungen und ist im Videotracking aufgrund dieser Latenz kaum zu gebrauchen. Nur bei Speicherlisten von sehr wenigen Elementen ist eine annähernde Reaktionszeit erzielbar, was allerdings auf Kosten der Glättung geht.

Der gewichtete gleitende Mittelwert, im folgenden WMA für „*Weighted Moving Average*“ genannt, funktioniert ähnlich wie der SMA, erzielt aber eine Kompensierung der Latenz, indem die gespeicherten Funktionswerte gewichtet werden. Lange zurückliegende Werte werden dabei schwä-

cher bewertet als erst kürzlich gespeicherte. Die übliche Wichtung ist dabei linear. Wenn man als Beispiel eine Datenreihe über die letzten 10 Werte heranzieht, wird z.B. der neueste Wert mit der Wichtung 10 multipliziert, während der älteste Wert die mit dem Index 10 die Wichtung 1 erhält. Dazwischenliegende Werte erhalten jeweils die Zahl ihrer Indexnummer als Wertung, so wird etwa der Datensatz an der Stelle 7 mit 7 multipliziert. Die so gewichteten Werte werden addiert und durch die Summe der Wichtungsmultiplikatoren dividiert – in unserem Beispiel also  $10+9+8+7+6+5+4+3+2+1$  (= 55). Der WMA bietet eine gute Reaktionszeit bei guter Glättung.

Der exponentielle gleitende Mittelwert, im folgenden EMA für „*Exponential Moving Average*“ genannt, funktioniert ähnlich wie der WMA, allerdings werden die Wichtungen exponentiell gesetzt. Als solcher stellt er in seinem Wesen eigentlich einen klassischen Tiefpassfilter dar. In der praktischen Umsetzung bedeutet das, dass der EMA mit weitaus geringerem Aufwand umzusetzen ist, da es nicht zwingend notwendig ist, einen Speicher aus historischen Datensätzen zu betreiben. Stattdessen reicht es aus, immer mit dem letzten Datensatz zu arbeiten und diesen mit dem aktuellen Datensatz rückzukoppeln. Hierbei wird der aktuelle Datensatz (mit einem Multiplikationsfaktor  $n \leq 1$ ) zum letzten Datensatz (mit einem Multiplikationsfaktor  $n-1$ ) addiert. Die Summe der Multiplikationsfaktoren muss dabei gleich 1 sein. In empirischen Versuchen habe ich herausgefunden, dass sinnvolle Werte in diesem Anwendungsgebiet bei  $n = 0,7 - 0,8$  liegen.

In der Praxis führten allerdings die Ergebnisse all dieser Glättungsverfahren zu enttäuschenden Ergebnissen. Auf der Suche nach alternativen Methoden stiess ich immer wieder auf Internet-Seiten aus der Wirtschaftsinformatik. Darum begann ich, nach Techniken zu suchen, die bei Aktienhändlern Anwendung finden, mit der einfachen Erklärung, dass dort, wo viel Geld im Spiel ist, mit besonderem Augenmerk auf verlässliche Vorhersagen von Kurs-Ausschlägen und Trends zu rechnen ist. In der Folge stieß ich auf eine spezielle Variante des doppelten Durchschnitts, die von einem Manager namens Allen Hull propagiert wird.

Der gleitende Mittelwert nach Hull, im folgenden HMA für „*Hull Moving Average*“ genannt, stellt eine Kombination aus zwei parallel laufenden gewichteten gleitenden Mittelwerten dar, und zwar einem „kurzen“ WMA über die letzten  $n$  Daten und einem zweiten „längeren“ über die letzten  $n \times 2$  Daten. [ $n$  muss hierbei eine gerade Zahl sein.] Somit stellt der HMA einen doppelt gewichteten WMA dar. (Hull, 2015) Der HMA hat die Eigenschaft, dass er nahezu ohne Latenz auf Auslenkungen reagiert, bei geringer Auslenkung eine gute Glättung erreicht, allerdings auch zum Übersteuern neigt. Dies kann etwa bei Auslenkungen mit starker Flankensteilheit, die sich plötzlich beruhigen, beobachtet werden. Allerdings überwiegen die Vorteile der schnellen Reaktionszeit die Nachteile bei der Genauigkeit.



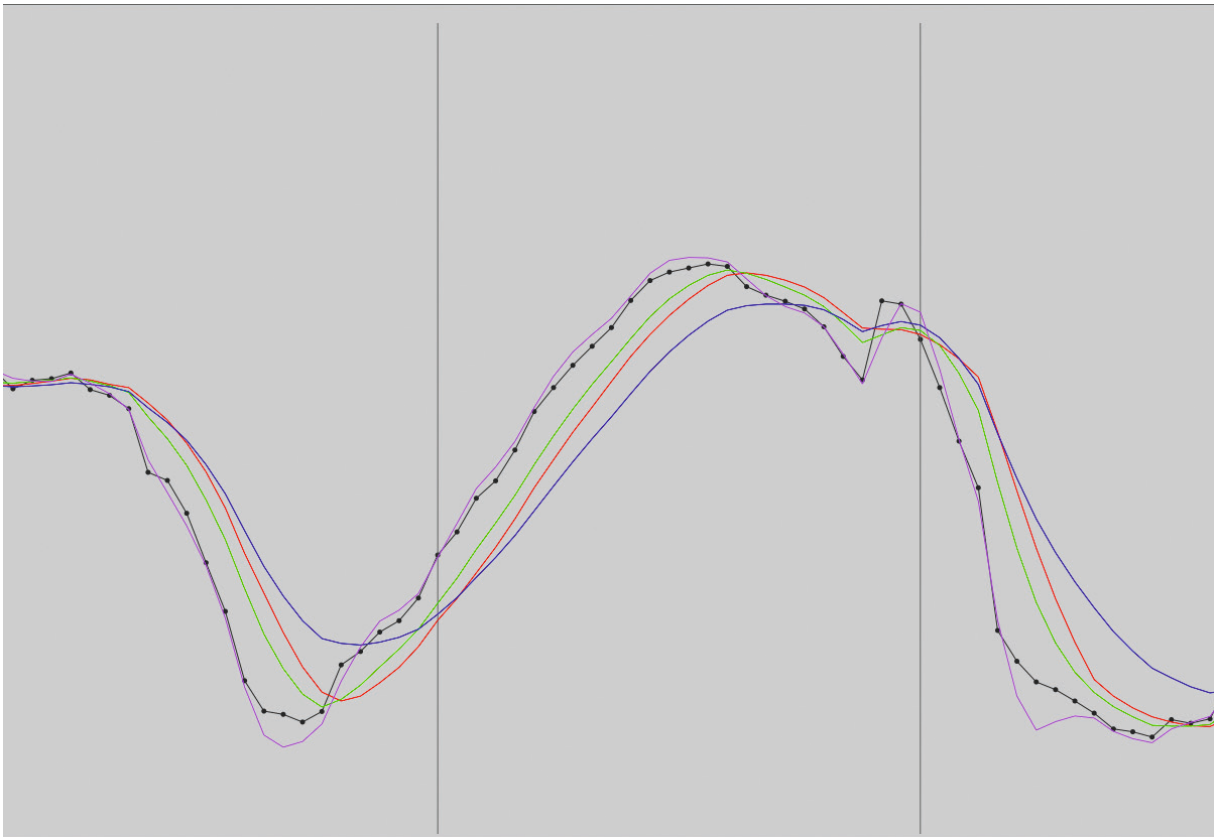


Abbildung 3.1: Vergleich verschiedener Glättungsmethoden: SMA (Rot), WMA (Grün), EMA (n=0.8) (Blau), HMA (Violett)

Anmerkung: Die originale Microsoft Kinect SDK benutzt eine „Holt-Winters Double Exponential Smoothing“ genannte Methode. (Graeme, 2011) Diese wurde jedoch von mir bei *Sonic Manifolds* deaktiviert, da der HMA meines Erachtens nach bessere Ergebnisse liefert, und die Anwendung zweier konsekutiver Glättungsmethoden auf den selben Datensatz die Latenz zu sehr erhöht.

### 3.2 Trenderkennung

Die Glättung der Koordinatendaten ist jedoch nur der erste Schritt einer komplexen Kette von Datenauswertungsmethoden. Da die Impulse nur bei Bewegung hörbar sein sollten, musste diese erfaßt und mit der Lautstärke gekoppelt werden. Weiters sollte die Frequenz der erzeugten Impulse mit der Bewegungsgeschwindigkeit gesteuert werden. Da die geglätteten Daten immer noch ein geringes Rauschen aufweisen bzw. leichtes Zittern ebenfalls aussortiert werden sollte, implementierte ich eine Trenderkennung, die eine kontinuierliche Bewegung elektronisch feststellt. Hierzu werden die Richtungs-Vektoren zwischen aufeinanderfolgenden Positionskoordinaten erfaßt und miteinander verglichen. Werden die aufeinander folgenden Bewegungen ungefähr in der selben Richtung ausgeführt, wird ein intern errechneter Bewegungstrend-Index erhöht. Bei rapiden Richtungswechseln kann von ungewollten Bewegungen oder Zittern ausgegangen werden und der Trend-Index wird verringert. Aufeinanderfolgende Winkel, die kleiner als  $60^\circ$  sind, bedeuten mit

hoher Wahrscheinlichkeit, dass es sich um gewollte Bewegung handelt. Es wird angenommen, dass bei geringem Winkel zwischen den Bewegungs-Vektoren eine flüssige Bewegung zugrunde liegt. Ist der Winkel zwischen mehreren aufeinanderfolgenden Winkel kleiner als 60°, wird ein Bewegungstrend erkannt und der Trend-Index erhöht. Ist der Winkel zwischen aufeinanderfolgenden Vektoren jedoch größer als 90 Grad, wird der „Trend-Index“ vermindert. Der Trend-Index wiederum steuert die Lautstärke der erzeugten Klänge. Der Grund, warum die erkannten Winkel zwischen den Bewegungsvektoren nicht direkt die Lautstärke steuern, sondern ein Umweg über den Trend-Index gegangen wird, ist, dass so etwaige unpräzise Bewegungen noch einmal geglättet werden sollen. Weiters können gegebenenfalls Löcher im Datenstrom überbrückt werden, die sich dadurch ergeben könnten, dass der Computer mit einer Überlast zu kämpfen hat, wodurch nicht in der zu erwartenden Rate neue Koordinaten eintreffen.

Formel für Berechnung der Winkel:

$$\alpha = \arccos(\text{dot}(v1 / \text{norm}(v1), v2 / \text{norm}(v2)))$$

wobei

$v1$  = aktueller Vektor zwischen aktuellem Punkt und letztem bekannten Punkt

$v2$  = vorheriger Vektor

$\alpha$  = resultierender Winkel zwischen  $v1$  und  $v2$

Aus dieser Berechnungsmethode ergibt sich ein Problem: Sind aufeinanderfolgende Punkte ident, führt die Formel eine Division durch 0 aus, was wiederum zu einer Fehlermeldung führt. Um diese abzufangen und Störungen in der Klangerzeugung zu vermeiden, wird in diesem Fall als Winkel 0 ausgegeben, was wiederum einen rückläufigen Trend zur Folge hat. Als weiterer Faktor wird die Geschwindigkeit der Bewegung durch Errechnung der Länge der Bewegungsvektoren ermittelt. Trend und Geschwindigkeit werden dann gemeinsam an den Klanggenerator weitergeleitet, wo die Parameter Eingang finden in die Klangerzeugung. (Vgl. Kapitel 5)

## Szenen

Um die künstlichen Raumkonfigurationen erstellen und abfragen zu können, habe ich für „sonic manifolds“ einen eigenen Editor geschrieben, der es mir ermöglicht, in einem dimensionslosen kubischen Bereich eine beliebige Anzahl primitiver Objekte mit beliebigen Dimensionen zu positionieren und mit verschiedenen Parametern zur Klangerzeugung zu verknüpfen. Die Szenen können in Echtzeit erstellt und modifiziert werden und werden in einem hierarchischen „JSON“ Dokument abgespeichert. JSON ist ein offenes, standardisiertes Dokumentformat. *„JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language.“* (Json.org, 2015)

Der Vorteil von JSON gegenüber anderen Formaten ist der hierarchische Aufbau sowie die Eigenheit, dass Datensätze in der Baumstruktur sowohl aus Texten, Nummern, Arrays oder auch weiteren Unterknoten bestehen können. Somit können auch gemischte Datensätze sehr gut erfaßt und gegliedert werden. Im Fall von „sonic manifolds“ wird jede Szene in einem „Preset“-Datensatz angelegt, der wiederum aus per Nummern indizierten Unter-Objekten besteht, die jeweils eine Raumzone beschreiben.

Diese Zonen werden durch folgende Eigenschaften beschrieben:

Shape – beschreibt die Form der Zone; graphisches Primitiv; kann Quader, Kugel, Zylinder, Konus oder „Kapsel“ sein (eine spezielle Form des Zylinders mit Halbkugeln an den Enden).

Dimension – Größe der Raumzone, abgeleitet von der maximalen Größe.

Position – Mittelpunkt der Raumzone innerhalb der Szene.

Rotate – Rotation im Raum.

Preset – verknüpfte Raumeigenschaften. Hierdurch werden die akustischen Eigenschaften der Zone festgelegt; referenziert eine weitere JSON-Einstellungsdatei in der diese getrennt abgelegt werden.

Solid – Fließkommawert zwischen 0 und 1; beschreibt ob es sich um einen Festkörper oder ein Volumen handelt.

Solidtype – Integerwert; beschreibt Art des Festkörpers / Volumens; entspricht einem Preset.

Priority – beschreibt die Hierarchie von Zonen; wichtig, um bei Überlagerungen von Zonen die momentan Gültige zu ermitteln.

Direct Level – (Fließkommazahl zwischen 0 und 1) – erlaubt nachträglich, den Direktschall abzudämpfen oder zu verstärken.

Reflections Level – (Fließkommazahl zwischen 0 und 1) – erlaubt nachträglich, die Reflexionen abzudämpfen oder zu verstärken.

Durch die Kombination mehrerer derartiger Grundobjekte, die jeweils eine Raumzone definieren, können komplexe Szenerien erstellt werden.

Das „Preset“ Feld wiederum verweist auf Einträge in einer weiteren JSON Datei, in der spezifische Einstellungen für die Audio Engine abgelegt werden. Der Grund für diese Auslagerung ist die effizientere Nutzung von Speicher sowie die schnellere Anpassung von Szenen. Wenn z.B. eine Szene mehrere Zonen mit identischen Einstellungen aufweist, müssen diesen nicht alle per Hand in die Szene selbst hineingespeichert werden

Die Parameter für die Raumsimulation lauten wie folgt, werden aber erst durch die Lektüre von Kapitel 7, welches der Gestaltung des Hallalgorithmus gewidmet ist, verständlich werden.

Vol – beschreibt das Volumen des simulierten Raumes in  $m^3$

Predelay – beschreibt eine Dauer in ms, um die Erstreflexionen manuell verzögern zu können.

Fdn – beschreibt die Delay-Zeiten der einzelnen Verzögerungsschleifen im Feedback-Netzwerk. (Vgl. Kapitel 7.3.2)

Abs – beschreibt die Absorptionseigenschaften der Wände.

Fdn\_mod – die Modulationsfrequenz, mit der die einzelnen Verzögerungsschleifen „belebt“ werden können.

Fdn\_modamp – Der Grad, zu dem die Modulation die Verstärkung der einzelnen Schleifen beeinflusst.

Fdn\_filter – Ein Filter, mit dem das Klangbild weiter verfeinert werden kann.

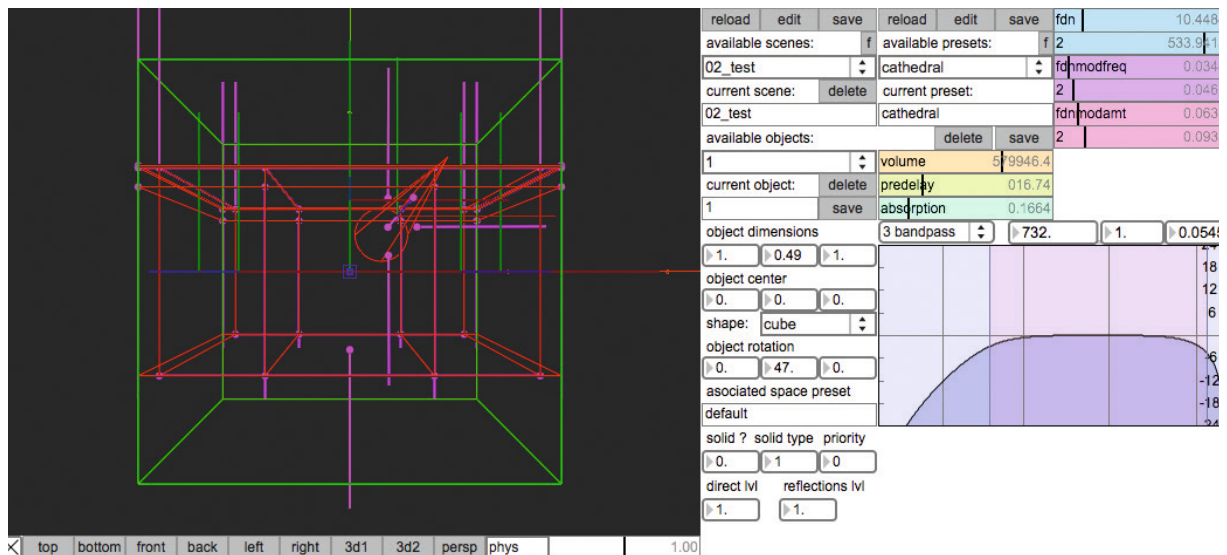


Abbildung 4.1: Screenshot des Szenen-Editors von "Sonic Manifolds"

Der mittels Videotracking überwachte Raumbereich wird nun auf die künstliche Szene mit den Ausdehnungen  $(-1 < x < 1, -1 < y < 1, -1 < z < 1)$  übertragen. Die ermittelten Koordinaten des Benutzers werden benutzt, um die Szene abzutasten und festzustellen, welche Raumzone im künstlichen Bereich der Position der aktuellen Koordinaten entspricht. Bei mehreren sich überlagernden Zonen wird dabei die mit der jeweils höchsten Priorität ausgegeben. Liegen mehrere Zonen mit der gleichen Priorität vor, wird ein Mittelwert zwischen diesen interpoliert. Die resultierenden Werte, die der ermittelten „gültigen“ aktuellen Zone entsprechen, werden an die Audioerzeugung weitergeleitet. Dies bedeutet in der Praxis, dass eine kontinuierliche Bewegung des Benutzers im Realraum auch eine kontinuierliche Erzeugung von künstlichen Klangereignissen zur Folge hat, deren Position der jeweiligen Lage im Realraum entspricht, aber gleichzeitig Reflexionen in der Geometrie des künstlichen Raumes verursacht.

**Die Bewegung im Realraum findet also eine Antwort im künstlichen Raum.**



## Klangerzeugung

Der Kern von *Sonic Manifolds* ist also immer ein Frage-und-Antwort-Spiel des Benutzers mit dem Raum. Technisch gesehen kann dieser Dialog in zwei Komponenten aufgeteilt werden: Erreger-Signale und Schwingkreise. Der Benutzer als Erreger bringt beim Durchschreiten und Abtasten des Raumes selbigen zum Schwingen. Im Hinblick auf „Spaces Speak“ von Blesser und Salter können diese Erreger-Signale als akustische „Lichtquellen“ betrachtet werden, die den Raum aural „beleuchten“. In realen Räumen geschieht dies dadurch, dass unsere Schritte, Bewegungen, Berührungen und vokalen Laute im Raum widerhallen. Wir hören wie der Raum antwortet und können diese Antwort verstehen und so seine Struktur „erhören“. (Vgl. Teil 1, Kapitel 3) Der Mensch nutzt diese unbewusste Fähigkeit ständig, aber sie kann auch bewusst eingesetzt werden. Dafür ist es vonnöten, dass unser Gehirn in der Informationsverarbeitung das Erreger-Geräusch von der Antwort des Raumes trennen kann. Unsere körpereigenen Geräusche begleiten uns Tag für Tag ein Leben lang und sind uns dadurch vertraut, weshalb unser Gehirn sie automatisch auszublenden vermag. Betreten wir eine neue Umgebung, die wir aural erkunden wollen, fällt es uns so leicht, diese vertrauten selbst erzeugten Geräusche von denen unserer Umgebung zu trennen. Das beginnt bei unbewusst erzeugten Geräuschen, gilt aber auch für bewusst provozierte Laute. In der synthetischen Umgebung einer Computersimulation allerdings sind wir mit neuen, ungewohnten Klängen konfrontiert. Die Aufmerksamkeit wird vom Klang des Raumes selbst auf die Erregerklänge abgelenkt bzw. das Trennen von Erreger-Signalen und Schwingkreis wird insgesamt erschwert. Soll in einer solchen akustischen Umgebung der Fokus auf den künstlichen Raum gelenkt werden, muss entweder bei der Erzeugung der Erreger-Signale ein zufriedenstellendes Mass an Naturalismus erzielt oder der Erreger so weit wie möglich eliminiert werden, ohne das Schwingen des Raumes zu beeinträchtigen. Ersteres erfordert immensen Aufwand, bei dem nicht sichergestellt ist, dass das Ziel erreicht werden kann, denn mit fortschreitendem technologischen Entwicklungsstand werden Simulationen, die heute realistisch wirken, morgen bereits als hölzern und künstlich empfunden. Das Thema von *Sonic Manifolds* ist aber der Klang des Raumes und nicht der des Menschen. Der künstliche Raum ist es, der für den Menschen hörbar gemacht werden soll, nicht ein synthetisches



Abbild seiner selbst. Der klare Fokus der Arbeit auf den Raum und die Notwendigkeit, diesen möglichst leicht wahrnehmen zu können hat also zur Konsequenz, dass Erreger-Signale so weit wie möglich reduziert werden sollten. Die Suche nach einer adäquaten künstlerisch-akustischen Sprache führte schließlich zum Entschluss, auf der Seite der Erreger-Signale möglichst auf Abstraktion zu setzen und nur bei den Schwingkreisen, also den künstlichen Räumen selbst, Realismus zuzulassen. Obwohl es sich bei *Sonic Manifolds* um eine aurale Installation handelt, spielt also paradoxerweise die Stille eine große Rolle.

Wie kann man nun die „analogen“ akustischen Äußerungen eines Menschen im Raum in ein digitales Äquivalent übersetzen, das als natürlich empfunden wird und gleichzeitig in den Hintergrund tritt? In der Einleitung zu Teil 1 habe ich Murray Schafer zitiert, der das Hören als „Tasten aus der Ferne“ bezeichnete. Dieses tastende Hören ist die sinnliche Grundlage für das Erkunden von *Sonic Manifolds*. Unsere Hände mit diesem „tastenden Hören“ zu verbinden, ist also logische Konsequenz des Versuchs, dieses Zitat von der metaphorischen in die praktische Ebene zu übersetzen. Indem das Erreger-Signal mit Position und Bewegung unserer Hände verbunden wird, wird aus einer vermeintlichen Metapher eine erlebbare und spürbare Realität. Der Klangraum wird so körperlich zum Schwingen gebracht – das Tasten im Raum macht ihn hörbar. Die Reduktion des Fokus auf die bewusste Bewegung der Hände macht die Absicht hinter deren Vertonung klar. Einem guten Teil von „Spaces Speak“ ist die Beschreibung navigatorischer Aspekte der „Auditory Spatial Awareness“ gewidmet. So verwenden Tiere wie z.B. Fledermäuse oder Zahnwale Ultraschall, um ihre Umgebung aural erfassen zu können. Aber auch der Mensch kann diese Fähigkeit bewusst einsetzen. So hat der Blindenstock nicht nur motorisch-taktile, sondern auch aurale Funktion und auch das Schnalzen mit der Zunge, um Raumantworten zu erhalten, ist eine Navigationshilfe, die immer mehr Verbreitung findet unter sehbehinderten Menschen. (Vgl. Kapitel 3) Diesen Methoden ist allesamt gemein, dass, um Raumantworten zu erhalten, sehr kurzer akzentuierte Schallwellen ausgestoßen werden: Impulse.

## 5.1 Impuls

Als Impuls bezeichnet man in der Akustik einen überaus kurzen, nadelförmigen Amplitudenstoß mit hoher Flankensteilheit, hoher Auslenkung und einer gegen Null gehenden Dauer. Man könnte ihm die Charakteristika eines Anti-Klanges zuschreiben: Ein beinahe körperloser Druckstoß im Schallfeld ohne greifbare Form oder Klangfarbe. Als kurze Unterbrechung der Raum-Zeit macht er das Davor und Danach erst erlebbar: Ein kurzer Stoß, gefolgt von Stille, in der die Antwort der Umgebung in den Vordergrund tritt. Als nahezu abstrakter Klang ist er gleichzeitig aurales Symbol für Energie und Aktion an sich.

Die Wahl eines Klangs zur Vertonung des „Ertastens“ des Raumes fiel also auf den Impuls. Er steht in *Sonic Manifolds* als Platzhalter für das Klatschen und Schnippen, das Menschen verwenden, um einen Raum antworten zu hören. Auch die Wissenschaft nutzt Impulse, um Räume akustisch zu vermessen. (Vgl. Kapitel 7) Impulse bieten den Vorteil, Frequenzen im ganzen Spektrum anzuregen, was wiederum erlaubt, den Klang eines Raumes mit all seinen charakteristischen Färbungen erfassen zu können. Klänge, die eine geringere Flankensteilheit aufweisen, können Teile der Raumantwort verdecken. Der Impuls als Auslenkung mit gegen Null gehender Dauer aber tritt ob seiner Formlosigkeit in den Hintergrund. Seine abstrakte Natur macht seine Bedeutung innerhalb der Simulation offensichtlich: Der Klangstoß als Sinnbild für die menschliche Aktion. Eine Kette von erzeugten Impulsen wiederum kann eine komplexe Bewegung des Benutzers in rhythmische Muster übersetzen. Dabei ist das Timing der Klangerzeugung kritisch dafür, ob die Vertonung als natürlich empfunden werden kann oder nicht.

Technisch gesehen wird der Impuls in Max dadurch erzeugt, dass die Amplitude genau ein Sample lang auf den Maximalwert gesetzt wird. Bei einer Sample-Rate von 44100 Hz bedeutet dies, dass die Dauer des Impulses 0,0000227 Sekunden ist. Über reale Lautsprecher wiedergegeben würde eine Auslenkung der Membran (von Stille auf Maximum und wieder zurück) in diesem Zeitraum aufgrund von physischer Trägheit nicht möglich sein. Darum wird stattdessen auch bei der Messung von Impulsantworten in Räumen eine Pistole verwendet. In der Raumsimulation innerhalb der Software (d.h. noch vor der Ausgabe über Lautsprecher) allerdings ist die korrekte Berechnung dieses 1-Sample Impulses natürlich kein Problem und die Berechnung der Raumantwort daher möglich.

Die Anzahl und Frequenz der generierten Impulse folgt dabei der Dynamik der Bewegungen des Benutzers und wurde in empirischen Versuchen aus den Geschwindigkeiten der Bewegungen ermittelt. Zurückgelegte Distanzen werden zwischen 0 und 100 cm logarithmisch auf Impulslängen von 250 Millisekunden bis 50 Millisekunden skaliert. 50 Millisekunden entsprechen dabei einer Frequenz von 20 Hz, was als maximaler Wert erachtet werden kann bei dem der Impuls als eigenständiges Klangereignis identifiziert wird. In der Praxis sind durchgehende Koordinaten-Deltas von 100 cm auch bei wildem Armrudern im Zeitfenster eines Frames ( $1/25 \text{ sek} = 40 \text{ ms}$ ) praktisch unmöglich. Damit bleiben die generierten Impulse im gewünschten Bereich der schon durch die menschliche Physiognomie selbst begrenzt wird. Die logarithmische Skalierung wiederum bedeutet, dass die Steigerung der Impuls-Frequenz im unteren Bereich der Bewegungen-Deltas sehr langsam ist. Ein langsame Drehung des Handgelenks kann so einen einzelnen Impuls erzeugen, eine langsame Bewegung des Unterarms eine Impulsfolge, während ein Stoß mit der Hand nach vorne schnelle Impuls-Cluster bewirken kann. Die Dynamiken menschlicher Gestik werden so effektiv

mittels eines abstrakten Signals simuliert, das doch organische Eigenschaften aufweist und so glaubwürdig das Eintauchen in die künstliche Welt ermöglicht.

Ein weiterer wichtiger Faktor, um den erzeugten Impulsen eine organische und menschliche Dynamik zu verleihen, ist die Lautstärke. Die im Kapitel 3.2 erläuterte Trenderkennung regelt diese, um bei einem erkannten Bewegungs-„Trend“ mit zunehmender Geschwindigkeit die Klangintensität zu skalieren. Dabei bewerkstelligt die Trenderkennung allerdings nicht nur die Anhebung oder Abschwächung der erzeugten Impulse, sondern entscheidet, ob überhaupt etwas zu hören ist. Steht ein Benutzer still mit hängenden Armen innerhalb der Installation, werden auch keine Impulse erzeugt. Erst die einsetzende Aktion des Benutzers führt, je nach Intensität, zu entweder vereinzelt tastenden Spitzen oder einer ganzen Reihe von lauten Knallen. Durch die Koppelung der Faktoren Frequenz und Lautstärke können so komplexe Dynamiken mit diesen sonst hoch abstrakt wirkenden artifiziellen Impulsen ausgedrückt werden.

## 5.2 Puls

Über das Tasten mittels Impulsen hinaus gibt es allerdings noch eine weitere Klanggattung in *Sonic Manifolds*. Ein dumpfes Pulsieren kündigt die Gegenwart des Benutzers an. Es handelt sich dabei um einen abstrahierten Herzschlag, dessen Lautstärke nur knapp über der Hörschwelle bleibt. Musikalisch gesehen sollen sie nicht als aurale Protagonisten der Szenerie von *Sonic Manifolds* in Erscheinung treten, sondern einen Grundton bilden, der dem Tasten des Besuchers einen Spannungskontrast entgegensetzt. Das Pulsieren ertönt im Gegensatz zum Tastgeräusch auch im Zustand der Entspannung. Sobald der Besucher die Installation betritt, setzt es ein. Tontechnisch gesehen soll es die Impulse, die durch das Tasten der Hände erzeugt werden nicht maskieren oder überlagern, darum bleibt es in einem sehr tiefen Frequenzbereich. Im Gegensatz zu den Tastimpulsen wird es auch nur gering moduliert. So erhöht sich dieser virtuelle Puls geringfügig, wenn die Datenauswertung in der Pose des Besuchers Anspannung zu erkennen glaubt. Darüber hinaus verbleibt es aber im Hintergrund der Interaktion.

In der englischen Literatur wird zwischen Impuls und Puls nicht unterschieden. An dieser Stelle wollen wir die beiden Schallereignisse dadurch voneinander abgrenzen, dass der Impuls als isoliertes Ereignis betrachtet wird, während der Puls einen periodischen Strom an Anspannungs- und Entladungsmomenten beschreibt, im Sinne des englischen Begriffs „Pulse Wave“ (manchmal auch „Pulse Train“), der in der Klangerzeugung mittels Synthesizern Verwendung findet. Puls und Impuls bilden so eine konzeptuelles und ästhetisches Gegensatzpaar.

Tontechnisch gesehen wird der Puls mittels zweier phasenverschobener Dreieckswellen erzeugt, die

voneinander subtrahiert werden. Die sich ergebende Pulswelle wird dann mit einer Sinusgrundschwingung multipliziert. Die resultierende Pulswelle wird räumlich im Torso des Besuchers (Vgl. Kapitel 2.1) verortet und der Raumklangsimulation zugeführt.

Zwar ist *Sonic Manifolds* an sich ohne Benutzereingabe stumm, aber die meisten wirklichen Räume werden bereits von einem diffusen Grundgeräusch „beleuchtet“ (Vgl. Teil 1, Kapitel 3). Dieses diffuse Grundgeräusch kann maschinellen oder natürlichen Ursprungs sein, es hat jedenfalls zumeist die Eigenschaft, eine vage Atmosphäre zu erzeugen, ohne die Struktur an sich genau hörbar zu machen. Einer strengen Logik zu Folge müsste jedes derartige diffuse Geräusch mit einer bewusst gesetzten Klangquelle gekoppelt sein und erst im Widerhall des Raumes entstehen. In manchen Szenen bricht Blunzgröstl jedoch mit dieser Logik, um besondere Kontraste herzustellen. In solchen Szenen wird der Algorithmus, der den diffusen Nachhall erzeugt mit Rauschen gespeist.



## Raumklang

### 6.1 Kurze Geschichte des Raumklangs

Technisches Grunderfordernis für *Sonic Manifolds*, um eine räumliche aurale Immersion zu erzielen, ist die Erzeugung von Raumklang. Wie im Kapitel 1 beschrieben benutzt unser Gehör verschiedene Techniken, um Klang zu lokalisieren. Diese beruhen auf dem Vergleich des von beiden Ohren gehörten auf Laufzeit- und Pegelunterschiede, sowie Frequenz-Unterschiede im ganzen Spektrum, die durch die HTRF (engl. *head-related transfer function*) verursacht werden. Die dominante Technik zur Erzeugung von Raumklang ist bis heute Stereo. Vorläufer der Stereophonie existieren schon seit dem späten 19. Jhdt., wobei die erste belegte Verwendung das Théâtrophone war (1881). Als geistiger Vorläufer des Rundfunks war es dafür gedacht, Live-Übertragungen von Theater und Operaufführungen über Telephon-Leitungen zu übertragen. (Schwartz, 2011, S.454; Weinzierl, 2008, S.610) Die mittels Kohlemikrofonen aufgenommenen Audio-Signale wurden am Empfängerende zumeist über Münzautomaten mittels Headsets mit je einem Lautsprecher pro Ohr wiedergegeben. Das System blieb vor allem in Frankreich bis in die 1930er Jahre in Betrieb, wurde später auch für Nachrichten- und Gottesdienst-Übertragungen verwendet, setzte sich aber insgesamt nicht weiter durch. (Lange, 2002)

Entscheidende Fortschritte sowie den Durchbruch bei der Stereophonie brachte aber nicht die Musikindustrie, sondern der Film. Im Speziellen kann Alan Dower Blumlein als Vater der modernen Stereophonie gesehen werden. (Weinzierl, 2008, S.610) Er entwickelte als erster ein auf zwei Lautsprechern basierendes System zur Erzeugung von Raumklang für den Film. Das erste bekannte Mehrkanalsystem steuerten dann William E. Garity and J. N. A. Hawkins mit „Fantasound“ für den 1940 erschienenen Disney-Film *Fantasia* bei. 3 Audiospuren wurden dabei mittels „Pan-Potting“<sup>1</sup> auf bis zu 65 Lautsprecher verteilt. Blesser und Salter schreiben hierzu in „*Spaces Speak*“: „*The requirements of surround sound, in the context of 1930s technology, placed unique demands on the*

1 Spatialisierung des Signals durch Intensitätsregelung mittels Potentiometern im Panorama.

*recording, mixing, and reproduction processes. Up to 33 microphone signals were recorded on nine master tracks, then mixed down to three optical sound tracks, and finally, expanded to feed dozens of loudspeakers throughout the theater. By allocating one of the four optical tracks to control the variable-gain amplifiers in real time, audio engineers could share a small number of sound tracks among a large number of loudspeakers, at one moment providing direct sound, and at another providing ambient. Similarly, the newly invented constant-power pan potentiometer (pan pot) allowed a single source to be smoothly cross-faded between neighboring loudspeakers, thereby creating the illusion of motion.*“ (Blessner, Salter, 2007, S.207)

Während in der Folge einzelne musikalische Projekte für Mehrkanal-Raumklangsysteme umgesetzt wurden, blieb das primäre Anwendungs- und Forschungsgebiet weiterhin das Kino. Musik wurde bis zum Ende der 1960er Jahre vorwiegend in Mono aufgenommen. Einzelne herausragende Live-Experimente mit Mehrkanalklang stellen die Kompositionen *Poème électronique* von Edgar Varese und *Concret PH* von Iannis Xenakis dar, die für den Phillips Pavillon auf der Weltausstellung 1958 über ein Array von 350 Lautsprechern verteilt wurden, sowie das Stück „Spiral“ von Karlheinz Stockhausen, das bei der Weltausstellung 1970 in Osaka im kugelförmigen deutschen Pavillon über 50 Lautsprecher wiedergegeben wurde. (Blessner, Salter, 2007, S.149, 171; Föllmer, 2015) Auf Seiten der Pop-Musik sei Pink Floyd erwähnt, die schon im Jahre 1967 begannen, Klang mittels speziellen Joystick-Controllern bei Live-Konzerten über mehrere Lautsprecher zu verteilen. (Peters, 2010, S.4)

Diesen Mehrkanal-Systemen war aber allesamt zu eigen, dass es sich um keine wahren Raumklang-Formate handelte. Stattdessen wurden wenige Spuren entweder manuell (Stockhausen) oder mittels aufgezeichneter Steuerdaten (Varese, Xenakis) über eine große Zahl von Lautsprechern verteilt. Dies hat zur Folge dass nur eine begrenzte Anzahl von Schallquellen spatialisiert werden können, da nur eine begrenzte Zahl an Originalsignalen vorhanden ist. Gemischte Spuren im Nachhinein mit analogen Methoden zu trennen war ein Ding der Unmöglichkeit. Eine weitere Einschränkung ergibt sich durch „Pan-Potting“ selbst: Hierbei wird mittels Lautstärken-Regulierung ein Lautsprecher beschickt, der in der gewünschten Hör-Richtung liegt, während die entgegen liegenden Lautsprecher keinen Anteil an der Erzeugung der Spatialisierung haben. Tatsächlicher Raumklang ist aber weitaus komplexer, und würde erfordern, dass der Klang aus allen Lautsprechern zu hören ist, inklusive seinen respektiven Modifikationen, die er durch seine Reise durch den Raum erfahren hat. Realistischer Raumklang hüllt den Hörer ein, denn die Druckwellen in einem echten Raum kommen aus verschiedenen Richtungen und werden nicht nur von einzelnen Punktquellen abgestrahlt, wie dies bei Lautsprecherbetrieb üblich ist. Die manuelle Methode der Spatialisierung mittels „Pan-Potting“ repräsentiert also eine zu starke Vereinfachung um je ein





Abbildung 6.1: Karlheinz Stockhausen bei der Weltausstellung Osaka 1970 (Stockhausen, 1978)



realistisches Raumklang-Erlebnis zu bieten.

Um realistischen Raumklang zu erzeugen, müsste also das ganze Schallfeld aufgenommen oder synthetisiert und später wieder im Raum abgebildet werden können. Verschiedene Forschungen in diese Richtung haben zu ausgezeichneten Ergebnissen geführt, darunter das Anfang der 1970er am NRDC entstandene „Ambisonics“ und die ab 1988 von einer Initiative aus verschiedenen Universitäten und Instituten entwickelte Wellenfeldsynthese. Trotzdem sind klassische „Pan-Pot“-Methoden zur Verteilung von Klang im Raum bis heute Standard geblieben. Mehr noch, Stereo wurde zumindest in der Musik bis heute nicht abgelöst. Eine kurze Exkurs der Musikindustrie in die Quadrophonie in den 70er Jahren scheiterte kolossal an technischen Unzulänglichkeiten. Im Film dominieren Surround Verfahren, die vom ursprünglichen Dolby Surround der 1970er Jahre abgeleitet sind. Diese beruhen auf einer Erweiterung von Stereo durch Effekt- bzw. Ambiente- sowie Subwoofer-Kanälen.

Stereo ist dem menschlichen Ohr nachempfunden und beruht auf dem Grundgedanken, dass unser linkes und rechtes Ohr jeweils ein leicht unterschiedliches Klangbild wahrnehmen, woraus unser Gehirn die räumliche Szene rekonstruiert. Allerdings funktioniert dies nur gut, wenn wir die Signale mittels Kopfhörer zugeführt bekommen, denn selbst bei gut aufbereitetem Audiomaterial kann bei Lautsprecherbetrieb nicht gewährleistet werden, dass unser linkes Ohr nur den linken Kanal und unser rechtes Ohr nur den rechten Kanal wahrnimmt. Eine realistische Abbildung des Schallfeldes ist ohne Tricks in der Stereophonie nicht möglich. Stattdessen werden wir an beiden Ohren jeweils eine Mischung aus beiden Signalen wahrnehmen. Durch diese „Übersprecher“ (engl. cross talk) wird das Stereobild verwischt und der Raumeindruck reduziert, wodurch es bei ungünstigen Bedingungen es auch zu Auslöschungen kommen kann. Das wird besonders deutlich, wenn sich die Hörer nicht im „Sweet Spot“ befinden und/oder den Kopf nicht nach der Abhörsituation ausgerichtet haben. Der „Sweet Spot“ ist die Zone, in der ideale Bedingungen zur Wahrnehmung des Wiedergegebenen herrschen und befindet sich im Schnittpunkt der gedachten Achsen der Lautsprecher.

Aber diese Probleme charakterisieren nicht nur Stereo, denn auch Dolby Surround und seine Derivate (z.B. 5.1, 7.1, etc.) leiden unter diesen Limitationen. Einer der Gründe ist die vorausgesetzte Direktionalität der Abhöranalagen und -Situationen. Die historische Ursache hierfür ist in den Ursprüngen dieser Formate zu suchen, denn sie alle entstammen nicht der Musikindustrie, sondern sind Entwicklungen der Filmindustrie. So soll Blumlein der angeblich während einer Filmvorführung auf die Idee gekommen sein, sich mit Raumklang zu beschäftigen, als ein Schauspieler rechts auf der Leinwand den Mund öffnete, aber seine Stimme von links ertönte. Die von ihm im Jahre 1935 erzeugten Demonstrationsfilme „Trains At Hayes“ und „Walking And Talking“ gelten als die

ersten Filme mit Stereo-Klang. Im Weiteren wurde Stereo auch fürs erste eine vorwiegend im Kino Verwendung findende Technologie. (Blumlein, 1935; 1935b)

Der Fokus auf einen frontalen Bereich außerhalb des Auditoriums ist dem Medium Film inhärent. Das Kino als treibende Kraft der technischen Innovationen am Audiosektor hat diese Standards gesetzt, die Musikindustrie zog erst später nach. Darum spielt sich die Räumlichkeit von Stereo in einem sehr beschränkten Bereich vor den Boxen ab und funktioniert am Besten, wenn der Winkel vom Besucher zu den Boxen  $60^\circ$  nicht überschreitet. Ist der Winkel größer, zerfällt das räumliche Bild und ein akustisches Loch tut sich auf. Klangereignisse, die eigentlich in der Mitte lokalisiert werden sollten, hängen dann links und rechts. Weiters zerfällt das Stereobild auch, wenn der Benutzer den Kopf dreht. Da die Dolby Surround Formate der selben Tradition entstammen, gelten auch für sie diese Beschränkungen. Die Niederlage der quadrophonischen Formate der Sechziger und Siebziger Jahre kann unter diesem Gesichtspunkt leicht erklärt werden: Mit einem Winkel von  $90^\circ$  zwischen den Lautsprechern waren sie technologische Totgeburten, die niemals glaubwürdige Schallfelder reproduzieren konnten. (Gerzon, 1974) Weiters handelte es sich bei der Mehrzahl dieser Formate um keine echten Mehrkanal-Formate. Stattdessen wurden aus 2 Kanälen mittels Matrixverfahren 4 Kanäle gewonnen.

Die besten Ergebnisse im Raumklang erzielen sogenannte binaurale Kopfhörer-Systeme, bei denen die linken und rechten Signale mit sogenannten HTRF (engl. head related transfer function) – Filtern behandelt werden. (Kutruff, 2009, S.25ff.) Da die Form des Kopfes und unserer Schultern den Schall, der unsere Ohren beim Hören erreicht, verändert, kann durch Simulation dieser Effekte ein nahezu perfektes Schallfeld erzeugt werden. Dies kann entweder durch Aufnahme mittels eines sogenannten Kunstkopf-Mikrophons oder durch künstliches Filtern des Audiosignals mittels DSP erzielt werden. Das reproduzierte Schallfeld entsteht allerdings ausschließlich im Kopf und wird bei Bewegungen desselben mitgedreht, was (psychologisch gesehen) die Immersion im künstlichen Schallraum beeinträchtigt. Methoden bei denen mit Hilfe von Head Tracking die Position und Drehung des Kopfes erfasst wird, um die Veränderung des Schallfeldes in Echtzeit zu berechnen, funktionieren zwar außerordentlich gut, isolieren den Hörer aber von der Umgebung und den Mitmenschen und kamen deshalb für *Sonic Manifolds* nicht in Frage.

Alan Blumlein gilt zwar als Vater der Stereophonie, sein besonderes Interesse galt aber Methoden der Reproduktion binauralen Hörens. Sein Ziel war den hohen räumlichen Realismus von binauralen Aufnahmen auch auf Lautsprechern reproduzieren zu können. Er entwickelte eine Mikrophonierungstechnik, die einen außerordentlich guten räumlichen Eindruck bei Wiedergabe über

Kopfhörer erzielte.<sup>2</sup> Allerdings stellte er fest, dass diese Aufnahmen auf Lautsprechern unnatürlich klangen, da die Laufzeitunterschiede, die beim Hören durch Kopfhörer die Räumlichkeit herstellen, bei der Reproduktion über Lautsprecher nicht mehr den gewünschten Effekt erzielen, da beide Ohren beide Signale wahrnehmen (Vgl. Cross Talk). „*Directional hearing sense is due to phase and intensity differences between sounds reaching the two ears, phase differences being more effective for the lower frequencies and intensity differences for higher frequencies. As phase differences in two loudspeakers (both heard by both ears) do not produce the required effect and normally reproduced intensity differences are not sufficiently marked, the modifying arrangements translate low frequency phase differences into intensity differences and amplify the higher frequency intensity differences.*“ (Blumlein, 1933)

Als Lösung entwarf er den sogenannte Blumlein-Shuffler, der dazu diente, die ITDs im tieferfrequenten Bereich in ILDs umzuwandeln. „*Blumlein conceived stereo not just as a left (L) and right (R) speaker signal but also in terms of a sum signal M (= L+R) and a difference signal S (=L-R).*“ (Gerzon, 1986) Dabei wird also ein „Mitten-Signal“ aus der Summe des linken und rechten Kanals generiert, während die Differenz der beiden Kanäle ein „Seiten-Signal“ erzeugt. Durch Veränderung der Pegelbalance zwischen Mitten- und Seiten-Kanal kann nun die Breite des Stereobildes verändert werden. Bei Einsatz eines Equalizers kann auch die fehlende Breite von binauralen Aufnahmen in tieferen Frequenzen (z.B. < 600 Hz) kompensiert werden.<sup>3</sup>

Konversionsformeln MS ↔ Stereo

$$M = L + R$$

$$S = L - R$$

$$L = \frac{1}{2} \times (M + S)$$

$$R = \frac{1}{2} \times (M - S)$$

Somit entdeckte Blumlein, ohne sich dessen bewusst zu sein, das Prinzip der Mitte-Seite-Stereophonie, welche später die Basis für Stereoklang im Rundfunk werden sollte. Darüber hinaus nahm er die Zerlegung des Schallfelds in Schalldruck und -Schnelle-Komponenten vorweg, die vierzig Jahre später die Basis für Ambisonics werden sollte.

Ambisonics ist einer der zwei momentan fortschrittlichsten Methoden, um Raumklang in einer Umgebung mittels mehrerer Lautsprecher umzusetzen. Dabei wird das Schallfeld durch ein dichtes

<sup>2</sup> Blumlein wird mit der bis heute verwendeten XY-Stereophonie (auf Englisch auch als „Blumlein Pairs“ bekannt) in Verbindung gebracht. Tatsächlich unterscheidet sich Blumleins Methode allerdings von jener, für die er heute bekannt ist. Blumlein verwendete zwei Kugelmikrophone, die durch einen Separator getrennt waren, um die Abschattung des Klangs durch den menschlichen Kopf (HRTF) zu simulieren. In der XY-Stereophonie werden zwei 90° zueinander gedrehte bidirektionale („8er“) Mikrophone verwendet, die allerdings im Jahr 1931 noch gar nicht erfunden waren.

<sup>3</sup> Dabei sollte mit dem Anheben des Pegels in einem Kanal eine Abschwächung im anderen einhergehen, um das gesamte Frequenzbild nicht zu stören. (Gerzon, 1986)

Array von Lautsprechern synthetisiert. Während ihm in seinen analogen Anfangstagen kein kommerzieller Erfolg beschieden war, erlebt es seit der kostengünstigen Verfügbarkeit leistungsfähiger Computer und der Verbreitung digitaler Musikproduktion ein Wiederaufleben. Noch vielversprechender als Ambisonics ist die Wellenfeldsynthese. Erfunden in den späten 1980er Jahren, stellt sie ebenfalls eine revolutionäre Entwicklung dar, die aber abseits akademischer Umgebungen bisher kaum eingesetzt worden ist. Während Ambisonics in der Tradition von Blumlein steht und das Schallfeld aus einem Punkt heraus abbildet, erzeugt die Wellenfeldsynthese mittels eines zweidimensionalen Arrays aus einer sehr hohen Zahl an Lautsprechern ganze Wellenfronten. Diese werden im Raum erst durch die Überlagerung von Einzelwellen erzeugt und erlauben deshalb einen sehr hohen Realismus, da das Schallfeld unabhängig von einem Bezugspunkt synthetisiert werden kann. Der Nachteil ist, dass eine extrem hohe Dichte (und damit Anzahl) an Lautsprechern vonnöten ist, wodurch das System unflexibel und teuer ist. Weiters ist das reproduzierbare Frequenzspektrum durch die Abstände der Lautsprecher begrenzt. (Weinzierl, 2008, S.669) Ambisonics hingegen hat sich durch die vielfältigen Möglichkeiten, die die digitale Signalverarbeitung bietet, zu einer kostengünstigen Möglichkeit entwickelt um exzellenten dreidimensionalen Raumklang zu erzeugen.<sup>4</sup>

## 6.2 Ambisonics

Ambisonics beruht auf dem Grundprinzip, dass Aufnahme- und Wiedergabesituation unabhängig voneinander sind. Während Stereo und 5.1-Surround diskrete Verfahren sind, bei denen jede Tonspur exakt einem Lautsprecher zugeordnet ist, kann bei Ambisonics deren Anzahl und Position frei gewählt werden. Das wird dadurch möglich, dass das ganze Schallfeld in Druck- und Schnelle-Komponenten zerlegt vorliegt und der Schalldruck durch Matrizen-Rechnung an jeder beliebigen Stelle im Schallfeld im Nachhinein berechnet werden kann. Somit kann auch erst bei der Wiedergabe die Position des Lautsprechers festgelegt und der für diese Position korrekte Schalldruck ermittelt werden. Das Schallfeld wird somit unabhängig von der Abhörsituation immer korrekt berechnet – vorausgesetzt, dass die Positionen bei der Dekodierung korrekt eingegeben werden. Bei Stereo- oder Surround-Systemen ist dies nicht möglich. Entspricht die Aufstellung der Lautsprecher nicht der bei der Abmischung vorgesehenen, zerfällt das Schallfeld. Ein weiterer wichtiger Aspekt ist, dass erst im Zusammenspiel aller Lautsprecher das räumliche Schallfeld rekonstruiert wird. Dabei werden in verschiedenen Tonspuren die Anteile des Schallfeldes in je einer Raumachse aufgenommen. Ambisonics ist dabei in der räumlichen Auflösung skalierbar, was durch die Bezeichnung der Ordnung ausgedrückt wird. Ambisonics nullter Ordnung beschreibt den omnidirektionalen

---

<sup>4</sup> Als hohe Kostenfaktoren bleiben jedoch weiterhin der Raumbedarf und die Ausstattung mit Verstärkern und Lautsprechern für die Wiedergabe-Situation.

Schalldruck in einem (Mono-)Kanal (traditionellerweise genannt W) im Raumursprung. Bei Ambisonics erster Ordnung kommen drei Kanäle hinzu, die jeweils die Schallschnelle in x-, y- und z-Achse beschreiben und die Namen X, Y, Z tragen. In zweiter Ordnung kommen 5 weitere Kanäle hinzu (R, S, T, U, V), mit dritter 7 (K, L, M, N, O, P, Q), etc. wobei in einem Kanal immer eine Raum-Achse kodiert wird. Die Bezeichnungen der Kanäle höherer Ordnung beruhen auf der nachträglichen Erweiterung des ursprünglichen W-X-Y-Z Systems und können aufgrund fehlender semantischer Logik Verwirrung stiften. Ambisonics-Formate ab der 3. Ordnung verzichten deshalb weitgehend auf diese Tradition und werden gemeinhin durchnummeriert. Die Kanäle von Ambisonics werden auch als „sphärische Harmonische“ bezeichnet und können als Mittel verstanden werden, um jeweils die räumliche Auflösung der jeweils niedrigeren Ordnungsstufe zu erhöhen. B-Formate höherer Ordnung sind abwärtskompatibel, d.h. ein B-Format dritter Ordnung kann mit einem Dekoder erster Ordnung dekodiert werden. Dabei werden dann nur die Kanäle W, X, Y und Z verwendet, was einem Verlust an räumlicher Auflösung gleichkommt. Bei der Wiedergabe werden mittels der Position des Lautsprechers im Verhältnis zum Ursprung Schalldruck und -Schnelle an der gewünschten Stelle errechnet. Hierzu dient eine Reihe von Formeln.

Entwickelt wurde Ambisonics um 1970 an der britischen „National Research Development Corporation“ unter der Leitung von Peter Fellgett und in der Folge von Michael Gerzon perfektioniert. Gerzon entwickelte im Jahr 1975 das sogenannte „Soundfield“-Mikrophon, das zur Aufnahme von Ambisonics-kompatiblen Signalen verwendet wurde. Es besteht aus vier Mikrophonkapseln mit Nierencharakteristik, die an den Eckpunkten eines Tetraeders angeordnet sind und nach außen zeigen. Fellgett beschreibt das Soundfield Mikrophon als „[...] *an omnidirectional microphone in the true sense, which is the opposite of non-directional; it characterises in a symmetrical manner the waveform and directionality of sound arriving from any direction (including vertical components).*“ (Fellgett, 1975, S.10) Das Produkt einer Aufnahme mit dem Soundfield-Mikrophon liegt im sogenannten A-Format vor, welches nicht direkt zur Verwertung gedacht ist, sondern in der Folge in das B-Format umgerechnet wird. A-Format und B-Format erster Ordnung können jederzeit verlustfrei ineinander umgewandelt werden. B-Formate höherer Ordnung allerdings verlieren an Information bei der Rückwandlung ins A-Format. Während im B-Format z.B. die Phasenlage eines Kanals gleichzeitig die räumliche Richtung beschreibt (ähnlich wie in der Mitte-Seite Stereophonie), was eine Verzerrung der Räumlichkeit bei der Behandlung mit Effekten zur Folge haben kann, können die vier Kanäle des A-Formats theoretisch als unabhängige Monosignale behandelt werden. Während also das B-Format das ganze Schallfeld vom Ursprung aus enkodiert, beschreibt das A-Format den Schalldruck an vier Einzelpositionen mit gleichem Abstand zum Ursprung in jeweils 45° Winkel (links vorne oben, rechts vorne unten, links hinten unten, rechts hinten oben). Da die Klänge von *Sonic Manifolds* allesamt synthetisch erzeugt werden, sollte man annehmen, dass das A-Format im

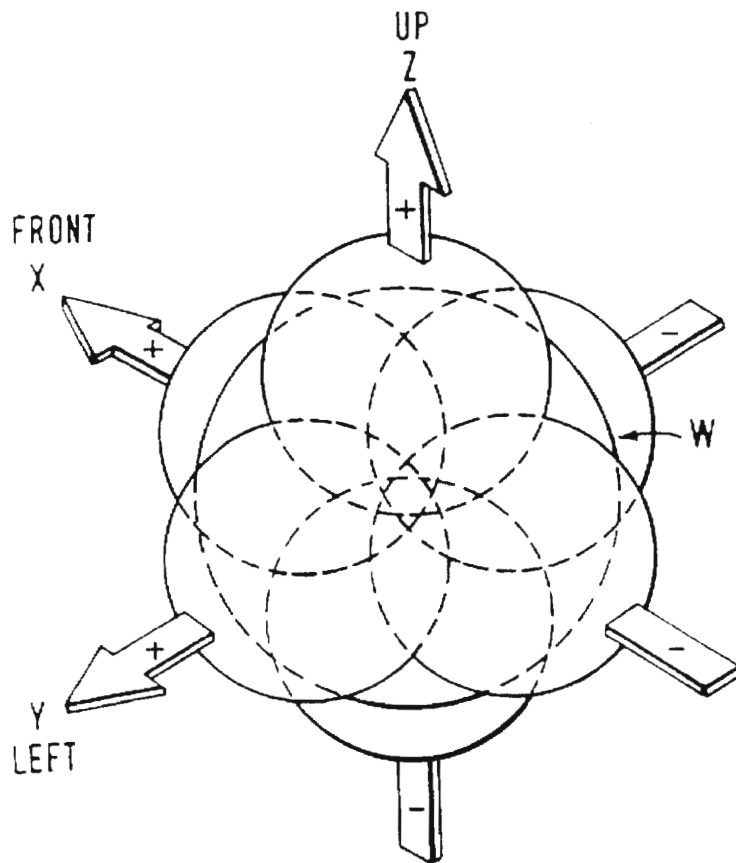


Abbildung 6.2: Ambisonics B-Format 1. Ordnung (Gerzon, 1998)



Abbildung 6.3: Soundfield Microphone (TSL, 2013)



Signalfluss gänzlich vermieden werden kann. Es gibt jedoch einen wichtigen Grund, warum es auch in *Sonic Manifolds* zum Einsatz kommt: Der Hall-Algorithmus bzw. im Speziellen der diffuse Nachhall. (Vgl. Kapitel 7) Da die Phasenlage der B-Format-Kanäle die Räumlichkeit beschreiben, können Klangeffekte nicht ohne weiteres angewandt werden, ohne das Schallfeld zu verzerren. Durch die Rückwandlung in das A-Format hingegen können gefahrlos beliebige Effekte eingesetzt werden ohne die Räumlichkeit zu beeinträchtigen.

## 6.3 Implementierung, Formeln

Konversion von A-Format ↔ B-Format

$$W = 0.5 \times (LF + RF + LB + RB)$$

$$X = 0.5 \times ((LF - LB) + (RF - RB))$$

$$Y = 0.5 \times ((LF - RB) - (RF - LB))$$

$$Z = 0.5 \times ((LF - LB) + (RB - RF))$$

Konversion von B-Format ↔ A-Format

$$LF = 0.5 \times (W + X + Y + Z)$$

$$RF = 0.5 \times (W + X - Y - Z)$$

$$LB = 0.5 \times (W - X + Y - Z)$$

$$RB = 0.5 \times (W - X - Y + Z)$$

Aufnahmen des Soundfield-Mikrophones können nur im B-Format der nullten und ersten Ordnung kodiert werden. Bei der synthetischen Erzeugung von B-Format Signalen empfiehlt es sich aber besonders bei höherer Dichte von Lautsprechern eine höhere Ordnungsstufe zu verwenden, denn dadurch erhöht sich gleichzeitig auch die räumliche Auflösung. Bei *Sonic Manifolds* erfolgt die Kodierung im Ambisonics B-Format der 2. Ordnung. Jan Schacher empfiehlt in der Dokumentation zu seinen ICTS Ambisonics-Externals (d.h. einem Erweiterungspaket) für MAX/MSP, sich bei Wahl der Ordnung an der zu erwartenden Anzahl an Wiedergabe-Lautsprechern zu orientieren. (Schacher, Kocher, 2014) Da die mir zur Verfügung gestellte Anlage acht Kanäle aufwies, wählte ich wie erwähnt die zweite Ordnung, die neun Kanäle aufweist. Eine höhere Ordnung würde erst bei einer höheren Anzahl von Lautsprechern Sinn ergeben. Ausserdem steigt die Belastung des verwendeten Computersystems aufgrund der wesentlich höheren Anzahl an zu verarbeitenden Signalen drastisch. So sind in der dritten Ordnung bereits 16 Spuren pro Klangquelle in Echtzeit zu berechnen. Besonders die trigonometrischen Funktionen in den Kodierungsformeln schlagen

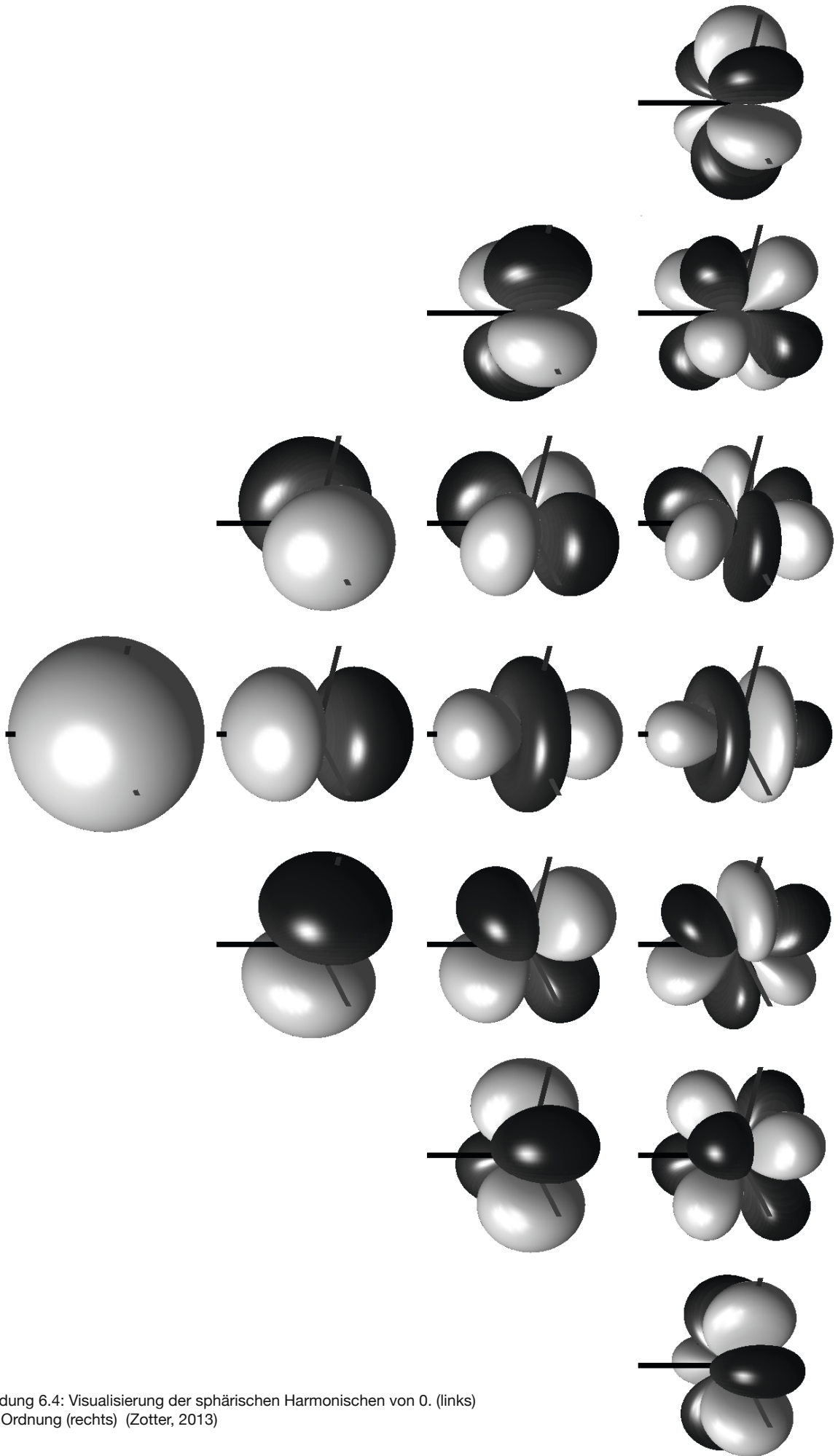


Abbildung 6.4: Visualisierung der sphärischen Harmonischen von 0. (links) bis 3.Ordnung (rechts) (Zotter, 2013)

hierbei sehr zu Buche.

Formeln zur Kodierung eines Mono Signals in das B-Format:

### 0. Ordnung

$$W = \text{Signal} \times 1 / \text{sqrt}(2)$$

### 1. Ordnung

$$X = \text{Signal} \times \cos(\varphi) \times \cos(\theta)$$

$$Y = \text{Signal} \times \cos(\varphi) \times \sin(\theta)$$

$$Z = \text{Signal} \times \sin(\varphi)$$

### 2. Ordnung

$$R = \text{Signal} \times \left(\frac{1}{2}\right) \times (3 \times \sin(\varphi) \times \sin(\varphi) - 1)$$

$$S = \text{Signal} \times \cos(\theta) \times \sin(2 \times \varphi)$$

$$T = \text{Signal} \times \sin(\theta) \times \sin(2 \times \varphi)$$

$$U = \text{Signal} \times \cos(2 \times \theta) \times \cos(\varphi) \times \cos(\varphi)$$

$$V = \text{Signal} \times \sin(2 \times \theta) \times \cos(\varphi) \times \cos(\varphi)$$

### 3. Ordnung

$$K = \text{Signal} \times \left(\frac{1}{2}\right) \times \sin(\varphi) \times (5 \times \sin(\varphi) \times \sin(\varphi) - 3)$$

$$L = \text{Signal} \times \text{sqrt}(135/256) \times \cos(\theta) \times \cos(\varphi) \times (5 \times \sin(\varphi) \times \sin(\varphi) - 1)$$

$$M = \text{Signal} \times \text{sqrt}(135/256) \times \sin(\theta) \times \cos(\varphi) \times (5 \times \sin(\varphi) \times \sin(\varphi) - 1)$$

$$N = \text{Signal} \times \text{sqrt}(27/4) \times \cos(2 \times \theta) \times \sin(\varphi) \times \cos(\varphi) \times \cos(\varphi)$$

$$O = \text{Signal} \times \text{sqrt}(27/4) \times \sin(2 \times \theta) \times \sin(\varphi) \times \cos(\varphi) \times \cos(\varphi)$$

$$P = \text{Signal} \times \cos(3 \times \theta) \times \cos(\varphi) \times \cos(\varphi) \times \cos(\varphi)$$

$$Q = \text{Signal} \times \sin(3 \times \theta) \times \cos(\varphi) \times \cos(\varphi) \times \cos(\varphi)$$

(Neukom, 2009, S.28f.)

Die Schalldruck-Komponente  $W$  könnte auch mit einem anderen Faktor skaliert werden als der traditionellen  $W$ -Wichtung von  $1/\text{sqrt}(2)$ . In der Praxis würde das (analog zur Mitte-Seite Stereo-

phonie) eine Veränderung der Breite des Schallfeldes bedeuten. (Vgl. Kapitel 7.3.4) Die Schallquellen in *Sonic Manifolds* werden mittels obiger Formeln in das B-Format kodiert. Hierzu müssen die Koordinaten in polarem Format vorliegen, d.H. es müssen Distanz (d), Azimuthwinkel ( $\theta$ ) und Höhenwinkel  $\phi$  („Elevation“) bekannt sein. Diese werden mittels folgender trigonometrischer Formeln aus den XYZ-Koordinaten gewonnen, die die Videotracking- und Datenglättungsalgorithmen liefern:

$$\theta = \text{atan2}(x, y)$$

$$\phi = \text{atan2}(z, \sqrt{x^2 + y^2})$$

$$d = \sqrt{x^2 + y^2 + z^2}$$

Da aber jedes im B-Format kodierte Signal auf einem Punkt auf der Schallfeld-Kugel mit gleichem Abstand zum Mittelpunkt liegt, muss ferner die Distanz mittels der Lautstärke simuliert werden. Für Punkte außerhalb der Ambisonics-Kugel gilt dabei eine Abschwächungsformel. (Vgl. Kapitel 7.2) Direktschall wird in *Sonic Manifolds* aber immer an der Position des Benutzers spatialisiert und liegt damit zwangsläufig innerhalb des Schallfeldes. Um die Direktionalität auch im Bereich zwischen den Lautsprechern zu simulieren, kommt eine Skalierungsmethode zum Einsatz, die die Schallschnelle-Komponenten mit zunehmender Nähe zum Ursprung verringert. Das bedeutet, dass nahe am Ursprung nur lediglich die Schalldruckkomponente im Monosignal W herangezogen wird. Im Ursprung kommt damit der Schall gleichmäßig aus allen Lautsprechern. Die Formel hierfür stammt von Jan Schacher und wurde von ihm in seinen für das ICST entwickelten Ambisonics-Externals implementiert. (Schacher, Kocher, 2014)

$$w\_gain = \text{center\_size} - ((1.0 - \text{fmh\_factor}) \times (d^2));$$

$$g\_gain = (\text{pow}((d \times (1 / \text{center\_size})), \text{center\_curve}) \times (1 - \text{dbunit})) + \text{dbunit};$$

$w\_gain$  = Skalierungsfaktor für den W-Kanal

$g\_gain$  = Skalierungsfaktor für die Schallschnellekanäle X,Y,Z,...

$\text{center\_size}$  = Größe der Ambisonics-Kugel in lokalen Einheiten im Koordinatensystem

$\text{fmh\_factor}$  = Ambisonics-Wichtung (Furse-Malham)

$d$  = Distanz zum Ursprung

$\text{center\_curve}$  = Abschwächungskurve (0 = keine Abschwächung, 1 = lineare Abschwächung)

$\text{dbunit}$  = Abschwächungsfaktor in Dezibel



## Raumsimulation

Wie ich bereits in Kapitel 3 (Teil 1) erläutert habe, ist das zentrale Element, welches architektonische Strukturen akustisch erfahrbar macht, der Nachhall. Nachhall ist die Antwort, die die Oberflächen einer geschlossenen Struktur zurückgeben, wenn man in deren Inneren ein Signal erzeugt. Nachhall kann dominant sein und als eigenständiges akustisches Phänomen wahrgenommen werden, wie dies z.B. in einer Kathedrale passiert. Er kann aber auch so kurz sein, dass er für uns Menschen mit dem direkten Signal verschmilzt. In der Akustik unterscheidet man zwischen diffusem Nachhall und Erstreflexionen. Schall wird an harten Oberflächen reflektiert, wobei der Ausgangswinkel der Reflexion bei glatten Oberflächen dem Eingangswinkel entspricht, bei zunehmender Rauheit von Oberflächen aber diffuser wird. In einem geschlossenen Raum werden diese ersten Reflexionen immer weiter an neuen Oberflächen reflektiert, wobei diese durch verschiedene Verzögerungen und die exponentielle Vermehrung der Reflexionen „verschmiert“ werden. Das Resultat ist diffuser Nachhall.

Die ersten Reflexionen, die in einer geschlossenen Struktur auf den Besucher zurückgeworfen werden, sind noch klarer als Reflexionen erkennbar als der diffuse Nachhall und dienen dem Besucher gleichzeitig als Navigationshilfe. So wird z.B. ein Klatschen in einem Raum von allen Wänden Erstreflexionen zurückwerfen, deren klangliche Eigenschaften Auskunft geben über Entfernung und Oberflächenbeschaffenheit der Wände. Dabei werden die Reflexionen von einer näherliegenden Wand unser Ohr früher erreichen als die einer weiter entfernten. Unser Gehirn kann diese Laufzeitunterschiede auswerten und mittels unseres binauralen Hörvermögens ein räumliches Abbild unserer Umgebung auf Basis akustischer Daten errechnen. Der diffuse Nachhall wiederum ist dieser Fähigkeit eigentlich abträglich, da er die Verständlichkeit von Signalen erschwert. Er weist kaum direktionale Informationen auf und kann bei anhaltender Beschallung Direktschall und Erstreflexionen überlagern. Damit ist diffuser Nachhall zwar ein Hindernis für die räumliche Orientierung, enthält aber andere Informationen über die Räumlichkeit. Mit ihm kann das Volumen eines Raumes erhört werden, überdies wie stark er mit Menschen oder Gegenständen gefüllt ist,

sowie die Beschaffenheit seiner Oberflächen. Beim Hören architektonischer Strukturen kommen nicht nur die navigatorischen Eigenschaften in der Dreidimensionalität zum tragen, sondern auch die Klangfarbe dieses diffusen Nachhalls. (Vgl. Teil 1, Kap. 3)

## 7.1 Kurze Geschichte des künstlichen Nachhalls

Heutzutage ist künstlicher Nachhall in klangbasierten Medien allgegenwärtig. Simulationen von Nachhall hatten aber bereits eine lange Tradition im 20. Jahrhundert. Verbreitung gefunden haben sie in erster Linie durch das Aufkommen von Tonträgern zur Musikreproduktion. Fortschritte in der Studioteknik haben sehr bald gezeigt, dass bei der Aufnahme von Musik ein möglichst „trockener“ Klang am meisten Flexibilität zur Nachbearbeitung im Studio bietet. Gleichzeitig ist der gänzlich „raum-lose“ Klang von aufgenommenen Instrumenten steril und macht Spielfehler hörbar. Schon früh wurden deswegen künstliche Hallgeräte entwickelt.

Alles begann mit der Möglichkeit, mit elektrischen Mitteln Musik für die spätere Wiedergabe aufzunehmen.<sup>1</sup> Frühe Tonaufnahmen vor der Erfindung des elektromagnetischen Tonbandes waren sogenannte „Direct-To-Disc“-Aufnahmen, bei denen die Musik live eingespielt und nach Durchlauf einer Mischkonsole direkt in die Oberfläche einer Master-Schallplatte geschnitten wurde. Dabei mussten Musiker und Tontechniker synchron arbeiten und der gesamte Produktionsprozess live durchgeführt werden – eine Methode, die keinerlei Möglichkeit bot, nachträglich Spielfehler zu korrigieren oder tontechnische Anpassungen vorzunehmen. Mit dem Aufkommen der Bandmaschinen ab Mitte der 1930er Jahre schließlich konnten Aufnahme-, Mischungs- und Endfertigungsprozesse entkoppelt werden. (Schoenherr, 2005) Multi-Track-Recording und Fortschritte in der Mikrophonierungstechnik erlaubten es plötzlich, Instrumente getrennt von einander aufzunehmen und zu bearbeiten. Schallgedämmte Aufnahmekabinen hielten in den Studios Einzug. Aber die Suche nach einer sauberen und möglichst „trockenen“ Aufnahme für maximalen Bearbeitungsspielraum führte auch zum Verlust des „Raumes“ auf der Aufnahme – Raum, der die einzelnen Stimmen einer Band oder eines Orchesters zusammenhält. Somit wurden bald Techniken zur Erzeugung künstlichen Nachhalls zum essentiellen Bestandteil jedes Tonstudios. Auch der Tontechniker erlebte eine Aufwertung als die Produktion eines Albums zum kreativen Akt wurde. Der Fokus in der Entwicklung von Hallgeräten lag eindeutig bei der kommerziellen Anwendung in der Musik, weswegen eine akkurate Raumsimulation kein primäres Ziel darstellte. Nachhall, der als Ersatz für den verlorengegangenen Klang des Aufnahme- und Wiedergaberaumes eingeführt wurde, verselbständigte sich und folgte dabei nicht immer realistischen Parametern. Die Simulation künstlicher Räume war

---

<sup>1</sup> Obwohl im 19. Jhdt. neben dem Edison-Wachszylinder noch einige andere Methoden zur Aufnahme und Wiedergabe von Musik erfunden worden waren, entwickelte sich eine Musikindustrie erst mit der Verbreitung der widerstandsfähigeren Schallplatte.



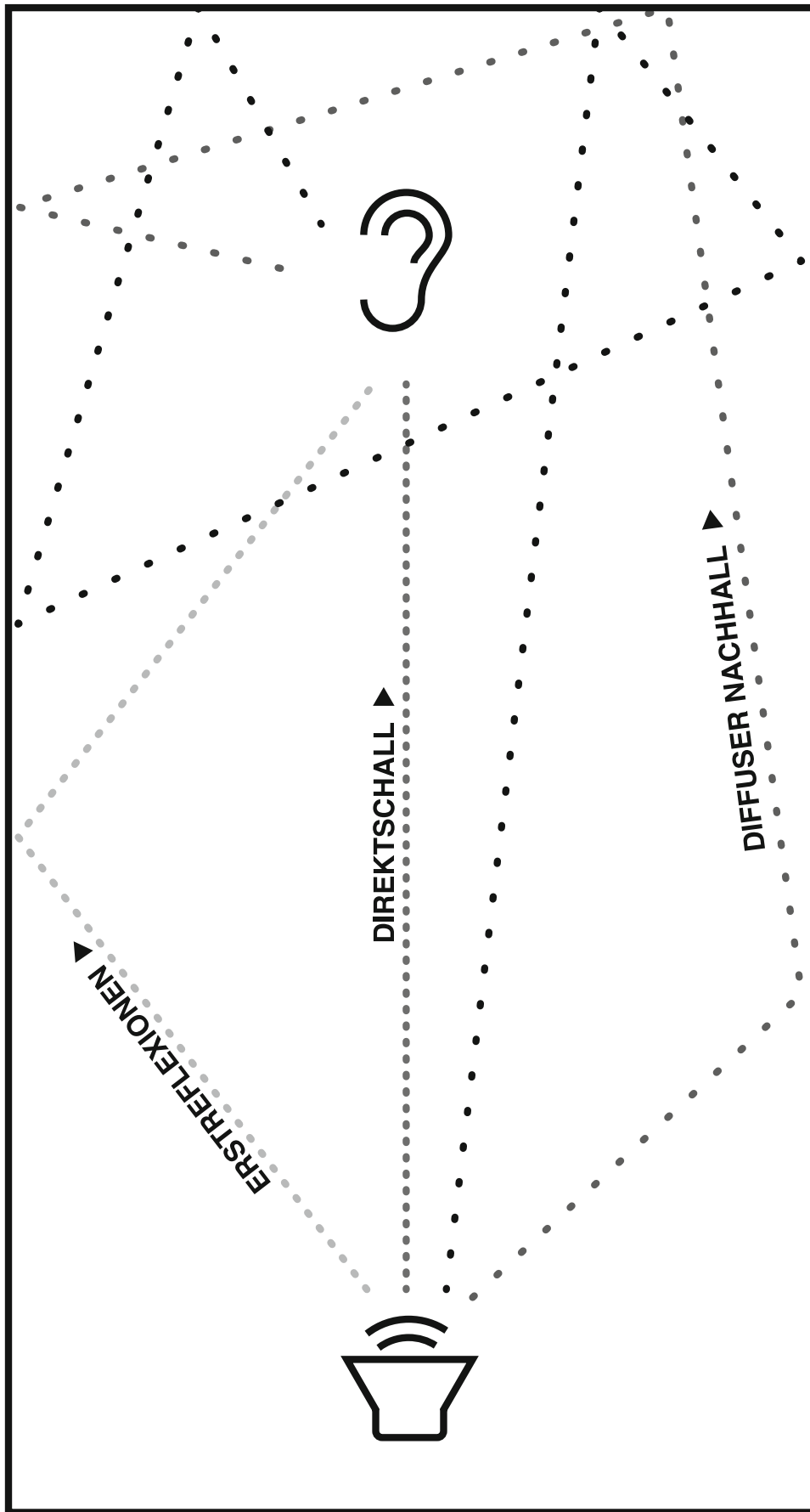


Abbildung 7.1: Direktschall, Erstreflexionen und diffuser Nachhall.

lange Zeit nur von Bedeutung für eine kleine Gruppe von Personen mit akademischen Interessen.

Frühe Reverb-Effekte benutzten raffinierte Mittel, um die diffuse Zirkulation von Reflektionen im Raum zu simulieren. Die wahrscheinlich älteste Methode, um künstliches Reverb zu gewinnen, stellen sogenannte „Echo Chambers“ dar. Dabei handelte es sich um Räume innerhalb des Studio-Komplexes, die speziell eingerichtet wurden, um einen angenehmen Nachhall zu haben. Über Kabelleitungen konnte ein Tontechniker vom Mischpult aus Klang zum Hallraum schicken, wo dieser über Lautsprecher wiedergegeben wurde. Ein (Satz) Mikrophon(e) nahm das verhallte Signal auf und speiste es zurück ins Mischpult. Manche dieser Echo Chambers hatten auch verschiebbare Wände, um die Hallcharakteristik zu verändern. Diese Technik lieferte je nach Ausgestaltung des Hallraumes exzellente Ergebnisse, war aber äußerst unflexibel und vor allem teuer. 1939 erfand schließlich Laurens Hammond, Entwickler der nach ihm benannten elektromechanischen Orgel, den Federhall. (Hammond, 1939) Dabei wird mit einem Klang eine locker gespannte mechanische Feder in Schwingung versetzt. Ein Kontaktmikrophon nimmt das Signal nach Durchlauf am Ende der Feder wieder auf. Federhall ist eine günstige Methode, um Reverb zu erzeugen und kommt auch heute noch zum Beispiel in Verstärkern für elektrische Gitarren zum Einsatz. Der resultierende Hall kann allerdings metallisch und unnatürlich klingen, weswegen manche Hersteller bewusst Unterbrechungen und Dellen in die Federn einarbeiten, um die Eigenresonanz der Feder zu reduzieren, so z.B. im berühmten AKG BX25, einem der beliebten Federhall-Geräte der 60er Jahre. Eine weitere Erfindung, die auf einem ähnlichen Prinzip wie der Federhall basiert, aber einen weitaus dichteren und realistischeren Nachhall liefern kann als dieser, ist der Plattenhall. Beim Plattenhall kommt eine abgehängte Metallplatte zum Einsatz, die zum Schwingen gebracht und dann mikrophoniert wird. Er wurde 1954 von dem deutschen Techniker Dr. Walter Kuhl entwickelt. Das erste Gerät dieser Art war das EMT 140. (Kühne, S.22) Dem Plattenhall folgte mit dem Folienhall noch eine weitere Evolutionsstufe in den 1970ern, dann allerdings ebte die Entwicklung dieser analogen Verfahren zur Hallerzeugung ab. Das digitale Zeitalter hatte begonnen.

Der erste bekannte digitale Hallalgorithmus wurde bereits im Jahre 1961 von Manfred Schroeder für die „Bell Labs“ entwickelt. Er basiert auf vier sequentiell geschalteten Kammfiltern, deren Output in parallele Allpass-Filter gefüttert wird, die das Original-Signal frequenzabhängig unterschiedlich verzögern. (Schroeder, Logan, 1961). In den 1970ern verfeinerte James A. Moorer das Schroedersche Modell, welches zuvor eher Demonstrationszwecken in Laborsituationen gedient hatte und praktisch keine Anwendung gefunden hatte. Er erweiterte dieses um geometrisch korrekt simulierte Erstreflexionen, was den Realismus des erzeugten Halls stark steigerte, sowie um Tiefpassfilter innerhalb der Rückkopplungsschleifen der Kammfilter, womit die Absorption hoher Frequenzen in Luft und an Oberflächen simuliert wurde. (Moorer, 1979; Floisand, 2012)

Ein weiterer, weniger bekannter Hallalgorithmus von Manfred Schroeder wird in seinem Artikel „Natural Sounding Artificial Reverberation“ beschrieben. Dieser basiert auf einem fundamental anderen Prinzip. Anstatt eine Serie paralleler Allpass-Filter zu treiben, wird ein Allpass-Filter in eine doppelte Rückkopplungsschleife eingebettet. Bei jedem Durchlaufen der Schleife wird das Signal so verdoppelt. Das „Schroeder-Reverb“, mit dem er üblicherweise in Verbindung gebracht wird, weist eine konstante Dichte von Reflexionen von Beginn des Reverbs bis zu dessen Ausklang auf. Der zweite hier beschriebene Algorithmus allerdings steigert die Reflexionsdichte mit fortschreitender Zeit, was dem tatsächlichen Verhalten von Nachhall in realen Räumen entspricht. (Schroeder, 1962; Costello, 2009)

Das zugrundeliegende Prinzip dieses Algorithmus bildet auch die Basis für die Ende der 1970er aufkommenden, kommerziell erhältlichen digitalen Hallgeräte. Rückkopplungsschleifen beginnen parallele Delay-Bänke zu ersetzen, wobei in den meisten dieser Designs auch die obligaten Allpass-Filter in die Schleifen integriert wurden. Die bekanntesten dieser Systeme stellen das „EMT 250“ von Barry Blesser<sup>2</sup> und das „Lexicon 224“ von David Griesinger dar. (Blesser, Bader, 1978; Griesinger, 1989; Dattorro, 1997) Letzteres war jahrzehntlang der Standard-Nachhall in grossen Studios und galt bis ins 21. Jahrhundert als Krönung der Hallgeräte. In den 80er und 90er Jahren kam dann ein neuer Typus auf, dessen Design im Allgemeinen Miller Puckette und John Stautner zugesprochen wird. Wiewohl Michael Gerzon bereits im Jahre 1971 einen vergleichbaren Algorithmus publiziert hatte (Gerzon, 1976), fand dieser aus Gründen des mangelnden Distributionsnetzes des Magazins in den Vereinigten Staaten kaum Verbreitung (Costello, 2011). Puckettes und Stautners Algorithmus stellt somit eine Parallelentwicklung dar, die später durch Jean-Marc Jot verfeinert wurde: Rekursive Netzwerke von Delays, auf Englisch Feedback Delay Networks, die ich im folgenden mit FDN abkürze. „FDN Reverbs“ haben eine Reihe von Vorteilen gegenüber anderen Systemen. Einer dieser Vorteile ist die oben erwähnte Steigerung der Dichte mit zunehmendem Aufbau des diffusen Nachhalls. Ein anderer ist, dass sie sehr einfach in Mehrkanal-Systeme integriert werden können.<sup>3</sup>

Dies sind die bedeutendsten Methoden, um künstlich Nachhall zu erzeugen. Der Vollständigkeit halber sein noch die zweifelsfrei realistischste Methode, um Nachhall zu erzeugen, genannt: „Convolution Reverbs“. Diese unterscheiden sich von den oben genannten Methoden dadurch, dass der Hall nicht im strikten Sinne künstlich erzeugt wird. Stattdessen werden sogenannte Impulsantworten genutzt, die zuvor in real existenten Räumen aufgezeichnet werden. Hierzu wird ein Knall im Raum erzeugt und die daraus resultierenden Impulsantworten aufgenommen. Anhand

---

2 Autor von „Spaces Speak – Are you listening?“

3 Aus diesem Grund besteht das Reverb in *Sonic Manifolds* auch aus einem FDN, das ich mit einem System zur Simulation von Erstreflexionen gekoppelt für Ambisonics adaptiert habe.

des Referenzknalls und der Antworten kann mittels Faltung jedes beliebige Signal mit dem aufgezeichneten Hall versehen werden. (Griesinger, 1989) Da das Ziel bei *Sonic Manifolds* aber die Erzeugung künstlicher Räume war und Convolution Reverbs die Aufzeichnung real existenter Räume voraussetzen und darüberhinaus nicht einfach in Mehrkanal-Systeme integrierbar sind, kam diese Methode der Hallerzeugung für mich nicht in Frage. Allerdings beeinflusste diese Technik die Wahl der Klangerzeugung des Erreger-Schalls in *Sonic Manifolds*: Mit einem Impuls wird der künstliche Raum „abgetastet“. (Vgl. Kapitel 3)

## 7.2 Erstreflexionen

Erstreflexionen sind der Teil eines Reverbs, der uns die Navigation mithilfe akustischer Mittel erlaubt. Im Gegensatz zu diffusem Nachhall müssen sie also auch geometrisch korrekt lokalisiert sein. Unser Gehirn kann von ihnen auf die Geometrie des Raumes schliessen. Sie sind in fast jeder Umgebung präsent. Ausnahmen sind z.B. schalltote Räume oder große weite Wiesen. Sie werden von uns zumeist nicht bewußt wahrgenommen. Für eine realitätsnahe Simulation einer akustischen Architektur sind sie essentiell. Aus Gründen der Komplexität kommt in *Sonic Manifolds* ein „Shoebox“-Modell als Näherungsmethode zur Simulation von Raumakustik zum Einsatz. Das bedeutet, dass alle erzeugbaren Raumtypen quaderförmig sind. Die Berechnung der Positionen der Erstreflexionen ist in diesem Fall einfach: Durch Spiegelung der Position der Klangquelle an einer umgebenden Wand kann die Position einer „Spiegelquelle“ ermittelt werden. Diese wird um die Laufzeit der Distanz verzögert, die der Schall bei aktueller (angenommener) Lufttemperatur zwischen originaler Quelle, Spiegelquelle und Mikrophon (=Ursprung) zurücklegen muss. In der einfachsten Variante hat eine Schallquelle in einem „Shoebox“-Modell sechs Spiegelquellen 1. Ordnung. In weiterer Folge können Spiegelquellen höherer Ordnung berechnet werden. Diese ergeben sich aus den ganzzahligen Vielfachen der Raumdimensionen. In realen Räumen ist allerdings die Anzahl der Erstreflexionen wesentlich höher, da jede Oberfläche, die in einen rechten Winkel zur Klangquelle steht, eine derartige Erstreflexion erzeugt. Aufgrund der Rechenintensität der verwendeten trigonometrischen Funktionen, die in Echtzeit sowie auf per-Sample Basis als Signal berechnet werden (44100 mal pro Sekunde bei einer Audio-Abtastrate von 44kHz), beschränkt sich die Simulation in *Sonic Manifolds* auf Erstreflexionen der 1. Ordnung.

Die zeitlichen Verzögerungen der Erstreflexionen werden mit folgender Formel berechnet:

$$\Delta t = r / c$$

$\Delta t$  = Verzögerung in Sekunden

$r$  = Distanz in Metern

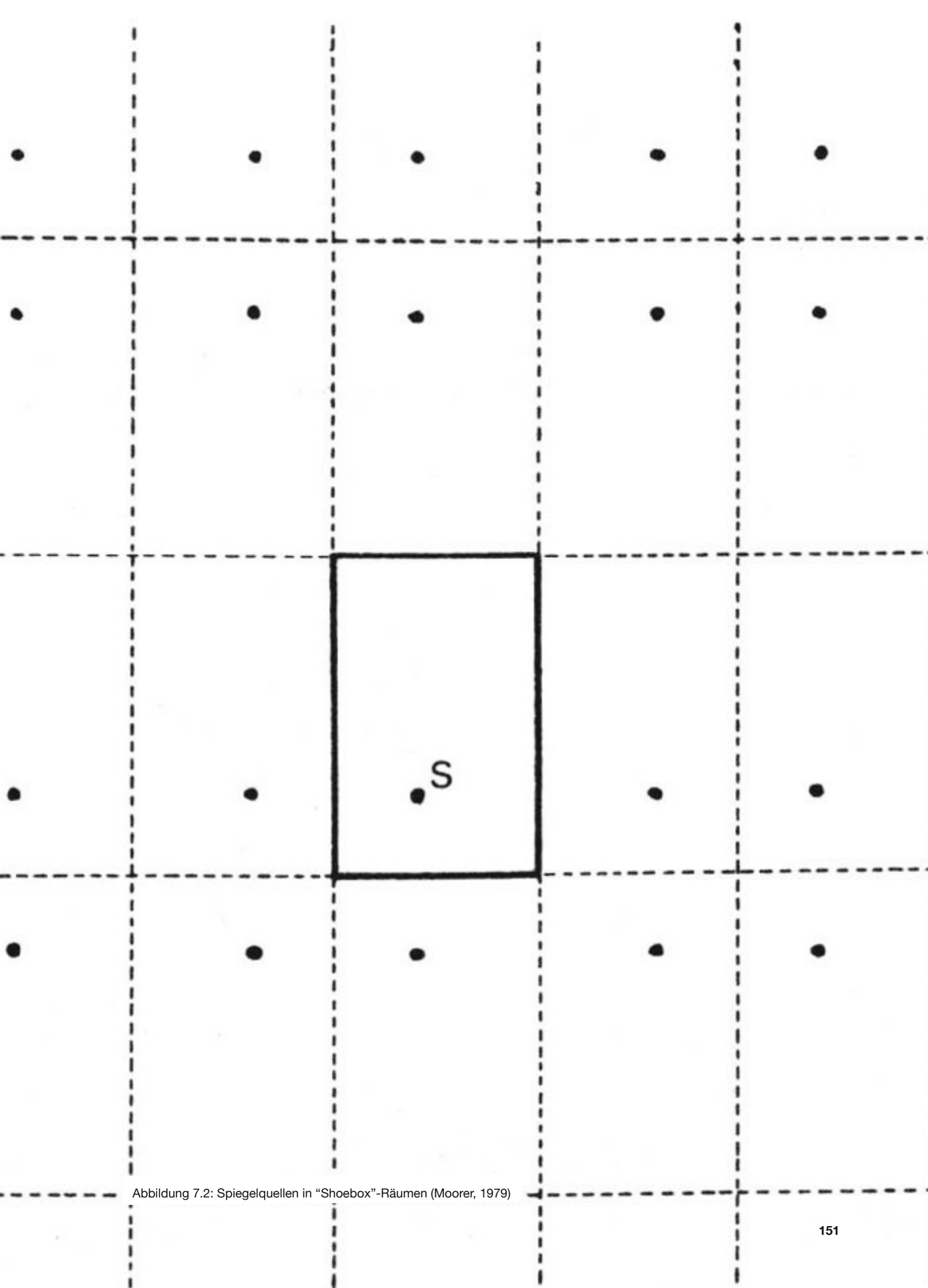


Abbildung 7.2: Spiegelquellen in "Shoebbox"-Räumen (Moorer, 1979)

$c$  = Schallgeschwindigkeit in Metern pro Sekunde

Die näherungsweise Schallgeschwindigkeit in Luft ist abhängig von der Temperatur und beträgt

$$c = 331,3 \times \sqrt{1 + \vartheta / 273,15}$$

$c$  = Schallgeschwindigkeit in Meter pro Sekunde

$\vartheta$  = Temperatur in Grad Celsius

Aufgrund praktischer Überlegungen erfolgt die mit der Distanz einhergehende Lautstärken-Abschwächung nicht natürlichen Bedingungen. Da Schall sich kugelförmig ausbreitet, sinkt der Druckpegel mit Verdopplung der Distanz auf die Hälfte ab, was einer Lautstärkenreduktion um 6dB gleichkommt. *Sonic Manifolds* simuliert jedoch künstliche Räume, die teils bis zu einhundert Meter auseinanderliegende Wände aufweisen. Im Kontrast zu kleinen Räumen, die direkt daneben liegen können, wären die Erstreflexionen also kaum hörbar oder würden im diffusen Nachhall untergehen, welcher mit zunehmendem Raumvolumen länger und dominanter wird. Deshalb habe ich mich entschlossen, die Invers Exponentielle Skalierungsmethode anzuwenden, die Jan Schacher bei der Programmierung seiner ICST Ambisonics Externalis angewandt hatte. (Schacher, Kocher, 2014) Diese erweitert die Abschwächungsformel  $1/r$  um den Abschwächungsfaktor  $a$ .

$$g = 1 / r a$$

$g$  stellt hierbei die Lautstärke (linear,  $1$  = ursprüngliche) bei der Entfernung  $r$  dar.

Bei einer Distanz von  $2$  und einem Abschwächungsfaktor von  $1$  ergibt sich daraus die „realistische“ Abschwächung von  $0.5$ , während bei  $a=2$  die neue Lautstärke quadratisch reduziert würde auf  $0.25$ , sowie bei  $a=0,5$  ein Ergebnis von  $0,707$  (das entspricht  $\sqrt{2}$ ) vorliegen würde. Der Abschwächungsfaktor erlaubt somit individuelle Anpassungen bei verschiedenen Raumtypen, je nach künstlerischem Erfordernis.

In real existierenden Räumen werden die Erstreflexionen weiters frequenzabhängig abgeschwächt. Das kommt einerseits von den reflektierenden Oberflächen, die je nach Material bestimmte Frequenzen schwächer zurückwerfen als andere. Zudem kommt es in der Luft zu Absorptionen durch die Atmosphäre, die abhängig von Frequenzbereich, Temperatur und Feuchtigkeit sind. Diese Faktoren werden in *Sonic Manifolds* näherungsweise durch einen distanzbasierten Tiefpassfilter simuliert. Die Distanz der Spiegelquelle bestimmt dabei die Filterfrequenz, die sich zwischen  $10\text{kHz}$  (bei  $1.7$  m Distanz) und  $1\text{kHz}$  (ab  $20$  m Distanz) bewegt. Die Absorption durch die reflektierenden Oberflächen wird durch einen Bandpass-Filter simuliert, dessen Charakteristik in den Szenen-Pre-sets festgelegt werden kann. Die Werte lehnen sich dabei an den Absorptions-Koeffizienten nach

Sabine an und sind einer Liste im Internet entnommen. (Sengpiel, 2003) Aufgrund der Komplexität der Beschreibung der Szenerien kommt aber in der gegenwärtigen Fassung von *Sonic Manifolds* nur ein Materialtyp pro Raumzone zum Einsatz.

Zusätzlich zu dieser distanzbasierten Lautstärke-Skalierung existiert noch ein weiterer szenengebundener Skalierungsfaktor, mit dem die Erstreflexionen unabhängig von ihrer Entfernung reduziert oder verstärkt werden können. Hiermit ist es möglich, die Simulation der Erstreflexionen auch vollständig zu unterbinden. Dies soll ermöglichen, in Szenen Bereiche zu definieren, die in der Wahrnehmung als geschlossene Volumina erscheinen sollen, die zwar durch „Hineingreifen“ zum Schwingen gebracht werden können, aber nie „betreten“. Das Resultat ist zwar physikalisch inkorrekt, ist aber als Näherung akzeptabel, um gewisse Effekte zu erzielen wie beispielsweise den einer Blechtonne, die zum Schwingen gebracht wird.

### 7.3 Diffuser Nachhall

Diffusen Nachhall mit elektronischen Mitteln zu simulieren ist eine komplexe Aufgabe. Während bei den Erstreflexionen Direktionalität und Erkennbarkeit ausschlaggebend dafür sind, dass unser Gehirn sie richtig verarbeiten kann, ist bei diffusem Nachhall eher die Klangfarbe und die „Form“ entscheidend, um als natürlich wahrgenommen zu werden. Diffuser Nachhall entsteht durch das fortschreitende Reflektieren des Direktschalls bis zur völligen „Verwischung“ und zeichnet sich durch eine immens hohe Dichte an Reflexionen aus, die als solche nicht mehr einzeln in der „Hallfahne“ identifiziert werden können. Die Klangfarbe eines idealisierten diffusen Nachhalls ist dabei neutral und nahe am weißen Rauschen. Entscheidend für den natürlichen Charakter sind außerdem der Einsatz und das Abklingen des Halls. So betont etwa David Griesinger, dass die ersten 200 Millisekunden des „Late Reverb“ entscheidend für die wahrgenommene Klangqualität des Halls sind. (Griesinger, 1989)

Manfred Schroeder gilt als Vater des elektronischen Reverbs. Zwar stehen inzwischen weitaus fortgeschrittenere Methoden zur Erzeugung von Nachhall zur Verfügung, als die von ihm entwickelten, allerdings hat er mit seinen Forschungen den Grundstein für eine Hall-Simulation mit elektronischen Mitteln gelegt. Noch heute sind klassische „Schroeder Reverbs“ weit verbreitet und oft im Einsatz. Als Pionier der elektronischen Hallerzeugung formulierte er bereits 1961 in seinem Artikel „Colorless“ Artificial Reverberation folgende Anforderungspunkte als Bedingung für einen natürlich klingenden Nachhall, die heute noch Gültigkeit haben: „1) *The frequency response must be flat when measured with narrow bands of noise, the bandwidth corresponding to that of the transients in the sound to be reverberated.* 2) *The normal modes of the reverberator must overlap and cover the entire audio frequency range.* 3) *The reverberation times of the individual modes must be equal or nearly equal so*



*that different frequency components of the sound decay with equal rates. 4) The echo density a short interval after shock excitation must be high enough so that individual echos are not resolved by the ear. 5) The echo response must be free from periodicities (flutter echos). 6) The amplitude-frequency response must not exhibit any apparent periodicities. Periodic or comb-like frequency responses produce an unpleasant hollow, reedy, or metallic sound quality and give the impression that the sound is transmitted through a hollow tube or barrel.*“ (Schroeder, Logan, 1961, S.211) Als projektspezifische Anforderung kommt bei *Sonic Manifolds* noch die Adaption auf die Mehrkanal-Situation hinzu.

### 7.3.1 Dichte & Klangfarbe

Um diese Anforderungen umzusetzen, greife ich auf einen bewährten Algorithmus zurück. Dabei handelt es sich um ein FDN-Reverb nach Ideen von Jean-Marc Jot. FDN-Reverbs habe ich bereits in Kapitel 7.1 erwähnt. Miller Puckette beschreibt in seinem Buch „Theory and Technique Of Electronic Music“ ein 4-Kanal Reverb auf Basis eines 4×4 Feedback Delay Networks so: *„The central idea is to idealize any room (or other reverberant space) as a collection of parallel delay lines that models the memory of the air inside the room. At each point on the walls of the room, many straight-line paths terminate, each carrying sound to that point; the sound then reflects into many other paths, each one originating at that point, and leading eventually to some other point on a wall. Although the wall (and the air we passed through to get to the wall) absorbs some of the sound, some portion of the incident power is reflected and makes it to another wall. If most of the energy recirculates, the room reverberates for a long time; if all of it does, the reverberation lasts forever. If at any frequency the walls reflect more energy overall than they receive, the sound will feed back unstably; this never happens in real rooms (conservation of energy prevents it), but it can happen in an artificial reverberator if it is not designed correctly.*“ (Puckette, 2007, S.193)

Anstatt also eine Bank von getrennten, parallelen Delay-Schleifen zu betreiben, wie dies bei den klassischen Schroeder und Moorer Reverbs der Fall war, kommt ein rekursives Netzwerk von Delays zum Einsatz. Die verzögerten Outputs werden zurück zu den Inputs geleitet. Essentieller Bestandteil von Puckettes Algorithmus, der zuvor bereits von Puckette und John Stautner in ihrem Artikel „Designing Multi-Channel Reverberators“ aus dem Jahre 1982 beschrieben wurde (Stautner, Puckette, 1982), ist eine Streumatrix, die den Output jeder „Delay-Line“ gleichmäßig an die anderen Delay-Lines verteilt. Laut Jean-Marc Jot, einem Pionier der DSP-Programmierung von FDN-Reverbs, kann dies eine „Householder“-Matrix, eine Zyklische Matrix oder eine „Hadamard“-Matrix sein. (Jot, 1997, S3)

### 7.3.2 FDN

Die konkrete Implementierung des in Blunzgrössl zum Einsatz kommenden FDN-Reverbs nach Jot basiert auf dem für das Jamoma DSP-Paket konzipierte Reverb von Nils Peters. (Peters, 2012). Der Algorithmus wurde von mir nahezu unverändert übernommen und erfüllt die von Schroeder aufgestellten Regeln. Allerdings erlauben uns geringfügige Eingriffe, diese auch gezielt zu verletzen, um besondere Effekte innerhalb der künstlichen Szenen von *Sonic Manifolds* zu erzielen bzw. auch mit dem naturalistischen Eindruck zu spielen und Verfremdungseffekte bewirken zu können. Als Streumatrix kommt in dem Algorithmus eine 16×16 Hadamard-Matrix zum Einsatz. Hadamard-Matrizen sind nach dem französischen Mathematiker Jacques Hadamard benannt, der sie im Jahre 1893 entdeckte, wurden allerdings ebenfalls schon 1867 vom englischen Mathematiker James Joseph Sylvester unter dem Bezeichnung „anallagmatic pavement“ erfunden. (Weisstein)

Dabei handelt es sich um sogenannte unitäre Matrizen, die verlustfrei und orthogonal sind. Das bedeutet, dass die Anzahl der Zeilen- und Spaltenvektoren identisch und die Matrix jederzeit umkehrbar ist. Weil alle Koeffizienten die gleiche Magnitude (1 bzw. -1) haben, erreicht man eine maximal Streuung, da jeder Koeffizient gleichmäßig auf die anderen aufgeteilt wird. So werden die Signale, nachdem sie die Delay-Lines verlassen haben, mittels der Hadamard-Matrix gleichmäßig auf alle jeweils anderen Delay-Lines verteilt, wodurch der Output jeder Delay Line pro Durchlauf versechzehnfacht wird. Nach nur 6 Durchläufen durch das FDN kann so die Zahl der Reflexionen auf 16 Millionen erhöht werden.

$$16 \times 16 \rightarrow 256 \times 16 \rightarrow 4096 \times 16 \rightarrow 65536 \times 16 \rightarrow 1048576 \times 16 \rightarrow 16777216 \times 16 \rightarrow \text{etc}$$

Dabei ist allerdings zu beachten, dass diese Steigerung nicht in „Blöcken“ erfolgt, sondern die Anzahl der Reflexionen nach und nach anwächst, da bei der Wahl eines zeitlich größeren Bereichs von Delay-Zeiten die kürzeste Delay-Line längst abgearbeitet sein kann, während die längste noch durchlaufen wird. Zu beachten ist deshalb auch, dass die Wichtung der Verzögerungszeiten der einzelnen Delay-Lines zueinander zur Form des diffusen Nachhalls beiträgt. Die oben angeführte Reihe ist deshalb als Vereinfachung zu verstehen.

Zu guter Letzt findet sich innerhalb jeder Delay-Line noch jeweils ein Filter, um die Klangfarbe zu regulieren sowie einen Verstärker, mit dem die Rückkoppelungen abgeschwächt werden, um die gewünschte T60 Dauer zu erreichen. Wie ich bereits in Kapitel 3 (Teil 1) beschrieben habe, ist die T60 Dauer die Zeit, die die Hallfahne in Anspruch nimmt, um 60 dB gegenüber dem Ursprungspegel zu verlieren. Der Abschwächungsfaktor wird mit folgender Formel, abhängig von der Verzögerungszeit der jeweiligen Delay-Line, berechnet und lautet  $10^{(-3 \times \text{delaylength}/t60)}$ .

Die Punkte (1) bis (3) von Schroeders Anforderungsliste für künstlichen Nachhall werden in einem FDN schon alleine dadurch erfüllt, dass die einzelnen Feedback-Schleifen durch die Matrix miteinander verbunden sind. Punkt (4) wird ebenfalls leicht dadurch erreicht, dass der Output jeder Feedback-Schleife an alle anderen Schleifen weitergegeben wird. Dadurch kommt es beim Durchlauf jeder Schleife zu einer Multiplikation mit 16. Da die Schleifen verschiedene Verzögerungsdauern aufweisen, erfolgt die Vervielfachung der Reflexionen zeitlich gesehen nicht in periodischen Blöcken, wenn man die Delay Zeiten exponentiell staffelt und ganzzahlige vielfache Werte meidet. So können frequenzabhängige Färbungen des Nachhalls vermieden werden. Während bei einer isolierten Feedback-Schleife bei ausreichendem Pegel früher oder später exakt die Frequenz verstärkt wird, die der Wellenlänge der Delay-Zeit entspricht (nach der Formel  $f = 1 / t$ , also zb. 50 Hz bei einem Delay von 20 ms), wird dem im FDN durch die Verteilung an andere Delays mit unterschiedlichen Delay-Zeiten entgegengearbeitet. Spitzen im Frequenzspektrum können vermieden werden, womit auch die Punkte (5) und (6) erfüllt sind.

In der Praxis zeigt sich jedoch, dass das alleine nicht ausreicht, um einen natürlichen Nachhall zu erzeugen. Reale Räume weisen keinen derart „sauberen“ diffusen Nachhall auf, sondern haben Schwankungen, weshalb das Reverb die Möglichkeit bietet, sowohl die Delay-Zeiten als auch die Verstärkung jedes Feedback-Kanals zu modulieren. Wenn der Modulationsfaktor gering ist, trägt das zusätzlich dazu bei, einzelne Resonanz-Spitzen zu eliminieren. Ist die Modulation der Delay-Zeiten sehr stark, kommt es allerdings zu hörbaren Frequenzmodulationen, wodurch die den Delay-Zeiten entsprechenden Frequenzen doch herausstechen, während eine starke Modulation der Amplituden das resultierende Reverb ungleichmäßig schwanken lässt. Beide Effekte sind zwar nachteilig für das Ziel der Erzeugung eines natürlich klingenden Nachhalls, können aber bei *Sonic Manifolds* genutzt werden. So kann das starke Schwanken der Frequenzen akustische Ähnlichkeiten mit der Eigenresonanz einer in Schwingung versetzten Blechtonne haben. Um wiederum die Darstellung irrealer Volumina zu ermöglichen, erlaubt hier *Sonic Manifolds* ebenfalls, die Delay-Zeiten der einzelnen FDN-Spuren entgegen den Regeln zur Erzielung eines natürlichen Nachhalls periodisch zu gestalten, um Resonanzspitzen doch noch zu ermöglichen.

### 7.3.3 Räumlichkeit

Als besondere Anforderung kommt bei *Sonic Manifolds* die Einbindung des Reverb-Algorithmus in Ambisonics hinzu. Bereits 1962 hat Manfred Schroeder ein Mehrkanal-Reverb beschrieben (er bezeichnet dieses als „Ambiophonic“), sein Ansatz beschränkte sich aber auf die Anwendung auf Monosignale. (Schroeder, 1962, S.223) Um eine wirklich räumliche Einbindung des Schallfeldes in den diffusen Nachhall zu erzielen, müssen in Ambisonics aber zumindest 4 Input-Kanäle (für

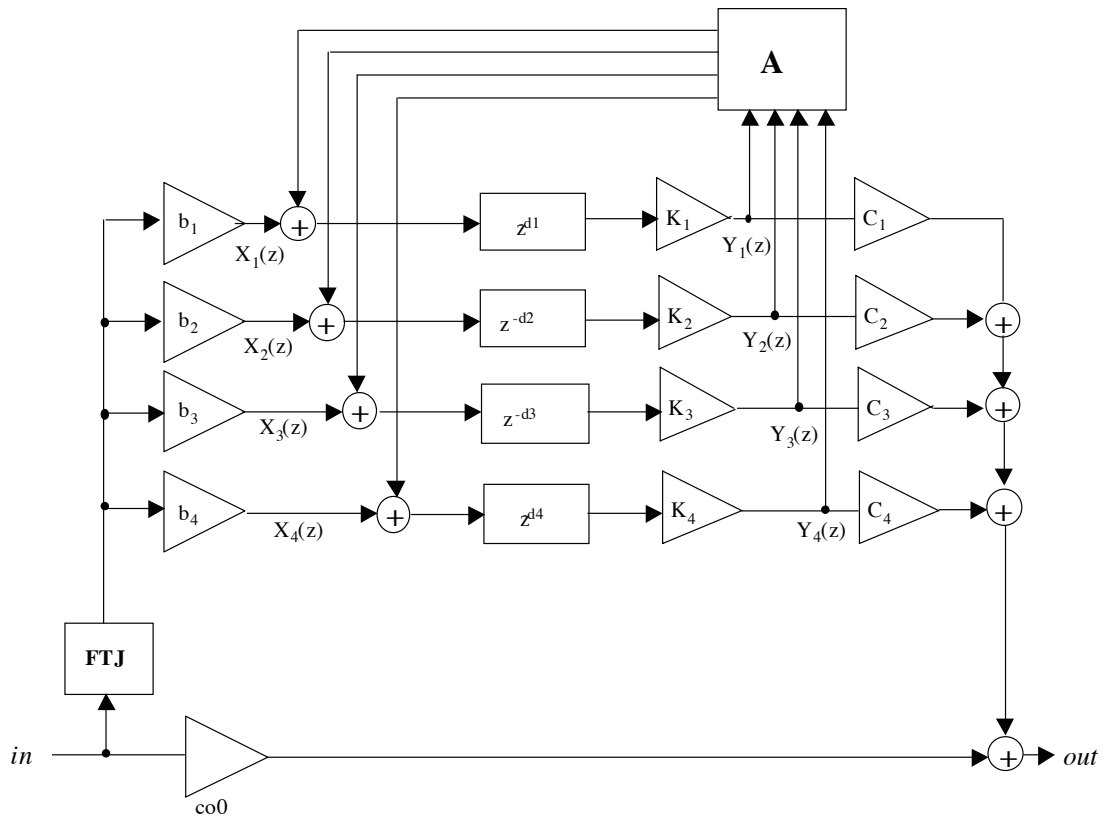


Abbildung 7.3: 4x4 FDN-Reverb nach Jot. (Beltránd et.al.,2002)

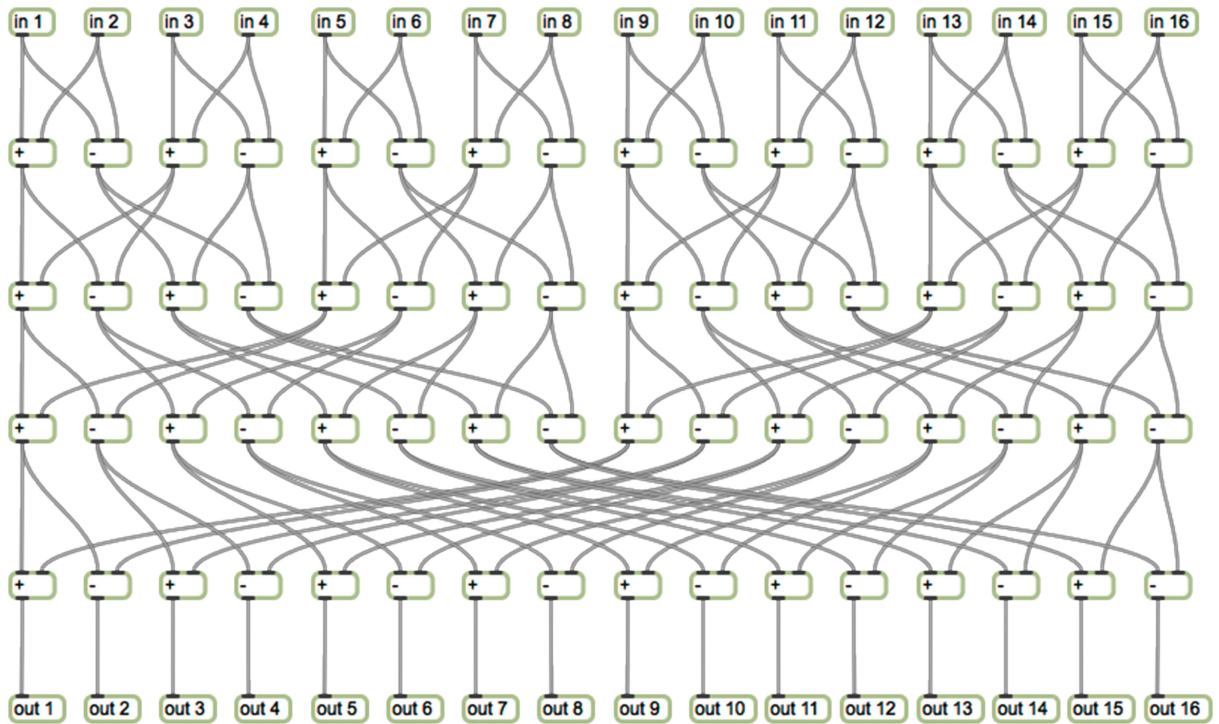


Abbildung 7.4: "Butterfly Processing" einer 16x16 Hadamard Matrix. (Peters, 2012)

Ambisonics 1. Ordnung) vorhanden sein. Der Einfachheit halber könnte man natürlich auf Basis der Schalldruckkomponente  $W$  ein Reverb mit Mono-Input und Mehrkanal-Output speisen, um die Illusion eines räumlichen diffusen Nachhalls zu erzeugen. Ich wollte jedoch einen Schritt weiter gehen, und die Druckschwankungen in einem Raum auf realistischere Art und Weise simulieren. Hierfür muss das Mehrkanal-Reverb auch mehrere Kanäle als Input erlauben.

Obwohl eine schier endlos anmutende Anzahl an (Late) Reverb-Effekten als Software-Implementierungen im Internet verfügbar ist, sind diese fast ausschließlich für Mono- oder Stereo-Anwendungen konzipiert. Die wenigen Ausnahmen, die Surround-Nachhall simulieren, sind zwar auf die Ausgabe von Surround spezialisiert, akzeptieren aber nur Mono- oder Stereo-Eingangssignale. Echte Surround-Reverbs, die ein glaubwürdiges diffuses Reverb auf Basis eines Mehrkanal-Inputs erzeugen, sind aber faktisch inexistent. Der Grund hierfür ist wohl vor allem in der gängigen Praxis der Musikproduktion zu suchen, mit Monoquellen zu arbeiten, die per „Hand“ in der Endmischung spatialisiert werden. Die mangelnde Nachfrage bestimmt also das Angebot. Technisch gesehen ist es auch wesentlich einfacher, ein Signal mit beschränkter Rauminformation um künstlichen räumlichen Nachhall zu erweitern, als die Rauminformation in einem vorhandenen räumlichen Signal zu verändern.

Diffuser Nachhall zeichnet sich durch Richtungslosigkeit aus. Er entsteht überhaupt erst durch das Diffundieren gerichteten Schalls im Raum, jedoch bedeutet dies nicht, dass er Mono-Charakteristika aufweist. Selbstverständlich gibt es auch beim diffusen Nachhall Druckschwankungen im Raum. Das heißt, die Simulation muss eine Räumlichkeit bewahren, jedoch gleichzeitig direktionale Informationen verwischen. Bei der Anwendung von Reverb-Effekten auf herkömmliche Stereo-Signale oder jedes andere Format, welches nicht das ganze Schallfeld abbildet, kann die Diffusion des Schalls nur zwischen den beiden Lautsprechern erfolgen. Die tatsächliche Diffusion in einem realen Raum erfolgt aber nicht nur zwischen zwei Fixpunkten (den Lautsprechern), sondern im ganzen Volumen. Es musste also eine Lösung gefunden werden, die erlaubt, das diffuse Reverb im Ambisonics-Schallfeld zu simulieren.

Einer der Gründe für die Wahl von Nils Peters FDN Reverb-Algorithmus war, dass er bereits darauf ausgelegt ist, Mehrkanal-Input zu verarbeiten. Allerdings wurde er für die Anwendung auf diskrete (lautsprecherabhängige) Kanäle konzipiert und nicht für die Anwendung auf Ambisonics-kodierte Schallfelder. Wie ich im Kapitel 6.2 beschrieben habe, können klangverändernde Effekte nicht ohne weiteres auf die Schallschnelle-Komponenten angewendet werden, ohne ungewünschte Verzerrungen des Schallfeldes zum Resultat zu haben. Es sollte aber das Hin-und-her-Wandern des Schalls zwischen den verschiedenen Quadranten des Schallfelds glaubwürdig simuliert werden. Wie sollte also die Einbindung eines Algorithmus, der für die Anwendung auf diskrete Kanäle gedacht

war, in Ambisonics bewerkstelligt werden? Der entscheidende Hinweis zur Lösung dieses Problems findet sich im Paper „Adapting Artificial Reverberation Architectures For B-Format Signal Processing“ von Sean Costello und Joseph Anderson. Diese schlagen die Umwandlung des Ambisonics B-Format Schallfeldes in das üblicherweise bei der Aufnahme mittels Soundfield-Mikrophon verwendete A-Format vor. (Costello, Anderson, 2009)

### 7.3.4 Einbettung FDN in Ambisonics

Der im Amisonics B-Format kodierte Direktschall wird also in das A-Format konvertiert und von dort in das FDN-Reverb weitergereicht. Ein Pre-Delay verzögert das diffuse Reverb noch weiter, um das verzögerte Einsetzen nach den Erstreflexionen in realen Räumen zu simulieren. Während Nils Peters in seiner Implementierung eine Formel verwendet, um das Pre-Delay zu errechnen, nutzt *Sonic Manifolds* manuell festgelegte Predelay-Zeiten. Der Grund hierfür ist wiederum im Anspruch des Projektes zu finden, zwar natürliche Klangbedingungen zu ermöglichen, sie aber gleichzeitig umgehen zu können, um den natürlichen Räumen jederzeit – physikalisch gesehen – surreale Pendanten entgegen stellen zu können.

Costello und Anderson empfehlen bei der Umwandlung ins A-Format eine W-Wichtung von  $1/\sqrt{3}$  und bezeichnen dies als W-Normalisiertes B-Format. *„Additionally, we'll need to observe that the canonical scaling on the W channel of B-format is  $1/\sqrt{2}$ , which is not ideal in this instance. This canonical scaling is usually referred to as an engineering consideration. For horizontal only soundfields (pantophonic), the scaling is a suitable choice, and when one considers B-format was initially intended as a successor of the 'quad' formats of the 1970s, this choice is credible. However, for a full periphonic soundfield (3D), a normalised scaling on W of  $1/\sqrt{3}$  is more appropriate.“* (Costello, Anderson, 2009, S.2)

Das in *Sonic Manifolds* zum Einsatz kommende Reverb nutzt intern das ebenfalls von Costello und Anderson vorgeschlagene W-normalisierte A-Format. Das bedeutet in der Praxis, dass die Schalldruckkomponente W gegenüber der üblichen „Furse-Malham“ Wichtung von  $1/\sqrt{2}$  abgeschwächt wird, wodurch das räumliche Bild „breiter“ wird. Somit finden verstärkt die Schallschnelle-Komponenten Eingang in die Erzeugung des diffusen Nachhalls, was den räumlichen Eindruck verstärkt. Die Formeln hierzu lauten:

Konversion von B-Format → W-Normalisiertes A-Format

$$LF = 1/\sqrt{6} \times W + 0.5 \times X + 0.5 \times Y + 0.5 \times Z$$

$$RF = 1/\sqrt{6} \times W + 0.5 \times X - 0.5 \times Y - 0.5 \times Z$$

$$LB = 1/\sqrt{6} \times W - 0.5 \times X + 0.5 \times Y - 0.5 \times Z$$

$$RB = 1/\sqrt{6} \times W - 0.5 \times X - 0.5 \times Y + 0.5 \times Z$$

Konversion von W-Normalisiertes A-Format → B-Format

$$W = \sqrt{3/8} \times (LF + RF + LB + RB)$$

$$X = 0.5 \times ((LF - LB) + (RF - RB))$$

$$Y = 0.5 \times ((LF - RB) - (RF - LB))$$

$$Z = 0.5 \times ((LF - LB) + (RB - RF))$$

(Costello, Anderson, 2009, S.2)



## Perspektiven

Der architektonische Raum ist in gewisser Weise ein schizophrener, denn er wird aus der abstrakten, kartesisch gemessenen Fläche geboren, aber von innen heraus erfahren. Während die Planer den Quadratmeter heranziehen, ihn zerteilen und gliedern, um zu einer Struktur zu gelangen, die als Behältnis, Ort oder Behausung dienen soll, ist es der erfahrbare Raum, der für die Bewohner einer solchen Struktur zählt. Freilich bieten etwa numerische Angaben zu Grundfläche und Zimmeranzahl ein gewisses Verständnis davon, wie der Raum beschaffen ist – und darüber hinaus Aufschluss darüber, wie teuer dieser erkaufte werden muss –, aber in der Praxis sagt er nur wenig darüber aus, wie sehr sich ein Raum für dessen potentielle Bewohner eignet. Der Wert von Raum erschliesst sich vielmehr aus topologischen Faktoren, die gemeinhin erst durch den Standpunkt bzw. die Perspektive zu ihrer Bedeutung gelangen. Wege, Relationen und Konstellationen sind also meist von grösserem Gewicht als die Eckdaten, die etwa aus einer Wohnungsanzeige gelesen werden können. Raum erschliesst sich somit für die Menschen immer erst von innen.

Natürlich ist auch im architektonischen Entwurfsprozess eine Ausgestaltung des Raumes von innen heraus ein gängiger Entwurfsansatz, allerdings ist dieser immer den unsichtbaren Grenzflächen unterworfen, die sich durch Grundfläche, Flächenwidmungsplan und Gesetze ergeben. Eine wirklich freie Gestaltung von innen heraus ist somit unmöglich. Anders sieht dies im künstlich simulierten Raum des Computers aus, denn dieser ist allerhöchstens durch Speicher- oder Softwarebeschränkungen begrenzt. Volumina, Verbindungen und Geometrien können hier völlig frei sein. Solch ein Raum müsste also nicht einmal den Gesetzen der Physik gehorchen.

Dieser künstliche Raum bietet also das ungeheure Potential, von innen heraus geformt zu werden, wie dies in der Realität niemals möglich wäre. Es ist dieses Potential, das etwa Marcos Novak dazu bewegen hat, eine Architektur des „Cyberspace“ zu begründen. Die Erfahrung solch eines virtuellen Raumes mittels sensorisch immersiver Technologien, wie etwa „Head Mounted Displays“, Datenhandschuhen und binauralen Kopfhörersystemen war am Ende des 20. Jhdts. für kurze Zeit

der grosse Hoffnungsschimmer der Präponenten einer digitalen Revolution. Wie ich aber in Kapitel 4 von Teil 1 dieser Arbeit beschrieben habe, hat diese Art von virtuellem Raum zwar kurzzeitig eine hohe Popularität erreicht, verbreitete sich aber aufgrund verschiedener Faktoren nicht weiter. Stattdessen trat das Virtuelle mittels internetfähiger Geräte in den Realraum ein.

Mit diesem „Herüberschwappen des Virtuellen“ (Vgl. Teil 1, Kapitel 4) in den Realraum, also dem, was Marcos Novak als „Eversion“ bezeichnet, vollzog sich aber auch ein Perspektivwechsel. Die Immersion in diesem „evertierten“ Cyberspace ist keine sensorische mehr, sondern eine mentale. Das Virtuelle, das uns sensorisch zu umhüllen begonnen hatte, wurde bei seinem Eintritt in den Realraum objektiviert. Es wurde also zu einem Ding, von dem wir ausserhalb positioniert sind, und das wiederum in einzelnen Objekthaften Knotenpunkten Zugang zu einer unsichtbaren digitalen Parallelwelt bietet, die sich über unsere reale Welt gelegt hatte. Auch Novak, dessen erste Projekte noch immersiver Natur im gänzlich digitalen Raum waren, hat diesen Perspektivwechsel mit dem Rückeintritt des digitalen in den Realraum mitvollzogen. Zu beobachten ist dies etwa an den physischen Repräsentationen seiner fluiden digitalen Geometrien bei den Projekten „Invisible Architectures“, „AlienWithin“ oder „Echinoderm“, die spezifisch unter dem Gesichtspunkt der späteren Objektwerdung entstanden sind. (Novak, 2002b, S.69ff.) Diese Versuche, das Digitale in den Realraum „zurückzuholen“ tragen somit viel eher skulpturale denn architektonische Züge.

Um den virtuellen Raum als von innen erfahrbaren in den Realraum zu holen, bedarf es einer sensorischen Immersion. Ansätze zur Verwirklichung einer „Augmented Reality“ durch Brillensysteme, die die Realität mit künstlichen Daten erweitern, könnten hierbei eine Perspektive bieten, um die Eversion von „innen“ zu vollziehen. Diese Immersion bleibt allerdings eine optische, und der Raum in den bisherigen Konzeptionen zumeist ein abstrakter, der sich durch Überblendungen von „Meta-Informationen“ über das Abbild der realen Welt auszeichnet, somit aber die geometrischen Beschränkungen des Realraums übernimmt.<sup>1</sup>

*Sonic Manifolds* nimmt eine andere Position ein, denn es erzielt eine sensorische Immersion in einem evertierten virtuellen Raum. Im Realraum entfalten sich also virtuelle Räume, die erhört werden können. Dies geschieht durch mehrere Schritte. Zum Ersten wird die Akustik eines künstlichen Raumes geometrisch simuliert. Zum Zweiten werden die Bewegungen der Benutzer im Realraum vermessen. Zum Dritten werden diese Bewegungen im virtuellen Raum nachvollzogen, wo sie in der künstlichen Geometrie widerhallen. Zum vierten wird dieser virtuelle akustische Raum mittels einer Tonanlage in den Realraum evertiert. Der Umstand, dass diese künstlichen Räume akustisch repräsentiert werden, macht diese physisch erfahrbar. So ist es möglich, dass der virtuelle

---

<sup>1</sup> Eine Ausnahme stellt hierbei die Microsoft HoloLens dar, die einen anderen Weg zu gehen verspricht, der allerdings weiterhin ein in erster Linie optischer bleibt.

Raum körperlich erfahrbar – ja „ertastbar“ – wird. Die digitale Natur dieser künstlichen Räume erlaubt, dass diese völlig frei positioniert sein und sich überlappen können. So werden Raumkonstellationen möglich, die eigentlich physikalisch unmöglich sind, aber dennoch sensorisch erfahrbar werden. Die Anordnung dieser künstlichen Räume im Realraum wiederum offenbart eine Form der räumlichen Komposition, die erst durch die Bewegung ihre Metrik und Kontraste zeigt.

Dennoch ist, obgleich die gegenwärtige Situation vielversprechend ist, das Ziel noch nicht erreicht. Die künstlichen Räume von *Sonic Manifolds* erscheinen zwar glaubwürdig, unterliegen aber noch vielen Limitationen. Neben technischen Aspekten der Klangerzeugung und Raumsimulation sind es vor allem Aspekte der Interaktion mit dem künstlichen Raum sowie die Dynamisierung der auralen Architekturen, die nach weiterer Entwicklung verlangen würden, damit sich der virtuelle Raum tatsächlich in der Form aural manifestieren kann, wie es mir vorschwebt.

## 8.1 Technische Perspektiven in der Klangerzeugung

Wiewohl die simulierten Klangräume in *Sonic Manifolds* bereits einen hohen Grad an Realismus zulassen, ließe sich dieser mit einigem Aufwand noch weiter steigern. Wie in Kapitel 7.2 beschrieben simuliert *Sonic Manifolds* sogenannte „Shoobox“-Räume, also quaderförmige Volumen mit parallelen Wänden. Unregelmässige Räume zu simulieren würde eine komplexere Programmierung erfordern. Wie ich in Kapitel 4 von Teil 1 beschrieben habe, ist die Simulation von Schallreflexionen eine komplizierte Angelegenheit, die eigentlich eine „Wave-Tracing“-Engine erfordern würde. Nun gibt es meines Wissens nach kein Softwareprojekt, das sich derzeit einer solchen annimmt. Allerdings könnte zumindest die Simulation von Erstreflexionen innerhalb einer unregelmässigen geometrischen Struktur mithilfe einer „Raytracing“-Engine approximiert werden. Um dies zu bewerkstelligen würde es sich also anbieten, ein bestehendes Projekt, das eigentlich der Graphikberechnung dient, zu adaptieren. Dies wäre meines Erachtens nach der nächste logische Schritt, um die künstlichen Räume in *Sonic Manifolds* zu verfeinern.

Ein weiteres Feld, in dem eine Verbesserung der Simulation sinnvoll wäre, liegt in der Berechnung von Raumresonanzen. In Kapitel 7.3 habe ich beschrieben, dass zur Simulation eines natürlich klingenden künstlichen Nachhalls ein möglichst neutrales Frequenzbild ohne spektrale „Färbung“ zum Einsatz kommen sollte, also alle Frequenzbereiche gleichmäßig verstärkt werden und abklingen sollten. Nun haben reale Räume keineswegs ein gleichmäßiges Frequenzbild im Nachhall. Stattdessen weisen reale Räume Resonanzfrequenzen auf, die auf der Wellenlänge der Raumabmessungen und deren Vielfachen basieren. Abhängig von der Regelmässigkeit bzw. Symmetrie des Raumes, der Glätte bzw. Rauheit der Oberflächen und etwaigen Hindernissen (Möbel etc.) sind diese unterschiedlich stark ausgeprägt. Manfred Schroeder schreibt zu dieser paradoxen Anforderung

rung folgendes:

„One might ask -oneself: “Why insist on perfectly flat frequency responses for an artificial reverberator if real rooms have highly irregular frequency responses?” In fact, for many sounds the human ear cannot distinguish between a flat response and the irregular response of a room, which fluctuates on the average about 10 db and has extreme variations of 40 db or more. Psychoacoustic experiments now in progress at Bell Telephone Laboratories indicate that such extreme response irregularities are imperceptible when the density of peaks and valleys on the frequency scale is high enough. There are about 15 large response peaks in every 100 cps interval for a room with 1 sec reverberation time. Thus, one might hope that if an artificial reverberator has a comparable number of response peaks it might sound just as good as a real room.“ (Schroeder, 1962, S.222)

Nun muss angemerkt werden, dass in den über 50 Jahren seit Schroeder diese Begründung formuliert hat, keine merkbaren Anstrengungen in die Simulation von Raum-Resonanzen gesetzt wurden. Der Grund für die mangelnde Implementierung von Raumresonanzen in Reverb-Algorithmen ist aber höchstwahrscheinlich nicht im „good enough“-Fakt zu suchen, sondern eher darin, dass Resonanzen, obgleich natürlich vorkommend, ein zu vermeidendes Übel in der Tontechnik und Raumakustik sind. Bei der Ausgestaltung von Studioräumen z.B. werden parallele Wandelemente wenn möglich vermieden und beim Ausbau wird alles daran gesetzt, Nachhall und Resonanzen zu minimieren, um eine möglichst neutrale Raumakustik zu erhalten. Da die Musikproduktion aber die dominante Kraft für den Fortschritt der Hallalgorithmen darstellt, und Resonanzen nicht von Interesse für diese sind, sind Reverbs welche, diese implementieren, faktisch inexistent. Eine Ausnahme hierbei stellen Convolution-Reverbs dar, da diese ja real existente Räume abbilden, womit auch eine nichtlineare Frequenzantwort des jeweiligen Raumes mit aufgenommen und reproduziert wird. Für Künstler oder Wissenschaftler, die versuchen realistische Räume akustisch zu simulieren ist allerdings die Simulation von Raumresonanzen sehr wohl von Bedeutung. Schließlich macht es einen großen Unterschied, ob man eine Hallsimulation benutzt, um einen gegebenen akustischen Inhalt in Szene zu setzen, oder ob das Ziel der Präsentation der künstliche Raum selbst ist. Aus diesem Grund wäre es für den Raumeindruck wünschenswert, Resonanzfrequenzen zu simulieren. Hierfür käme etwa eine Bank paralleler IIR(engl. *infinite impulse response*)-Filter in Frage.

Die größte technische Herausforderung für eine Steigerung der Raumerfahrung in *Sonic Manifolds* ist aber keine der Klangerzeugung, sondern der Klangunterdrückung. Um nämlich die künstlichen Klangräume wirklich perfekt realisieren zu können, müsste die Raumakustik des Realraums in dem die Installation aufgebaut wird „entfernt“ werden. Die Problematik ist die Folgende: Jeglicher künstlicher Klangraum, der im Computer simuliert wird, wird bei der Wiedergabe über eine Tonanlage innerhalb eines realen Raumes selbst Opfer von Reflexionen in selbigem. Der künstliche

Raum hallt also im Realraum wider. Da wir Menschen diese Reflexionen wahrnehmen, wird der künstliche Raum also sensorisch immer dem realen „untergeordnet“, denn die Färbung des Wiedergaberaumes überträgt sich unweigerlich auf alles, was in diesem erklingt. Um dieses Problem zu beseitigen gibt es zwei Wege. Erstens kann der Realraum akustisch gedämmt werden, wodurch Reflexionen gestreut, und Nachhall unterbunden wird. Ein solches Unterfangen torpediert aber die gewünschte „Eversion“, denn je mehr der Realraum präpariert werden muss, um den virtuellen Raum erlebbar zu machen, desto weniger „real“ wird der Realraum. Je weniger Bedeutung und Funktion abseits der Simulation der Realraum innehat, desto weniger ist er tatsächlicher Raum. Stattdessen wird solch ein Wiedergaberaum ein Apparat zum Einstieg in den virtuellen Raum. Jede Massnahme, die also den Realraum anderer Funktionen beraubt, stellt also eine Ausrichtung des Projektes in Richtung einer isolierenden immersiven „Virtual Reality“ dar, und somit ein metaphorisches „Wiederaufsetzen des Datenhelmes“. Eine zweite Möglichkeit, um den Realraum akustisch zu entfernen, wäre, dessen akustischen Eigenheiten in der Wiedergabesituation entgegenzuwirken, indem etwa etwaige Resonanzfrequenzen durch einen Equalizer reduziert würden. Weiters könnte der Realraum durch gezieltes Einsetzen von Maskierungseffekten vom künstlichen Klangraum überdeckt werden. Hierbei handelt es sich um ein hochkomplexes Unterfangen, für welches Spezialisten der Psychoakustik herangezogen werden müssten. Gelänge aber diese „Entfernung“ des Realraumes, könnte der Grad der sensorischen Immersion im künstlichen Raum um ein Vielfaches gesteigert werden.

## 8.2 Konzeptionelle Perspektiven

Abgesehen von technischen Verbesserungen, die die Erfahrung der künstlichen Räume steigern würden, gibt es noch eine Reihe von konzeptionellen und funktionalen Aspekten, die man durch fortgesetzte Arbeit am Projekt zur Entfaltung bringen könnte. Würde man das Projekt beispielsweise mit einer logischen Scriptsprache koppeln, die Veränderungen unter bestimmten Konditionen in Gang setzt, könnten die Klangräume von *Sonic Manifolds* auch zeitlich „flüssig“ werden. Unter den sich ergebenden Möglichkeiten erachte ich hierbei die Schaffung dynamischer Geometrien als faszinierendste Perspektive, die sich durch das Projekt ergeben könnte. Es handelt sich hierbei um eine Eigenschaft eines simulierten Raumes, für die es keine Entsprechung in den realen Räumen unserer materiellen Realität gibt. Beispielsweise könnten sich Wände verschieben, Räume verdrehen und nach belieben wachsen oder schrumpfen. Auch könnten sich Oberflächen verändern oder Löcher auftun. Bewegliche Räume könnten auch nach einer musikalischen Metrik ihre Zustände ändern. All dies haben wir Menschen schon oft gesehen – in fiktionalen Filmen, Videoclips und Computerspielen etwa sind solche Veränderungen der Umgebung keine Seltenheit –, aber dies zu hören wäre eine neue faszinierende Erfahrung.

Einen derartigen Steuerungsmechanismus zu erzeugen ist kein triviales Unterfangen. Abgesehen von der programmiertechnischen Herausforderung wäre bereits die logische Frage, wie dieses System aufgebaut sein sollte, und was es ermöglichen sollte hochgradig komplex. Beispielsweise müsste es in der Lage sein, neben der Abfrage von Variablen, die über etwaige Konsequenzen entscheiden, auch Bewegungen kontrollieren können. Weiters müsste es in der Lage sein, sowohl „globale“ zeitliche Abläufe zu synchronisieren, als auch individuelle bzw. relative Abläufe zu steuern. Höchstwahrscheinlich würde es sich deshalb anbieten, ein bestehendes System aus der Videospiele-Branche zu „zweckentfremden“. Dies hätte ausserdem den Vorteil, dass etwa auch Algorithmen künstlicher Intelligenz weiterverwendet werden könnten, mittels derer die künstlichen Klangräumen von akustischen Lebensformen bevölkert werden könnten. Diese wären dann in der Lage, auf die Besucher zu reagieren, bzw. sogar zu interagieren. Der dynamische Raum einer von mir erträumten Reinkarnation von *Sonic Manifolds* ist also ein belebter.

Eine weitere Konsequenz solch eines konditionalen Steuerungssystems wäre die Möglichkeit, Interaktionen der Besucher mit dem künstlichen Raum selbst bereitzustellen. So könnten Raumzonen zu Portalen umfunktioniert werden, die eine Reise in einen neue künstliche Umgebung in Gang setzen. Dies könnten Knotenpunkte im Raum sein, die, wenn sie denn berührt werden, den Zustand des ganzen Systems verändern, ganz im Sinne der Deleuzschen Singularitäten, die den Zustandsraum restrukturieren, wenn sie einmal erreicht sind. Würde man nun einhergehen, und etwaige Gesten der Besucher interpretieren, würde dies zur Folge haben, dass, gekoppelt mit der Fähigkeit der Räume, auf diese Besucher zu reagieren, diese selbst zu Schöpfern werden könnten. Die Besucher könnten also nicht nur Veränderungen im künstlichen Raum auslösen – sie könnten diesen selbst gestalten. Die Vorstellung, durch eine Geste einen neuen Raum im Bestehenden aufzuspannen, diesen dann zu vergrössern, um ihn schliesslich zu betreten und erkunden zu können, entfacht mannigfaltige Fantasien von fremden Klangwelten. Vielleicht werden wir solche Räume dereinst besuchen können: Aurale Multiplizitäten von unendlicher Dimension.







# Quellenverzeichnis

## Bücher

- Aristoteles (1995): Metaphysik IX. In: Aristoteles, Bonitz, Hermann, Seidl, Horst (Hg.): Metaphysik, Hamburg, Meiner, 1995.
- Aristoteles (2005): Poetik. Griechisch/Deutsch. Stuttgart, Reclam.
- Attali, Jacques (1985): Noise: The Political Economy of Music. Minneapolis, University of Minnesota Press, Translation by Brian Massumi, 10. Auflage 2009.
- Berghaus, Alexander, Rettinger, Gerhard und Böhme, Gerhard (1996): Hals-Nasen-Ohren-Heilkunde. Stuttgart, Hippokrates Verl., 1996.
- Blauert, Jens (1997): Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, Cambridge, MA
- Blessner, Barry, Salter, Linda-Ruth (2007): Spaces Speak – Are You Listening?. Cambridge, Massachusetts, Mit Press, 2006.
- Cage, John (1961): Silence, Middletown, Conn., Wesleyan University Press.
- Chion, Michel et. al. (1990): Audio-vision, New York, Columbia University Press, 1994.
- Davis, Erik (2004): TechGnosis, New York, Harmony Books.
- Erlmann, Veit (2010). Reason and Resonance: A History Of Modern Aurality. Zone Books, New York.
- Everest, F. Alton and Pohlmann, Ken C. (2009): Master Handbook of Acoustics, 5, New York, McGraw-Hill.
- Flusser, Villém (1991): Räume. In: Dünne, Jörg, Günzel, Stephan (Hg.): Raumtheorie. Frankfurt am Main, Suhrkamp, 2006.
- Gibson, William (1984): Neuromancer. New York, Ace Books.
- Grimshaw, Mark (2012): Sound and Player Immersion in Digital Games. In: Pinch, Trevor, Bijsterveld, Karin (Hg.) The Oxford Handbook Of Sound Studies. New York, Oxford University Press.
- Günzel, Stephan (1998): Immanenz: Zum Philosophiebegriff von Gilles Deleuze. Essen, Die Blaue Eule.
- Heim, Michael (1993): The Metaphysics of Virtual Reality, New York, Oxford University Press.
- Kahn, Douglas (1999): „Noise, water, meat“. Cambridge, Mass., MIT.
- Kluge, Friedrich; Bürgisser, Max; Gregor, Bernd u. a. (2002): Etymologisches Wörterbuch der deutschen Sprache. Berlin: De Gruyter.
- Koelsch, Stefan (2012): Brain and Music. Wiley-Blackwell.
- Kuttruff, Heinrich (2009): Room acoustics. London & New York, Spon Press/Taylor & Francis.
- Lem, Stanisław (1964): Summa technologiae. Frankfurt am Main, Suhrkamp, 1981.

- Mathews, Max (1999): *The Ear And How It Works*. In: Cook, Perry (Hg.): *Music, cognition, and computerized sound*, Cambridge, Massachusetts [etc], Mit Press, 2001.
- Mathews, M. (1999b): *The Auditory Brain*. In: Cook, Perry (Hg.): *Music, cognition, and computerized sound*, Cambridge, Massachusetts [etc], Mit Press, 2001.
- Metzger, W. (1975): *Was ist Gestalttheorie?* In: Guss, K. (Hrsg.): *Gestalttheorie und Erziehung*, Darmstadt, Steinkopff
- Novak, Marcos (1991): *Liquid Architectures in Cyberspace*. In: Benedikt, Michael (Hg.): *Cyberspace*, Cambridge, Mass., MIT Press, 1992.
- Novak, Marcos (2002a): *Invisible Architectures*. In: Jormakka, Kari (Hg.): *Absolute motion*, Tampere, Tampere University of Technology.
- Ong, Walter (1982). „Orality and Literacy: The Technologizing Of The World“, 2nd edition 2002, Routledge, New York, 2002
- Pierce, Jonathan (1999): *Hearing in time and space*. In: Cook, Perry (Hg.): *Music, cognition, and computerized sound*, Cambridge, Massachusetts [etc], Mit Press, 2001.
- Liberman, M. Charles, et. al. (2002): *Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier* In: *Nature*, Band 419, Ausgabe 6904, 2002300-304, DOI: 10.1038/nature01059.
- Puckette, Miller (2007): *The theory and technique of electronic music*. Hackensack, N.J., World Scientific Publishing Co., 2007.
- Russolo, Luigi (1916). „L'arte die rumori“, Akroama, The Soundscape Newsletter Europe Editions, Basel, 1999.
- Sabine, Wallace Clement (1923): *„Collected Papers on acoustics“*, Cambridge, Harvard Univ. Press.
- Schafer, R. Murray (1967): *„Ear Cleaning: Notes for an experimental music course“*, Toronto, Clark & Cruickshank.
- Schafer, R. Murray (1969): *The New Soundscape*. Don Mills, Ontario, BMI Canada.
- Schafer, R. Murray (1977). „The Soundscape: Our Sonic Environment And The Tuning Of The World“. Destiny Books, Rochester, 1994.
- Schwartz, Hillel (2011). „Making Noise: from Babel to the big bang and beyond“. Zone Books, New York.
- Serres, Michel (1994): *Atlas*, Berlin, Merve-Verl., 2005.
- Shepard, Roger (1999): *Cognitive Psychology and Music*. In: Cook, Perry (Hg.): *Music, cognition, and computerized sound*, Cambridge, Massachusetts [etc], Mit Press, 2001.
- Sterne, Jonathan, et. al. (2012). „The Sound Studies Reader“, New York, Routledge.
- Sterne, Jonathan (2012). „MP3: The Meaning Of A Format“. London, Duke University Press.
- Truax, Barry (1984). „Acoustic communication“, New Jersey, Ablex Publishing Co..
- Van Bergeijk, Willem André Maria, Pierce, John und David, Edward. (1960): „Die Schallwellen und wir“, München, Verlag Kurt Desch.
- Voegelin, Salome (2010), *Listening To Noise And Silence*, Continuum, London.
- Weibel, Peter (1989): *Territorium und Technik*. In: *Ars Electronica* (Hg.): *Philosophien der neuen Technologie*, Berlin, Merve-Verl.
- Weinzierl, Stefan (2008): *Handbuch der Audiotechnik*, Berlin, Heidelberg, Springer-Verlag.
- Westphal, Wilhelm H. (1958): *Physik*, Berlin, Heidelberg, Springer-Verlag.

## Dissertationen

- Hollerweger, Florian (2011): *The revolution is Hear! Sound Art, the Everyday and Aural Awareness*. Queen's University Belfast, 2012.
- Necciari, Thibaud (2010): *Auditory time-frequency masking: Psychoacoustical measures and application to the analysis-synthesis of sound signals*. Marseille, L'Université de Provence Aix-Marseille

Peters, Nils (2010): Sweet [re]production: Developing sound spatialization tools for musical applications with emphasis on sweet spot and off-center perception. Toronto, McGill University, 2010.

## Artikel

Bell, Andrew (2004): Hearing: Travelling Wave or Resonance? In: PLoS Biology, Band 2, Ausgabe 10, 2004e337, DOI: 10.1371/journal.pbio.0020337.

Beltrán, Fernando et.al. (2002): Matlab Implementation of Reverberation Algorithms In: Journal of New Music Research, Band 31, Ausgabe 2, 2002153-161, DOI: 10.1076/jnmr.31.2.153.8096.

Cohen, Yale E.; Knudsen, Eric I. (1999): „Maps versus clusters: different representations of auditory space in the midbrain and forebrain“. In: Trends in Neurosciences, Band 22, Ausgabe 3, 1999128-135, DOI: 10.1016/s0166-2236(98)01295-8.

Costello, Sean, Anderson, Joseph (2009): Adapting Artificial Reverberation Architectures For B-Format Signal Processing. Graz, Ambisonics Symposium 2009, Abgerufen am 2.10.2015, <http://freeverb3vst.osdn.jp/doc/AmbiSym09-JosephAnderson-SeanCostello-AmbiReverbArch.pdf>

Cryptologic Quarterly (1993): Virtual Reality. In: Cryptologic Quarterly, Band 12, Ausgabe 3-4, 1993, 21-50, Abgerufen am 13. 10. 2015, [https://www.nsa.gov/public\\_info/declass/cryptologic\\_quarterly.shtml](https://www.nsa.gov/public_info/declass/cryptologic_quarterly.shtml).

Cruz-Neira, Carolina, Sandin, Daniel J., DeFanti, Thomas A. and Hart, John C. (1992): The CAVE®: Audio Visual Experience Automatic Virtual Environment In: Communications of the ACM, Band 35, Ausgabe 6, 1992, 64-72, Abgerufen am 13. 10. 2015, <https://www.evl.uic.edu/entry.php?id=1299>.

Dahl, Luke, Jot, Jean-Marc (2000): A reverberator based on absorbent all-pass filters. Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7-9, 2000.

Dattorro, Jon (1997): Effect Design: Part 1: Reverberators and Other Filters. Journal of the Audio Engineering Society, Band 45, Ausgabe 9, 1997, 660-684

Felgett, Peter (1975): „Ambisonics. Part one: General system description“, Reproduced from Studio Sound, Vol. 17, pp 20-22, 40 (August 1975), by permission of IPC Media Ltd, publishers of Hi-Fi News, Abgerufen am 1.10.2015, <http://www.michaelgerzonphotos.org.uk/articles/Ambisonics%201.pdf>

Fowler, Michael (2015): Sounds in space or space in sounds? Architecture as an auditory construct In: Arq, Band 19, Ausgabe 01, 2015, 61-72, DOI: 10.1017/s1359135515000226.

Funkhouser, Thomas et. al. (1998): A beam tracing approach to acoustic modeling for interactive virtual environments. In: Proceedings of the 25th annual conference on Computer graphics and interactive techniques, p.21-32, July 1998

Gerzon, Michael (1974): What's wrong with quadraphonics? In: Studio Sound, Ausgabe 5, 1975.

Gerzon, Michael (1976): Unitary (energy-preserving) multichannel networks with feedback. In: Electronics Letters , Band 12, Ausgabe 11, 1976, 278 - 279

Gerzon, Michael (1986): Stereo Shuffling: New Approach – Old Technique, Reproduced from Studio Sound, July 1986, [http://www.audiosignal.co.uk/Resources/Stereo\\_shuffling\\_A4.pdf](http://www.audiosignal.co.uk/Resources/Stereo_shuffling_A4.pdf), Abgerufen am 1.10.2015

Griesinger, David (1989): Practical Processors and Programs for Digital Reverberation. In: Audio in Digital Times, Toronto, Proceedings 7th International AES Conference, 1989, 187-195

Gunther, Max (1967): The Sonics Boom. In: Playboy, Ausgabe 5, 1967, S 112-116,183-186.

Jonas, Jost B. et al. (1992): Human optic nerve fiber count and optic disc size. In: Investigative Ophthalmology & Visual Science May 1992, Vol.33, 2012-2018.

Jones, Steven (1993): A sense of space: Virtual reality, authenticity and the aural. In: Critical Studies in Mass Communication, Band 10, Ausgabe 3, 1993, 238-252, DOI: 10.1080/15295039309366866.

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

King, Andrew J. (1999): „Sensory experience and the formation of a computational map of auditory space in the brain“. In: BioEssays, Band 21, Ausgabe 11, Wiley & Sons, New York

King, Andrew J. (2002): „Neural Plasticity: How the Eye Tells the Brain about Sound Location“. In: Current Biology, Band 12, Ausgabe 11, 2002R393-R395, DOI: 10.1016/s0960-9822(02)00883-7.

- Lewald, Jörg (1997): „Eye-position effects in directional hearing“. In: Behavioural Brain Research“ Band 87, Ausgabe 1, 199735-48, DOI: 10.1016/s0166-4328(96)02254-1.
- Lyon, R. (2003), Product Sound Quality – from Perception to Design, Sound & Vibration Magazine, Vol. 3/2003, Abgerufen am 1.6.2015, <http://www.sandv.com/downloads/0303bate.pdf>
- Ludovico, Alessandro (2013): Marcos Novak Interview. In: Neural, Band 46, 74-76; Reprint from: Ludovico, Alessandro (2001), Neural, Band 18; Bari, Neural.
- Moorer, James A. (1979): About This Reverberation Business In: Computer Music Journal, Band 3, Ausgabe 2, 1979, 13, DOI: 10.2307/3680280.
- Novak, Marcos (2002b): Speciation, Transvergence, Allogenes: Notes on the Production of the Alien. In: Architectural Design, Band 72, Ausgabe 3, 64-71; New York, John Wiley & Sons.
- Rees, Adrian (1996): Sensory maps: Aligning maps of visual and auditory space. In: Current Biology, Band 6, Ausgabe 8, 1996955-958, DOI: 10.1016/s0960-9822(02)00637-1.
- Schafer, R. Murray (1976). Exploring The New Soundscape: Pioneer Research Into The Global Acoustic Environment. In: The UNESCO Courier, November 1976, Unesco, Paris.
- Schroeder, M.R., Logan, B.F. (1961): “Colorless” Artificial Reverberation, Journal of the Audio Engineering Society, Band 9, Juli, 193-194.
- Schroeder, M. R. (1962): Natural Sounding Artificial Reverberation In: Journal of the Audio Engineering Society, Band 10, Ausgabe 3, 219-223.
- Spoendlin, H., Schrott, A. (1989): Analysis of the human auditory nerve In: Hearing Research, Band 43, Ausgabe 1, 198925-38, DOI: 10.1016/0378-5955(89)90056-7.
- Stautner, John, Puckette, Miller (1982): „Designing Multi-Channel Reverberators“, Computer Music Journal, Vol. 6, No. 1, 1982
- Stockhausen, Karlheinz (1978): Texte zur Musik 1970-1977. Bd. 4, Köln
- Sturman, D.J. and Zeltzer, D. (1994): A survey of glove-based input. In: IEEE Comput. Grap. Appl., Band 14, Ausgabe 1, 1994, 30-39, DOI: 10.1109/38.250916.
- Yokosawa, Koichi; Pamilo, Siina; Hirvenkari, Lotta u. a. (2013): „Activation of Auditory Cortex by Anticipating and Hearing Emotional Sounds: An MEG Study“. In: PLoS ONE. 8 (11), S. e80284, DOI: 10.1371/journal.pone.0080284.

## Patente

- Abel, Jonathan S.; Foster, Scott H. (1994): Method and apparatus for efficient presentation of high-quality three-dimensional audio. US Patent #5,596,644
- Abel, Jonathan S.; Foster, Scott H. (1997): Method and apparatus for efficient presentation of high-quality three-dimensional audio including ambient effects. US Patent #5,802,180
- Blessner, Barry, Bader, Karl-Otto (1978): Electric Reverberation Apparatus, Pat. Nr. US 4181820, USA, 1980.
- Blumlein, Alan (1933): Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing systems. GB patent 394325, 1933
- Gerzon, Michael (1998): . US Patent Nr. US5757927
- Hammond, Laurens (1939): Electrical musical instrument, Pat. Nr. US 2230836 A, USA, 1941.
- Heilig Morton (1961): Sensorama Stimulator. US Patent #3,050,870, USA, 1962.

## Online-Verzeichnis

- Ablinger, Peter (2009): „Speaking Piano“. Abgerufen am 24.9.2015 von [http://ablinger.mur.at/speaking\\_piano.html](http://ablinger.mur.at/speaking_piano.html)
- Abomb1.org, (2003): „The atomic bombings of Hiroshima and Nagasaki“. Abgerufen am 07. 10. 2015 von [http://www.abomb1.org/hiroshim/hiro\\_med.html](http://www.abomb1.org/hiroshim/hiro_med.html).

Abrash, Michael (2012): Latency – the sine qua non of AR and VR, 2012. Abgerufen am 16. 10. 2015, <http://blogs.valvesoftware.com/abrash/latency-the-sine-qua-non-of-ar-and-vr/>.

Adweek (2012): Volkswagen Jetta Commercial Door Thunk. Abgerufen am 1.6.2015, <http://www.adweek.com/video/volkswagen-jetta-commercial-door-thunk-volkswagen-deutsch-los-angeles-142363>.

Ambler, M. (2013): This is what a Nissan Leaf REALLY sounds like. Abgerufen am 1.6.2015, <https://www.youtube.com/watch?v=iGwa44w158A>.

Ausfahrt.tv (2013): Probefahrt und Fahrbericht zum 2013 Audi R8 e-tron. Abgerufen am 1.6.2015, <https://www.youtube.com/watch?v=h7vfcPTZX7U&feature=youtu.be&t=5m39s>

Auto-News.de (2012): Der e-Sound von Audi: Akustische Innovation für künftige e-tron-Modelle, Abgerufen am 1.6.2015, [http://www.auto-news.de/greencars/anzeige\\_Audi-e-Sound-Auch-Elektro-Renner-sollen-nach-Sport-klingen\\_id\\_32031](http://www.auto-news.de/greencars/anzeige_Audi-e-Sound-Auch-Elektro-Renner-sollen-nach-Sport-klingen_id_32031).

Avalon Project, (2008): „The Atomic Bombings of Hiroshima and Nagasaki“. Abgerufen am 08. 10. 2015 von [http://avalon.law.yale.edu/20th\\_century/mp11.asp](http://avalon.law.yale.edu/20th_century/mp11.asp).

Bimmertoday.de, (2011): Mehr Details zum BMW M5: aktives Sounddesign und Keramikbremse. Abgerufen am 1.6.2015, <http://www.bimmertoday.de/2011/09/23/mehr-details-zum-bmw-m5-aktives-sounddesign-und-keramikbremse>.

Blumlein, Alan (1935): 'Trains At Hayes' - the world's first stereo film, made in 1935 (clip), 1935, Abgerufen am 01.10.2015, <https://www.youtube.com/watch?v=O8hOOuVtcX4>.

Blumlein, Alan (1935b): Invention of stereo sound: Alan Blumlein - 'Walking and Talking', 1935, Abgerufen am 01.10.2015, <https://www.youtube.com/watch?v=rqaMiDqE6QQ>.

Byron, E. (2012): The Search for Sweet Sounds That Sell. The Wall Street Journal, Abgerufen am 1.6.2015, <http://www.wsj.com/articles/SB10001424052970203406404578074671598804116>.

Costello, Jean (2009): Schroeder Reverbs: the forgotten algorithm, 2009, Abgerufen am 1.10.2015, <https://valhalladsp.wordpress.com/2009/05/30/schroeder-reverbs-the-forgotten-algorithm/>

Costello, Sean (2011): DSP Heros: Miller Puckette, 2011, Abgerufen am 20. 10. 2015, <https://valhalladsp.com/2011/06/01/dsp-heros-miller-puckette/>.

Crecente, B. (2015): For The Love Of Mechanical Keyboards. Polygon, Vox Media, Abgerufen am 1.6.2015, <http://www.polygon.com/2015/2/23/7372837/mechanical-keyboards>.

Davis, Erik (2000): Acoustic Cyberspace, 2000, Abgerufen am 12. 10. 2015, <http://techgnosis.com/acoustic-cyberspace/>.

De.muvs.org, (2015): „Museum für Verhütung und Schwangerschaftsabbruch“. Abgerufen am 07. 10. 2015 von [http://de.muvs.org/abbruch/instrumente/schallwaescher-id2021/?media\\_id=4601](http://de.muvs.org/abbruch/instrumente/schallwaescher-id2021/?media_id=4601).

Dechelle, Francois (2015): A brief history of MAX, Abgerufen am 19. 10. 2015, <http://jmax.sourceforge.net/history.html>.

Der Standard, (2005), „Blopp!“ soll Ottakringer-Umsatz um zehn Prozent steigen lassen. Abgerufen am 1.6.2015, <http://derstandard.at/2011492/Blopp-soll-Ottakringer-Umsatz-um-zehn-Prozent-steigen-lassen>

Dietrich, Dieter (1968): „Neue Waffe: Infraschall?“. ZEIT ONLINE, Abgerufen am 08. 10. 2015. <http://www.zeit.de/1968/05/neue-waffe-infraschall/komplettansicht>.

Duden (2015): Duden | Virtualität | Rechtschreibung, Bedeutung, Definition, Synonyme, Herkunft. Abgerufen am 01. 11. 2015, <http://www.duden.de/node/732996/visions/1382699/view>.

Duden (2015b): Duden | virtuell | Rechtschreibung, Bedeutung, Definition, Synonyme, Herkunft. Abgerufen am 01. 11. 2015, <http://www.duden.de/node/655108/visions/1613964/view>.

Edwards, Benj (2015): Unraveling The Enigma Of Nintendo's Virtual Boy, 20 Years Later, 2015, Abgerufen am 12. 10. 2015, <http://www.fastcompany.com/3050016/unraveling-the-enigma-of-nintendos-virtual-boy-20-years-later>.

Evl.uic.edu, (1996): Ars Electronica Center Opening '96, 1996, Abgerufen am 13. 10. 2015, <https://www.evl.uic.edu/entry.php?id=319>.

Floisand, Christian (2012): Algorithmic Reverbs: The Moorer Design, 2012, Abgerufen am 01. 10. 2015, <https://christianfloisand.wordpress.com/2012/10/18/algorithmic-reverbs-the-moorer-design/>.

Föllmer, Golo (2015): Stockhausen, Karlheinz: Spherical Concert Hall, 2015, Abgerufen am 01. 10. 2015, <http://www.medienkunstnetz.de/works/stockhausen-im-kugelauditorium/>.

Graeme, Malcolm (2011): Kinect SDK: Smoothing Skeleton Data, 2011, Abgerufen am 19. 10. 2015, <http://cm-bloggers.blogspot.co.at/2011/07/kinect-sdk-smoothing-skeleton-data.html>.

Gromala, Diane, Sharir, Yacov (1994): Virtual Bodies: Travels Within, 14. 10. 2015, Abgerufen am 14. 10. 2015, <http://www.art.net/~dtz/vb.html>.

Harwell, D. (2015): America's best-selling cars and trucks are built on lies: The rise of fake engine noise. Washington Post, Abgerufen am 1.6.2015, [http://www.washingtonpost.com/business/economy/americas-best-selling-cars-and-trucks-are-built-on-lies-the-rise-of-fake-engine-noise/2015/01/21/6db09a10-a0ba-11e4-b146-577832eafcb4\\_story.html?hpid=z2](http://www.washingtonpost.com/business/economy/americas-best-selling-cars-and-trucks-are-built-on-lies-the-rise-of-fake-engine-noise/2015/01/21/6db09a10-a0ba-11e4-b146-577832eafcb4_story.html?hpid=z2).

Hielscher.com, (2015): „Ultraschall in der Margarineproduktion“. Abgerufen am 07. 10. 2015 von <http://www.hielscher.com/de/ultrasonication-and-its-manifold-applications-in-food-processing.htm>.

Hodgkinson, Tim (1987): „Pierre Schaefer interviewed by Tim Hodgkinson“, ReRecords Quarterly Magazine 2, no. 1 (March 1987): 4-10. Abgerufen am 26.9.2015, [http://www.ele-mental.org/ele\\_ment/said&did/schaeffer\\_interview.html](http://www.ele-mental.org/ele_ment/said&did/schaeffer_interview.html).

Hull, Alan (2015): How to reduce lag in a moving average, 2015, Abgerufen am 01.10.2015, <http://alanhull.com/hull-moving-average>.

Json.org, (2015): JSON, Abgerufen am 19. 10. 2015, <http://www.json.org/>.

Lem, Stanislaw (1998): Probleme mit der Phantomatik, 1998, Abgerufen am 13. 10. 2015, <http://www.heise.de/tp/artikel/2/2363/1.html>.

Kurier (2013): Die Fehlzündung gibt's auf Knopfdruck. 13.3.2013, Abgerufen am 3.5.2015, <http://kurier.at/lebensart/motor/die-fehluendung-gibts-auf-knopfdruck/4.726.638>.

Kühne, Christian: Die Entstehung von Hall und seine künstliche Reproduktion, Abgerufen am 1.10.2015, <http://www.tontempel.de/docs/hallgeraete.pdf>

Lange, André (2002): LE THEATROPHONE, LE PREMIER MEDIUM ELECTRIQUE DE DIFFUSION CULTURELLE, 2002, Abgerufen am 19. 10. 2015, <http://histv2.free.fr/theatrophone/theatrophone.htm>.

Mannell, Robert (2015): Waveforms: Adding waveforms and Phase, 02. 11. 2015, Abgerufen am 02. 11. 2015, [http://clas.mq.edu.au/speech/acoustics/waveforms/adding\\_waveforms.html](http://clas.mq.edu.au/speech/acoustics/waveforms/adding_waveforms.html).

Mercedes-Benz TV (2009): Designing the interior sound. Abgerufen am 1.6.2015, <https://www.youtube.com/watch?v=d53zBU-jCwNI>.

Nissanlefaustralia (2012): the new sound - Nissan LEAF, Abgerufen am 1.6.2015, <https://www.youtube.com/watch?v=yJtC6Z-rODOY>.

Microsoft (2012): Kinect for Windows SDK v1.0, 2012, Abgerufen am 02. 11. 2015, <http://www.microsoft.com/en-us/download/details.aspx?id=28782>.

Moser, J. (1996): The Art of Noise: Designing Sound Characteristics Into a Motorcycle Engine, University of Wisconsin, Abgerufen am 1.6.2015, <https://web.archive.org/web/20120225234742/http://tc.engr.wisc.edu/uer/uer96/author8/index.html>.

NASA (2015): Shock and Awesome Pictures, 2015, Abgerufen am 02. 11. 2015, [http://www.nasa.gov/centers/armstrong/features/shock\\_and\\_awesome.html](http://www.nasa.gov/centers/armstrong/features/shock_and_awesome.html).

Neukom, Martin (2009): Ambisonics. Zürich, ZHdK / Institute for Computer Music and Sound Technology., Abgerufen am 26.10.2015, [https://www.zhdk.ch/fileadmin/data\\_subsites/data\\_icst/Publikationen/2009\\_Ambisonics\\_Einfuehrung.pdf](https://www.zhdk.ch/fileadmin/data_subsites/data_icst/Publikationen/2009_Ambisonics_Einfuehrung.pdf)

Nobelprize.org, (2015): The Nobel Prize in Physiology or Medicine 1961, 2015, Abgerufen am 18. 08. 2015, [http://www.nobel-prize.org/nobel\\_prizes/medicine/laureates/1961/](http://www.nobel-prize.org/nobel_prizes/medicine/laureates/1961/).

Noble, Jonathan (2014): F1 bosses agree to look into ways to increase engine noise. autosport.com, Abgerufen am 1.6.2015, <http://www.autosport.com/news/report.php/id/113317>

Novak, Marcos (1996): transArchitektur, 1996, Abgerufen am 13. 10. 2015, <http://www.heise.de/tp/druck/mb/artikel/6/6068/1.html>.

de Oliveira, Rafael, Duarte, Alexsander, Lima, Patrícia (2014): Interview with Murray Schafer (Corfu, 2011), 2014, Abgerufen am 31. 10. 2015, [https://www.youtube.com/watch?v=Hu4au\\_4Jlfo](https://www.youtube.com/watch?v=Hu4au_4Jlfo).

Olivetti, Justin (2012): The Game Archaeologist: Maze War, 2012, Abgerufen am 20. 10. 2015, <http://www.engadget.com/2012/06/12/the-game-archaeologist-maze-war/>.

Pade, Jochen (1993): „Zutiefst erschütternd“. ZEIT ONLINE. Abgerufen am 08. 10. 2015 von <http://www.zeit.de/1993/22/>



zutiefst-erschuetternd/komplettansicht.

Peters, Nils (2012): Nilson/ViMiC-and-friends, 2012, Abgerufen am 02. 10. 2015, <https://github.com/Nilson/ViMiC-and-friends/tree/master/spatialization/sur.reverb~>.

Russolo, Luigi (1913): Veglio Di Una Città, 1913, Abgerufen am 24. 09. 2015, <https://www.youtube.com/watch?v=VHL-mitA3o6g>.

Santer, B. (2009), KOMPAKTWAGEN: Absichtliche Fehlzündungen, Focus Vol. 15/2009, Abgerufen am 1.6.2015, [http://www.focus.de/auto/neuheiten/kompaktwagen-absichtliche-fehluendungen\\_aid\\_387251.html](http://www.focus.de/auto/neuheiten/kompaktwagen-absichtliche-fehluendungen_aid_387251.html).

Schacher, Jan, Kocher, Philippe (2014): Ambisonics External for MaxMSP, 2014, Abgerufen am 27. 10. 2015, [https://www.zhdk.ch/index.php?id=icst\\_ambisonicsexternals](https://www.zhdk.ch/index.php?id=icst_ambisonicsexternals).

Schafer, R. Murray (2010): La Semaine du Son 2010, Abgerufen am 26. 09. 2015, <https://www.youtube.com/watch?v=JX9V-zlCmKpA>.

Schneider, Toni (2009): A3D, 2009, Abgerufen am 20. 10. 2015, <http://toni.org/a3d/>.

Schleufe, Markus (2010): Auf der Suche nach dem "Wouuffff". Zeit Online, Abgerufen am 1.6.2015, <http://www.zeit.de/auto/2010-02/auto-sound/>.

Schoenherr, Steve (2005): Recording Technology History, 2005, Abgerufen am 19. 10. 2015, <http://www.aes.org/aeshc/docs/recording.technology.history/notes.html>.

Sengpiel, Eberhard (2003): Absorption coefficients, 2015, Abgerufen am 26. 10. 2015, <http://www.sengpielaudio.com/calculator-RT60Coeff.htm>.

South, Phil (2015): The Amiga 3000 – a VR pioneer, 2015, Abgerufen am 12. 10. 2015, <http://www.vrexperience.net/Newsroom.aspx?id=120>.

Sterling, Bruce (2009): Augmented Reality: "The Ultimate Display" by Ivan Sutherland, 1965, 2009, Abgerufen am 12. 10. 2015, <http://www.wired.com/2009/09/augmented-reality-the-ultimate-display-by-ivan-sutherland-1965/>.

Sterne, Jonathan (2010). „The World Soundscape Project“, <http://superbon.net/?p=1638> , Abgerufen am 15.7.2015

Sutherland, Ivan E. (1965): The Ultimate Display, Proceedings of IFIP Congress, pp. 506-508, 1965. Abgerufen am 11.10.2015, <http://worrydream.com/refs/Sutherland%20-%20The%20Ultimate%20Display.pdf>

Thompson, Emily (2002): „The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America“, 2002, Abgerufen am 30. 09. 2015, <http://video.mit.edu/watch/the-soundscape-of-modernity-architectural-acoustics-and-the-culture-of-listening-in-america-1900-1-9861/>.

Truax, Barry, „The World Soundscape Project“, <http://www.sfu.ca/~truax/wsp.html>, Abgerufen am 13.9.2015

Tsay, C. (2014): Why Harley-Davidson's trademark sound is its biggest challenge. Strategist Magazine, Abgerufen am 1.6.2015, <http://strategistmagazine.co/chuantsay/harley-davidsons-trademark-sound-biggest-challenge-133>

TSL (2013): SPS200 Software Controlled Microphone User Guide. Abgerufen am 1.11.2015, <http://www.tslproducts.com/wp-content/uploads/2013/05/SPS200.pdf>

Vrs.org.uk, (2015): VPL Research Jaron Lanier | Virtual Reality Site, 2015, Abgerufen am 11. 10. 2015, <http://www.vrs.org.uk/virtual-reality-profiles/vpl-research.html>.

Weisstein, Eric W.: "Hadamard Matrix." From MathWorld--A Wolfram Web Resource. Abgerufen am 02. 10. 2015, <http://mathworld.wolfram.com/HadamardMatrix.html>

Westerkamp, Hildegard (1991). „The World Soundscape Project“, Abgerufen am 15.9.2015, [http://wfae.proscenia.net/library/articles/westerkamp\\_world.pdf](http://wfae.proscenia.net/library/articles/westerkamp_world.pdf).

viamichelin.de (2013): Jaguar F-Type V6, V6S und V8S. Abgerufen am 1.6.2015, <http://www.viamichelin.de/web/Report-age?id=35baba13101169eec126cf56cac231b9>.

„World Forum for Acoustic Ecology“ Webiste, abgerufen. 13.9.2015, <http://wfae.proscenia.net/>.

The World Technology Network, (2005): 2004 World Technology Awards Winners & Finalists: Marcos Novak, 2005, Abgerufen am 14. 10. 2015, <http://www.wtn.net/2004/bio212.html>.

Zotter, Franz (2013): Spherical Harmonics up to degree 5, as used in fifth-order Ambisonics. Abgerufen am 1.11.2015, [https://commons.wikimedia.org/wiki/File:Spherical\\_Harmonics\\_deg5.png](https://commons.wikimedia.org/wiki/File:Spherical_Harmonics_deg5.png)

