

# Masterarbeit

Objektorientierte Audioumgebungen für Surround-Sound

im Studiengang Elektronische Medien Master - Schwerpunkt Medientechnik  
an der Hochschule der Medien Stuttgart  
Wintersemester 2012/2013

vorgelegt von: Stephan Stahl

Matrikel-Nr.: 23303

Prüfungdatum: 28.02.2013

Erstprüfer: Prof. Jens-Helge Hergesell

Zweitprüfer: Prof. Oliver Curdt



Hiermit versichere ich, *Stephan Stahl*, an Eides Statt, dass ich die vorliegende Masterarbeit mit dem Titel: „*Objektorientierte Audioumgebungen für Surround-Sound*“ selbständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der eidesstattlichen Versicherung und die prüfungsrechtlichen Folgen (§ 19 Abs. 2 Master-SPO der HdM) sowie die strafrechtlichen Folgen (gem. § 156 StGB) einer unrichtigen oder unvollständigen eidesstattlichen Versicherung zur Kenntnis genommen.

Ort, Datum \_\_\_\_\_

Unterschrift \_\_\_\_\_



### *Objektorientierte Audioumgebungen für Surround-Sound*

Die konstante Weiterentwicklung neuer Audio-Wiedergabesysteme ermöglicht eine zunehmend qualitative Verbesserung der Ton-Produktion, der Ton-Reproduktion und somit der Klangwahrnehmung. Dabei werden entlang der neuen Wiedergabetechnologien auch neue Anforderungen an die Audioumgebungen gestellt: Angefangen bei Mono über Stereo und 5.1-Surround bis hin zu aktuellen dreidimensionalen Surround-Wiedergabeverfahren, haben sich in den letzten Jahrzehnten immer komplexer werdende Systeme entwickelt, die nicht zwingend kompatibel zueinander sind und stets eine Anpassung der Arbeitsprozesse und der technischen Peripherie erfordern.

Im Rahmen der vorliegenden Master-Arbeit wird untersucht, ob eine maximale Kompatibilität zwischen Audioinhalten und existierenden Wiedergabeverfahren realisiert, und inwiefern eine Aufwärtskompatibilität zu zukünftigen Verfahren gewahrt werden kann. Zunächst werden die Arbeitsschritte Produktion, Speicherung und Wiedergabe als Bestandteile einer Audioumgebung definiert. Um Sie in den Kontext des aktuellen Stands der Technik zu setzen und um die Anforderungen aktueller Wiedergabesysteme darzulegen, werden die Audioumgebungen im ersten Schritt entlang ihrer Entstehungsgeschichte beleuchtet und die Bestandteile der Verfahren charakterisiert.

Entlang der ermittelten Anforderungen wird anschließend aufgezeigt, ob und inwiefern die Teilbereiche durch einen objektorientierten Ansatz optimiert werden können. Da bisher kaum Lösungen zur übergreifenden Integration der Objektorientierung in Audioumgebungen existieren, werden die dafür notwendigen Anpassungen der Bestandteile aufgeführt. Im Rahmen der Abschlussarbeit wird zudem die Entwicklung eines Prototyps für die Audioproduktion aufgezeigt, der den objektorientierten Ansatz exemplarisch darlegt.

Für die Speicherung objektorientiert produzierter Audioinhalte ist ein Trägermedium zwingend notwendig. Aus diesem Grund werden bestehende Dateiformate aufgeführt, auf einen objektorientierten Ansatz untersucht sowie deren technische Potenziale und Restriktionen dargelegt.

Abschließend werden objektorientierte Wiedergabeverfahren aufgezeigt, welche die unmittelbare Schnittstelle zum Zuhörer abbilden. Sie dienen letztlich der Wiedergabe der in der Audioproduktion erzeugten und über Dateiformate gespeicherten Audioinhalte.

In der vorliegenden Masterarbeit wird die Notwendigkeit objektorientierter Lösungsansätze in Bezug auf aktuelle und zukünftige Audioumgebungen aufgezeigt und weiterhin technische und wirtschaftliche Herausforderungen in Bezug auf den Erfolg solcher Systeme dargelegt.



*Object-oriented audio environments for surround-sound*

The constant development of new audio playback systems allow increasing qualitative improvements concerning audio-production, re-production and hence the perception of audio content in general. Along the new playback technologies, new requirements towards audio environments arise: From mono to stereo across 5.1-surround up to the latest three-dimensional surround techniques, systems with increasing complexity were developed over the last decades. Compatibility between these systems is not coercively possible and therefore workflows and peripheral equipment need to be adapted to fulfill new requirements.

The presented master thesis therefore contains a research that reveals if a maximum compatibility between audio content and existing playback techniques can be realized, and in how far upward compatibility can be provided between present and future systems. At first the processes production, storage and playback are defined as components of a general audio environment. In order to clarify the recent state of technology and in order to present the requirements of recent playback systems, the audio environments of the last decades are presented according to their historic appearance.

Along the determined requirements an object-oriented approach for improving the components of audio environments is presented. As hardly any solutions for comprehensive integration of object-oriented approaches exist, the necessary customizations of the components of audio environments are shown. Additionally, the development of a prototype for audio production is presented, in order to demonstrate the object-oriented approach.

A carrier medium is absolutely mandatory in order to store the object-oriented audio content. On this account, existing file formats are listed and analyzed concerning the object-oriented approach as well as their technical potentials and restrictions.

Finally object-oriented playback systems are specified. They are used to decode the object-oriented files and consequently represent the immediate interface to the listener, since they transmit audio content.

The presented master thesis shows that object-oriented approaches in audio environments are necessary for future developments. Furthermore technical and economic challenges for the success of these systems are discussed.



## Inhaltsverzeichnis

1	Objektorientierte Audioumgebungen.....	13
1.1	Gegenstand der vorliegenden Arbeit .....	14
2	Die Schall(re)produktion.....	17
2.1	Räumliches Hören .....	17
2.1.1	Horizontalebene .....	17
2.1.2	Vertikalebene.....	18
2.1.3	Horizontal- und Vertikalebene .....	19
2.2	Evolution der Schallspeicherung und -Wiedergabe .....	19
2.2.1	Monophonie.....	20
2.2.2	Stereophonie .....	21
2.2.3	Surround .....	26
2.2.4	Auro 3D .....	29
2.3	Motivation und Vorteile von Surround-Formaten.....	31
2.4	Probleme bei der Wiedergabe von Surround .....	32
2.4.1	Psychophysiologische Betrachtung.....	32
2.4.2	Wirtschaftlich-technische Betrachtung .....	33
2.4.3	Anatomische Betrachtung .....	34
2.4.4	Unausgewogene Abhör-Räumlichkeiten .....	35
2.4.5	Trägermedien und Wiedergabehardware.....	36
3	Objektorientierung in Audioumgebungen .....	39
3.1	Reflexion .....	39
3.1.1	Invariable Faktoren.....	39
3.1.2	Variable Faktoren.....	39
3.2	Objektorientierter Ansatz .....	40
3.2.1	Objektorientierung – Begriffserklärung .....	40
3.2.2	Theorie: Objektorientierte Audioumgebungen .....	41
3.2.3	Bestandteile einer objektorientierten Audioumgebung .....	43

3.3	Zielgruppenanalyse .....	44
3.3.1	Produktion.....	44
3.3.2	Speicherung.....	44
3.3.3	Wiedergabe.....	45
4	Objektorientierte Audioproduktion (OOAP) .....	47
4.1	Workflowanalyse einer Audioproduktion .....	47
4.2	Theoretischer Lösungsansatz .....	48
4.2.1	Rückbesinnung auf das räumliche Hören .....	49
4.3	Konzept & Entwicklung eines Prototyps .....	50
4.3.1	Anforderungen eines objektorientierten Kanals.....	50
4.3.2	Untersuchung: Einbindung in Produktionsumgebungen .....	51
4.3.3	Programmierungsumgebung Max for Live: Kombination von Ableton Live und Max/MSP.....	51
4.3.4	Max for Live-Patch .....	52
4.4	Ansteuerung des Renderers.....	54
4.4.1	Übermittlung der Audio-Daten .....	55
4.4.2	Übermittlung der Panorama-Daten .....	61
4.5	Panorama-Eingabemethoden.....	62
4.5.1	Zweidimensionale Positionierung.....	63
4.5.2	Dreidimensionale Positionierung.....	64
4.6	Vorteile der OOAP.....	66
4.6.1	Lautsprechersituationen .....	66
4.6.2	Panorama-Versionen .....	67
5	Objektorientierte Datenspeicherung (OODS).....	69
5.1	Dateiformate bei der OOAP.....	69
5.2	Konzept: Verbindung der Audioinhalte und Panoramainformation .....	70
5.3	Multichannel interleaved WAVE.....	70
5.4	Containerformate .....	74

5.5	MPEG-4.....	75
5.5.1	MPEG-4 AAC (Audiostandard).....	76
5.5.2	MPEG-4 ALS (Audiostandard).....	77
5.5.3	MPEG-4 SA (Audiostandard).....	77
5.5.4	MPEG-4 BIFS (Szenenbeschreibung).....	77
5.6	MPEG-SAOC.....	78
6	Objektorientierte Wiedergabeverfahren (OOWV).....	81
6.1	Rendering – Vorüberlegungen.....	81
6.2	Optimiertes 5.1-Surround durch OO.....	82
6.3	Vector Base Amplitude Panning.....	83
6.4	Ambisonics.....	86
6.5	Wellenfeldsynthese.....	88
6.6	IOSONO.....	91
6.7	Shure Atmosphaea.....	94
6.8	Dolby Atmos.....	96
7	Objektorientierte Audioumgebungen – Epilog.....	99
7.1	Risiken und Probleme der OOAU.....	99
7.1.1	Produktion & Stakeholder der OOAP.....	99
7.1.2	Dateispeicherung & Stakeholder der OODS.....	101
7.1.3	Wiedergabe & Stakeholder der OOWV.....	102
7.2	Chancen und Potenziale der OOAU.....	103
7.2.1	Raum-unabhängiger Lautsprecheraufbau.....	103
7.2.2	Eingriff in das Klanggeschehen.....	103
7.3	Zukünftige Forschungsarbeiten.....	104
7.4	Perspektiven der OOAU.....	105
8	Danksagung.....	107

## Inhaltsverzeichnis

---

9	Quellenverzeichnis.....	109
9.1	Abbildungsverzeichnis.....	109
9.2	Tabellenverzeichnis .....	110
9.3	Literaturverzeichnis .....	111

# 1 Objektorientierte Audioumgebungen

Ob Konzert, Oper, Interview, einfaches Gespräch oder Tonträgerwiedergabe – der darin verankerte Kommunikationsprozess erfolgt immer nach dem Modell von Shannon-Weaver [1, 2]: Ein Sender verschickt eine Nachricht über ein Übertragungsmedium zu einem Empfänger, der dieses aufnimmt und verarbeitet. Auf die Audiotechnik übertragen entspricht der Sender einer Schallquelle und der Empfänger dem Zuhörer. Der Übertragungskanal kann dabei verschiedene Ausprägungen annehmen: So kann ein Straßenmusiker allein mit seinem Instrument über das Medium Luft seine Nachricht zu seinen Zuhörern transportieren. Die technischen Entwicklungen der letzten zwei Jahrhunderte haben jedoch dazu beigetragen, den Übertragungsweg zu erweitern: So können künstlerische Aufführungen auf verschiedenste Art und Weise gespeichert, transportiert und somit über räumliche und zeitliche Grenzen hinweg zeitversetzt miterlebt werden.

Gerade das Festhalten der Schallquelle auf Datenträgern hat sich auf beiden Seiten – sowohl bei Sender als auch Empfänger – über Jahrzehnte hinweg entwickelt und etabliert und verschiedene Technologiestadien durchlaufen. Angefangen bei der Aufzeichnung auf wachsbeschichteten Zinkplatten oder verschiedenen Magnetbandträger, bis hin zu digitalen Formaten mit unterschiedlichen haptischen Trägermedien, sind wir heute bei einem nicht mehr greifbaren Trägermedium angelangt: Digitale Medieninhalte, die aus verschiedensten Quellen (sowohl legal als auch illegal) bezogen werden können, stellen neue Herausforderungen an die Geschäftsmodelle der Musik- und Medientreibenden dar, ermöglichen aber auch den Fortschritt technischer Entwicklungen, klanglich-qualitative Verbesserungen zwischen Sender und Empfänger und eine schnelle Distribution von produzierten Medieninhalten.

Der Ablauf der Audioproduktion, Speicherung und Wiedergabe folgt immer demselben Prinzip: mit einer definierten Anzahl Mikrofone wird die Schallquelle eingefangen und in elektrische Signale umgewandelt. Im Produktionsprozess werden diese elektrischen Signale aufgezeichnet und aufbereitet. Für die Wiedergabe beim Zuhörer werden die Signale am Ende des Produktionsprozesses auf einem Trägermedium gespeichert. Ein Wiedergabesystem interpretiert die gespeicherten elektrischen Signale und reicht sie an ein Lautsprechersystem weiter, das die elektrischen Signale zurückwandelt und somit die Inhalte über das Medium Luft zum Zuhörer transportiert.

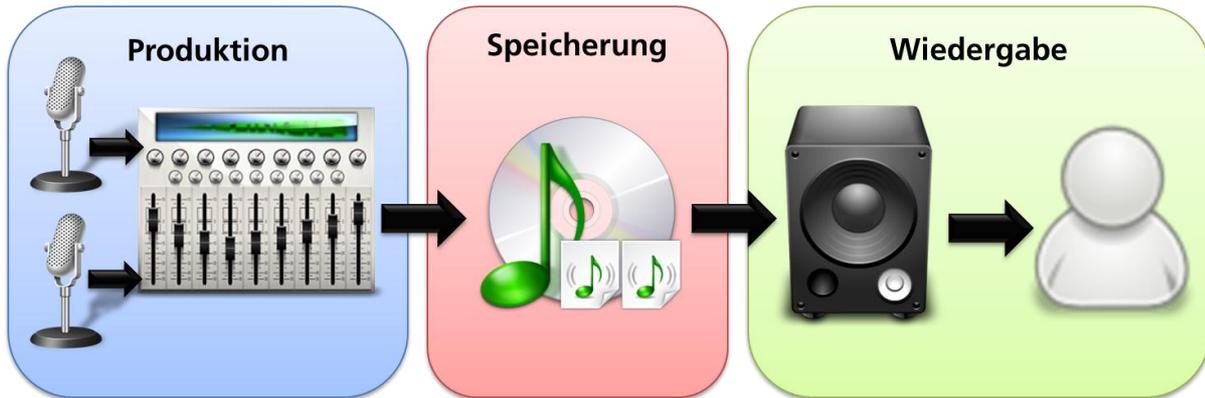


Abb. 1.1 Traditionelle Audioumgebung: Produktion, Speicherung und Wiedergabe

## 1.1 Gegenstand der vorliegenden Arbeit

Mit steigender Lautsprecheranzahl werden neue Anforderungen an die Audioproduktion, Datenspeicherung und Wiedergabe gestellt. Die aufgeführten Teilbereiche lassen sich unter dem Begriff Audioumgebungen zusammenfassen und werden in den folgenden Kapiteln genauer untersucht. Dabei sollen folgende Fragestellungen beantwortet werden:

- Wie sind aktuelle mehrkanalige Audioumgebung aufgebaut und was sind deren Bestandteile?
- Wie kann eine maximale Kompatibilität zwischen den Bestandteilen der Audioumgebungen und zukünftigen variablen Wiedergabeformaten erreicht werden?
- Was sind dabei die Anforderungen an die Audioumgebungen?
- Wie können Audioumgebungen modifiziert werden, um diese Anforderungen zu erfüllen?

Um die grundsätzliche Funktion von Audioumgebungen zu verdeutlichen, werden in Kapitel 0 zunächst die Grundlagen der Schallerzeugung, der Schallwahrnehmung und der Schallreproduktion dargelegt. Dies beinhaltet sowohl das räumliche Hörvermögen des Menschen als auch Strategien zur Übertragung von Schall zum Gehör über verschiedene Wiedergabesysteme. Weiterhin werden für das Verständnis notwendige Begriffe und Verfahren erläutert und die sich ständig ändernden Anforderungen an Audioumgebungen durch einen geschichtlichen Anriss dargestellt. Ferner werden Chancen und Risiken bestehender Wiedergabesysteme aufgeführt.

Anhand der gewonnenen Erkenntnisse wird in Kapitel 3 dargelegt, wie Audioumgebungen durch einen objektorientierten Ansatz modifiziert werden können, um die aufgeführten Anforderungen zu erfüllen. Dabei wird die Objektorientierung in den Kontext der Audiotechnik gestellt und die Chancen und Risiken der bestehenden Wiedergabesysteme reflektiert, um die

technische Relevanz der objektorientierten Audioumgebungen zu definieren. Um die Nutzergruppen von objektorientierten Audioumgebungen darzustellen, wurde zudem eine Zielgruppenanalyse durchgeführt.

Um den objektorientierten Ansatz auf Audioumgebungen zu übertragen, wurden die einzelnen Bestandteile ausführlich untersucht und entsprechende Anpassungen vorgenommen. Beginnend mit der Produktion werden in Kapitel 4 bisherige Produktionsprozesse anhand einer Workflowanalyse veranschaulicht. Entlang den ermittelten Anforderungen an objektorientierte Produktionsumgebungen wird anschließend die Entwicklung eines Prototyps beschrieben, der zur Aufzeichnung von Rauminformationen dient und im Rahmen der Abschlussarbeit entstand. Zudem werden alternative Eingabemethoden für die Positionierung von Schallquellen im Raum skizziert.

Die produzierten Audioinhalte müssen zunächst auf ein Trägermedium gespeichert werden, um Sie zum Zuhörer zu transportieren. In Kapitel 5 werden Dateiformate aufgeführt, die eine objektorientierte Speicherung der Inhalte ermöglichen. Eine kurze Analyse stellt dabei die Möglichkeiten und Grenzen der aufgeführten Dateiformate dar.

Objektorientierte Wiedergabeverfahren sind in der Lage, die gespeicherten Audioinhalte aufzubereiten und sie an Lautsprechersysteme zu übertragen. In Kapitel 6 werden theoretische Modelle zur Berechnung der Lautsprechersignale beschrieben und bestehende objektorientierte Wiedergabeverfahren vorgestellt.

Abschließend werden die Erkenntnisse der vorliegenden Arbeit präsentiert und mögliche Chancen und Risiken der objektorientierten Audioumgebungen dargestellt. Dabei werden zukünftige Potenziale und Risiken der objektorientierten Audioumgebungen charakterisiert und die Bindung der Nutzergruppen und unmittelbaren Beteiligten verdeutlicht.



## 2 Die Schall(re)produktion

### 2.1 Räumliches Hören

Jeder Empfänger bzw. Zuhörer kann über sein Gehör Schallquellen im Raum wahrnehmen. An dieser Stelle soll zunächst zwischen Schallereignissen und Hörereignissen unterschieden werden [3]: Schallereignisse sind die tatsächlich von einer Schallquelle ausgehenden Signale. Hörereignisse sind die von unseren Ohren an das Gehirn weitergeleitete und interpretierte Signale. Bei dieser Interpretation können verschiedene Aussagen über das Hörereignis getroffen werden: So können laut und leise unterschieden, Tonhöhe und Tondauer des Hörereignisses bestimmt, Rückschlüsse auf Entfernung getroffen und schließlich auch die räumliche Position – die Richtung der Schallquelle zum Zuhörer – annähernd genau bestimmt werden. Diese räumliche Wahrnehmung von Schallquellen wird als Lokalisierung bezeichnet und lässt sich über ein kopfbezogenes Koordinatensystem darstellen:

Ein lokalisiertes Schallereignis lässt sich über die Polarkoordinaten Azimuth (Winkel in der Horizontalebene), Elevation (Winkel in der Medianebene) und der Entfernung (Radius zur Schallquelle) in ein mathematisches Modell übertragen [4]. Durch die Verbindung dieser drei Polarkoordinaten lässt sich ein Schallereignis in einem dreidimensionalen Raum bzw. in einem Koordinatensystem mit X-, Y- und Z-Achse darstellen.

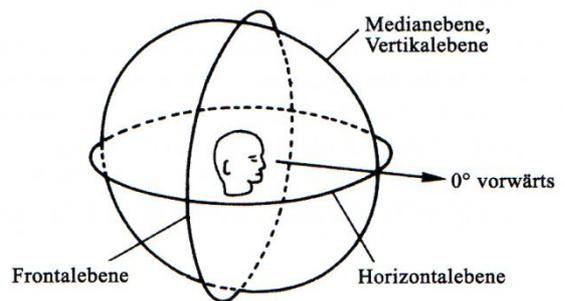


Abb. 2.1 Kopfbezogenes Koordinatensystem  
(Quelle: [5] S. 109)

Die Lokalisation eines Schallereignisses erfolgt binaural („zwei-Ohrig“) [6] – für eine annähernd genaue Lokalisation sind also immer zwei Ohren notwendig. Die Lokalisation selbst erfolgt auf Horizontalebene und Vertikalebene, allerdings auf unterschiedliche Weise:

#### 2.1.1 Horizontalebene

Trifft Schall seitlich auf den Zuhörer ein, so kann dieser über Laufzeit- und Pegeldifferenzen auf die Richtung in der Horizontalebene schließen. Laufzeitdifferenzen ergeben sich, wenn ein Ohr der Schallquelle zugewandt (und somit näher) und das andere Ohr der Schallquelle abgewandt (und somit weiter entfernt) ist. Der Schall trifft entsprechend zuerst am zugewandten Ohr und anschließend am abgewandten Ohr ein.

## 2 Die Schall(re)produktion

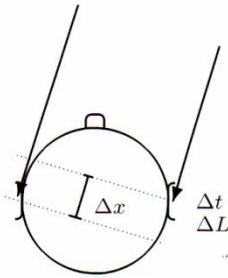


Abb. 2.2 Seitlich eintreffende Schallereignisse in der Horizontalebene (Quelle: [6] S. 11)

Pegeldifferenzen entstehen ebenfalls bei seitlich eintreffendem Schall: am Ohr, das der Schallquelle zugewandt ist, trifft ein höherer Pegel als am abgewandten Ohr ein. Am zugewandten Ohr entsteht ein Druckstau, am abgewandten Ohr entsteht (je nach Einfallswinkel des Schallereignisses) ein Schallschatten durch den Kopf des Zuhörers [6].

Laufzeitdifferenzen und Pegeldifferenzen bilden die Grundlage der Stereophonie und werden in diesem Kontext auch als Laufzeitstereophonie und Intensitätsstereophonie bezeichnet (siehe dazu Kapitel 2.2.2 Stereophonie). Da sowohl Laufzeit- als auch Pegeldifferenzen einen Effekt beschreiben, der „zwischen den Ohren des Zuhörers“ entsteht, werden ihnen die Begriffe Interaural Level Difference (ILD) für Pegeldifferenzen und Interaural Time Difference (ITD) für Laufzeitdifferenzen zugewiesen.

### 2.1.2 Vertikalebene

Die Lokalisation einer Schallquelle in der Vertikalebene erfolgt nicht wie bei der Horizontalebene über Pegel- und Laufzeitdifferenzen: Trifft Schall in der Horizontalebene „mittig“ (also in einem Winkel von  $0^\circ$  auf den Zuhörer) ein, gibt es keine Laufzeit- oder Pegelunterschiede – das eintreffende Signal ist auf beiden Ohren gleich.

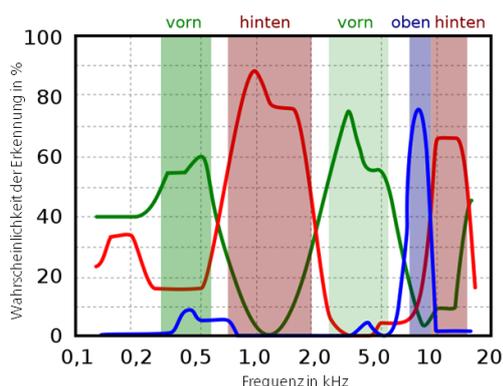


Abb. 2.3 Richtungsbestimmende Frequenzbänder bei Schalleinfall aus der Vertikalebene (Quelle: [7])

Die Lokalisation in der Vertikalebene erfolgt über die (Deutung der) Frequenzanteile eines Signals. Das Außenohr bildet dabei zusammen mit dem Gehörgang ein Resonatorsystem [7]: Der eintreffende Schall unterscheidet sich je nach eintreffendem Winkel in seiner Klangfarbe, die vom Resonatorsystem erzeugt wird, und kann je nach Ohrform eine individuelle Prägung aufweisen. Die Lokalisation der Schallquelle ist demnach ungenau und von den sogenannten richtungsbestimmenden Frequenzbändern abhängig [8] [5].

Diese richtungsbestimmenden Frequenzbänder wurden u.a. von J. Blauert genauer untersucht [8]. Seinen Untersuchungen zu Folge kann das menschliche Gehör bei verschiedenen Frequenzen eine Aussage über die Lokalisation einer Schallquelle treffen kann. Wie in Abb. 2.3 dargestellt, kann je nach Frequenzbereich zu einer gewissen Wahrscheinlichkeit bestimmt werden, ob das Signal von vorne, oben oder hinten auf den Zuhörer eintrifft. In diesem Zusammenhang gilt es zu beachten, dass es Frequenzbereiche gibt, die nur mit sehr geringer Wahrscheinlichkeit lokalisiert werden können. Signifikant ist die schlechte Lokalisation von oben eintreffenden Schallquellen, die lediglich im Frequenzbereich zwischen 5.000 Hz und 11.000 Hz annähernd genau lokalisiert werden können.

### **2.1.3 Horizontal- und Vertikalebene**

Durch die Lokalisation über Horizontal- und Vertikalebene können demnach Aussagen über die Position einer Schallquelle im dreidimensionalen Raum getroffen werden. Das daraus resultierende räumliche Hören dient als Grundvoraussetzung für die Wahrnehmung und Beurteilung von Raumklang. Die teils schlechte Lokalisationsfähigkeit in der Vertikalebene muss bei dreidimensionalen Wiedergabeverfahren zwingend beachtet werden. Ein Schallereignis setzt sich in der Regel aus Grund- und verschiedenen Obertönen zusammen. Daher ist es denkbar, dass verschiedene Schallereignisse – je nach Obertonspektrum – besser oder schlechter lokalisiert werden können.

## **2.2 Evolution der Schallspeicherung und -Wiedergabe**

Die Idee, eine Schallquelle über ein zwischengeschaltetes (technisches) Medium zum Empfänger zu transportieren, ist unter anderem auf die Erfindung des Telefons zurückzuführen. Anknüpfend an diese Idee wurde 1881 auf der „La Lumière Électrique“ ein interessanter Versuchsaufbau dargeboten: In einem Opernhaus wurden 8 Mikrofone nahe eines Orchesters aufgestellt. Die Signale der Mikrofone wurden paarweise an vier Zuhörerräume weitergeleitet, in denen mögliche Probanden das dargebotene Konzert mit je zwei Telefonhörern verfolgen konnten. Einer der Zuhörer bemerkte bereits damals folgenden Effekt: „[...] listening with both ears at the two telephones, the sound takes a special character of relief and localization which a single receiver cannot produce.“ [9] Auch wenn Clement Ader, der Verantwortliche für den damaligen Versuchsaufbau, sich nicht zwangsweise darüber bewusst war, wurde in diesem Versuchsaufbau der Grundstein für das weit später folgende Stereophonie-Verfahren gelegt (siehe Kapitel 2.2.2 Stereophonie). Der vom Zuhörer beschriebene Effekt lässt sich auch heute noch beim Direktvergleich zwischen monophonen und stereophonen Aufnahmen feststellen und eindrucksvoll reproduzieren.

## 2 Die Schall(re)produktion

---

Auch wenn dies bereits die Anfänge stereophoner Wiedergabe und Klangübertragung darstellte, dauerte es doch bis 1933 bis Blumlein das Patent für Stereo anmeldete [10] [11]. Auch lange nach der Anmeldung des Stereopatents war die monophone Wiedergabe das etablierte Wiedergabeverfahren auf dem Markt der Konsumenten.

### 2.2.1 Monophonie

Die Wortkombination aus Mono (griech. „einzig“) und phonie (griech. „Klang“) beschreibt bereits die Strategie hinter dem Wiedergabeverfahren: aufgenommene Schallereignisse werden über einen einzigen Lautsprecher zum Hörer übertragen. Zu Beginn der Monophonie war der Weg der Übertragung noch 1 zu 1: Ein Mikrofon diente als einziges Aufnahmegerät und lieferte das Signal für die folgende Wiedergabe über einen einzigen Lautsprecher. Mit der Entwicklung von Mischpulten war es schließlich möglich die Signale mehrerer Mikrofone zusammenzuführen und das Verhältnis der abgenommenen Schallereignisse zueinander anzupassen. Die maximale Anzahl an möglichen Mikrofonen war entsprechend nur noch an die Ausstattung der Mischpulte gebunden. Die Relation verschob sich zu einer n:1-Beziehung.

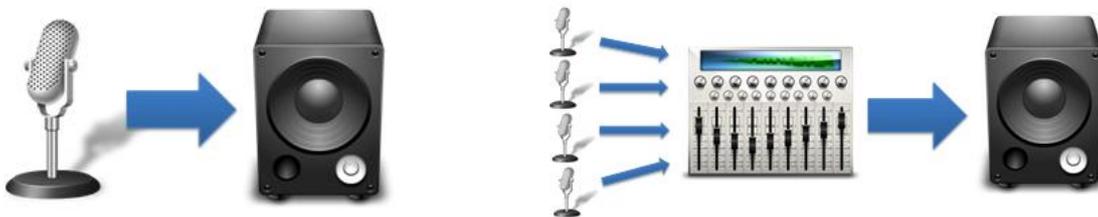


Abb. 2.4 Links: 1:1-Monophonie, rechts: n:1 Monophonie durch die Verwendung eines Mischpults

Bezogen auf Raumklang und räumliches Hören lassen sich bei der Monophonie folgende Aspekte betrachten: Der Zuhörer kann zwar die Schallquelle – bzw. im konkreten Fall den Lautsprecher – im (realen) Raum lokalisieren, jedoch keine Aussage über den „virtuellen“ Raum, also den Raum in dem die Schallereignisse aufgenommen wurden, treffen. Es kann lediglich bestimmt werden, ob das aufgenommene Schallereignis nahe oder weiter entfernt ist. Somit ist prinzipiell schon eine eindimensionale Tiefenstaffelung möglich. (Dieser Effekt lässt sich über einen künstlichen Hall simulieren: wird der Hallanteil auf dem Schallereignis erhöht, erscheint die Schallquelle weiter entfernt und vice versa.) Auch wenn somit eine eindimensionale Lokalisation in der Tiefe des Raums (und somit der Y-Achse) möglich ist, so ist die Monophonie für die Wiedergabe eines virtuellen Raumklanges ungeeignet, da sie nur eine einzige Dimension abbilden kann.

## 2.2.2 Stereophonie

Der Begriff Stereo hat ebenfalls einen griechischen Wortstamm und bedeutet in etwa „fest“ oder „starr“. Der Name Stereophonie entstand analog zur Namensgebung der Stereoskopie. Die Kurzform Stereo hat sich jedoch für die Wiedergabe von Schallquellen auf zwei Lautsprechern etabliert und ist auch heute noch das am meisten verbreitete Verfahren bei der Tonwiedergabe. Dabei werden die in Kapitel 2.1 beschriebenen Pegel- und Laufzeitdifferenzen ausgenutzt, um eine Schallquelle im Raum zu positionieren. Der Hörer kann dabei differenzieren, ob ein Signal aus Lautsprecher 1 oder Lautsprecher 2 erklingt, oder ob es aus beiden Lautsprechern zeitgleich wiedergegeben wird. Erfolgt eine zeitgleiche Wiedergabe auf beiden Lautsprechern, bildet sich eine sogenannte Phantomschallquelle. Das Prinzip der Phantomschallquellen wird im Folgenden genauer erläutert, zunächst soll jedoch der technische Aufbau eines stereophonen Wiedergabesystems untersucht werden.

### 2.2.2.1 Technischer Aufbau stereophoner Wiedergabesysteme

Bei stereophonen Wiedergabesystemen werden zwei Lautsprecher in einem definierten Abstand (der sogenannten Basisbreite) zueinander aufgestellt. Der Zuhörer bildet zusammen mit den beiden Lautsprechern ein gedachtes Dreieck. Er befindet sich daher nicht auf einer Linie mit den beiden Lautsprechern, sondern steht nach hinten versetzt mit Blick zu den Lautsprechern. So trivial diese Aufstellung zwischen Lautsprechern und Zuhörer auch sein mag, konnte sich in der Entwicklung von Stereo eine exakte Aufstellungsempfehlung herausstellen: Dabei bildet der Zuhörer mit den Lautsprechern ein gedachtes gleichseitiges Dreieck, der Öffnungswinkel beträgt bei dieser Aufstellung  $60^\circ$ , der Zuhörer steht also im Winkel von  $30^\circ$  bzw.  $-30^\circ$  zu den Lautsprechern.

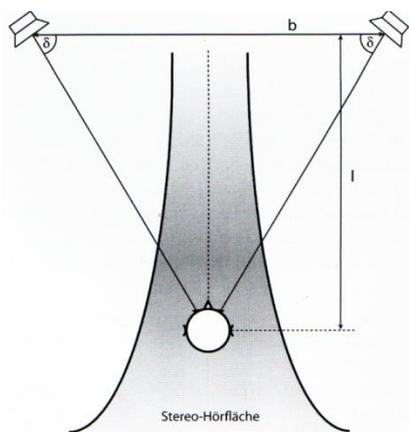


Abb. 2.5 Hörfläche bei der Stereophonie  
(Quelle: [11] S. 611)

Je weiter sich der Zuhörer von den Lautsprechern entfernt oder sich ihnen nähert, desto größer bzw. kleiner muss in der Theorie die Basisbreite der Lautsprecher gewählt werden. Sofern die Basisbreite der Lautsprecher konstant bleibt, lassen sich folgende Effekte feststellen: Wird der Abstand verkleinert (und somit der Öffnungswinkel vergrößert), vergrößert sich das Stereobild (und die wahrgenommene akustische Breite). Dabei kann es jedoch zu Instabilität bei der Wahrnehmung der Hörereignisorte kommen.

Weiterhin kann die Lokalisation der Schallquellen bei Kopfbewegungen des Zuhörers ungenau werden. Bei zu großem Abstand der Lautsprecher zueinander können sogar zwei Schallquellen getrennt oder fälschlicherweise „von oben“ wahrgenommen werden.

Wird dagegen der Abstand zu den Lautsprechern bei konstanter Basisbreite vergrößert (und somit der Öffnungswinkel verkleinert), wird das Stereobild eingeschränkt, das wahrgenommene Schallereignis „rückt zusammen“. Für die stereophone Wiedergabe existiert allerdings ein Toleranzbereich, der im Bereich zwischen  $45^\circ$ - $80^\circ$  des Öffnungswinkels liegt. Wird die Hörposition nicht über diesen Toleranzbereich hinaus verändert, treten nur bedingt Lokalisationsfehler auf [11].

### 2.2.2.2 Phantomschallquellen

Das menschliche Gehör kann – wie bereits beschrieben – über Laufzeit- und Pegelunterschiede Schallquellen lokalisieren. Wird nun auf Lautsprecher 1 oder Lautsprecher 2 ein Schallereignis wiedergegeben, kann dieses vom Zuhörer dem jeweiligen Lautsprecher genau zugeordnet werden. Wird ein Signal auf beiden Lautsprechern zeitgleich und in gleicher Lautstärke wiedergegeben, so erreichen den Zuhörer zwei Signale gleicher Lautstärke zur selben Zeit. Das menschliche Gehirn kann die beiden eintreffenden Signale nicht getrennt verarbeiten, wodurch der Eindruck entsteht, dass das Schallereignis von einer Schallquelle in der Mitte der beiden Lautsprecher ausgeht. Die so entstandene virtuelle Schallquelle wird als Phantomschallquelle bezeichnet.

#### **Platzierung der Phantomschallquelle über Laufzeitunterschiede:**

Wird das Signal nun auf dem rechten Lautsprecher zeitverzögert wiedergegeben, so verschiebt sich die Phantomschallquelle in Richtung des linken Lautsprechers. Die zeitliche Verzögerung des Signals wird auch Laufzeitstereophonie genannt.

#### **Platzierung der Phantomschallquelle über Pegelunterschiede:**

Wird der Pegel des Signals auf dem linken Lautsprecher erhöht und auf dem rechten gesenkt, verschiebt sich die Phantomschallquelle ebenfalls in Richtung des linken Lautsprechers. Das Anordnen von Phantomschallquellen über Pegeldifferenz wird als Intensitätsstereophonie bezeichnet.

**Mischung der beiden Verfahren:**

Wird eine Phantomschallquelle sowohl über Laufzeiten (Laufzeitstereophonie) als auch über Pegeldifferenzen (Intensitätsstereophonie) reproduziert, wird von Äquivalenzstereophonie gesprochen [6].

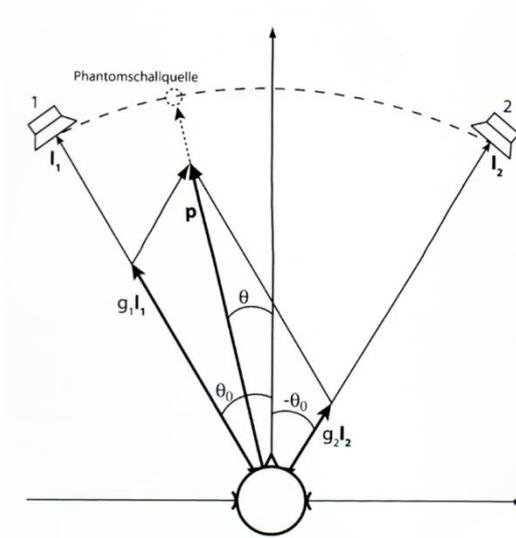


Abb. 2.6 Phantomschallquelle durch Kombination der Lautsprechersignale (Quelle: [11] S. 658)

Über Laufzeit-, Intensitäts- und Äquivalenzstereophonie lassen sich beliebige virtuelle Schallquellen auf der gedachten Achse zwischen den beiden Lautsprechern realisieren. Somit können verschiedene Signale auf der X-Achse des Raumes positioniert und vom Zuhörer lokalisiert werden. Über die Zuhilfenahme von Hall kann erneut eine Tiefenstaffelung geschaffen werden, die Stereophonie bildet also eine zweidimensionale Darstellung von (Phantom-) Schallquellen im Raum ab.

**Probleme bei der Lokalisierung von Phantomschallquellen:**

Die korrekte Abbildung einer Phantomschallquelle setzt allerdings voraus, dass sich der Zuhörer im vorgegeben Hörbereich – dem sogenannten Sweet Spot – der Lautsprecher befindet. Er muss sich dabei auf der gedachten Mittelachse (der Winkelhalbierenden des Öffnungswinkels) zwischen den beiden Lautsprechern befinden. Wird dieser Hörbereich durch eine Abweichung nach links oder nach rechts verlassen, verschieben sich die Phantomschallquellen gemäß dem Gesetz der ersten Wellenfront: Eine mittig angelegte Phantomschallquelle wandert bei einer Verschiebung der Hörposition nach links zum linken Lautsprecher, bei einer Verschiebung nach rechts wandert sie zum rechten Lautsprecher [11].

Weiterhin kann es zur verfälschten Wahrnehmung von Phantomschallquellen kommen, wenn der Zuhörer seinen Kopf dreht, bzw. die Lautsprecher seitlich (bspw. um 90° gedreht) zum Zuhörer stehen. Bei kleineren Drehungen und entsprechend kleinem Winkel zur gedachten Mittelachse treten kaum Lokalisationsfehler auf.

## 2 Die Schall(re)produktion

Je größer der Winkel zur Mittelachse wird, desto ungenauer wird die Lokalisation der Phantomschallquelle. Sobald der Zuhörer einen Winkel von 90° zur Lautsprecheranordnung einnimmt, ist eine Lokalisation nicht mehr sicher möglich. Die wahrgenommene Lokalisation der Phantomschallquelle „kippt“ zu einem der beiden Lautsprecher [12] [13].

### 2.2.2.3 Panning Law

Eine weitere Voraussetzung für die Bildung von Phantomschallquellen ist das sogenannte Panning Law. Dieses besagt, dass die wahrgenommene Lautstärke einer virtuellen Schallquelle beim Zuhörer stets gleich laut erscheinen soll, unabhängig davon wo sich die virtuelle Schallquelle befindet [14]. Technisch ausgedrückt muss die Summe der beiden akustischen Leistungen (und somit der Lautsprecher) stets konstant bleiben. Um dies zu erreichen, entstanden verschiedene mathematische Modelle zur Berechnung der Pegel der beiden Lautsprecher [12].

Die akustische Leistung verhält sich proportional zum Quadrat der elektrischen Spannung. Ein cosinus- bzw. sinusförmiger Verlauf der Spannungsfaktoren der Lautsprecher erfüllt ein solches Kriterium, daher lässt sich folgende Formel aufstellen:

$$\text{Formel 2.1: } g_1^2 + g_2^2 = \sin(\alpha)^2 + \cos(\alpha)^2 = 1$$

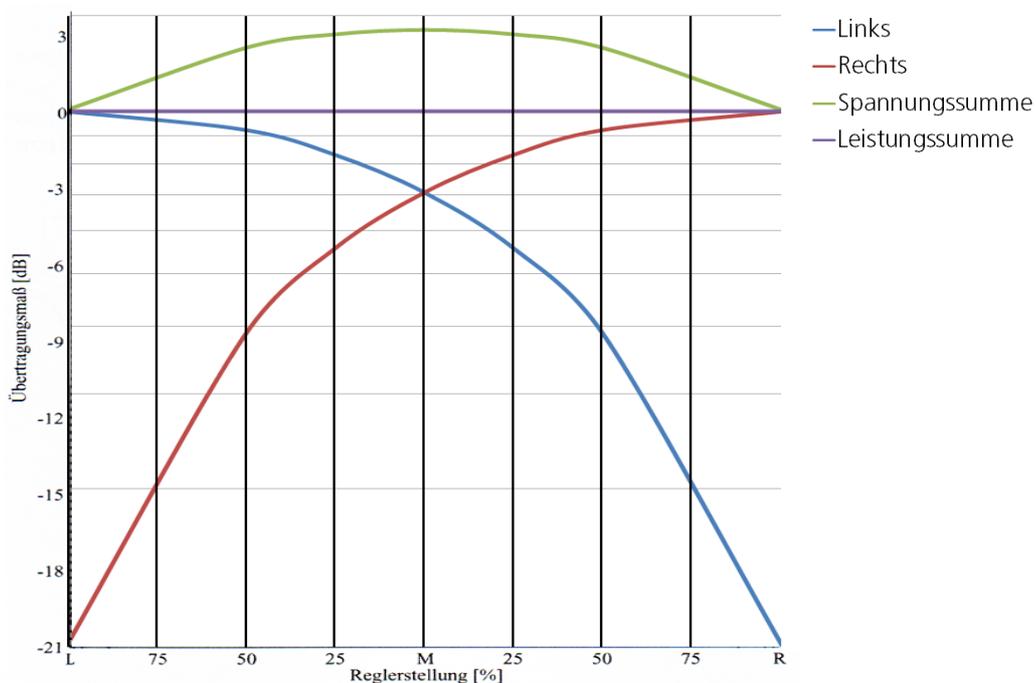


Abb. 2.7 Mathematische Approximation Panning Laws

Im konkreten Beispiel: Panoramaregler bei -3 dB in Mittelposition (Quelle: angelehnt an [15])

Die meisten Mischpulte haben zur Positionierung von Phantomschallquellen einen sogenannten Panorama-Regler. Mit diesem lassen sich Schallquellen nahtlos zwischen Lautsprecher 1 und 2 (inklusive den Zwischenpositionen, also den Phantomschallquellen) hin- und her-positionieren. Die Panorama-Regler verwenden dabei oftmals den Ansatz der Intensitätsstereophonie und folgen strikt den Regeln des Panning Laws. Hier gilt allerdings festzuhalten, dass Mischpulthersteller unterschiedliche Berechnungen des Panning Laws verwenden. So bevorzugen die einen eine Absenkung um 3 dB in der Phantom-Mitte, andere bevorzugen eine Absenkung um 6 dB. Manche Hersteller haben sich auf den Mittelwert von -4,5 dB festgelegt.

#### **2.2.2.4 Stereo-Aufnahmeverfahren**

Die Entwicklung der stereophonischen Wiedergabe hat ebenso verschiedene Stereo-Mikrofonierungsverfahren hervorgebracht. Das Prinzip hinter den Aufnahmeverfahren ist dabei immer dasselbe: Zwei Mikrofone werden so platziert, dass Laufzeit- und/oder Intensitätsunterschiede aufgezeichnet werden. Wird das Signal des ersten Mikrofons auf den ersten Lautsprecher und das Signal des zweiten Mikrofons auf den zweiten Lautsprecher gegeben, bilden sich die Phantomschallquellen entlang des gewählten Mikrofonierungsverfahren (Intensitäts-, Laufzeit- oder Äquivalenzstereophonie). Im Sinne der stereophonen Aufnahme und Wiedergabe kann man von einem 2:2-Verfahren sprechen: N Schallquellen werden über zwei Schallwandler eingefangen, festgehalten und anschließend auf zwei Lautsprechern reproduziert.

Die verschiedenen Mikrofonierungsverfahren sollen an dieser Stelle nicht weiter erläutert werden, sind aber im Hinblick auf die geschichtliche Entwicklung der Aufnahme- und Klangreproduktion nicht zu vernachlässigen. (Ein ausführlicher Überblick über die verschiedenen Mikrofonierungsarten (bspw. A/B, X/Y und M/S) findet sich u.a. in [16].)

Werden – wie bei der Monophonie – separate Mikrofone für einzelne Schallquellen verwendet und an einem Mischpult zusammengeführt, ergibt sich analog dazu ein n:2-Verfahren.

#### **2.2.2.5 Historie**

Obwohl die Stereophonie eine neue räumliche Dimension aufspannen kann, dauerte es eine geraume Zeit, bis sich die Vorteile des Wiedergabeverfahrens auf Seiten der Audioproduktion kenntlich machten. Dies belegen unter anderem frühe Aufnahmen der Beatles, in denen die Stereophonie nicht dazu genutzt wurde einen virtuellen Raum in Breite und Tiefe abzubilden, sondern lediglich um eine Kanaltrennung zwischen den verschiedenen Instrumenten zu erreichen [17].

## 2 Die Schall(re)produktion

---

So wurde beispielsweise das Schlagzeug komplett auf den linken Kanal (und somit den linken Lautsprecher) positioniert, der Gesang auf den rechten Kanal (und somit rechten Lautsprecher). Die restlichen Instrumente wurden in verschiedenen Ausprägungen als Phantomschallquellen zwischen den Lautsprechern verteilt.

Aus wirtschaftlich-technischer Sicht war die Stereophonie jedoch nicht unproblematisch. So erfolgt in der kompletten Signalkette eine Verdoppelung der benötigten Peripherie, angefangen bei den Lautsprechern bis hin zum Übertragungsmedium, das zwei unabhängige Kanäle und somit die doppelte „Bandbreite“ bzw. den doppelten Speicherplatz bereitstellen muss. Dennoch konnte sich die Stereophonie auch durch den erzielten Effekt – das Aufspannen eines zweidimensionalen Raumes – beim Konsumenten durchsetzen.

Auch wenn sich mit der Stereophonie quasi unendlich viele Schallquellen im zweidimensionalen Raum darstellen lassen, ist es nicht möglich, eine Schallquelle hinter oder seitlich des Zuhörers zu positionieren. Um den Zuhörer weiter in das Klanggeschehen zu integrieren, wurden im Laufe des letzten Jahrhunderts verschiedene Versuche mit mehr als zwei Lautsprechern evaluiert, einige solcher sogenannten Surround-Verfahren konnten sich auf dem Konsumenten-Markt etablieren.

### 2.2.3 Surround

Der Begriff Surround lässt sich frei mit „umgebend“ übersetzen und beschreibt am besten die dabei verfolgte Wiedergabestrategie: Bei Surround-Wiedergabeverfahren ist der Zuhörer von Lautsprechern und den von ihnen erzeugten Schallereignissen umgeben.

Die Idee Schallquellen rund um den Zuhörer anzuordnen entstand jedoch nicht erst im 20. Jahrhundert. Bereits um 1830 experimentierte Hector Berlioz mit der Positionierung des Orchesterensembles: Er platzierte in der Aufführung von „Symphonie Fantastique“ die Bläser hinter dem Publikum, in „Requiem“ ordnete er vier Orchester den Himmelsrichtungen Nord, Ost, Süd und West zu [18]. Es bleibt zu vermuten, dass nicht nur Berlioz schon früh mit Zuhörer-umgebenden Schallquellen experimentierte.

Als maßgeblicher Meilenstein in der Geschichte von Surround-Wiedergabe wird in der Literatur meist die Vorführung des Films „Fantasia“ im Jahr 1940 genannt. Für die Filmvorstellung wurde eigens ein Soundsystem mit dem Namen „Fantasound“ [18] entwickelt und kann als Vorgänger-Verfahren des heute bekannten 5.1-Surround-Verfahrens betrachtet werden. Dabei wurden drei Frontlautsprecher (Links, Center, Rechts), zwei Front-Ecklautsprecher (LCo und RCo), zwei Seitenlautsprecher (L und R), zwei Rücklautsprecher (LS und RS) sowie ein Deckenlautsprecher verbaut [10]. Leider zeigte die Filmindustrie nur geringes Interesse am Fantasound-Konzept,

aufgrund dessen es nicht weiter verfolgt wurde. Anfang der 70er Jahre entstand das Quadrofonie-Verfahren, das erstmals vier unabhängige Kanäle für vier unterschiedliche Lautsprecher bereitstellen konnte. Aufgrund der mangelnden Kompatibilität zu Stereo und der erneuten Verdoppelung der Systemanforderungen (doppelte Lautsprecheranzahl, Endstufenanzahl und Bandbreite) konnte sich das Verfahren allerdings nicht am Markt etablieren [19].

Auch wenn die Anfänge von Surround-Wiedergabe von Misserfolgen durchzogen waren, gab die Kinoindustrie nicht auf. Von Fantasound über Quadrofonie ausgehend entwickelten sich verschiedene Formate seitens der Kanalübertragung und der Lautsprecheranordnung. Bei der Entwicklung dieser Verfahren waren und sind die Firmen Dolby und Digital Theater Systems maßgeblich beteiligt und zeigen auch aktuell neue und innovative Ansätze (siehe auch Kapitel 6.8 Dolby Atmos). Die einzelnen Verfahren und (Speicher-)Formate sollen an dieser Stelle nicht aufgeführt werden, festzuhalten bleibt allerdings, dass sich entlang der verschiedenen Verfahren explizit das 5.1-Surround Wiedergabeverfahren durchsetzen und auch auf dem Konsumentenmarkt etablieren konnte: 2006 verfügten ca. 35% der Haushalte in Deutschland, Österreich und der Schweiz über einen DVD-Spieler [11]. Die DVD bietet dabei als Trägermedium von Filmmaterial häufig die Möglichkeit 5.1-Surround-Sound wiederzugeben. Da der Besitz eines DVD-Spielers aber nicht auf das Vorhandensein eines 5.1-Surround-Systems schließen lässt, besteht die letzte Hürde im Zweifelsfall in der Anschaffung eines Lautsprechersystems. (Das Problem des Trägermediums DVD wird in Kapitel 2.4.5 aufgeführt.)

### 2.2.3.1 Technischer Aufbau von 5.1-Surround

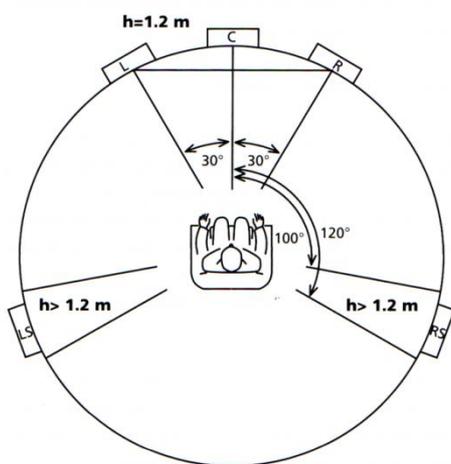


Abb. 2.8 Aufstellungsempfehlung für 5.1 Surround nach ITU-R BS.775 (Quelle: [19] S. 127)

Bei 5.1-Surround wird die stereophone Lautsprecheranordnung um einen Center-Lautsprecher, zwei weitere Rücklautsprecher (Links Surround und Rechts Surround) und einen separaten Lautsprecher zur Tieffrequenzwiedergabe („Low Frequency Effect“ – LFE) ergänzt. Die Namensgebung 5.1 bezieht sich also in diesem Fall auf die fünf den Zuhörer umgebenden Lautsprecher (bzw. die Lautsprecherkanäle) und den Subwoofer (respektive den Subwoofer-Kanal). Für die Positionierung der Lautsprecher konnte sich die Empfehlung ITU-R BS.775 durchsetzen.

## 2 Die Schall(re)produktion

---

Bei 5.1-Surround werden Schallquellen wiederum als Phantomschallquellen zwischen den Lautsprechern dargestellt, daher hat sich für 5.1-Surround der Begriff Mehrkanalstereophonie etabliert. Der Center-Kanal zwischen linkem und rechtem Lautsprecher dient der Stabilisierung des Mittensignals beim Filmtone: Wird bei der Stereophonie die Hörzone nach links oder rechts verlassen, verschiebt sich die Phantomschallquelle (siehe Kapitel 2.2.2.2). Der Center-Kanal ermöglicht – auch bei verlassen der Hörzone – eine Lokalisation der Schallquelle in der Mitte, bezogen auf den Filmtone also unmittelbar am Bild ausgerichtet. Gerade in Kinosälen würden sich Phantomschallquellen bei herkömmlichen Stereophonieverfahren für einen Großteil der Zuschauer verschieben. Der Center-Kanal findet daher meist Anwendung bei Dialog und explizit bildbezogenen Inhalten, die der Videofläche zugeordnet werden müssen.

Der LFE-Kanal darf nicht mit einem reinen Subwoofer im Heimgebrauch verwechselt werden: Er dient als reiner Effektkanal für die Tieftonwiedergabe im Kinobereich, und gibt nur ein frequenzbandbegrenztes Signal wieder. Subwoofer in Consumer-Anlagen kombinieren dagegen den LFE-Kanal mit Signalanteilen der anderen fünf Lautsprecher, die unter Umständen bauartbedingt nur ein gewisses Frequenzspektrum abdecken können und keine tieffrequenten Signalanteile wiedergeben können. Ein Subwoofer im Heimbereich kombiniert also die sechs Signale bis zu einer definierten Grenzfrequenz, ein Betrieb ohne Subwoofer wäre unter Umständen nicht möglich.

Ähnlich wie bei der herkömmlichen Stereophonie kann es bei 5.1-Surround zu Lokalisationsproblemen kommen: Bei seitlich eintreffendem Schall, also Phantomschallquellen zwischen L und LS bzw. R und RS ist eine exakte Lokalisation nicht möglich (siehe Kapitel 2.2.2.2). Dieses Problem lässt sich ansatzweise durch die Erweiterung des Systems auf 7.1-Surround durch zwei weitere Lautsprecher zwischen L und LS bzw. R und RS beheben. Da das Endformat allerdings lediglich 5 Lautsprechersignale bereitstellt, müssen die Signalanteile für die unterstützenden Kanäle berechnet werden: Sind die Signale der Lautsprecher L (bzw. R) und LS (bzw. RS) gleich, sprich befindet sich eine Phantomschallquelle exakt in der Mitte der beiden Lautsprecher, wird das Signal ausschließlich über den unterstützenden Lautsprecher wiedergegeben. Analog werden Schallquellen in den Zwischenpositionen berechnet. Dies kann allerdings nur als Kompromisslösung aufgeführt werden, da sich die Frequenzanteile der Phantomschallquellen in den Zwischenpositionen nicht exakt bestimmen lassen. (Bei Verfahren wie dem Center-Canceling werden die Lautsprechersignale auf gleiche Signalanteile untersucht. So können Phantomschallquellen in der Mitte der Lautsprecher (bspw. Gesang bei einer Audioproduktion) extrahiert werden.

Die Zwischenpositionen zwischen den Lautsprechern können nur ungenau extrahiert werden, da die Signalanteile auf den Lautsprechern in der Regel nicht bekannt sind bzw. nicht isoliert vorliegen, und sich mit weiteren Signalen bzw. Schallquellen mischen.)

Weiterhin ist es nicht immer möglich, Lautsprecher entlang der ITU-Empfehlung in Wohnräumlichkeiten zu positionieren. Der dabei entstehende Effekt soll separat in Kapitel 2.4.4 betrachtet werden.

## 2.2.4 Auro 3D

Auro 3D ist ein zum Zeitpunkt der Ausarbeitung aktuelles dreidimensionales Wiedergabeverfahren, das als noch nicht finalisiert betrachtet werden kann. Während der Erstellung der vorliegenden Arbeit haben sich die Bezeichnung des Systems und der entsprechende Aufbau von 9.1 über 10.1 bis hin zu 11.1 und 13.1 für Kinowiedergabe verändert. Der im Folgenden aufgeführte Aufbau soll daher exemplarisch für das Modell stehen.

### 2.2.4.1 Technischer Aufbau

Der technische Aufbau von Auro 3D erweitert ein bestehendes 5.1-Surround-Setup nach ITU-R BS.775 um weitere fünf Höhenlautsprecher. Vier davon sind kreisförmig jeweils über L, R, LS und RS positioniert, ein weiterer Lautsprecher wird dabei unmittelbar über dem Kopf des Zuhörers platziert [20].

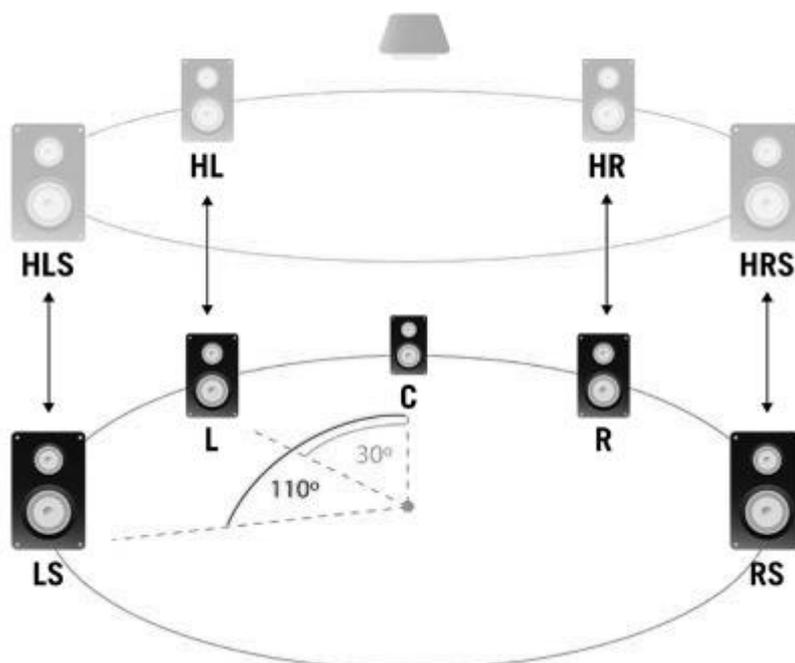


Abb. 2.9 Auro 3D 10.1, exemplarischer Aufbau (Quelle: [20])

## 2 Die Schall(re)produktion

---

Das Verfahren kann ebenfalls wie 5.1-Surround als mehrkanaliges Stereophonieverfahren betrachtet werden, da Schallquellen ebenfalls über Phantomschallquellen zwischen den Lautsprechern abgebildet werden. Hier gilt allerdings festzuhalten, dass diese nicht nur zwischen L und LS, sondern auch zwischen L und HL bzw. dreidimensional zwischen L, HL und LS positioniert werden können. Die Berechnung der Lautsprechersignale eines dreidimensionalen Modells wird in Kapitel 0 genauer aufgeführt.

Bisherige Produktionsumgebungen ermöglichen jedoch bisher keine nahtlose Positionierung zwischen den Lautsprechern und dem dreidimensionalen Aufbau: Bisher können lediglich zwei separate 5.0-Kanäle erzeugt und Signalanteile über Send-Wege (siehe auch Kapitel 4.2) auf die Lautsprecher (bzw. 5.0-Kanäle) gemischt werden. Diese Problemstellung soll an dieser Stelle nicht weiter erläutert werden, sie ist zentraler Bestandteil des Kapitels 4 Objektorientierte Audioproduktion (OOAP).

Prinzipbedingt ergeben sich bei Auro 3D dieselben Lokalisationsprobleme wie bei 5.1-Surround und der Stereophonie: Schallquellen, die seitlich auf den Zuhörer eintreffen, können im Zweifelsfall nicht korrekt lokalisiert werden (siehe Kapitel 2.2.2.2). Weiterhin muss bei Auro 3D das räumliche Hören des Menschen in der Vertikalebene beachtet werden (siehe Kapitel 2.1.2 Vertikalebene). Auch wenn Auro 3D eine dreidimensionale Abbildung von Audioinhalten ermöglicht, müssen diese – bedingt durch die Fähigkeiten des menschlichen Gehörs – nicht zwangsweise dreidimensional wahrgenommen werden.

Eine wirtschaftliche Analyse von Auro 3D wurde unter anderem in [21] durchgeführt. Hörversuche belegen dabei wahrnehmbare Unterschiede zwischen 5.1-Surround und Auro 3D. Benjamin Reichert führt dabei auf:

„[...] anhand] der durchgeführten empirischen Untersuchungen [ist] ein Nutzenzuwachs durch „Auro-3D“ unter den Teilnehmern auszumachen [...]. [...] zumindest im Kinobereich [resultiert daraus] eine vorhandene Preisbereitschaft der Teilnehmer.“ ([21] S. 105)

Eine weitere wirtschaftliche Einschätzung soll an dieser Stelle nicht erfolgen. Auro 3D führt dennoch exemplarisch die Probleme von und die neuen Anforderungen an dreidimensionale Wiedergabeverfahren auf: Einerseits müssen Produktionsumgebungen angepasst werden, damit eine Positionierung der Schallquellen im Raum möglich ist, andererseits müssen die Verfahren auf technische und perzeptive Vorteile untersucht werden.

## 2.3 Motivation und Vorteile von Surround-Formaten

Anhand der Evolution der Schallspeicherung- und Wiedergabe wurden über Ihre historische Entwicklung hinweg verschiedene Motive und Strategien der Wiedergabeverfahren dargelegt. Die explizite Motivation hinter Surround-Verfahren soll in ihren Grundzügen angerissen und kritisch betrachtet werden. Eine der dabei elementaren Fragen ist exemplarisch: Was ist die Motivation für und der Vorteil von Surround-Strategien gegenüber Stereo?

Surround-Sound kann generell bei jeder erdenkbaren Tonproduktion eingesetzt werden. Als konkrete Fallbeispiele lassen sich die Kategorien Musik, Filmtone und Hörspiel festlegen. Die Surroundlautsprecher können dem Toningenieur dabei als besonderes Mittel zu einer künstlerischen Klangmanipulation und –darstellung dienen. Unabhängig davon ob es sich um Musik, Filmtone oder Hörspiel handelt, sind folgende Faktoren für eine Surround-Produktion charakteristisch:

Der Zuhörer wird komplett von Schall umgeben und nicht nur von „vorne“ beschallt. Somit kann eine detailliertere räumliche Tiefe, ein expliziter Raumeindruck und eine wirkliche „Klang-Umhüllung des Zuhörers“ erschaffen werden. Demnach kann eine Schallquelle rund um den Zuhörer positioniert oder der Zuhörer in „einen anderen virtuellen Raum transportiert“ werden – den gestalterischen Möglichkeiten sind diesbezüglich keine Grenzen gesetzt.

Ein Vorteil von Surround gegenüber Mono und Stereo lässt sich zudem durch den Cocktailparty-Effekt beschreiben: Das menschliche Gehör besitzt die Fähigkeit bestimmte Schallquellen zu fokussieren und diese aus Umgebungsgeräuschen zu extrahieren. Verschiedene Experimente [22] [23] zeigen, dass diese Fokussierung nur dann uneingeschränkt möglich ist, wenn die Schallquelle dabei mit einem Räumlichkeitseindruck verbunden werden kann. Wird die Schallquelle inklusive Umgebungsgeräuschen über ein Mikrofon aufgenommen und anschließend wiedergegeben, ist eine Unterdrückung der Störgeräusche bzw. der Umgebungsgeräusche nicht mehr möglich. Je mehr Schallquellen also auf einem Lautsprecher abgebildet werden, desto schwieriger wird die Fokussierung auf eine einzelne Schallquelle. Über 5.1-Surround können Schallquellen daher präziser abgebildet werden und bietet dem Zuhörer somit eine bessere Trennung der Schallereignisse.

Ein weiterer Vorteil dieser Trennung ist, dass ein Toningenieur bei der Musikproduktion die Schallquellen explizit räumlich trennen kann und nicht über die Bearbeitung der Frequenzspektren für genügend „Platz“ für die einzelnen Instrumente sorgen muss. Laut Everett Porter sei ein „Surround [Mix] einfacher als 2-Kanal Stereo [anzufertigen]“ (frei ins Deutsche übersetzt nach [24]). Er sagt: „Surround is a good way to get people involved in recorded

music“. Seiner Meinung nach ermögliche es Surround, dass Zuhörer sich „keine Gedanken über die Reproduktionsqualität machen müssen und sich stattdessen auf die Musik konzentrieren können“ (frei ins Deutsche übersetzt nach [24]). Eine Gegendarstellung zum Effekt der besseren Trennbarkeit wird in Kapitel 2.4.3 im Kontext des auditiven Fokus dargelegt.

Auch wenn die aufgeführten Argumente zunächst für Surround sprechen, soll im Folgenden betrachtet werden, wie und inwiefern die beschriebenen Effekte auf den Zuhörer wirken. Dabei sollen sowohl psychophysiologische und anatomische, als auch wirtschaftlich-technische Aspekte beleuchtet werden.

## 2.4 Probleme bei der Wiedergabe von Surround

### 2.4.1 Psychophysiologische Betrachtung

Die Frage, ob Surround-Verfahren stets die besseren Ergebnisse beim Zuhörer erzielen, bleibt eine Gretchenfrage. Ein gut dargebotenes Musikstück kann auch auf einem qualitativ schlechten einkanaligen (Mono-)Küchenradio beim Hörer positive Emotionen auslösen. Ein solches Erlebnis kann weitaus mehr begeistern als die Wiedergabe eines Stücks auf der besten Surround-Anlage der Welt, sofern der Zuhörer davon emotional und musikalisch unberührt bleibt. Ein gutes Musikstück bleibt ein gutes Musikstück und funktioniert auch so, unabhängig von technischen Wiedergabebedingungen. Dass sich auf einem besseren Wiedergabesystem auch das qualitative Hörerlebnis verbessert, ist offensichtlich, aber nicht immer maßgebend für den Erfolg eines solchen Systems.

Surround-Sound spielt aber nicht nur bei Musikproduktionen eine Rolle: Das Genre Stummfilm ausgenommen, setzen heutzutage alle Spielfilme auf eine ausgefeilte Tonproduktion – samt Synchronisierung, Sounddesign, Positionierung der Schallquellen im Raum als auch der dazu passenden Filmmusik. Diese Faktoren spielen im Detail entscheidende Rollen bei der Rezeption von Spielfilmen (siehe u.a. [25]). Zu beachten gilt hier besonders: Fällt einer der genannten Faktoren aus – egal ob Sprache oder Filmmusik – kann sich die Aussage des Films bzw. die zu übermittelnde Nachricht entscheidend verändern. Wird auf die Tonübertragung verzichtet, kann zwar über Bilder möglicherweise der Handlung gefolgt werden, die eigentliche Aussage (bzw. Nachricht) eines Films kann sich aber entscheidend verändern. Im Zweifelsfall wird der Film aufgrund mangelnder Toninformation gänzlich „abgeschaltet“. In diesem Zusammenhang ist interessant, dass bereits die Mono-Wiedergabe ein Verfolgen des Films möglich macht, und nicht zwangsweise ein Surround-Setup notwendig ist. Surround kann also fast als eine Art „Add-on“ zur qualitativen Wahrnehmung betrachtet werden, und ist somit weder obligatorisch noch notwendig.

## 2.4.2 Wirtschaftlich-technische Betrachtung

Bei der Anschaffung eines Surround-Systems kann der Zuhörer auf potentielle Stolpersteine stoßen. Bereits beim Wechsel von Mono auf Stereo ist eine Verdopplung der kompletten Signalkette notwendig, angefangen bei der Bandbreite (doppelte Kanalzahl), bis hin zur doppelten Anzahl an Endstufen und Lautsprecher. Heutzutage werden Mono-Abspielgeräte eher rar, weswegen sich ein objektiver (wirtschaftlicher) Vergleich zwischen Mono- und Stereo-Wiedergabesystemen nicht aufstellen lässt. Dennoch lässt sich die Aussage treffen, dass die doppelte Anzahl auch das Doppelte kostet. Bei der Verwendung eines 5.1-Systems kommt es im Direktvergleich zu Mono sogar zu einer Versechsfachung der Signalkette, im Verhältnis zu Stereo zu einer Verdreifachung. Die einfache Formel, die Kosten entsprechend zu addieren, lässt sich in diesem Beispiel nicht zwangsweise anwenden, weil Surround-Systeme – rein technisch gesehen – anders aufgebaut sind als Stereo-Systeme: Wird bei Stereo versucht über 2 zwei Lautsprecher das komplette (in der Audioproduktion verwendete) Frequenzspektrum wiederzugeben, so teilen sich die Lautsprecher der meisten 5.1-Surround-Systeme das Frequenzspektrum mit dem dazugehörigen Subwoofer. Die fünf Lautsprecher lassen sich im Zweifelsfall also kostengünstiger und kompakter konstruieren, wenn ein entsprechender Frequenzanteil auf den Subwoofer ausgelagert wird (siehe Kapitel 2.2.3 Surround). Dennoch: Wird ein 5.1-Setup mit den gleichen Lautsprechern wie im Stereo-Setup realisiert, so ist – zumindest auf Seiten der Lautsprecher, Endstufen und Bandbreiten – eine Versechsfachung (zur Monophonie) zu erwarten. Auch wenn bei einem solchen Aufbau prinzipiell auf einen Subwoofer verzichtet werden könnte, ist er dennoch obligatorisch.

Zusammenfassend bleibt festzuhalten: Für 5.1-Surround sind höhere Kosten notwendig als bei Stereo. Dies lässt sich nicht nur auf die Wiedergabe, sondern letztendlich auch auf die Produktion übertragen: Hier muss ebenfalls – entsprechend des Endformats – für eine Abhörlösung gesorgt werden, auf der produziert werden kann. Wird versucht ein Konzept mit mehr als 6 Kanälen durchzusetzen, erhöhen sich entsprechend die Kosten, sowohl auf Produzenten- als auch auf Konsumentenseite.

Die Verdoppelung und die Verdreifachung der Kosten muss aber nicht zwangsweise konstant zur Verbesserung der wahrgenommenen Klangqualität sein. Bei der Erweiterung von Mono auf Stereo sind – durch die entstehende X-Y-Räumlichkeit – bereits signifikante Unterschiede im Klang und der Räumlichkeit festzustellen (siehe Kapitel 2.2.2 Stereophonie). Die erneute Verdreifachung auf Surround liefert zwar weiterhin Möglichkeiten zur besseren Raumwahrnehmung, ob sich die Qualität des Klangerlebnisses jedoch ebenso verdreifacht, sei an dieser Stelle dahingestellt (siehe Psychophysiologische Betrachtung). Es gilt dennoch: Um einen

## 2 Die Schall(re)produktion

---

explizit wahrnehmbaren Unterschied zu erreichen, ist im Zweifelsfall jeweils eine Verdoppelung der Lautsprecheranzahl notwendig. Die wahrgenommene qualitative Klang-Verbesserung trifft (besonders bei ungeübten Hörern) früher auf eine natürliche Sättigung.

Abb. 2.10 stellt diesen Effekt schematisch dar. Auf die Angabe von Maßeinheiten für Lautsprecheranzahl und perzeptive Qualität wurde bewusst verzichtet, da diese Werte von psychophysiologischen Faktoren abhängen und – je nach Zuhörer – unterschiedlich ausfallen können. Exemplarisch seien zwei Beispiele aufgeführt: So kann es einerseits sein, dass ein Zuhörer keinen merklichen Unterschied zwischen einer Stereo- und 5.1-Produktion wahrnimmt, und daher die Mehrkosten nicht aufwenden möchte.

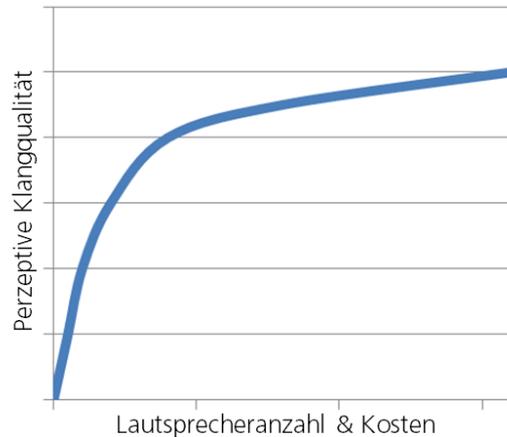


Abb. 2.10 Kosten und zu erwartende Klangqualität bei Verdopplung der Lautsprecheranzahl

Andererseits wäre es möglich, dass ein Zuhörer auch das Abspielen einer reinen Mono-Produktion auf einem 5.1-System präferiert, obwohl das abgespielte Signal auf jedem Lautsprecher dasselbe ist – schließlich steht bei einer Mono-Produktion keine zusätzliche Rauminformation bereit. (Hier kann beim ungeübten Zuhörer der Eindruck entstehen, dass durch den Zugewinn an Lautstärke – unter Verwendung einer höheren Lautsprecheranzahl – zeitgleich die Qualität verbessert wird.)

### 2.4.3 Anatomische Betrachtung

Surround-Produktionen stehen unabhängig von der in Abb. 2.10 beschriebenen Sättigungskurve weiteren Problemen gegenüber: Wird die primäre Attraktion des Klanggeschehens (bspw. Gesang bei einer Popproduktion) auf die Surround-Kanäle gegeben, ist der Zuhörer geneigt sich der Schallquelle zuzuwenden – der Zuhörer dreht sich also in Richtung der Surround-Lautsprecher. Die Front-Lautsprecher mutieren dabei zu Surround-Lautsprechern. Bei der Audioproduktion muss also stets darauf geachtet werden, dass der auditive Fokus nicht zu Ungunsten der eigentlichen Hörposition gesetzt wird. Zwar ist es – im Falle einer Musikproduktion – durchaus möglich auch andere Instrumente auf die Surround-Kanäle zu legen, das Prinzip bleibt jedoch immer dasselbe: rückt der akustische Fokus zu weit nach hinten, ist der Zuhörer geneigt seine Hörposition anzupassen. Die Positionierung von Schallquellen auf Surround-Kanäle muss entsprechend dezent gewählt werden, und kann in diesem Zusammenhang eher als Effekt oder als auditives Add-on betrachtet werden.

Weiterhin ist die unmittelbare Positionierung einer Schallquelle zwischen den Front- und den Surround-Lautsprechern als problematisch einzustufen. Da hier keine präzise Lokalisation der Schallquelle mehr möglich ist, kann die Schallquelle nach vorne oder nach hinten „kippen“. Die Surround-Kanäle können im Zweifelsfall tatsächlich zu reinen Effekt-Kanälen verkommen, oder lediglich zu einem verbesserten Raumeindruck beitragen. Gerade in Bezug auf die Nutzung als Effekt-Kanäle hängt der Erfolg von einer Surround-Produktion sowohl vom kreativen Ansatz des Künstlers als auch des Produzenten und Toningenieurs ab.

Die bereits aufgeführte bessere Trennbarkeit von Schallereignissen (siehe dazu Cocktailparty-Effekt in 2.3) kann sowohl vorteilhaft, aber auch nachteilhaft sein: Bei einer Stereo-Musikproduktion ist es durchaus denkbar, dass Schwächen eines Musikers durch die Mitmusiker maskiert werden. Bei einer 5.1-Surround-Produktion und damit verbundener besserer Trennung der Schallereignisse (und damit der Musiker) könnten somit auch Fehler unfreiwillig aufgedeckt werden. Dadurch steigt die musikalische Anforderung an die Musiker, was generell im Sinne einer guten Performance gesehen werden kann. Dennoch gilt dies – besonders bei der Musikproduktion – zu beachten. Durch die bessere Trennung der Schallquellen entsteht zudem die Gefahr, dass einzelne Instrumente losgelöst vom Gesamtklang des Ensembles wirken und das (Klang-)Konzept zusammenbricht.

Das Prinzip des auditiven Fokus und der besseren räumlichen Trennbarkeit lässt sich auf den Filmtone übertragen: Die Parameter Dialog, Filmmusik und Effekte wie Klangräume und Atmosphäre müssen entsprechend des Sinngehalts auf die Kanäle verteilt werden. Dialog wird in den seltensten Fällen von hinten erscheinen, da sich die Schauspieler meist (bildlich gesehen) vor dem Zuhörer / Zuschauer befinden. Andere Anordnungen des Dialogs sind als Effekt zu betrachten. Die Filmmusik kann von einer solchen Surround-Anordnung profitieren und auf die übrigen Kanäle ausweichen, womit dem Dialog zugleich akustisch Platz bereitet wird. Bei Filmmusik gilt allerdings erneut, dass der akustische Fokus nicht beeinträchtigt werden sollte. Die Surround-Kanäle können aber durchaus sinnvoll für Sounddesign und Effekte verwendet werden (ein Beispiel sind „passierende Autos“). Dennoch bleibt die Frage, wie häufig ein solcher Effekt in einem Film auftritt und ob dieser die Komplexität eines solchen Systems rechtfertigt.

#### **2.4.4 Unausgewogene Abhör-Räumlichkeiten**

Audioinhalte werden meist nach ITU-R BS.775 produziert. Wiedergabeseitig ist bei Einhaltung der Aufstellungsempfehlung eine originalgetreue Reproduktion der Signale möglich. Ein Problem stellen dabei aber meist die Räumlichkeiten der Zuhörer dar: Leider ist es nur in wenigen Fällen auf der Konsumentenseite realisierbar, die eigenen räumlichen Gegebenheiten mit den vorgegebenen Normen in Einklang zu bringen.

## 2 Die Schall(re)produktion

---

So kann es sein, dass sich die Sitzmöbel unmittelbar an einer Wand befinden, und es faktisch nicht möglich ist die Surround-Lautsprecher hinter dem Zuhörer zu positionieren.

Weiterhin kann sich der Fernseher für die Filmwiedergabe in einer Raumecke befinden. Da der Center-Kanal sinngemäß mittig zum Bild ausgerichtet sein sollte, entsteht eine unvorteilhafte Abweichung zwischen dem Verhältnis L zu Center und Center zu R. Diese abweichenden Positionierungen äußern sich entsprechend im zu erwartenden Klangerlebnis und tragen negativ zur räumlichen Darstellung bei. Dies kann bei der Produktion nicht berücksichtigt werden.

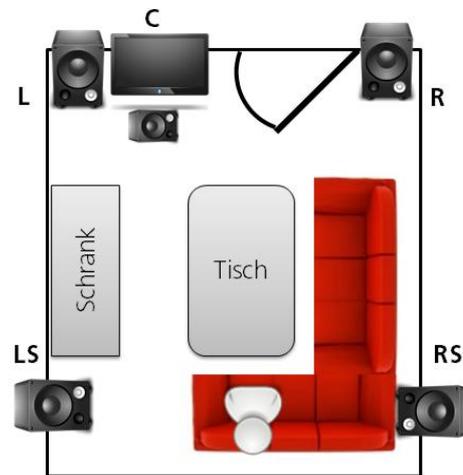


Abb. 2.11 Exemplarischer Lautsprecheraufbau entlang räumlichen Gegebenheiten

Zudem kann es durch die Überlagerung von Schallwellen der Lautsprecher zu Interferenzen und Kammfiltereffekten kommen. Dieser Effekt tritt allerdings bereits bei der stereofonen Wiedergabe auf, steigt aber mit einer größer werdenden Lautsprecheranzahl. Ein möglicher Lösungsansatz ist in Kapitel 6.5 aufgeführt.

### 2.4.5 Trägermedien und Wiedergabehardware

Die genannten Aspekte – psychophysiologisch, wirtschaftlich-technisch und anatomisch – grenzen bereits die Anwendung von Surround-Formaten – zumindest hinsichtlich der zu erwartenden Benutzerakzeptanz – erheblich ein. Ein weiterer bisher unbeachteter Faktor bzgl. der eingeschränkten Verbreitung von Surround in Bezug auf Musikproduktion lässt sich aber historisch durch die Bindung von Trägermedium an die dazugehörige Hardware darlegen: Entlang der Wiedergabeverfahren und –Strategien entstanden neben technischen Konventionen und Normen auch immer dazu passende Trägermedien. Angefangen bei der Aufnahme von Schall auf Walzen und der Wiedergabe mit dem Phonograph, hinweg über die Schallplatte und Magnetbandträger bis schließlich zu den digitalen Varianten CD, DVD, Blu-Ray und den digitalen (Datei-)Speicherformaten. Die meisten dieser Trägermedien waren lange Zeit auch an ein entsprechendes Abspielgerät gebunden, schließlich ist es technisch unmöglich eine Schallplatte mit einem CD-Spieler wiederzugeben. Diese Trägermedien waren – auch aufgrund ihrer Kapazität – allerdings auf eine vom System fest definierte Kanalanzahl beschränkt. So waren fast alle Trägermedien, die sich bis zur CD hin auf dem Konsumentenmarkt durchgesetzt haben, maximal zweikanalig vorbespielt und konnten entsprechend nur für eine

Stereowiedergabe verwendet werden. (Ausnahmen bilden Formate wie bspw. die SACD. Bei insgesamt ca. 4.500-5.000 SACD-Veröffentlichungen [26] kann man allerdings nicht von Marktdichte oder -relevanz sprechen.) Eine herkömmliche Audio-CD kann demnach nur zwei Kanalspuren beinhalten und die Endgeräte lediglich zwei Tonspuren abspielen.

Mit der Entwicklung der DVD war es schließlich möglich auch größere Datenmengen – und somit mehr Audiokanäle – auf ein Trägermedium zu bannen. Letztendlich hat sich 5.1-Surround mit insgesamt 6 Audiokanälen als kleinstmögliche sinnvolle Surround-Variante auf dem Konsumentenmarkt durchgesetzt. Die DVD hat sich jedoch hauptsächlich für Filme- bzw. Filmtone durchgesetzt. Für reine Tonproduktionen werden auch nach wie vor Audio-CDs verwendet, mitunter begründbar durch die Kompatibilität zu alten CD-Spielern, die wiederum das nachfolgende Trägermedium – die DVD – nicht abspielen können. Weiterhin gibt es kaum Autoradios, die technisch in der Lage sind, DVDs (oder gar SACDs) abzuspielen. In diesem Bereich konnte sich die CD als Trägermedium konstant halten.

Mit jeder Neuentwicklung war daher auch eine Neuanschaffung entsprechender Abspielhardware notwendig und lieferte der Käuferschicht ein entsprechendes Kontra-Argument. Die fehlende Aufwärtskompatibilität der Endgeräte für zukünftige Abspiel-Medien führte daher zwangsläufig zu einer gegenläufigen Tendenz zwischen technischen Möglichkeiten (Mehrkanalaudio) und Akzeptanz seitens der Konsumenten (Hardwaregebundenheit).

Durch die Entwicklung digitaler Formate und der Nutzung von Computern, mp3-Playern und neuerdings auch Smartphones als Wiedergabegeräte wird jedoch die Problematik der Vor- und Rückwärtskompatibilität von Hardware ausgehebelt. Ein Computer ist problemlos in der Lage sowohl Stereo als auch 5.1-Surround-Dateien abzuspielen sowie eine große Anzahl an Kanälen zeitgleich zu verarbeiten. Einzig die Software und die damit verbundene Rechenleistung stellt Grenzen an die maximale Kanalanzahl. Das Format ist somit nicht mehr an einen haptischen Datenträger gebunden, sondern einzig von der Software-Spezifikation abhängig. Wird ein neuer Audio-Codec entwickelt, muss keine neue Hardware gekauft werden oder gar das (im Zweifelsfall haptische) Trägermedium ersetzt werden – eine Aktualisierung der Software genügt meist, um die Kompatibilität mit dem neuen (digitalen und nicht-haptischen) Trägermedium zu gewährleisten. Weiterhin können Formate abseits der Stereo- oder 5.1-Norm auf einem beliebigen Endgerät abgespielt werden.

## 2 Die Schall(re)produktion

---

Ein konkretes Beispiel soll dies verdeutlichen: Ein Smartphone wird zur Wiedergabe von mp3s verwendet. Ändert sich die Spezifikation von mp3 auf ein anderes Format – sei dies ogg, mp4 oder andere – muss lediglich der Software-Player angepasst werden, um die digitale Umwandlung (die sog. Decodierung) vornehmen zu können, sprich aus einem digitalen Datenspeicher ein analoges (Audio-)Signal zu generieren. Unter diesen Voraussetzungen ist es daher möglich sowohl Stereo, 5.1-Surround als auch weitere denkbare Formate (7.1, 22.2 und so weiter) auf einem Smartphone wiederzugeben. (In diesem Beispiel sei angemerkt, dass Smartphones – mangels einer passenden Hardwareschnittstelle – in der Regel die Möglichkeit fehlt, Surround-Anlagen auf analogem Weg anzusteuern. Über eine digitale Übertragung zu einem Digital-Analog-Wandler – bspw. über Bluetooth, WLAN o.a. – ließe sich aber auch dieses Problem umgehen.) Es bleibt festzuhalten: das Trägermedium entscheidet nicht mehr über die nötige Hardware-Technologie, sondern lediglich über die Software-Spezifikation bei der Wiedergabe.

Da diese „neuen Endgeräte“ an sich beliebig viele Kanäle zeitgleich wiedergeben können, müssen geeignete Formate für die Wiedergabe mehrkanaliger Inhalte entwickelt werden. Natürlich wäre es möglich, einen Standard wie bspw. 22.2 einzuführen, der 24 separate Kanäle für bspw. 22 Lautsprecher und 2 Subwoofer bereitstellt. Ändert sich aber die Anzahl der Lautsprecher erneut nach oben oder unten, müssen wiederum Umrechnungen der Kanäle stattfinden und – bspw. bei Verringerung der Lautsprecheranzahl – der entsprechende Kanal als Phantomschallquelle zwischen zwei Lautsprechern positioniert werden. Ein solches Problem lässt sich mit einer herkömmlichen Standard-Definition daher nicht lösen. Ein möglicher Ansatz wird im folgenden Kapitel genauer beleuchtet.

## **3 Objektorientierung in Audioumgebungen**

Ungeachtet der im vorigen Kapitel aufgeführten Kritikpunkte – seien diese nun psychophysiologisch, wirtschaftlich-technisch oder anatomisch begründet – gibt es einen Markt für Surround-Produkte [21] [24]. Die im vorigen Kapitel aufgeführte Momentaufnahme des derzeitigen Stands der Computertechnik zeigt, dass die Potenziale – sowohl von Seiten der Surround-Produktion als auch der Wiedergabe – nicht gänzlich ausgeschöpft sind. Zunächst soll rekapituliert werden, welche Probleme und Nachteile bisherige Surround-Wiedergabeverfahren mit sich bringen, um diese auf Verbesserungspotenziale zu untersuchen. Anschließend wird ein Lösungsvorschlag in Form der objektorientierten Audioumgebung aufgezeigt.

### **3.1 Reflexion**

#### **3.1.1 Invariable Faktoren**

Psychophysiologische Faktoren lassen sich nur schwer bis gar nicht ändern oder beeinflussen. Sofern von Seiten des Zuhörers kein Interesse an Surround-Sound besteht, bzw. dieser die Vorteile des Systems nicht wahrnimmt oder wahrnehmen will, ist es schwer, Surround-Produkte an diese Zielperson zu vermarkten. Der technisch-wirtschaftliche Aspekt – in Bezug auf die Material- und Anschaffungskosten – bleibt auch nach wie vor bestehen: sechs Lautsprecher inklusive Endstufe sind letzten Endes teurer als zwei. Mit sinkenden Herstellungskosten und sinkenden Preisen für Technik und somit auch für Lautsprecherprodukte wird das Angebot für potentielle Käufer jedoch zunehmend attraktiver. Die wirtschaftlich-technischen Bedenken sind dahingehend – zumindest auf Konsumentenseite – als weniger bedeutend einzustufen. Das Phänomen des akustischen Fokus bleibt dauerhaft bestehen, unabhängig von einer Systemanpassung. Dies muss im Hinblick auf Surround-Produktionen immer beachtet werden.

#### **3.1.2 Variable Faktoren**

Die Analyse von Trägermedium und Hardware hat gezeigt, dass die Grenzen zwischen audiowiedergebenden Endgeräten und der Computertechnik immer weiter verschwimmen. Wird davon ausgegangen, dass die Abspielhardware nicht mehr an ein haptisches Medium wie CD, DVD oder Blu-Ray gebunden ist, entsteht ein klarer Vorteil auf der Wiedergabeseite: Ändert sich die Spezifikation des Trägermediums, so ändert sich nur die Software des Endgeräts und nicht die Hardware selbst. Ein Ansatzpunkt ist daher die Verwendung von digitalen (Datei-) Formaten als Trägermedium und die Verwendung von (Computer-ähnlicher) Hardware, die auf verschiedene Software-Player zurückgreifen kann.

## **3 Objektorientierung in Audioumgebungen**

---

Ein weiteres Problem stellt die nachteilige Aufstellung von Surround-Anlagen in den Räumlichkeiten des Zuhörers dar: Das Verlassen der Aufstellungsempfehlung führt wie aufgeführt zu Fehllokalisationen und zur inkorrekten Darstellung des wiedergegebenen Raumes bzw. der virtuellen Schallquellen. Zudem ist es für den Zuhörer nicht möglich, eine – auf die Räumlichkeiten angepasste – Lautsprecheranzahl zu realisieren. Die Anzahl ist stets an das festgesetzte Format und die damit einhergehenden Normen und Empfehlungen (wie bspw. ITU-R BS.775) gebunden. Ein Lösungsansatz ist die Verwendung eines digitalen Dateiformats in Verbindung mit entsprechender Abspielhardware, die sowohl die Lautsprecher-Anzahl als auch -Aufstellung im Raum kennt. Die Lautsprechersignale werden dann erst zum Zeitpunkt der Wiedergabe berechnet.

Die Verwendung von digitalen Dateiformaten, einer geeigneten (Computer-)Hardware zum Abspielen der Dateien und die Berücksichtigung eines variablen Lautsprecheraufbaus werden in den folgenden Kapiteln genauer betrachtet.

### **3.2 Objektorientierter Ansatz**

Die variablen Faktoren dienen als Grundlage für die Erstellung eines Konzepts der objektorientierten Audioproduktion und –wiedergabe. Zunächst soll aber der Begriff Objektorientierung im Kontext der Audioproduktion erklärt und definiert werden.

#### **3.2.1 Objektorientierung – Begriffserklärung**

Die Objektorientierung bezeichnet die Veranschaulichung komplexer Systeme anhand kooperierender und zueinander in Verbindung stehender Objekte und findet zunehmend in der Softwareentwicklung Verwendung. Dabei werden explizit Eigenschaften und Funktionen eines Objekts zusammengefasst und architektonische Baupläne konstruiert. Diese sind in der Regel verständlicher und einfacher auf reale Beispiele zu übertragen als in der funktionsorientierten Programmierung. Ein kurzes Beispiel soll die primäre Grundidee erläutern:

Jeder Gegenstand, jedes Medium bzw. jede konkrete Form lässt sich als Objekt bezeichnen respektive darstellen. So kann ein Stuhl ein Objekt sein, ein Schreibtisch, ein Lautsprecher, ein Mensch und so weiter. Jedes Objekt weiß dabei eigene Eigenschaften und Verhaltensweisen auf. Ein Mensch kann beispielsweise gehen, stehen, sitzen, sprechen (Verhaltensweisen, genauer: Funktionen). Er kann groß, klein, dick, dünn sein (Eigenschaften, genauer: Attribute). Eine Schreibtischlampe kann dagegen nicht gehen und auch nicht sprechen – sie leuchtet (Funktion) und hat meist eine konkrete Form und Farbe (Attribute bzw. Eigenschaften). Folglich kann eine Klassifizierung des Mediums über die Funktionen und Attribute geschehen und zu einer nächstgrößeren Einheit – dem Objekt – zusammengefasst werden.



Abb. 3.1 Vereinfachtes UML-Diagramm für die Objekte „Mensch“ (Auszug) und „Schreibtischlampe“ (Auszug)

### 3.2.2 Theorie: Objektorientierte Audioumgebungen

Die Theorie der Objektorientierung lässt sich nun auf Audioumgebungen übertragen. Um die Zusammenhänge genauer zu verstehen, ist jedoch ein Einblick in die bisherigen Produktionsverfahren notwendig: Ob Mono-, Stereo- oder 5.1-Surround-Produktion – das Endformat bzw. die Kanalanzahl entspricht immer genau der Lautsprecheranzahl des jeweiligen (Wiedergabe-) Verfahrens. Bei der Produktion werden daher Lautsprechersignale generiert, die sich einzig auf die starre Anordnung der Wiedergabenorm übertragen lassen. Eine solche Produktionsweise ist in Bezug auf einen vorher definierten (und optimierten) Lautsprecheraufbau zwar zielführend, führt jedoch wie bereits aufgeführt zu Problemen, wenn die Aufstellungsempfehlung nicht eingehalten werden kann oder sich die Lautsprecheranzahl ändert.

Die Lautsprechersignale werden dabei wie folgt generiert: Bei Stereo wird eine Schallquelle anteilig auf das Lautsprecherpaar gegeben – je nachdem, ob Sie sich weiter links oder weiter rechts befinden soll, ändert sich der Pegel der Signalquelle auf dem Lautsprecher. Bei 5.1-Surround ist die Vorgehensweise ähnlich: die Schallquellen werden auf einem zweidimensionalen Feld zwischen Lautsprechern positioniert. Die Lautsprecher können dabei paarweise Phantomschallquellen abbilden. Vereinfacht ausgedrückt: Links-Center, Center-Rechts und LS-RS bilden ein Stereopaar, ebenso L-LS und R-RS. Wird eine Schallquelle also auf dem zweidimensionalen Feld positioniert, werden die Signalanteile für die einzelnen Lautsprecher entsprechend berechnet.

Die bei der Produktion durchgeführte Positionierung der Schallquellen im zweidimensionalen Feld (bzw. Raum) kann nun für ein objektorientiertes Modell genutzt werden: Jede Schallquelle besitzt bereits an dieser Stelle eine Position in einem virtuellen, gedachten Raum. Diese Rauminformationen, bzw. die Koordinaten der Schallquelle im Raum, stellen die erste Ausprägung der objektorientierten Audioumgebungen dar: Eine Schallquelle (Objekt) hat die

### 3 Objektorientierung in Audioumgebungen

---

raumbezogenen Koordinaten (Attribute) in einem virtuellen Raum (Objekt). Die anschließende Berechnung der Lautsprechersignale und die darauffolgende Wiedergabe stellt lediglich die Manifestierung der Schallquellen an einem Ort im „realen“ Raum dar. Abb. 3.2 verdeutlicht diesen Ansatz:

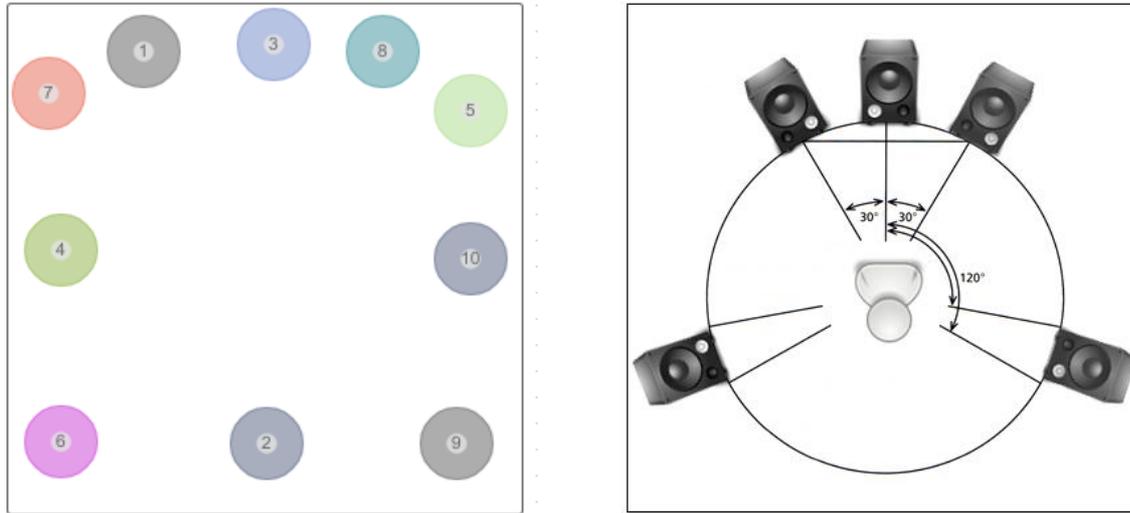


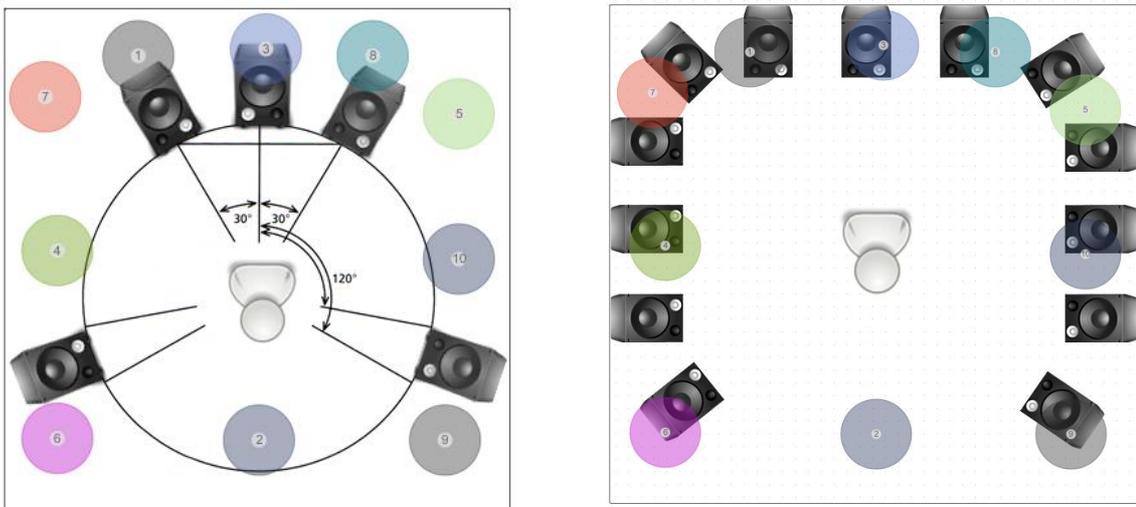
Abb. 3.2 Links: Schallquellen mit Ihrer Position im abstrahierten virtuellen Raum, rechts: „realer“ Raum mit Lautsprechern

Der virtuelle Raum ist bei der konventionellen Produktionsweise jedoch eng an die Lautsprecheraufstellung gekoppelt: Wird eine Schallquelle virtuell positioniert, so ist sie unmittelbar an einen (und als Phantomschallquelle an zwei) Lautsprecher gebunden. Verschiebt man die Lautsprecheraufstellung, so verschiebt man ebenfalls die Schallquelle. Da aber ohnehin jede Schallquelle im virtuellen Raum platziert wird – im Beispiel 5.1 zweidimensional – können die dabei entstandenen Meta-Informationen (bzw. die Attribute) der Schallquelle für ein angepasstes Wiedergabesystem genutzt werden: Wenn die (virtuelle) Rauminformation eine Aussage über die Position der Schallquelle gibt, kann das Lautsprechersignal anhand dieser Information generiert werden, je nachdem, wie die Lautsprecheraufstellung im realen Raum des Wiedergabesystems gewählt wird. Die Positionierung der Schallquelle im virtuellen Raum erfolgt also nach wie vor auf Produktionsseite, die Berechnung der Lautsprechersignale im realen Raum des Zuhörers jedoch erst bei der Wiedergabe: sie werden erst dann berechnet, wenn die Lautsprecheraufstellung bekannt ist. Somit lassen sich alternative Lautsprecheranzahlen und -positionen realisieren, da der Anteil einer Schallquelle für jeden Lautsprecher dynamisch generiert wird. Das Aufzeichnen der Rauminformation wird in Kapitel 0 ausführlich erläutert. Wie die aufgezeichnete Rauminformation wieder mit den Audioinhalten verknüpft wird ist Gegenstand des 6. Kapitels.

Im Zuge der Objektorientierung seien erneut die Objekte zusammengefasst, die für das Verständnis des Modells notwendig sind:

- Der virtuelle Raum
- Die Schallquellen im virtuellen Raum mit den Parametern x-, y- (und z-) Position
  
- Der reale Raum
- Die Lautsprecher im realen Raum mit den Parametern x-, y- (und z-) Position

Der objektorientierte Ansatz ermöglicht die Trennung zwischen virtuellem Raum (und somit den Positionen der virtuellen Schallquellen) und dem realen Raum und seinem Lautsprecheraufbau. Die Beschaffenheit des realen Raumes und der damit verbundene Lautsprecheraufbau sind für die Positionierung der Schallquellen im virtuellen Raum unerheblich. Anders ausgedrückt wird die Beziehung zwischen Schallquelle und Lautsprecher ausgehebelt: Es gibt keine n:2 (Stereo) oder n:5.1 (Surround) Beziehung mehr – das Schallereignis jeder Schallquelle kann dynamisch auf die jeweiligen Lautsprecher, abhängig von Ihrer Position, berechnet werden.



*Abb. 3.3 Projektion des virtuellen Raums auf den realen Raum  
links: 5 Lautsprecher, rechts: exemplarisch 13 Lautsprecher*

### 3.2.3 Bestandteile einer objektorientierten Audioumgebung

Das vorhergehende Kapitel erläutert die theoretische Trennung von virtuellem und realem Raum. Um dieses Konzept umzusetzen, ist es notwendig Audioumgebungen in ihre Bestandteile aufzugliedern: Zunächst erfolgt die Produktion der Audioinhalte (Musik, Filmtone, Hörspiel etc.), dann die Speicherung auf einem (digitalen) Trägerformat, abschließend die Wiedergabe auf dem Endgerät des Konsumenten. Einzig die Logik des Systems verlagert sich von der Produktion zur

## **3 Objektorientierung in Audioumgebungen**

---

Wiedergabe: erst bei der Wiedergabe wird festgelegt, was für ein Endformat (bzw. Lautsprecheraufbau) tatsächlich genutzt wird. Die einzelnen Bestandteile – Produktion, Dateispeicherung auf Trägermedium und Wiedergabe – werden in den Kapiteln 4 Objektorientierte Audioproduktion (OOAP), 5 Objektorientierte Datenspeicherung (OODS) und 6 Objektorientierte Wiedergabeverfahren (OOWV) genauer erläutert. Zunächst soll jedoch eine Zielgruppenanalyse darlegen, wer von einer objektorientierten Audioumgebung profitieren kann.

### **3.3 Zielgruppenanalyse**

Die Zielgruppenanalyse ist nicht nach sozioökonomischen Faktoren aufgegliedert. Sie soll vielmehr die Hauptinteressentengruppe bzw. die maßgeblich involvierten Parteien aufführen. Prinzipiell gliedert sich die Zielgruppe analog zu den Bestandteilen einer Audioumgebung.

#### **3.3.1 Produktion**

Bei der Audioproduktion sind primär Toningenieure, Produzenten, aber auch Musiker als Zielgruppe betroffen. Der Vorteil der objektorientierten Audioproduktion (OOAP) stellt für sie der geringere Aufwand bezüglich der Produktion auf Endformate dar: Sie müssen sich keine Gedanken über das tatsächliche Endformat machen, da sie lediglich im „virtuellen Raum“ produzieren und ihr Produkt bereits auf variable Raumgegebenheiten angepasst ist. Digitale Distributionswege und das Fehlen der Hardwaregebundenheit auf Seiten der Konsumenten ermöglichen zudem, dass die Surround-Produkte – egal ob Filmtone, Musik oder Hörspiel – auch tatsächlich vom Konsument rezipiert werden können. Dadurch kann die Surround-Produktion auch für Muskschaffende interessant werden (und steht in diesem Zusammenhang konträr zu Kapitel 2.4).

#### **3.3.2 Speicherung**

Die Speicherung auf ein Trägerformat dient als Schnittstelle zwischen Produktion und Wiedergabe. Letztlich sind beide Parteien vom Trägermedium abhängig, da die produzierten Audioinhalte zum Konsumenten vertrieben werden müssen. Dennoch soll das Trägerformat an dieser Stelle lediglich als „Mittel zum Zweck“ betrachtet werden. Das eigentliche Produkt ist schließlich der Inhalt des Trägermediums und nicht das Trägermedium selbst. (Die Stakeholder der Datenspeicherung auf das Trägerformat sollen im abschließenden Kapitel 0 erläutert werden, da zunächst die eigentliche Funktionsweise des Verfahrens erklärt werden soll.)

### **3.3.3 Wiedergabe**

Auf Seiten der Wiedergabe bleibt die Zielgruppe prinzipiell dieselbe wie bei allen Surround-Verfahren: Jeder Konsument, der die Vorteile von Surround-Produktionen zu schätzen weiß (siehe dazu Kapitel 2.4.1) kann als potentieller Interessent für die Wiedergabe von objektorientierten Audioproduktionen gesehen werden. Der Konsumentenmarkt bleibt allerdings immer abhängig vom Angebot der erhältlichen Produktionen: Zunächst müssen objektorientierte Produktionsverfahren angewandt und entsprechende Speicher- und Wiedergabeverfahren entwickelt werden, bevor ein Konsumentenmarkt entstehen kann. Letztlich ist es schwer einzuschätzen, ob die Konsumenten eine Surround-Wiedergabe von Musik bevorzugen, oder sich – gerade im Zeitalter von mp3s und in-ear-Kopfhörern – mit einem schlechteren Format zufrieden geben. Sollte sich durch die neuen Wiedergabemöglichkeiten jedoch auch das Angebot an Surround-Inhalten vergrößern und werden die Barrieren von Distribution und Wiedergabehardware niedrig gehalten, könnte sich ein objektorientiertes Format auch auf Konsumentenseite durchsetzen.



## 4 Objektorientierte Audioproduktion (OOAP)

Die objektorientierte Audioproduktion – im Folgenden mit OOAP abgekürzt – ähnelt der konventionellen 5.1- und ähnlichen Surround-Audioproduktion. In diesem Kapitel soll zunächst der generelle Workflow einer Audioproduktion untersucht werden, anschließend wird erläutert, wie die OOAP technisch in diesen Workflow und die Produktionsumgebungen eingebunden werden kann. Im Zuge der damit verbundenen Panorama-Aufzeichnung entsteht zudem der Bedarf nach neuen Eingabegeräten und -methoden, die in diesem Zusammenhang vorgestellt werden. Das Ziel des Kapitels ist herauszustellen, wie im Produktionsprozess Schallquellen mit eigenen Orts- bzw. Panorama-Informationen gespeichert werden können um sie somit von starren Lautsprecheraufstellungen und Endformaten zu entkoppeln.

### 4.1 Workflowanalyse einer Audioproduktion

Jede Audioproduktion startet unter denselben Voraussetzungen und mit den gleichen Arbeitsschritten: Jede für die Produktion geplante Schallquelle bekommt in der sogenannten Produktionsumgebung einen eigenen Kanal zugewiesen, dort werden die Schallquellen als „Audioinhalte“ grafisch dargestellt. Ob Schallquellen nun über ein Mikrofon aufgenommen werden oder ob die Audioinhalte künstlich (bspw. über virtuelle Instrumente) erzeugt werden, sei dahingestellt. Jeder Kanal verfügt am Ende des (Aufnahme-)Prozesses idealerweise über eigene Inhalte und kann somit wiederum eine eigene Schallquelle abbilden.



Abb. 4.1 Visuelle kanalbasierte Darstellung von Schallquellen samt Lautstärkemixer der Produktionsumgebung „Cubase“

## 4 Objektorientierte Audioproduktion (OOAP)

Diese Inhalte werden in der Lautstärke zueinander abgestimmt und können über verschiedene Effekte bearbeitet oder gar verzerrt werden. Eine Positionierung im (virtuellen) Raum gibt letztlich Aufschluss darüber, wo sich die Schallquelle – respektive der auditive Inhalt – befindet bzw. wo (und wie) dieser dargestellt werden soll. Sofern die Schallquellen nicht bereits über klangliche Rauminformationen verfügen (siehe bspw. 2.2.2.4 Stereo-Aufnahmeverfahren bzw. [13]), kann der „Klang des virtuellen Raumes“ über die Zuhilfenahme von künstlichen Hallalgorithmen beeinflusst und somit die Klangfarbe der Schallquelle beeinflusst werden.

Die Positionierung der Schallquellen im virtuellen Raum wird auch als Panorama-Verfahren bezeichnet und stellt den Ausgangspunkt der OOAP dar. Das Panorama-Verfahren ermöglicht einen „auditiven Blick“ über den gesamten Horizont – also 360° um den Zuhörer – und soll im Folgenden genauer betrachtet werden.

### 4.2 Theoretischer Lösungsansatz

Bei herkömmlichen Produktionsumgebungen (wie bspw. Pro Tools, Cubase etc.) hat jeder Kanal in der Regel ein eigenes Panorama-Feld, in dem die Schallquellen entweder zwischen L und R, oder zweidimensional im virtuellen Raum positioniert werden können. Dieses zweidimensionale Panorama-Feld zeichnet dabei nicht nur die Position der Schallquelle auf, sondern sorgt vielmehr für die Berechnung der Signalanteile zu dem in der Produktionsumgebung gewählten Lautsprecher-Setup. Wird 5.1-Surround als Endformat gewählt, so erfolgt die Berechnung der Lautsprechersignalanteile jedes einzelnen Kanals entlang der Positionierung.

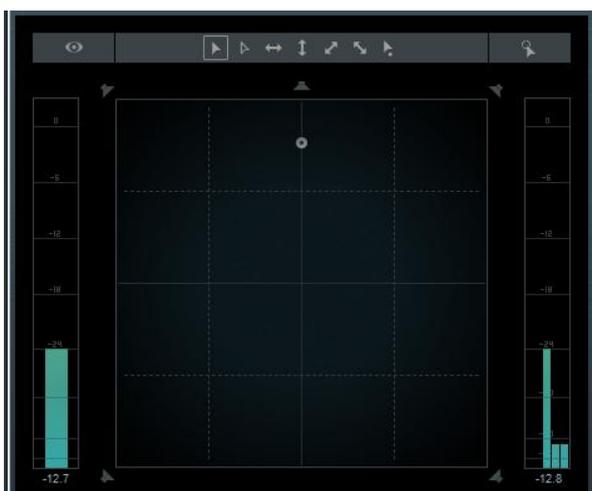


Abb. 4.2 Cubase-Surround-Panning-Tool  
links im Bild: Eingangssignal  
rechts im Bild: Berechnung der Aux-Signale

Vereinfacht gedacht kann dieses Verfahren wie folgt beschrieben werden: Jeder Kanalzug kann sein eigenes Signal pegelabhängig weiterreichen („send“). Als Anlaufstelle dient jedem Kanal ein sog. Aux-Weg (Aux für Auxiliary, in etwa „helfend“ oder „zusätzlich“). Auf einem Aux-Kanal können also Signale von Kanälen mit Audioinhalten zusammenlaufen und entsprechend – je nach Pegelanteile der „sendenden“ Kanäle – zusammengemischt werden

Bei einem Stereoprojekt wird also prinzipiell auf zwei Aux-Kanäle (Links und Rechts) gemischt, bei einem 5.1-Projekt entsprechend auf 6 Aux-Kanäle (L, C, R, LS, RS, LFE). Für die effiziente Positionierung im Raum wäre es nach diesem Verfahren notwendig, sechs Regler – also die „Sends“ auf jeden Aux-Kanal – immer parallel zu bedienen. Da dies zu aufwendig ist, bietet die zweidimensionale Darstellung einen guten Überblick und eine einfache Eingabemethode (siehe dazu Abb. 4.2).

Die Berechnung der Signalanteile erfolgt kongruent zu dem oben beschriebenen Verfahren und ist für festgelegte Lautsprecher setups – wie bspw. 5.1 – zwar zielführend; dennoch wird bei diesem Verfahren immer explizit auf ein definiertes Endformat produziert. An dieser Stelle soll die OOAP greifen und dieses Verfahren ersetzen.

### **4.2.1 Rückbesinnung auf das räumliche Hören**

Die zweidimensionale Darstellung des Panorama-Feldes deckt bereits die Grundzüge der räumlichen Positionierung ab. In Kapitel 2.1 wurden die Horizontalebene und Vertikalebene genauer betrachtet, diese sollen im Hinblick auf die OOAP als theoretische Grundlage dienen.

Die Positionierung auf einem zweidimensionalen Feld lässt sich auf ein x-y-Koordinatensystem übertragen. Die Hälfte der Breite und die Hälfte der Höhe des Feldes sei in diesem Zusammenhang der gedachte Mittel- bzw. Nullpunkt des Koordinatensystems bzw. der Koordinatenachsen. Dieser Mittelpunkt steht repräsentativ für die Abhörposition des Zuhörers. Über das Koordinatensystem lassen sich nun Schallquellen positionieren und die relativen Abstände zur Hörposition errechnen. Prinzipiell kann die Angabe der Schallquelle sowohl über kartesische X- und Y-Koordinaten oder aus Sicht des Hörers auch über ein Polarkoordinatensystem mit Winkel und Radius dargestellt werden. Der Winkel beschreibt dabei die Auslenkung der Schallquelle, der Radius den Abstand zum Zuhörer. Wird zu diesem System die dritte Dimension aufgespannt, erfolgt die Darstellung entsprechend mit X-, Y- und Z-Koordinaten oder über Horizontalwinkel (Azimuth), Vertikalwinkel (Elevation) und Abstand (Radius).

Die dreidimensionale Betrachtung ist im Kontext neuer Formate (wie bspw. Auro 3D in Kapitel 2.2.4) wichtig und spielt auch im Hinblick auf die in Kapitel 6 untersuchten Wiedergabeverfahren eine wesentliche Rolle. Zur Vereinfachung der visuellen Darstellung wurde an dieser Stelle allerdings mit einem zweidimensionalen Modell gearbeitet. Eingabemethoden für dreidimensionale Panoramadaten werden in Kapitel 4.5 exemplarisch aufgeführt.

Anstatt nun explizit Lautsprechersignale zu berechnen, werden Schallquellen also nur im Raum positioniert. Das eigentliche Lautsprechersignal soll erst bei der Wiedergabe berechnet werden.

### 4.3 Konzept & Entwicklung eines Prototyps

Eine gesonderte Aufzeichnung der Panoramadaten ist mit bisherigen Funktionen der gängigen Produktionsumgebungen (wie Pro Tools, Cubase und andere) derzeit nicht möglich. Die zur Verfügung stehenden Panoramafelder berechnen zwar den Signalanteil einer Schallquelle für die Lautsprecher, geben aber keine klare Auskunft über die Positionierung im Raum. Im Rahmen dieser Abschlussarbeit entstand daher ein Prototyp eines Panorama-Tools, welches als Plugin in die Produktionsumgebung eingebunden werden und die Position der Schallquelle im virtuellen Raum separiert aufzeichnen und wiedergeben kann. Der Prototyp sei im Folgenden kurz vorgestellt und soll repräsentativ für die Einbindung in mögliche Produktionsumgebungen stehen. Wie sich zeigen wird, lässt sich der Prototyp nicht in sämtliche Produktionsumgebungen integrieren, dennoch bildet er die Grundzüge der OOAP ab.

#### 4.3.1 Anforderungen eines objektorientierten Kanals

Wie bereits dargelegt, repräsentiert ein Kanal in der Produktionsumgebung jeweils eine eigene Schallquelle. Übertragen auf die Objektorientierung lässt sich ein objektorientiertes Modell eines solchen Kanals erstellen und wie folgt charakterisieren:

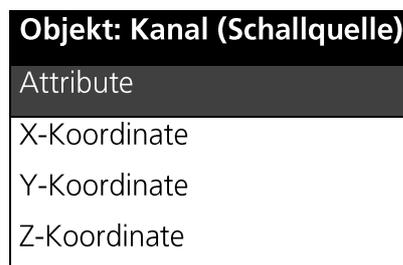


Abb. 4.3 Vereinfachtes UML-Diagramm eines Kanals mit Attributen

Anhand der Attribute des Kanals und der Untersuchung des bisherigen Workflows lassen sich die Anforderungen an den Prototyp definieren:

1. Visuelle Darstellung der Panorama-Information (zunächst zweidimensional)
2. Aufzeichnung der Panoramainformation (wahlweise als x-/y-Koordinate oder in Form von Radius und Winkel).
3. Übermittlung der Panoramainformation an ein Wiedergabe- bzw. Rendering-System

### **4.3.2 Untersuchung: Einbindung in Produktionsumgebungen**

Durch die Integration von Plugins kann der Funktionsumfang von Produktionsumgebungen erweitert werden. So können beispielsweise Effekte von Drittanbietern eingebunden werden, die vom Hersteller der Produktionsumgebung nicht berücksichtigt wurden. Historisch bedingt sind herstellerabhängige Schnittstellen für die Einbindung solcher Plugins entstanden. Die heutzutage gängigen Schnittstellen sind Real-Time Audio Suite (RTAS) von Avid für die Audioproduktionsumgebung Pro Tools und Virtual Studio Technology (VST) von Steinberg für die Audioproduktionsumgebung Cubase (und weitere Produkte dieses Herstellers).

RTAS von Avid ist eine proprietäre Schnittstelle und erfordert für die Entwicklung von Plugins Lizenzverträge mit dem Hersteller. Die Entwicklung eines Prototyps für diese Schnittstelle wurde dementsprechend zunächst ausgeschlossen. VST von Steinberg ist dagegen ein offener Standard, der auch von anderen Produktionsumgebungen implementiert und somit unterstützt wird. Diese Voraussetzungen machen die Entwicklung eines Plugins auf diese Schnittstelle erstrebenswert, da eine Entwicklung Hersteller-ungebunden durchgeführt werden kann und ein entsprechendes Plugin von vielen Produktionsumgebungen unterstützt wird. Die Programmierung eines Plugins erfolgt dabei in der Programmiersprache C++. Die Erfahrung des Autors mit der Programmiersprache C++ hat allerdings gezeigt, dass die Entwicklung eines Prototyps damit nur schwer zu erreichen ist: Um eine visuelle Darstellung auf einem zweidimensionalen Feld zu ermöglichen (erste Anforderung an den Prototyp) und die aufgezeichneten Panoramadaten für eine spätere Wiedergabe zu übermitteln (dritte Anforderung an den Prototyp), wäre ein extremer Entwicklungsaufwand notwendig. Der zu erwartende Entwicklungsaufwand war im Rahmen der Abschlussarbeit nicht in Einklang mit der Beweisführung des Prototyps (dem sog. Proof of concept) zu bringen, woraufhin alternative Programmiersprachen und Produktionsumgebungen evaluiert wurden.

### **4.3.3 Programmierumgebung Max for Live:**

#### **Kombination von Ableton Live und Max/MSP**

Im Zuge der Evaluation stellte sich die Kombination der Produktionsumgebung Ableton Live und der Programmierumgebung Max/MSP als zielführend für die Entwicklung des Prototyps heraus. Ableton Live stellt sowohl die VST- und mit „Max for Live“ zudem eine weitere Schnittstelle bereit: Plugins, die mit Max/MSP entwickelt werden, können unmittelbar in die Kanalzüge von Ableton Live integriert werden. Ableton Live hat – im unmittelbaren Vergleich zu den Produktionsumgebungen Pro Tools und Cubase – eine abweichende Bedienungsführung.

## 4 Objektorientierte Audioproduktion (OOAP)

---

Diese soll in diesem Zusammenhang allerdings eine untergeordnete Rolle spielen, da die Integration des Prototyps und die Verwendung dessen auf den jeweiligen Kanalzügen im Vordergrund stehen sollen. Sofern nötig, werden die Programmbestandteile im Kontext erklärt, dennoch ist ein Grundverständnis von Audioproduktionsumgebungen notwendig.

Max/MSP ist eine grafische Programmierumgebung des Herstellers Cycling 74, die – im Vergleich zu rein Schrift-basierten Programmiersprachen wie C++ oder Java – eine steile Lernkurve und eine explizite visuelle Darstellung des Programmgeschehens ermöglicht. In Max werden Objekte verschiedenster Funktion miteinander verbunden, diese bilden in Kombination einen sogenannten „Patch“ und damit im Idealfall ein funktionsfähiges, eigenständiges Programm. Die grafische Darstellung eines Max-Patches lässt sich annähernd mit (Elektronik-)Schaltplänen vergleichen, da jedes Objekt in Max einem gewissen Signalfluss unterliegt. Die Objekte erwarten dahingehend immer einen Impuls, der wiederum Funktionen der Objekte auslöst. Die Objekte wiederum können selbst Impulse an weitere Objekte weiterleiten. Max wird im Anhang „Max – eine grafische Entwicklungsumgebung“ auf der beigelegten DVD genauer beschrieben, dort werden zudem die Zusammenhänge der verschiedenen Objekte zueinander erläutert.

„Max for Live“ stellt die Schnittstelle zwischen Ableton Live und Max/MSP dar. Die Entwicklung dieser Schnittstelle entstand in Kooperation zwischen Ableton und Cycling 74 und ermöglicht eine enge Verzahnung der beiden Programme. So können Patches, die mit Max/MSP entwickelt wurden, sowohl in Ableton Live eingebunden werden, als auch unmittelbar auf Funktionen und Programmbestandteile zugreifen. Ableton Live wiederum kann die Steuerung verschiedener Parameter der Max-Patches übernehmen. Diese gegenseitigen Kontrollmöglichkeiten wurden entsprechend für den Prototyp ausgenutzt.

### 4.3.4 Max for Live-Patch

Es gibt viele Wege und Möglichkeiten, aus einer Problemstellung heraus einen Algorithmus zu entwickeln. Dies gilt sowohl für Programmiersprachen wie Java und C++, als auch für die Programmierung mit der grafischen Entwicklungsumgebung Max. Da Max auf eine große Bibliothek an Objekten zurückgreift, gibt es viele verschiedene Lösungsansätze für einen Prototyp, die alle sowohl eine Daseinsberechtigung als auch ihre Richtigkeit besitzen. Im Folgenden soll daher exemplarisch ein Ansatz aufgeführt sein, der für die Entwicklung des Prototyps als zielführend und einfach zu verstehen betrachtet wurde. Weitere Ansätze sind dabei durchaus denkbar.

Die Panorama-Oberfläche aus Pro Tools und Cubase (siehe Abb. 4.2) diene als Vorbild für die Positionierung der Audioinhalte eines Kanals im virtuellen Raum. Max stellt für diesen Zweck diverse Bausteine bereit, für den Prototyp wurde das Objekt „Nodes“ verwendet, das sowohl Kartesische- als auch Polarkoordinaten verarbeiten kann.



Abb. 4.4 Visuelle Darstellung der Raumposition über das Objekt Nodes

Dieses Objekt wurde so modifiziert, dass die Abhörzone gemäß ITU-R BS.755 durch einen Kreis repräsentiert wird. Der Mittelpunkt dieses Kreises und somit des Panoramafeldes stellt dabei zudem die Abhörposition dar. Die Position der Schallquelle im Raum wird durch einen gelben Punkt dargestellt, dieser lässt sich auf dem Panoramafeld und somit im virtuellen Raum verschieben. Das Objekt Nodes reicht dabei die Position dieses Punktes an Ableton Live-eigene Objekte weiter, diese wiederum dienen der Erfassung der Automationsdaten: Jedes Ableton-Objekt erzeugt automatisch eine eigene Automationsspur für jeden Kanal, in der die Veränderungen der Objektzustände gespeichert werden.

Wird nun der Punkt bzw. die Schallquelle während der Aufnahme auf dem Panoramafeld positioniert, kann der dadurch entstehende Bewegungsverlauf unmittelbar in der Automationsspur gespeichert werden. Bei der Wiedergabe werden die aufgezeichneten Automationsdaten ausgewertet und wiederum an die Ableton-Objekte übergeben, diese übermitteln die Automationsdaten an das Nodes-Objekt und ermöglichen so, die Position der Schallquelle samt Bewegungspfade auf dem XY-Feld darzustellen.

Die Dimensionen des Panoramafelds bzw. des Objekts Nodes reichen von 0 bis 1 jeweils in der vertikalen und der horizontalen Ebene. Die Angabe der Position erfolgt in Fließkommazahlen zwischen 0 und 1 und ist demnach für die Darstellung und Positionierung hinreichend genau. Da die Automationsspuren auf MIDI basieren, sind allerdings nur 128 unterschiedliche Abstufungen und somit Objektzustände erfassbar, daher werden die Werte zwischen 0 und 1 auf 128 Abstufungen umgerechnet.

## 4 Objektorientierte Audioproduktion (OOAP)

---

Gemäß des XY-Panoramafeldes bedeutet dies, dass 128x128 verschiedene Positionen erfasst werden können – die Stelle (0|0) bildet dabei die obere linke Raumecke, (127|127) die rechte untere. Insgesamt ergibt dies 16.384 verschiedene Positionen im Raum, was 14 Bit entspricht. (Um den Nullpunkt des Koordinatensystems wieder auf den Mittelpunkt und somit die Abhörposition zu setzen, ist eine Transposition des Koordinatensystems notwendig.)

Durch die Verwendung von Horizontalwinkel im Bogenmaß (Radiant) und Radius anstatt X- und Y-Koordinaten ist eine ähnlich präzise Darstellung möglich – durch die Einschränkung auf 128 verschiedene Abstufungen wäre allerdings auch hier eine Konvertierung der Winkel- und Radiusdaten auf die Abstufungen notwendig. Wie sich zeigen wird, kann die Darstellung in Polarkoordinaten für die spätere Wiedergabe bzw. das damit verbundene Rendering von Vorteil sein. (Siehe dazu Kapitel 6.3 Vector Base Amplitude Panning und die damit verbundene Vektorberechnung.)

Unter Verwendung dieses Prototyps können Schallquellen – zumindest visuell – im Raum positioniert werden. Um die visuelle Repräsentation der Schallquelle auch auditiv darzustellen, ist es zwingend notwendig, einen Renderer an die Produktionsumgebung zu koppeln, der die Audioinhalte samt Panoramainformation verarbeiten und auf ein Lautsprechersystem übertragen kann. Mögliche Renderer und Wiedergabesysteme samt ihrer exakten Funktionsweise werden im Kapitel 6 vorgestellt. An dieser Stelle sei ein theoretischer Renderer vorausgesetzt, der beliebige Lautsprechersignale anhand der Panoramainformationen und der Audioinhalte generieren kann. Die Ansteuerung eines solchen Renderers über die OOAP soll im Folgenden erläutert werden.

### 4.4 Ansteuerung des Renderers

Unabhängig davon, ob mit dem entwickelten Prototyp produziert wird oder ob ein alternatives Programm verwendet wird, müssen Audiodaten von der Produktionsumgebung zum Renderer übermittelt werden, um letztendlich Lautsprechersignale zu generieren. Exemplarisch soll erneut der Prototyp den Weg der Audiodaten erläutern: Wird eine Schallquelle im Raum positioniert, so besitzt jeder Kanal eigene Panorama-Informationen gemäß des aufgestellten objektorientierten Modells. Um diese objektweise an einen Renderer zu übermitteln, kann entweder auf eine Datenspeicherung (siehe Kapitel 5) zurückgegriffen oder die Produktionsumgebung direkt mit einem Renderer verbunden werden. Da Ableton Live bisher über keine interne Rendering-Möglichkeit in diesem Umfang verfügt, muss – um eine Abhörmöglichkeit für die Produktion zu schaffen – die Produktionsumgebung unmittelbar mit einem separaten Renderer verbunden werden.

### 4.4.1 Übermittlung der Audio-Daten

Da die Objekte von den Lautsprechersignalen entkoppelt werden sollen, kann nicht mehr auf einen Summenbus bzw. Aux-Kanal gemischt werden (siehe dazu Kapitel 3.2 und 4.2), die Lautsprechersignale werden erst beim Renderer generiert. Um die Objekte zu übermitteln, muss jedem Kanal ein separater (physikalischer) Ausgang zugewiesen werden.

Diese Übermittlung lässt sich durch ein 1:1 Routing auf die Ausgangskanäle von Ableton Live und ein 1:1 Routing von den Ausgangskanälen auf die Eingangskanäle des Renderers realisieren.

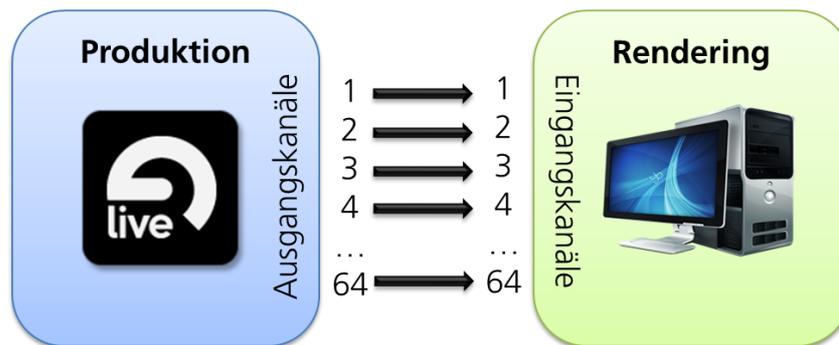


Abb. 4.5 Routing der Produktion zum Rendering, im Beispiel 1:1-Routing von 64 Kanälen

Abb. 4.5 führt dabei den schematischen Routing-Vorgang auf. Es gibt derzeit kaum Möglichkeiten, Audioinhalte von Software (Produktion) zu Software (Renderer) zu übermitteln. Demnach muss ein Umweg über die Hardwareschnittstelle der Produktionsumgebung gewählt werden, dabei wird jedem Objekt-Kanal ein expliziter Soundkartenkanal zugewiesen. Abb. 4.6 stellt die Konfiguration der Ausgänge in Ableton Live dar. Mit steigender Produktionsgröße und steigender Objektanzahl erhöht sich daher auch die Anzahl der benötigten Ausgangskanäle.

Ein entsprechendes Routing von der Produktionsumgebung zum Renderer lässt sich über verschiedene Hard- und Software-Ansätze realisieren und soll im Folgenden beschrieben werden:

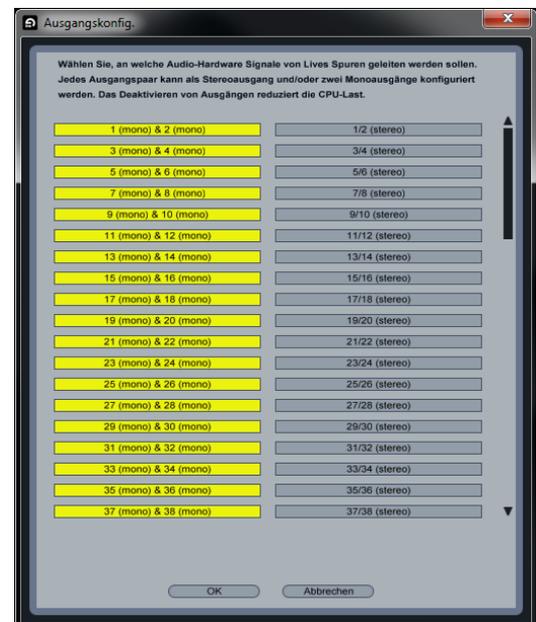


Abb. 4.6 1:1 Routing der (Objekt-)Kanäle von Ableton Live auf die physikalischen Ausgänge

## 4 Objektorientierte Audioproduktion (OOAP)

### 4.4.1.1 Hardware-Ansatz: Physikalischer Ausgang über MADI-Soundkarten

Beim Hardware-Ansatz befinden sich Produktionsumgebung und Rendering auf zwei physikalisch getrennten Maschinen (bspw. Computern). Diese werden über Soundkarten miteinander verbunden: jeder Soundkartenausgang der Produktionsumgebung wird mit einem separaten Soundkarteneingang des Renderers verbunden. Dieses Konzept verwendet für diesen Zweck mehrkanalige Soundkarten, die Produktion gibt dabei – je nach Umfang – die benötigte Anzahl der Ein- und Ausgänge vor. Ein geeigneter technischer Lösungsansatz ist die Verwendung der Hardware-Schnittstelle „Multi Channel Audio Digital Interface“ (MADI). Durch eine Soundkarte mit MADI-Schnittstelle lässt sich eine zeitgleiche Übermittlung von bis zu 64 Audiokanälen über ein einziges Kabel realisieren. (Details zur MADI-Spezifikation befinden sich in [27].) Die aktuelle Soundkartenentwicklung zeigt, dass bis zu 3 MADI-Schnittstellen in einer einzigen Soundkarte kombiniert werden können [28]. Somit stehen der Produktionsumgebung maximal 192 Ausgangskanäle und dem Renderer maximal 192 Eingangskanäle zur Verfügung.

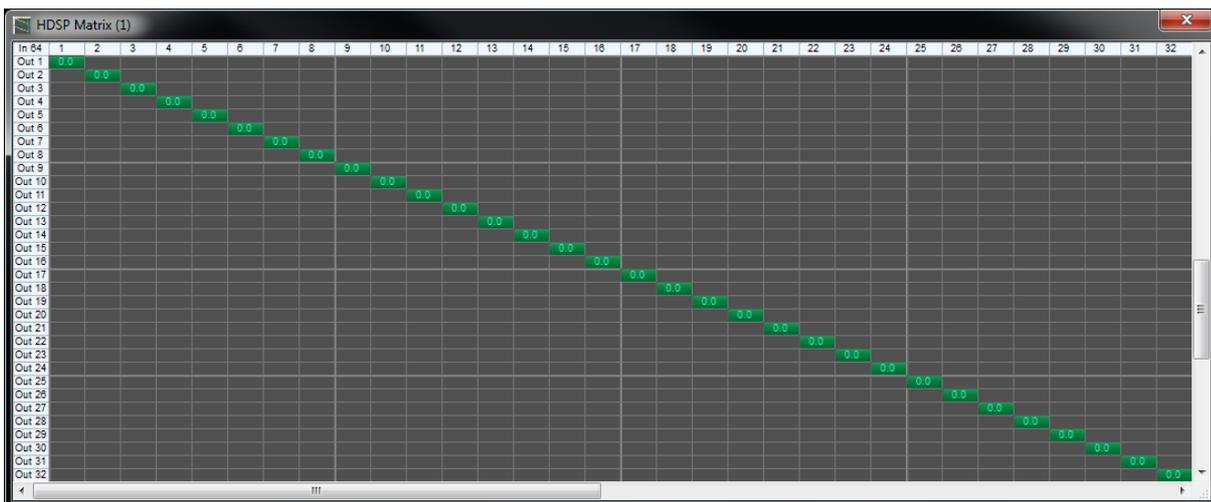


Abb. 4.7 Die Routing Matrix bei MADI-Soundkarten des Hersteller RME Audio (davon 32 Kanäle im 1:1 Routing)

### 4.4.1.2 Software-Ansatz: Simulierter physikalischer Ausgang über JACK

Es ist aber nicht immer zwingend notwendig die Produktionsumgebung vom Renderer zu trennen. Sofern die Rechenleistung der Produktionsumgebung genügt um auch das Rendering zu übernehmen, kann beides auf derselben physikalischen Maschine stattfinden. Dazu ist es notwendig über eine sogenannte Software-Soundkarte entsprechende Ein- und Ausgangskanäle zu simulieren. Eine solche Software-Soundkarte wird dabei sowohl in der Produktionsumgebung als auch im Renderer eingebunden. Das 1:1 Routing der Produktion zum Renderer erfolgt also nicht über Hardware, sondern lediglich über Software.

Ein Beispiel für eine solche virtuelle Software-Soundkarte stellt das Programm JACK Audio Connection Kit dar. Mit ihm lassen sich fast beliebig viele Programme untereinander vernetzen. Im exemplarischen Beispiel von Abb. 4.8 dient Jack als Soundkarte für Ableton-Live und einen gedachten Renderer (in Abb. 4.8 durch „Max“ repräsentiert). Jack stellt beiden jeweils 64 Ein- und Ausgänge zur Verfügung.

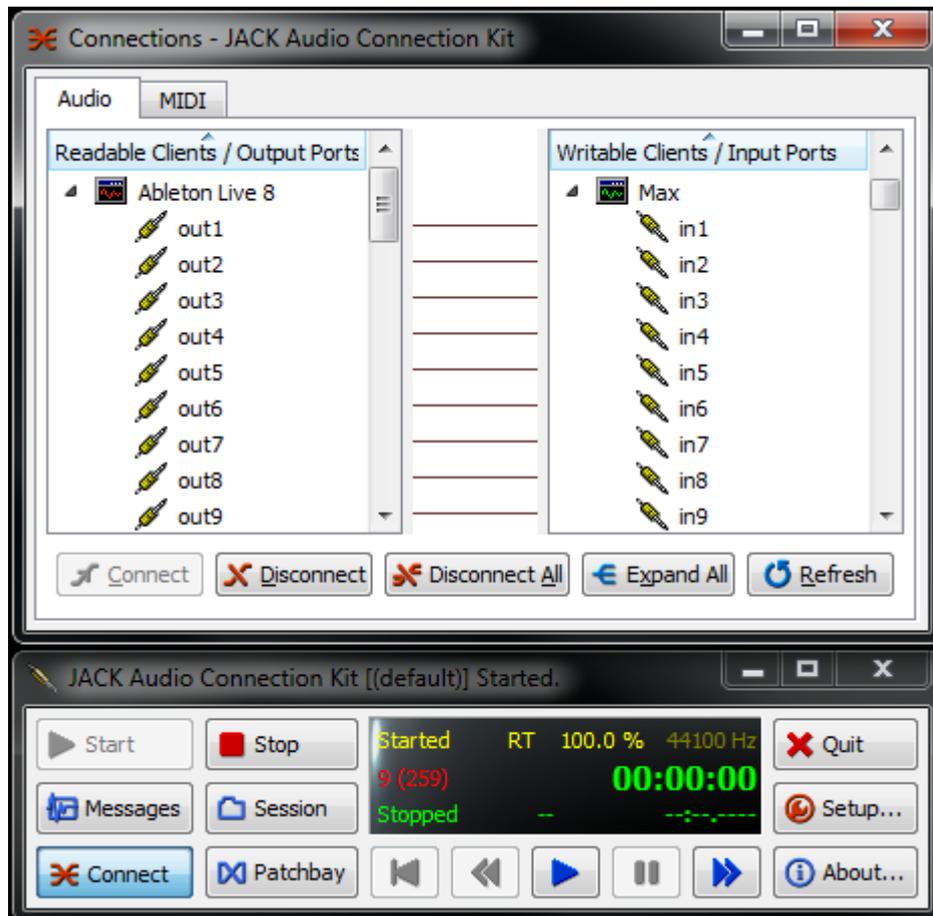


Abb. 4.8 Verknüpfung von Applikationen durch JACK

Generell sind Jack bei der Kanalanzahl nach oben hin keine Grenzen gesetzt – hier definiert lediglich die verwendete (Computer-)Hardware die Auslastungsgrenze. Da diese Faktoren in jeder Produktionsumgebung unterschiedlich sind, wurde im Rahmen der vorliegenden Arbeit kein Belastungstest durchgeführt. Die im Versuchsaufbau aufgeführten 64 Kanäle wurden in einer Stichprobe auf dem System des Autors verifiziert. Die Kontrolloberfläche von Jack definiert die maximale Anzahl auf 999 Kanäle, mehr sind im Zweifelsfall über alternative Konfigurationen der Software und entsprechenden Hardwarevoraussetzungen erreichbar.

## 4 Objektorientierte Audioproduktion (OOAP)

### 4.4.1.3 Gemischter Ansatz: Physikalischer Ausgang über Netzwerkschnittstelle

Neben reinen Soft- und Hardwarebasierten Ansätzen lässt sich auch eine gemischte Lösung realisieren. Der Hersteller Audinate vertreibt mit der Produktreihe Dante eine proprietäre IP-basierte Lösung für Audio over Ethernet. Der Hersteller vertreibt sowohl netzwerkfähige Soundkarten als auch eine Software-Lösung, die im Folgenden genauer untersucht wird.

Der technische Aufbau entspricht dem Hardware-Ansatz, bei dem die Produktionsumgebung und der Renderer auf zwei verschiedene Systeme ausgelagert sind. Sie kommunizieren dabei jedoch anstatt über Audioschnittstellen über die Netzwerkschnittstelle. Dabei wird die von Audinate entwickelte „Dante Virtual Soundcard“ (DVS) – ähnlich wie JACK – als Software-Soundkarte in der Produktionsumgebung eingebunden und dient als Sender der Audiodaten über die Netzwerkschnittstelle. Der Renderer als Empfänger auf der Gegenseite erhält die übermittelten Daten ebenfalls über die Netzwerkschnittstelle und muss DVS als Software-Soundkarte verwenden um die Kanäle weiterzuverarbeiten. Die Leistung der DVS lässt sich annähernd mit MADI vergleichen: Es ist möglich 64 Kanäle bei 48kHz und 24 Bit über Gigabit-Ethernet zu übertragen, weitere Kanäle sind über diese Software-Soundkarte nicht möglich.

Über DVS lassen sich somit Produktionsumgebung und Renderer über eine (im Zweifelsfall bestehende) Netzwerkinfrastruktur verbinden. Allerdings erfordert DVS den Betrieb eines sog. „Sync Masters“, eines Taktgebers – ohne dessen Betrieb die Software-Soundkarten keine Daten austauschen können. Ein solcher Taktgeber ist ausschließlich in Audinate- bzw. Dante-Hardware integriert, eine softwareseitige Lösung existiert nicht. Audinate wirbt auf der Herstellerwebseite mit der Unterstützung von jeweils 512 Sende- und Empfangskanälen (bei 48kHz Abtastrate und 24 Bit Systemauflösung) über eine Gigabit-Ethernet Verbindung bei Hardware-Soundkarten.

Werden Produktionsumgebung und Renderer mit je einer Hardware-Soundkarte ausgestattet, wäre also prinzipiell die zeitgleiche Übertragung von 512 Kanälen möglich.

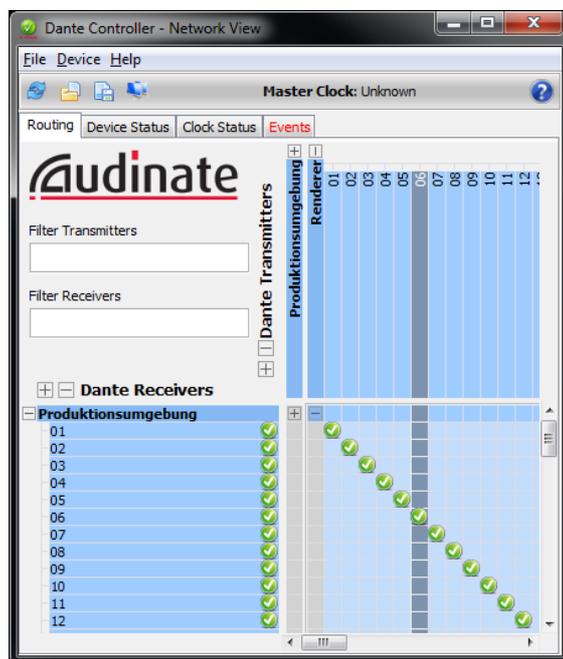


Abb. 4.9 Verknüpfung von Audiogeräten mit Audinates "Dante Virtual Soundcard"

#### 4.4.1.4 Alternativen und Ausblick

Für jeden der aufgeführten Ansätze existieren Alternativen, sie können dahingehend also nicht als einzige Lösungsmöglichkeit dargestellt werden. Zur rein hardwarebasierten Lösung mit Soundkarten, mit dem Ziel möglichst viele Kanäle zu übertragen, gibt es allerdings kaum Alternativen zur MADI-Schnittstelle. Eine Anbindung über die Schnittstelle „Alesis Digital Audio Tape Optical Interface“ (ADAT) wäre zwar denkbar, bietet allerdings – selbst bei Kombination mehrerer ADAT-Datenströme – nicht der zu erwartenden Kanalzahl bei einer OOAP.

Soundflower ist eine weitere softwarebasierte Soundkarte, ermöglicht ein internes Routing zwischen Applikationen und bildet damit das Pendant zu Jack und DVS, ist allerdings nur kompatibel mit Mac-Betriebssystemen. Die Entwicklung weiterer Software-Soundkarten ist durchaus denkbar, jedoch muss immer die zu Grunde liegende (Sound-) Architektur des Betriebssystems beachtet werden. Ein möglicher plattformübergreifender Ansatz ist die Verwendung einer Brückentechnologie wie PortAudio um auf die Soundarchitektur verschiedener Betriebssysteme zugreifen zu können [29].

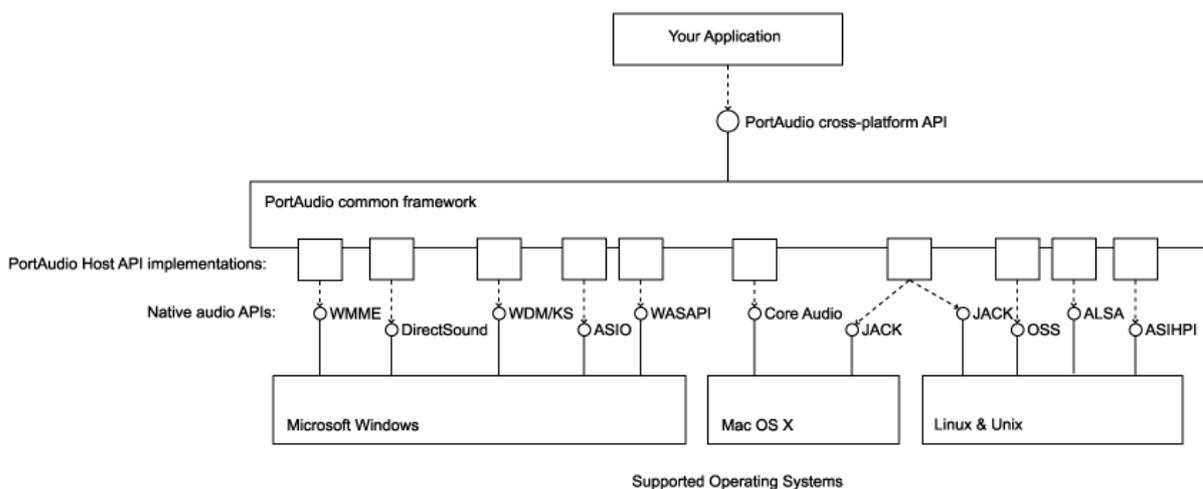


Abb. 4.10 Ansteuerungsmöglichkeiten von PortAudio auf Betriebssystemebene (Quelle: [29])

Die Nutzung der Netzwerkschnittstelle zur Übertragung von Audioinhalten (Audio over Ethernet) wird dagegen von immer mehr Herstellern genutzt und mit „Audio Video Bridging“ (AVB) auch vom „Insitute of Electrical and Electronics Engineers“ (IEEE) vorangetrieben [30].

Da die Computernetze jedoch über das ISO/OSI-Modell in verschiedene Schichten geteilt werden [31], gibt es auch entsprechend viele verschiedene Spezifikationen und Implementierungen: Audinates Dante arbeitet auf Layer 3 und versendet Audioinhalte über UDP [32]. Das von IEEE geförderte „Audio Video Bridging“ (AVB) arbeitet dagegen auf Layer 2 des ISO/OSI-Modells [30]. Weitere Implementierungen sind in Tabelle 4.1 aufgelistet, sollen an dieser Stelle aber nicht

## 4 Objektorientierte Audioproduktion (OOAP)

weiter erläutert werden. Dante und AVB sollen exemplarisch für das Potential der Netzwerkschnittstelle in der Audioübertragung stehen. Die steigende Leistung von Drahtlosnetzwerken (sog. Wireless-LANs) könnte zukünftig ebenso interessant für die Übermittlung von Audiodaten werden [33].

Technology	Transport	Network capacity	Max sampling rate
<b>AES47</b> <a href="http://www.aes.org">www.aes.org</a> <a href="http://www.ninetiles.com">www.ninetiles.com</a>	ATM	$\infty$	192 kHz
<b>AES50</b> <a href="http://www.aes50.com">www.aes50.com</a>	Ethernet Physical Layer	48 channels	384 kHz and DSD
<b>AudioRail</b> <a href="http://www.audiorail.com">www.audiorail.com</a>	Ethernet Physical Layer	32 channels	48 kHz (32 ch), 96 kHz (16 ch)
<b>Aviom Pro 64</b> <a href="http://www.aviom.com">www.aviom.com</a>	Ethernet Physical Layer	64 channels	208 kHz
<b>CobraNet</b> <a href="http://www.cobranet.info">www.cobranet.info</a>	Ethernet data-link layer	$\infty$	96 kHz
<b>Dante</b> <a href="http://www.audinate.com">www.audinate.com</a>	Any IP medium	700 channels	192 kHz
<b>EtherSound ES-100</b> <a href="http://www.ethersound.com">www.ethersound.com</a>	Ethernet data-link layer	64	96 kHz
<b>EtherSound ES-Giga</b> <a href="http://www.ethersound.com">www.ethersound.com</a>	Ethernet data-link layer	512	96 kHz
<b>HyperMAC</b>	Gigabit Ethernet	384+ channels	384 kHz and DSD
<b>Livewire</b> <a href="http://www.axiaaudio.com">www.axiaaudio.com</a>	Ethernet data-link layer	$\infty$	48 kHz
<b>mLAN</b> <a href="http://www.yamaha.co.jp">www.yamaha.co.jp</a>	IEEE-1394	63 devices (800 Mbps)	192 kHz
<b>Nexus</b> <a href="http://www.stagetec.com">www.stagetec.com</a>	Dedicated fiber	256 channels	96 kHz
<b>Optocore</b> <a href="http://www.optocore.com">www.optocore.com</a>	Dedicated fiber	512 channels at 48 kHz	96 kHz
<b>Rocknet</b> <a href="http://www.medianumerics.com">www.medianumerics.com</a>	Ethernet Physical Layer	160 channels (48 kHz/24bit)	96 kHz
<b>UMAN</b>	IEEE 1394 and Ethernet AVB	400 channels (48 kHz/24bit)	192 kHz

Tabelle 4.1: Überblick über den Transfer von Audioinhalten über Netzwerk, angelehnt an [33].

## 4.4.2 Übermittlung der Panorama-Daten

In Kapitel 4.3 wurde die Aufzeichnung der Panoramadaten in Ableton Live anhand des entwickelten Prototyps dargelegt. Die Panoramadaten sind dabei Attribute der Objektkanäle und liegen somit in Form von Meta-Daten vor. Sie sind dahingehend nicht auf komplexe Soundkarten-Infrastrukturen angewiesen.

Innerhalb des Prototyps erfolgt die Übertragung der Panoramadaten über das „User Datagram Protocol“ (UDP, siehe dazu [31]). Für die Übermittlung werden jeweils die Kanalnummer und die aktuellen Panoramadaten im virtuellen Raum an einen Empfänger mit spezifizierter IP-Adresse und Port gesendet. Der Empfänger kann diese Steuerdaten anschließend über UDP empfangen und die Panoramadaten entlang der Kanalnummer den entsprechenden Kanälen zuordnen.

In Kapitel 4.3.4 wurde ermittelt, dass sich die Panoramainformation im zweidimensionalen Raum mit 14 Bit abgebildet werden können. Panoramadaten benötigen allerdings keine Abtastrate von 44.1 kHz. Hier ist eine Skalierung zwischen 1 Sekunde und 10 Millisekunden denkbar, um eine Bewegung im Raum darzustellen. In Konsequenz ergibt sich eine Abtastrate zwischen 1 Hz und 100 Hz, die Bitrate liegt entsprechend zwischen 14 bit/s und 1.400 bit/s und ist für die Netzwerkbelastung – zumindest in Relation zu Audioinhalten über Netzwerk – als unbedeutend zu betrachten.

Die Steuerdaten lassen sich exemplarisch wie folgt abbilden:

*Formel 4.1 als OSC-Code: /panorama\_information/ch\_number x-value y-value*

Im direkten Beispiel für den Kanal 1:

*Formel 4.2 als OSC-Code: /panorama\_information/1 0.5 0.1*

Das Auslesen der Steuerdaten wird in Kapitel 0 & 6 genauer erläutert. Prinzipiell könnte auch jede Nachricht mit einem entsprechenden TimeCode versehen werden. Eventuelle zeitliche Varianzen sind aber zunächst zu vernachlässigen: Ob eine zeitliche Verschiebung von Panoramadaten zu den Audioinhalten auditiv wahrnehmbar ist, müsste in einem Hörversuch evaluiert werden.

Da jeder einzelne Kanal seine Panorama-Informationen via UDP übermittelt, können die Positionen bereits vor dem Rendering erfasst und so eine in eine gemeinsame Übersicht der Schallquellen übertragen werden. Eine solche Applikation ließe sich auch dezentral – also unabhängig von Max for Live – realisieren; aus Gründen der besseren Nutzbarkeit wurde diese

## 4 Objektorientierte Audioproduktion (OOAP)

Schallquellenübersicht aber explizit in Form eines eigenen XY-Feldes in den Prototyp integriert. (So kann unmittelbar in Ableton Live auf die Schallquellenübersicht zurückgegriffen werden, ohne auf eine separate Applikation zurückgreifen zu müssen.) Für die Schallquellenübersicht wurde erneut das Objekt Nodes verwendet: die Raumposition jedes einzelnen Kanals wird über seine Kanalnummer identifiziert und anschließend anhand der übermittelten Panoramadaten auf dem XY-Feld dargestellt.



Abb. 4.11 Darstellung aller Schallquellen im Prototyp

Bei Produktionsumgebungen wie Pro Tools und Cubase muss jeder einzelne Kanalzug angesteuert werden, um die aktuelle Position der Schallquelle zu prüfen. Durch das Abfangen der UDP-Datenströme und der gemeinsamen Darstellung der Panoramadaten innerhalb des Prototyps muss nicht separat in die entsprechenden Kanalzüge gewechselt werden. Als zukünftige Entwicklung wäre eine explizite Benennung der Schallquellen (anstatt der in Abb. 4.11 aufgeführten Nummerierung) denkbar – so könnten die Objekte bei steigender Kanalanzahl einfacher identifiziert werden.

### 4.5 Panorama-Eingabemethoden

In Kapitel 4.2 wurden bereits Eingabemethoden für Panoramadaten angerissen. Bei analogen Mischpulten erfolgt diese noch über sogenannte Panorama-Potis, die eine Verteilung auf den linken und rechten Lautsprecher ermöglichen (siehe Kapitel 2.2.2.3). Das im Prototyp entwickelte und den konventionellen Produktionsumgebungen nachempfundene XY-Feld kann dabei als aktueller Repräsentant der Schallquellen-Darstellung und -Positionierung betrachtet werden. Auch wenn sich die Positionierung der Objekte in der OOAP nicht zwangsweise von der Vorgehensweise in einer herkömmlichen Produktion unterscheidet, sollen an dieser Stelle

weitere Wege zur Eingabe der Panoramainformationen untersucht werden. Besonders in Bezug auf dreidimensionale Verfahren sind neue Eingabemethoden notwendig, da die Positionierung nicht mehr über ausschließlich zweidimensionale Methoden erfolgen kann – letztlich kann die Speicherung der dritten Dimension nicht in einem zweidimensionalen Feld erfolgen. Zunächst sollen aber zweidimensionale Verfahren vorgestellt werden, um die aktuellen technischen Möglichkeiten darzulegen. Anschließend werden durch die Verwendung alternativer Hardware dreidimensionale Eingabemethoden dargelegt.

### 4.5.1 Zweidimensionale Positionierung

Das im Prototyp entwickelte zweidimensionale XY-Feld erfüllt nicht nur den Zweck der Darstellung der Position, sondern letztlich auch der Platzierung der Schallquelle. Auch wenn die Eingabe über ein Endgerät wie eine Computermaus intuitiv ist, ist die Interaktion indirekt.



Abb. 4.12 Hardware-Surround-Panner für Pro Tools (Quelle: [34])

Eine Produktion mit viel bewegten Schallquellen ist so nur mühsam zu erreichen, da die Maus in ihrer Bewegung nicht auf das Panoramafeld begrenzt ist und daher – im Zweifelsfall – im Millimeterbereich verschoben werden muss. Ein Beispiel für eine direktere Interaktion sind sogenannte Surround Panner Joysticks wie in Abb. 4.12. Mit ihnen lassen sich Panoramafahrten intuitiver durchführen, da sich Ihre Dimensionen meist 1:1 auf die Höhe und Breite des Panoramafelds übertragen, und Schallquellen daher nicht versehentlich außerhalb des Panoramafeldes positioniert werden (siehe Mausbedienung).

Ein weiterer Ansatz ist die Verwendung von Touch-Geräten, wie beispielsweise Smartphones oder Tablets. Touchfähige Endgeräte haben den Vorteil, dass die Bedienoberfläche dynamisch an die jeweiligen Bedürfnisse angepasst werden kann.

## 4 Objektorientierte Audioproduktion (OOAP)

---

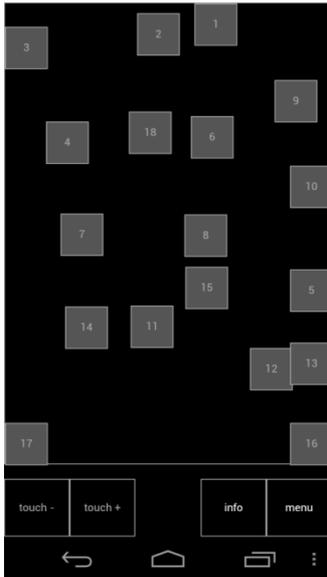


Abb. 4.13 Schallquellenpositionierung mit der Android-App Control

Im Zuge der Entwicklung des Prototyps wurde ein Android-Smartphone in Verbindung mit der Applikation „Control“ verwendet, um Schallquellen zu positionieren und die Panorama-Information an die Produktionsumgebung weiterzureichen. Control sendet dabei die Kanalnummer und XY-Koordinaten über UDP an einen Empfänger, im konkreten Fall den Prototyp. Dieser kann die Koordinaten empfangen, auswerten und auf das XY-Panoramafeld übertragen. Der klare Vorteil von Control ist die wesentlich direktere Bedienung als mit Maus und Joystick: Über eine einfache Fingereingabe können sämtliche Schallquellen positioniert werden. Control bietet ähnlich dem Prototyp eine gemeinsame Darstellung aller Objekte und ermöglicht so eine schnelle Übersicht und Positionierung.

### 4.5.2 Dreidimensionale Positionierung

Für die Positionierung im dreidimensionalen Raum genügt eine XY-Panoramafäche allein jedoch nicht. Zwar wäre es denkbar, die Elevationskoordinate (Z-Koordinate) über einen zusätzlichen Slider (einem „Software-Schiebereglern“) zu implementieren – die Handhabung wäre dabei aber mehr als umständlich: Ließe sich die XY-Position mit einer und die Z-Position dann mit der zweiten Hand auf einem Touch-Gerät durchführen, ist eine sinnvolle Eingabe über eine Computermaus damit nicht mehr möglich – jede Panoramafahrt müsste dabei doppelt ausgeführt werden, da die Maus zeitgleich nur eine Eingabe (XY-Position oder Z-Position) steuern kann.

Weiterhin könnte auf zwei Hardware Surround-Panner zurückgegriffen werden: Ein Surround-Panner steuert dabei die XY-Position, der andere die Z-Position – entsprechend wird beim zweiten Surround-Panner nur auf eine Achse zurückgegriffen. Im Zuge der Entwicklung von berührungslosen Technologien lassen sich aber weitere theoretische Ansätze entwickeln:

Nach dem Vorbild der Zeichnung des vitruvianischen Menschen von Da Vinci lassen sich auch Eingabegeräte mit natürlich wirkenden und intuitiven Eingabebewegungen entwickeln. So kann mit beiden Armen ein gedachter Kreis bzw. im dreidimensionalen Sinne eine gedachte Kugel aufgespannt werden. Ein solcher Bewegungsradius lässt sich für die dreidimensionale Eingabe unter Verwendung verschiedener Technologien nutzen.

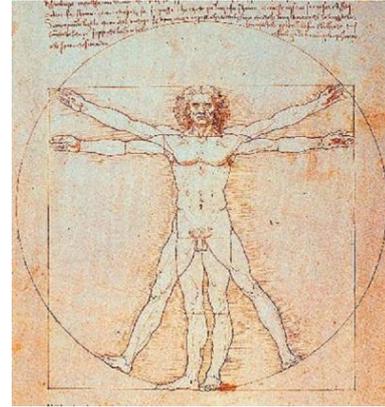


Abb. 4.14 Vitruvianischer Mensch  
(Quelle: [35])

Exemplarisch seien Smartphones mit eingebautem Lagesensor oder auch die von Nintendo entwickelte Wii-Mote aufgeführt. Über die Lagesensoren der Geräte lassen sich Angaben über X-, Y- und Z-Koordinate treffen, die sich wiederum auf ein dreidimensionales Panorama-Modell übersetzen lassen.

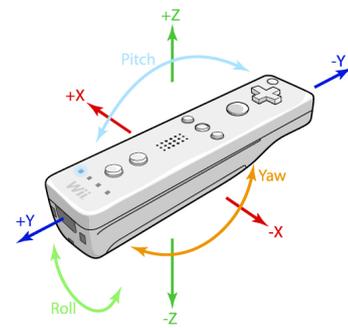


Abb. 4.15 Nintendo Wii-Mote mit  
Achsendarstellung (Quelle: [36])

Die von Microsoft entwickelte Kinect geht dabei noch einen Schritt weiter: Kinect ist ein 3D-Kamerasystem, das Bewegungen im Raum erfassen kann und keine zusätzlichen Eingabegeräte benötigt. Dabei wird der menschliche Körper vom System als Gittermodell bis hin zu den einzelnen Gelenken des Körpers erfasst, sodass explizit Körperbewegungen – bis hin zu einzelnen Gelenkbewegungen – eingefangen werden können.

## 4 Objektorientierte Audioproduktion (OOAP)

---



Abb. 4.16 Kinect mit Kameraerfassung und Erkennung der Körperachsen (Quelle: [37])

Ob Smartphone, Wii-Mote oder Kinect: unter Verwendung dieser Technologien ließe sich eine intuitive und direkte Eingabemethode für den dreidimensionalen Raum realisieren. Auch wenn es für einen Toningenieur ungewohnt sein mag, eine Schallquelle mit ganzem Körpereinsatz zu positionieren, könnte dies durchaus ein interessanter Ansatz sein, gleicht das Positionieren der Schallquelle dabei fast einem Dirigieren.

### 4.6 Vorteile der OOAP

Die vorhergehenden Kapitel haben einen objektorientierten Produktionsansatz und technische Weiterentwicklungen dargelegt. Im Folgenden sollen die daraus resultierenden Vorteile der OOAP zusammengefasst werden.

#### 4.6.1 Lautsprecher-situationen

Durch die Entkopplung der Produktion vom letztendlichen Lautsprecheraufbau können endformatunabhängige Produktionen durchgeführt werden. Weiterhin können bereits während der Produktion verschiedene Lautsprecher-situationen evaluiert werden– sofern ein entsprechender Renderer bereitsteht. Denkbar wäre bspw. auf einem zwei- oder sogar dreidimensionalen Wiedergabesystem zu produzieren und – durch Verändern des Renderer-Systems – unmittelbar zwischen verschiedenen Verfahren zu wechseln. So kann ein unmittelbarer Wechsel zwischen 5.1, Auro 3D, 22.2 oder sogar Verfahren wie der Wellenfeldsynthese (siehe dazu Kapitel 6) realisiert werden. Weiterhin können nicht-ideale Lautsprecher-situationen (siehe Kapitel 2.4.4) untersucht werden, um auch explizit auf Consumer-Bedingungen einzugehen. Entsprechende Rendering-Verfahren und Strategien werden im Kapitel 6 ausführlich vorgestellt.

### **4.6.2 Panorama-Versionen**

Durch die OOAP können zudem verschiedene Panorama-Versionen – sprich verschiedene Varianten der Positionierungen der Schallquellen – angefertigt und bei der Distribution integriert werden. Da die Panoramadaten getrennt von den Audioinhalten aufgezeichnet werden und im Hinblick auf Speicherkapazität unbedenklich sind, können verschiedene Varianten – jeweils angepasst auf spezielle Abhörbedingungen – angefertigt werden, ohne die Audioinhalte bearbeiten zu müssen. Somit stehen dem Produzenten verschiedene gestalterische Möglichkeiten offen. Umkehrt könnte der Konsument bei der Wiedergabe zwischen verschiedenen Panorama-Varianten wählen und so eventuell neue Eindrücke über die wiedergegebenen Audioinhalte gewinnen. (Beispielsweise könnte eine Version mit statischen Positionierungen von Instrumenten einer dynamischen bewegten Positionierung gegenübergestellt werden.)



## 5 Objektorientierte Datenspeicherung (OODS)

Um die in der Produktion entstandenen Audioinhalte auch auf anderen Systemen wiederzugeben, ist zwangsläufig ein Speicherformat notwendig. Wie in Kapitel 2.4.5 aufgeführt, lösen digitale Formate die herkömmlichen haptischen Datenträger immer weiter ab. Auch wenn Blu-Ray unter anderem wegen der relativ hohen Speicherkapazität aktuell einen Stellenwert bzgl. der Distribution von Bewegtbildinhalten besitzt, könnte auch dieses Medium in den nächsten Jahren abgelöst werden. Zwar gibt es durchaus Konzepte, die Blu-Ray als mehrkanaliges Trägermedium zu nutzen, da Blu-Ray-Player meist über Firmware-Updates an neue Spezifikationen angepasst werden können [38]; dennoch steigt der Konkurrenzdruck auf das Medium: Immer schnellere Breitbandinternetanbindungen – auch auf dem Mobilfunkmarkt – ermöglichen so hohe Transferraten, dass Bild- und Toninhalte quasi live übermittelt werden können. Ein solcher „on-demand“-Abruf von Medieninhalten ist schneller und im Zweifelsfall flexibler als die Beschaffung eines haptischen Datenträgers mit entsprechendem Inhalt. Auch wenn sich das Bewusstsein zum haptischen Datenträger noch nicht ganz gelöst hat (Umgangssprachlich formuliert in etwa: „Was ich in den Händen halte, gehört mir“), ist eine zunehmende Entwicklung zu nicht-haptischen Datenträgern festzustellen.

Unabhängig davon, ob zur Distribution ein haptisches oder rein digitales Trägermedium verwendet wird, müssen die produzierten Audioinhalte in einer solchen Form gespeichert werden, dass sie beim Rendering wieder reproduziert werden können. Die Anforderungen an ein solches Speicherformat sind – rein in der Theorie – relativ simpel: Jeder Kanal muss zusammen mit seinen Panoramainformationen in ein Containerformat übertragen werden.

### 5.1 Dateiformate bei der OOAP

Bei der OOAP werden wie in Kapitel 4 aufgeführt Audioinhalte und Panoramadaten (als Steuerdaten) aufgezeichnet. Im konkreten Fall werden die Audioinhalte als sogenannte WAVE-Dateien („Waveform Audio File Format“) aufgezeichnet und die Panoramadaten unter Verwendung des Prototyps als MIDI-Daten gespeichert. Jede Produktionsumgebung bietet die Möglichkeit Audioinhalte als WAVE-Dateien zu exportieren respektive zu speichern. Die aufgezeichneten Panoramadaten lassen sich als MIDI-Daten exportieren und können auch entsprechend weiterverarbeitet werden. Aus dem Export der Audioinhalte als WAVE-Dateien und den Panoramadaten als MIDI-Dateien resultieren allerdings immer eine Vielzahl an Dateien – idealerweise sollen diese aber kombiniert in einer einzelnen Trägerdatei repräsentiert werden. Ein Export auf eine einzige Datei mitsamt allen Audio- und Panoramainhalten ist derzeit mit den verfügbaren Funktionen der Produktionsumgebungen aber nicht möglich.

### 5.2 Konzept: Verbindung der Audioinhalte und Panoramainformation

Seit den 80er Jahren haben sich etliche Dateiformate entwickelt und einige der Audioformate sowohl in Studioumgebungen als auch beim Endverbraucher etabliert. Es liegt daher nahe, die beiden bereits genannten Formate – WAVE und MIDI – genauer zu betrachten und weiterhin alternative Dateiformate auf die gestellte Anforderung alle Inhalte in einer Datei zu vereinen zu untersuchen. Die Abb. 5.1 liefert einen Überblick über die Evolution ausgewählter Audio-Dateiformate.

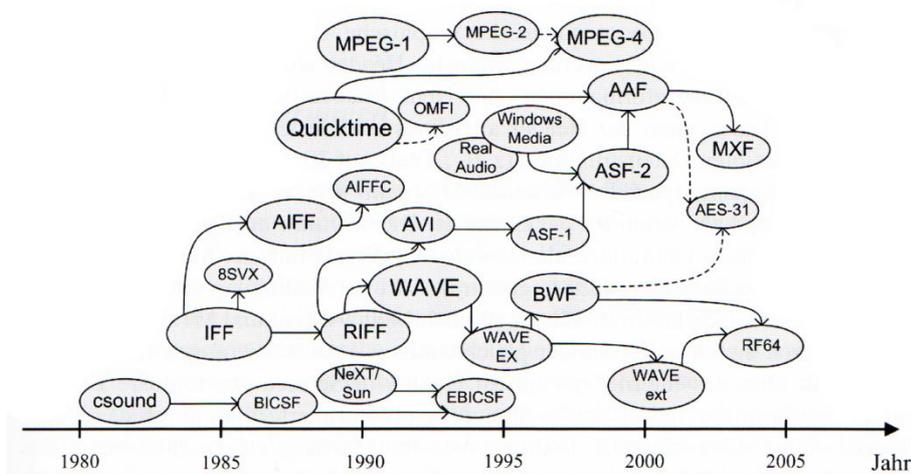


Abb. 5.1 Evolution der Audio-Dateiformate (Quelle: [39] S. 691)

Für die Fusion der Audioinhalte und Panoramainformation in einer Datei existieren dahingehend verschiedene Strategien: Einerseits könnte ein bestehendes Containerformat zur Vereinigung verwendet werden, andererseits könnten die Panoramainformationen auch in ein bestehendes (mehrkanales) Dateiformat eingeflochten werden.

### 5.3 Multichannel interleaved WAVE

Das WAVE-Format wurde in diesem Kapitel bereits angeschnitten. Es dient bei der Audioproduktion sowohl dem Aufzeichnen als auch dem anschließenden Export von Audioinhalten. Der Export folgt dabei immer auf das zu erwartende Endformat: Mono, Stereo, 5.1-Surround o.Ä. Ein Mono-Export repräsentiert eine einzelne Kanalspur, Stereo und Surround beinhalten entsprechend zwei bzw. sechs Kanalspuren. Die mehrkanaligen Dateien werden in diesem Zusammenhang als „interleaved-Datei“ (deutsch: „verschachtelt“) bezeichnet. Eine Stereo-WAVE-Datei ist also – ebenso wie die Surround-Vertreter – eine multichannel-interleaved-WAVE-Datei.

Die Verschachtelung von mehreren Kanälen in eine einzige interleaved-Datei begünstigt zudem die Nutzung von WAVE als Trägermedium für die OODP: Sämtliche Kanäle können dabei zusammengefasst werden und auch die Panorama-Informationen der Kanäle lässt sich separat abspeichern. WAVE und MIDI haben mit dem sogenannten „Interchange File Format“ (IFF) einen gemeinsamen Vorfahren, der die Spezifikation bzw. das Grundgerüst für Audio-, Video- und weitere Formate bereitstellt [39]. Da WAVE auf diesem Grundgerüst basiert, müssen WAVE-Dateien nicht zwangsweise für Audioinhalte verwendet werden; prinzipiell kann jede zeitbasierte Information in eine WAVE-Datei codiert werden. Somit können auch die X-, Y- und Z-Koordinaten der Panorama-Information in Form einer WAVE-Datei abgebildet werden.

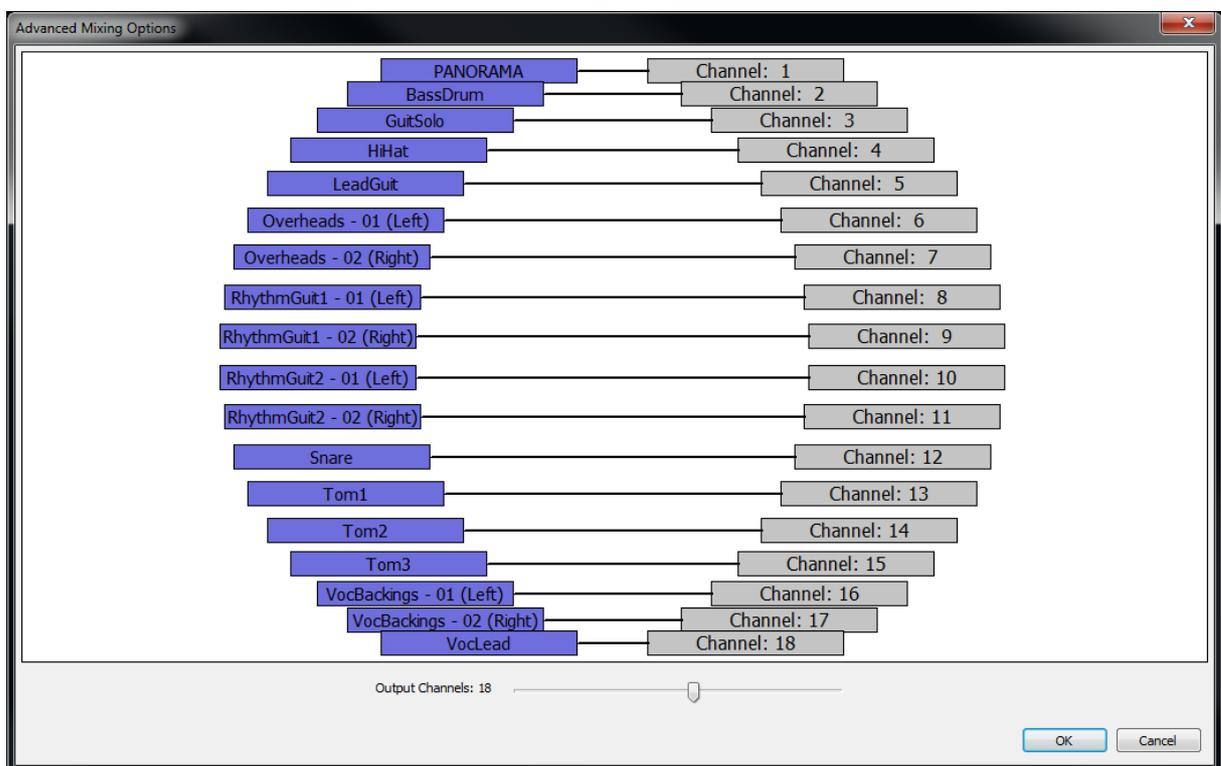


Abb. 5.2 Speicherung einer multichannel-Interleaved-Datei mit Audacity

Wie in Kapitel 4.4.2 aufgeführt, ist eine Abtastung der Panoramadaten entlang der gewählten Abtastrate (bspw. 44.100 Hz) nicht notwendig und würde für unnötiges Speicheraufkommen sorgen. Die Panoramainformationen der Kanäle können dagegen jeweils in eine WAVE-Datei codiert werden: Wenn die Panoramadaten alle 10 ms aktualisiert werden (siehe Kapitel 4.4.2), ergibt sich eine Abtastrate von 100 Hz. Es wäre daher denkbar, die Panoramainformation von mehreren Kanälen innerhalb einer einzelnen WAVE-Datei zu schachteln. Umgerechnet auf eine Abtastrate von 44.100 Hz bei einer WAVE-Audio Datei ergeben sich entsprechend:

$$\text{Formel 5.1: } \frac{44.100 \text{ Hz} * \text{Panoramainformationen}}{100 \text{ Hz}} = 441 \text{ Panoramainformationen}$$

## 5 Objektorientierte Datenspeicherung (OODS)

Bei 3 Panoramaspuren (X-, Y- und Z-Koordinate) je Kanal ergibt sich

$$\text{Formel 5.2: } \frac{441 \text{ Panoramainformationen}}{3 \text{ Panoramaspuren}} = 147 \frac{\text{Panoramainformationen}}{\text{Panoramaspur}}$$

Bei einer Abtastrate von 44.100 Hz könnten also – innerhalb einer einzigen WAVE-Datei – die Panoramainformationen (mit X-, Y- und Z-Koordinaten) von bis zu 147 Kanälen abgebildet werden.

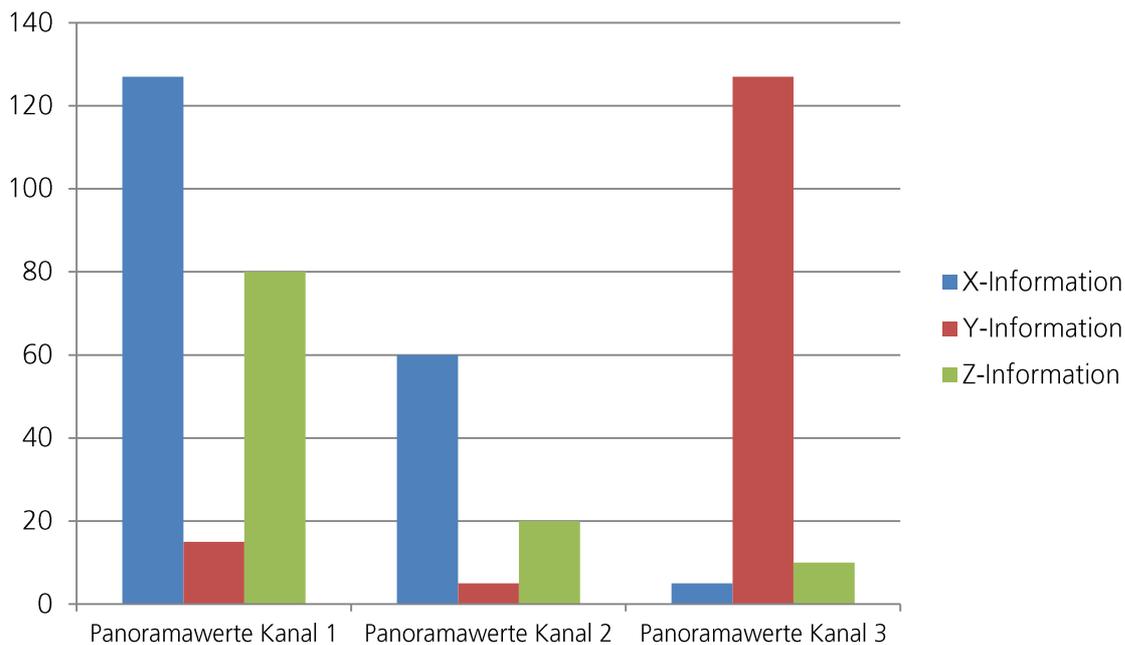


Abb. 5.3 Exemplarische Aneinanderreihung der Panoramawerte X-, Y- und Z in einer WAVE-Datei mit willkürlichen Werten. Bei einer Quantisierung von 16 Bit werden dabei lediglich 128 Abstufungen (=7 Bit) beansprucht.

Durch eine weitere Verschachtelung kann zudem eine noch höhere Effizienz erreicht werden: Der Prototyp verwendet für die Aufzeichnung der Panoramadaten gemäß MIDI-Spezifikation 128 Abstufungen und somit 7 Bit pro Koordinate ( $2^7 = 128$ , siehe u.a. Kapitel 4.3.4). Für die Abbildung von X- und Y-Koordinate (zweidimensional) sind demnach 14 Bit notwendig, für die Abbildung von X-, Y- und Z-Koordinate (dreidimensional) 21 Bit:

$$\text{Formel 5.3: } 2^7 * 2^7 = 2^{14} \text{ und } 2^7 * 2^7 * 2^7 = 2^{21}$$

Eine Wave-Datei mit einer Abtastrate von 44.100 Hz und einer Quantisierung von 16 Bit könnte demzufolge auf einer einzigen Spur die Panoramainformationen von 441 Kanälen abbilden. Wie Formel 5.3 zeigt, ist für die dreidimensionale Abbildung dagegen eine Quantisierung von mindestens 21 Bit notwendig. Die WAVE-Spezifikation erlaubt allerdings verschiedene

Quantisierungsstufen: Für die Weiterverarbeitung auf Audio-CDs wurde zwar die Abtastrate von 44.100 Hz bei einer Quantisierung von 16 Bit gewählt, diese Limitierung kann allerdings vernachlässigt werden, da die Audio-CD im Zusammenhang der OOAP keine Verwendung mehr findet. Denkbar sind daher durchaus auch höhere Abtastraten (bspw. 96.000 Hz) und eine höhere Quantisierung (bspw. 24 Bit).

Rein theoretisch könnten mit einer multichannel-interleaved-WAVE-Datei daher sowohl alle Audioinhalte als auch die dazugehörenden Panorama-Informationen gespeichert werden. Jedoch gibt es bei der WAVE-Datei-Spezifikation eine Limitierung auf 4 GB Datenblockgröße [39]. Dies ist im Hinblick auf die Verwendung als mehrkanaliges Trägerformat jedoch kritisch: Bei einer OOAP mit 50 Spuren bei einer Abtastrate von 44.100 Hz und 16 Bit (=2 Byte) Quantisierung ergibt sich in etwa folgende Laufzeit:

$$\text{Formel 5.4: } 50 * 2 \text{ Byte} * 44.100 \text{ Hz} * \text{LAUFZEIT} = 4.294.967.296 \text{ Byte}$$

Aufgelöst nach Laufzeit:

$$\text{Formel 5.5: } \text{Laufzeit} \cong 973 * \frac{1}{\text{Hz}} \cong 16,23 \text{ Minuten}$$

Eine Laufzeit von ca. 16 Minuten bei 50 Spuren könnte zwar im Hinblick auf Pop-Produktionen anwendbar sein; längere Konzerte oder gar eine Filmproduktion mit größerer Spuranzahl sind damit allerdings nicht zu realisieren. Weiterhin hat sich in der Praxis gezeigt, dass nur wenige aktuell verfügbare Programme tatsächlich entsprechende interleaved-WAVE-Dateien generieren können. Das Programm Audacity (siehe Abb. 5.2) ermöglicht zwar einen Export von 32 Kanälen auf eine einzelne Datei – dies kann aber nicht immer den Anspruch einer OOAP erfüllen. Als Alternativen zu WAVE können die Formate RF64 oder Wave64 gezählt werden, welche die genannte 4 GB-Grenze nicht aufweisen [39] und somit als mehrkanalige Speichermedien durchaus tauglich wären. Audacity implementiert diese beiden Dateiformate zwar, kann hier jedoch ebenfalls nur eine 32-kanalige Datei erstellen. Cubase als Vertreter der Produktionsumgebungen unterstützt Wave64 als Exportformat, limitiert die Kanalanzahl jedoch auf herkömmliche Surround-Formate.

Die Anpassung einer WAVE-Spur innerhalb einer interleaved-Datei ermöglicht die Datenspeicherung im Sinne der OOAP, erfordert allerdings eine kanalweise Codierung der Panoramadaten. Diese müssen zudem bei der Wiedergabe erneut decodiert werden. Ein alternativer Ansatz ist die Verwendung von Containerformaten.

### 5.4 Containerformate

Ein Containerformat dient primär als Behältnis für Dateien unterschiedlichen Datenformats. Mit wechselnden Anforderungen haben sich verschiedene Containerformate entwickelt, meist mit dem Hintergrund Video- und Audioinhalte zu vereinen. Einige der Verfahren ermöglichen dabei eine verlustfreie Datenkompression, andere eine verlustbehaftete Reduktion. Eine Auswahl ist in Tabelle 5.1 aufgeführt.

Containerformat	Extension	Objekte	Einsatz
<b>Quicktime</b>	.mov, .qt	Multimedia	Consumer
<b>MPEG-4</b>	.mp4	Multimedia	Consumer
<b>MPEG-4 Audio</b>	.m4a, .m4b, .m4p	AAC Audio	Consumer
<b>OMFI</b>	.omf	Video, Audio, Projektdaten	Studio
<b>AAF, MXF</b>	.aaf, .mxf	Multimedia	Studio
<b>AES 31</b>	Fat32-Dateien	Audio, Projektdaten	Studio
<b>WMA, ASF</b>	.wma, .asf	Multimedia	Consumer
<b>Real Media</b>	.ra, .rm	Multimedia	Consumer
<b>OGG</b>	.ogg, .ogm	Audio, Video, Text	Consumer
<b>FLAC</b>	.flac	Lossless-Audio	Consumer, Studio
<b>Matroska</b>	.mka, .mkv	Audio, Video, Text	Consumer
<b>OpenMG</b>	.oma, .omg	ATRAC Audio, DRM	Consumer

Tabelle 5.1: Übersicht Container-Formate (Quelle: [39] S. 705)

Ein Containerformat könnte sich im Zusammenhang auch ähnlich eines simplen Ordners – wie in Computerbetriebssystemen – verhalten und dabei die exportierten WAVE und MIDI-Dateien kapseln. Containerformate wie bspw. ZIP (das ebenso zur Datenkomprimierung verwendet wird) könnten für diesen Zweck durchaus in Frage kommen, da sich in ihnen eine quasi unbegrenzte Anzahl an Dateien mit verschiedenen Formaten zusammenfassen lässt [40]. Eine diesbezüglich gekapselte Datei müsste allerdings beim Rendering wieder entpackt werden – da Verfahren wie ZIP aber nicht explizit auf die Audiowiedergabe ausgelegt sind, könnte es hier zu unerwünschten Verzögerungen beim Abspielen der Inhalte kommen.

Die Tabelle 5.1 führt einige explizit für Audio- bzw. Multimedialinhalte entwickelte Containerformate auf. FLAC und Matroska stellen dabei prominente Vertreter dar, beide können auch mehrkanalige Audioinhalte kapseln. Während FLAC dabei maximal 8 Spuren [41]

unterstützt, gibt es bei Matroska prinzipiell keine Restriktionen in der Kanalzahl. Allerdings wird Matroska meist nur in direkter Verbindung mit Bildinhalten verwendet. Zwar wird mit der Dateiendung „.mka“ ein eigenes Containerformat für Audioinhalte zur Verfügung gestellt, es gibt jedoch kaum Software, die Audioinhalte tatsächlich zu einer mehrkanaligen Datei vereinen kann und kaum Software, die diese wiederum wiedergeben kann: Eines der wenigen Programme, das explizit Audioinhalte in MKA-Dateien bündeln kann, ist „mkvmerge“. Hiermit lassen sich mehrkanalige MKA-Dateien erstellen, die sowohl Audioinhalte (bspw. in Form von WAVE-Dateien) als auch MIDI-Dateien beinhalten. Eine anschließende Wiedergabe mit den Software-Playern „VLC“ oder „FooBar2000“ ließ sich im Zuge der Abschlussarbeit nicht realisieren. Auch wenn die Wiedergabefunktion der genannten Player die Audioinhalte und MIDI-Informationen prinzipiell decodieren kann, besitzt keines der beiden Programme eine entsprechende Decodier-Logik, um die Inhalte miteinander zu verknüpfen. Eine als MKA-gepackte Datei ließ sich mit „mkvmerge“ aber wieder auf die tatsächlichen Bestandteile zurückführen. Die Voraussetzung des theoretischen Codierens und Decodierens des Dateiformats soll dahingehend als Beweisführung genügen: Sofern ein Renderer das Matroska-Containerformat implementiert, kann er sowohl auf die Audio- als auch MIDI-Daten zugreifen und über eine geeignete Programmlogik eine Verbindung zwischen den Inhalten herstellen.

## 5.5 MPEG-4

Bislang wurden nur Formate erläutert, welche die Audioinhalte unverändert abbilden und in einer einzelnen Datei verpacken können. Der bereits erwähnte Container Matroska kann sowohl WAVE-Dateien als auch reduzierte Audioinhalte vereinen. Da Matroska aber in diesem Zusammenhang lediglich als Container bzw. Behälter dient, sollen an dieser Stelle weitere Containerformate mit integrierten Kompressionsverfahren und explizitem Bezug zur OODS beleuchtet werden.

Den Siegeszug der Formate der „Moving Picture Coding Experts Group“ (MPEG) hat sicherlich das Format „MPEG-1 Layer 3“ – auch unter der Abkürzung mp3 bekannt – eingeleitet. Mp3 selbst ist ein verlustbehaftetes Format zur Datenreduktion und kann Audioinhalte – unter Verwendung verschiedener psychoakustischer Modelle – mit einem Bruchteil der Datengröße des Originalformats abbilden. Digitale Distributionswege wie das Internet ermöglichen zudem den – durch die geringe Datengröße bedingten – schnellen Versand und Austausch der Audiodateien, auch zwischen einzelnen Endverbrauchern. Auch wenn das Format qualitative Nachteile mit sich bringt, da die Originaldatei durch die Reduktion nicht mehr komplett reproduziert werden kann, diente das Format sogar als Namensgeber für ganze Geräteklassen:

## 5 Objektorientierte Datenspeicherung (OODS)

---

Manche Audio-Hardware, die zum Abspielen von digitalen Formaten verwendet werden kann, wird heute noch als mp3-Player bezeichnet – auch wenn diese Geräte viele weitere (vorhergehende und nachfolgende) Dateiformate auslesen und abspielen können.

Auf das Format „MPEG-1 Layer 3“ folgten weitere Formate wie „MPEG-2“, „AC3“ und „DTS“ bis hin zu „MPEG-4“, das im Zusammenhang der OOAU untersucht wird. Die Bezeichnungen der MPEG-Dateiformate sind gemäß ihrer Entstehungsgeschichte stringent, es kann dennoch zu Verwirrungen durch die ähnliche Namensgebung kommen: MPEG-4 ist primär ein Containerformat, das die objektorientierte Speicherung von Medienströmen, Bildinhalten und weiteren Dateiformaten ermöglicht [39] und ist gemäß ISO-Spezifikation in einzelne Bestandteile aufgegliedert [42]. Eine MPEG-4-Datei kann – bedingt durch die Containerstruktur – viele verschiedene Ausprägungen annehmen: So ist durch die Dateiendung „.mp4“ nicht ersichtlich, ob es sich um eine Containerdatei mit reinen Video- oder um eine Datei mit reinen Audioinhalten handelt. Die für diesen Zweck eingeführte Dateiendungen „.m4a“ für MPEG-4-Dateien mit Audioinhalten ermöglicht zwar eine visuelle Abgrenzung zu „.mp4“ mit (möglichen) Videoinhalten – dennoch handelt es sich bei beiden Dateiendungen um dasselbe Format (MPEG-4) mit denselben Spezifikationen<sup>1</sup>.

### 5.5.1 MPEG-4 AAC (Audiostandard)

Die Spezifikation von MPEG-4 wird auf verschiedene Standards aufgegliedert und beschreibt dabei die einzelnen Bestandteile. Diese reichen von Video über Audio bis hin zum Dateiformat. ISO/IEC 14496-3 (Part 3) beschreibt in diesem Zusammenhang die verschiedenen Audio-Standards für MPEG-4. Einer dieser Audio-Standards stellt dabei „Advanced Audio Coding“ (AAC) dar. Dieser Standard entstand bereits bei der Entwicklung von MPEG-2 als Weiterentwicklung des MPEG-1 Layer 3-Formats und dient der datenreduzierten Speicherung von Audioinhalten. MPEG-4 AAC erreicht laut Herstellerangaben [42] bei einer Bitrate von 128 kb/s bereits annähernd CD-Qualität.

---

<sup>1</sup> Eine Differenzierung zwischen „.mp4“ für Video- und „.m4a“ für Audioinhalte ermöglicht dennoch die Nutzung unterschiedlicher Wiedergabesoftware für unterschiedliche Wiedergabezwecke: So kann das verwendete Betriebssystem die Dateiendung „.mp4“ einem Videoplayer (exemplarisch VLC) und die Dateiendung „.m4a“ einem Audioplayer (exemplarisch FooBar2000) zuweisen.

### 5.5.2 MPEG-4 ALS (Audiostandard)

Ein weiterer Standard ist „Audio Lossless Coding“ (ALS), welcher eine verlustfreie Datenkompression ermöglicht. Er wurde an der TU Berlin entwickelt und grenzt sich von AAC und MPEG-1 Layer 3 durch die verlustfreie Kompression ab, anhand derer das Ursprungssignal originalgetreu reproduziert werden kann. ALS ermöglicht zudem die (theoretische) simultane Speicherung von bis zu 65.536 Kanälen im MPEG-4-Container [43]. Die maximale Kanalanzahl unter Berücksichtigung einer verlustfreien Kompression macht MPEG-4-ALS daher interessant für die Datenspeicherung der Audioinhalte bei der OOAP. Prinzipiell könnten hier Audiokanäle – ähnlich wie bei interleaved-WAVE – für die Speicherung der Panoramadaten zweckentfremdet werden.

### 5.5.3 MPEG-4 SA (Audiostandard)

„MPEG-4 Structured Audio“ stellt einen weiteren Audio-Standard des MPEG-4-Containerformats dar. Mit ihm lassen sich unter anderem MIDI-Dateien speichern. Da die Panoramadaten im Prototyp explizit mit MIDI erfasst werden, wäre eine direkt Übertragung in MPEG-4 SA denkbar. In Kombination mit MPEG-4 ALS wäre so der explizite Zugriff auf die in der Produktion erzeugten Audioinhalte und die dazugehörigen Panoramainformationen möglich.

### 5.5.4 MPEG-4 BIFS (Szenenbeschreibung)

„MPEG-4 Binary Format for Scenes“ (BIFS) stellt einen eigenen Standard (Part 11) zur Szenenbeschreibung des MPEG-4-Containers dar [42]. In BIFS lassen sich Meta-Informationen für zwei- und dreidimensionale Video- und Audioinhalte speichern. In unmittelbaren Bezug auf die OODS können neben den Panoramainformationen weitere Parameter wie bspw. Lautstärke oder auch Automationsverläufe gespeichert werden. Da MPEG-4 SA bereits die Speicherung von MIDI-Kanälen ermöglicht, soll im Kontext nicht weiter auf die Spezifikation von MPEG-4 BIFS eingegangen werden. Dennoch ermöglicht BIFS als Standard für Szenenbeschreibung die Speicherung weiterer Meta-Daten, die für die OOAP nützlich sein könnten: Exemplarisch sei die bi-direktionale Kommunikation mit einem Eingabegerät (wie bspw. einem Smartphone) aufgeführt, mit der explizit Einfluss auf die gespeicherten Meta-Daten genommen werden kann. Denkbar wäre nicht nur das Verändern der Panorama-Informationen von einzelnen Schallquellen, sondern auch das Ändern der Lautstärke oder das Einflussnehmen auf Effektparameter oder Hallräume.

### 5.6 MPEG-SAOC

„MPEG Spatial Audio Object Coding“ (MPEG SAOC) reiht sich in die Familie der MPEG-Dateiformate ein. Mit der Verabschiedung des Standards im Jahre 2010 kann es als relativ junges Dateiformat bezeichnet werden [44]. MPEG SAOC entstand als Nachfolger für MPEG Surround, welches bereits den Ansatz einer datenreduzierten Speicherung von Surround-Inhalten implementiert. Während MPEG Surround einen kanalbasierten Ansatz bzgl. der zu erwartenden Lautsprecherkanäle (5.1 oder 7.1) verfolgt, ist MPEG SOAC rein objektbasiert: Es können N Audiokanäle gespeichert werden, die beim Decodieren auf M Lautsprecher übertragen werden können.

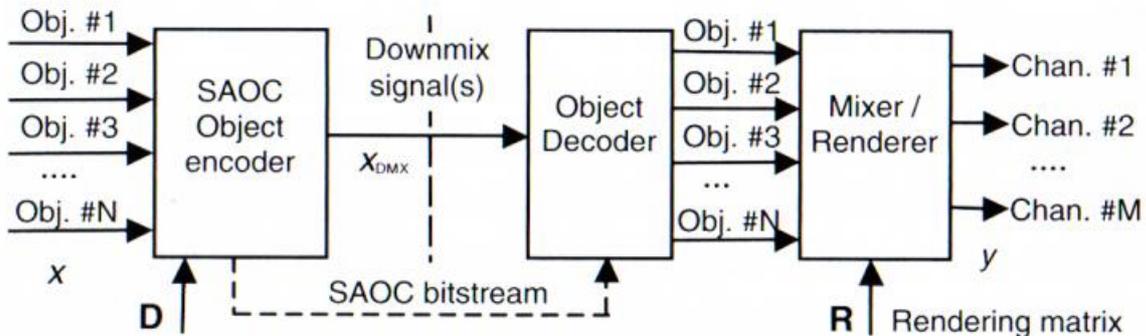


Abb. 5.4 Encoding- und Decoding-Prozess von MPEG SAOC (Quelle: [44] S. 657)

MPEG SAOC verfolgt dabei einen anderen Ansatz als MPEG-4: SAOC ermöglicht zwar – ähnlich wie MPEG-4 – das Speichern von Meta-Informationen; die Audiokanäle werden jedoch nicht separat gespeichert, sondern zu einem Mono- oder Stereodownmix zusammengeführt und mit Meta-Informationen versehen, die zu einer Rückgewinnung der ursprünglichen Audiokanäle dienen sollen. Weiterhin entsteht eine Kompatibilität zu bisherigen Mono- und Stereosystemen. Die Parameter zur Rückgewinnung lauten im Detail [44]:

- „Object Level Differences“ (OLC) beschreiben die relative Lautstärke zum Objekt (bzw. Audiokanal) mit der höchsten Lautstärke in Abhängigkeit von Zeit- und Frequenzband
- „Inter-Object Cross Correlations“ (IOC) beschreiben die Ähnlichkeit zwischen Objekten in Abhängigkeit von Zeit- und Frequenzband
- „Downmix Gains“ (DMG) beschreibt die Lautstärkepegel, die beim Downmix auf die einzelnen Objekte angewandt wurde
- „Object Energies“ (NRG) stellen das absolute Energieniveau des lautesten Objekts in Abhängigkeit von Zeit- und Frequenzband dar.

Anhand dieser Informationen können die Objektkanäle aus dem Downmix zurückgewonnen und für das anschließende Rendering verwendet werden. Beim Decodieren auf ein variables Lautsprecheresetup setzt SOAC auf MPEG Surround auf und dient in dem Zusammenhang nur als Transcoder – es wandelt die im SOAC-Format gespeicherten Inhalte in ein Format, das vom MPEG Surround-Decodierer interpretiert werden kann.

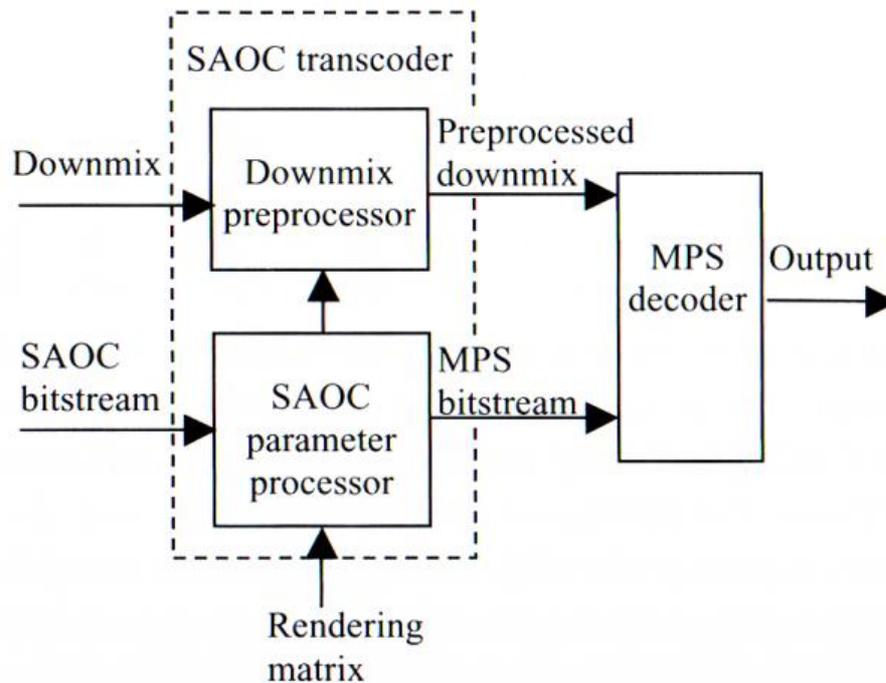


Abb. 5.5 MPEG SOAC als Transcoder für MPEG Surround (Quelle: [44] S. 659)

Der Vorteil gegenüber MPEG-4 ist das effiziente Zusammenführen der Audioinhalte zu einem Mono- oder Stereodownmix und der anschließenden Rückgewinnung über die aufgezeichneten Meta-Informationen. Allerdings verfügt SOAC auch über Restriktionen: es können lediglich 32 Objektkanäle gleichzeitig gespeichert und reproduziert werden. Dies disqualifiziert SOAC im Sinne der in Kapitel 3 und 4 aufgestellten Anforderungen an die OODS, falls die Produktion 32 Kanäle übersteigt. Zudem wird beim Codieren bzw. dem Downmix viel Verantwortung bzw. Logik auf das Format übertragen: Beim Codieren müssen zunächst komplexe Berechnungen erstellt werden, um die Objektkanäle zu kapseln und um anschließend beim Decodieren die entsprechenden Lautsprecheresetup zu generieren. Dieser Prozess erfolgt bei der OOAU in der Regel durch den Renderer, würde im konkreten Fall aber durch SAOC (unter Voraussetzung eines entsprechenden Decoders) vorweggenommen. Ob das Dekodieren über MPEG Surround Geschwindigkeitsvorteile gegenüber einer in MPEG-4 gespeicherten OOAP und separatem Rendering mit sich bringt, muss gesondert betrachtet werden und liefert Gegenstand für

## **5 Objektorientierte Datenspeicherung (OODS)**

---

zukünftige wissenschaftliche Untersuchungen. Weiterhin ermöglicht der Downmix auf Stereo zwar geringere Dateigrößen als eine 32-kanalige MPEG-4-Datei – dennoch lässt sich die These aufstellen, dass dies im Zuge der steigenden Festplattenkapazitäten und dem Ausbau von Breitbandinternetanbindungen nicht zwingend notwendig ist.

## 6 Objektorientierte Wiedergabeverfahren (OOWV)

Die vorhergehenden Kapitel haben dargelegt, wie in der OOAP die Objektkanäle von den Lautsprecher signalen getrennt werden, und wie diese in der OODS gespeichert werden können um sie zu einem entsprechenden Wiedergabesystem zu transportieren. Um die gekapselten Daten wieder aufzubereiten, oder auch um der OOAP eine unmittelbare Abhörmöglichkeit bereitzustellen, müssen die Audioinhalte wieder mit den Panoramadaten zu einer sogenannten Audioszene zusammengeführt werden.

### 6.1 Rendering – Vorüberlegungen

Das Darstellen einer Audioszene wird im folgenden Rendering genannt, das komplette Wiedergabesystem wird als Renderer bezeichnet. Zwischen den Schallquellen bzw. den Produktionskanälen und den Lautsprecher signalen entsteht dabei eine n:m-Beziehung: Jede Schallquelle kann durch eine beliebige Anzahl Lautsprecher repräsentiert werden, und jeder Lautsprecher kann eine beliebige Anzahl an Kanälen wiedergeben. Dies lässt sich anhand Abb. 6.1 verdeutlichen:

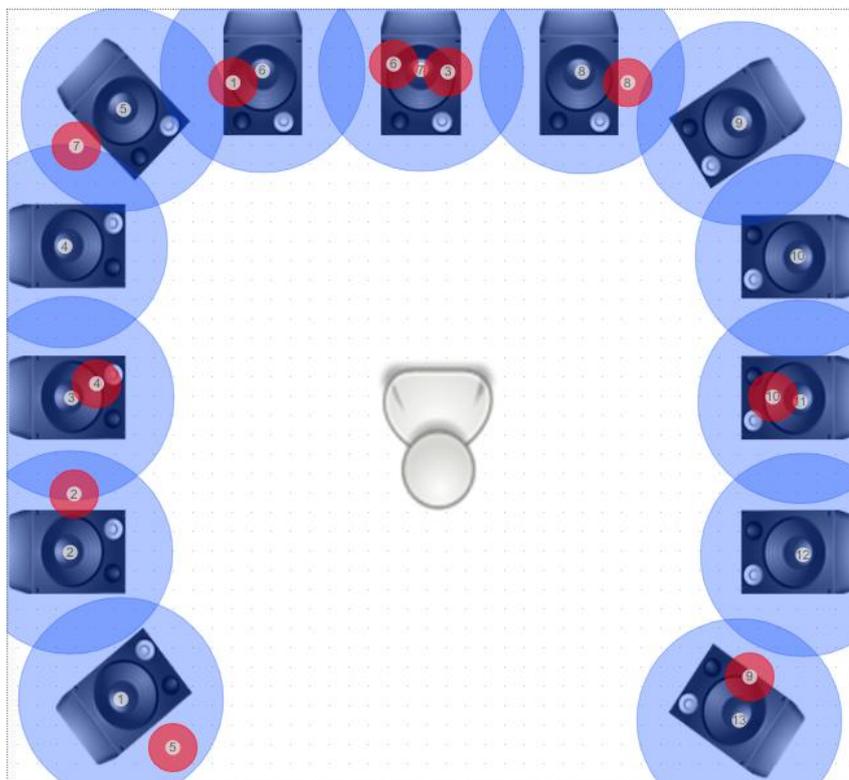


Abb. 6.1 n:m Beziehung zwischen Schallquelle und Lautsprecher (blau) und Schallquellen (rot)

Für die Berechnung der Lautsprecher-signale ist es dabei zwingend notwendig, dass der Renderer die tatsächliche Raumposition der Lautsprecher kennt. Nur so können die Kanäle auch entlang ihrer Panoramainformation im (realen) Raum positioniert werden. Durch diese Vorgehensweise entstehen Vorteile gegenüber dem in Kapitel 2.2.3 aufgeführten 5.1-Surround-Verfahren, die im Folgenden betrachtet werden.

### 6.2 Optimierte 5.1-Surround durch OO

Die in Kapitel 2.2.3 beschriebenen Probleme der Verschiebung der Phantomschallquellen, wenn die Lautsprecher nicht gemäß ITU-R BS. 775 angeordnet sind, wird durch den Renderer ausgehebelt: Wird die Lautsprecherposition im Raum verändert, kann der Renderer explizit auf die neue Position eingehen und somit die Phantomschallquellen berechnen. Dies kann bereits einige Darstellungsprobleme bei unausgewogenen Abhörräumlichkeiten (siehe 2.4.4) und damit verbundenen Lautsprecherpositionen lösen. Bei diesem Ansatz bleibt allerdings zu beachten, dass das aufgeführte Center-Problem bei der Filmwiedergabe damit nicht gelöst werden kann – dieses kann letztlich nur durch die passende Positionierung des TV-Geräts in der tatsächlichen Mitte zwischen L und R gelöst werden.

Dieses Verfahren ermöglicht zudem die Annäherung an weitere Surround-Standards, ohne das Trägermedium zu ändern: Beispielsweise kann ein dynamischer 7.1 Aufbau realisiert werden, indem zwischen L und LS bzw. R und RS weitere Lautsprecher positioniert werden – die Signale für diese Lautsprecher können vom Renderer unmittelbar berechnet werden. Sofern bei der OOAP auch dreidimensionale Panoramadaten aufgezeichnet wurden, können Höhenlautsprecher angebracht und an die Räumlichkeiten angepasst werden.

Die Lautsprecheranzahl kann daher dynamisch an die eigenen Räumlichkeiten und Bedürfnisse angepasst werden. Im Folgenden werden Ansätze aufgeführt, die eine annähernd genaue Raumklangwiedergabe anstreben.

## 6.3 Vector Base Amplitude Panning

„Vector Base Amplitude Panning“ (VBAP) ist ein Verfahren zur Berechnung von LautsprecherSignalen in zwei- und dreidimensionalen Wiedergabesituationen und wurde erstmals von Ville Pulkki im Jahre 1997 vorgestellt [45]. Dabei dient das bereits beschriebene Panning Law (siehe 2.2.2.3) als Grundlage zur Berechnung der Signalanteile und wird für den dreidimensionalen Raum angepasst.

Bei einem stereophonen Aufbau gilt, dass die Leistungssumme der Lautsprecherpaare stets konstant gehalten werden muss. Mit  $g$  für gain (Verstärkungsfaktor des LautsprecherSignals) gilt daher:

$$\text{Formel 6.1: } g_1^2 + g_2^2 = 1$$

In einem zweidimensionalen Modell kann unter Verwendung von  $l_1 = \begin{bmatrix} l_{11} \\ l_{12} \end{bmatrix}$  als Basisvektor zu Lautsprecher 1 und  $l_2 = \begin{bmatrix} l_{21} \\ l_{22} \end{bmatrix}$  als Basisvektor zu Lautsprecher 2 der zur Phantomschallquelle zeigende Vektor  $p$  berechnet werden.

$$\text{Formel 6.2: } p = g_1 * l_1 + g_2 * l_2$$

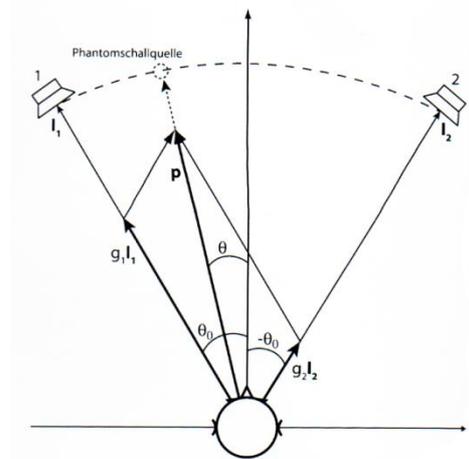


Abb. 6.2 Berechnung der Position der Phantomschallquelle anhand von Vektoren  
(Quelle: [45] S. 458)

Die Formel 6.2 stellt das Ergebnis des Skalarprodukts der Vektoren

$$g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \text{ und } L = \begin{bmatrix} l_1 \\ l_2 \end{bmatrix} \text{ dar.}$$

$$(L \text{ lässt sich alternativ darstellen durch: } L = [l_1 \ l_2]^T = \begin{bmatrix} l_{11} & l_{21} \\ l_{12} & l_{22} \end{bmatrix}^T = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix})$$

Die Formel lautet entsprechend:

$$\text{Formel 6.3: } p^T = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} * \begin{bmatrix} l_1 \\ l_2 \end{bmatrix} = g * L$$

## 6 Objektorientierte Wiedergabeverfahren (OOWV)

Dies lässt sich letztendlich zur Gewinnung der Lautsprecher Signale umformen zu:

$$\text{Formel 6.4: } g = p^T * L^{-1} = [p1 \ p2]^* \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}^{-1}$$

Unter Verwendung dieses mathematischen Modells lassen sich die Lautsprecheranteile berechnen, um eine Phantomschallquelle im Raum positionieren: Stehen Panoramadaten (wie bspw. vom Prototyp aus Kapitel 4) zur Verfügung, können diese so umgeformt werden, dass sie den Vektor  $p$  der Phantomschallquelle mathematisch abbilden. In Kombination des Vektors der Phantomschallquelle und der Vektoren der Lautsprecher lassen sich so die Lautsprecher Signale berechnen. (Eine Anpassung der aufgezeichneten Daten des Prototyps wäre für einen solchen Aufbau jedoch notwendig: Das Koordinatensystem müsste korrigiert und der Nullpunkt in die Mitte des Panoramafeldes verschoben werden, damit die Koordinaten unmittelbar auf den Vektor  $p$  der Phantomschallquelle übertragen werden können, siehe dazu Kapitel 4.3.4.)

Die Berechnung der Lautsprecher Signale lässt sich analog auf ein mehrkanaliges System anwenden: VBAP kann exemplarisch für 5.1-Surround verwendet werden, dabei bilden sich zwischen jedem Lautsprecherpaar eigene Koordinatensysteme anhand der Basiswinkel zu den Lautsprechern (siehe Abb. 6.3).

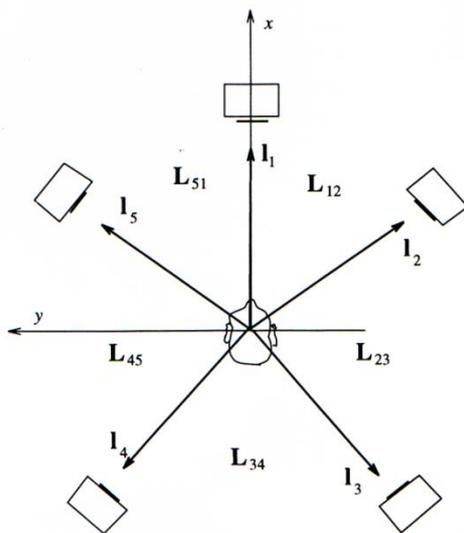


Abb. 6.3 Zweidimensionaler Surround-Aufbau mit VBAP (Quelle: [45] S. 459)

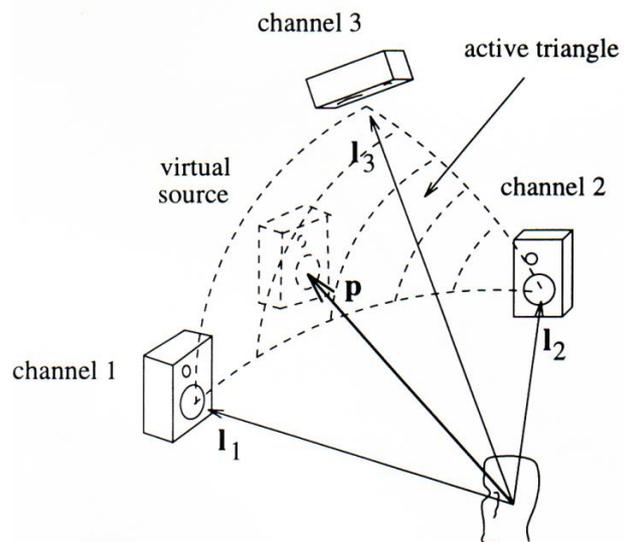


Abb. 6.4 Beispielkonfiguration 3D-Panning mit VBAP (Quelle: [45] S. 480)

VBAP lässt sich zudem auf den dreidimensionalen Raum übertragen, siehe Abb. 6.4. Dabei bilden jeweils drei Lautsprecher einen Lautsprechertripel, die Berechnung erfolgt analog im dreidimensionalen Raum.

Die Grundidee entspricht dabei dem Panning Law übertragen auf drei Lautsprecher: Wenn die Schallquelle unmittelbar auf einen Lautspechters platziert wird, ist nur dieser aktiv. Befindet sich die Phantomschallquelle zwischen zwei Lautsprechern, wird der Pegel entlang des Panning Laws berechnet. Der dritte Lautsprecher hat dabei den Verstärkungsfaktor 0. Befindet sich die Phantomschallquelle in der Mitte des Lautsprechertripels, sind alle Verstärkungsfaktoren der Lautsprecher gleich [45]. Mathematisch ausgedrückt bedeutet dies für die Lautsprechertripel:

$$\text{Formel 6.5: } p = g_1 l_1 + g_2 l_2 + g_3 l_3$$

Somit ergibt sich für die Berechnung der Pegel des Lautsprechertripels:

$$\text{Formel 6.6: } g = p^T * L^{-1} = [p_1 \ p_2 \ p_3] * \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix}^{-1}$$

Durch das Hinzufügen weiterer Lautsprecher ergeben sich neue Lautsprechertripel, für die wiederum die entsprechenden Lautsprechersignale generiert werden können.

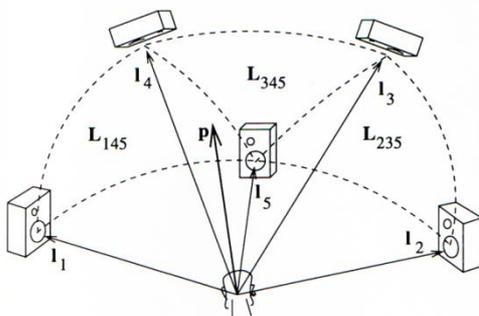


Abb. 6.5 Dreidimensionales VBAP mit fünf Lautsprechern (Quelle: [45] S. 461)

Durch jeden weiteren Lautsprecher entstehen so neue gedachte Dreiecke und Lautsprechertriples, innerhalb derer Phantomschallquellen positioniert werden können. Exemplarisch könnte der Aufbau einer Kuppel aus (Lautsprecher-)Dreiecken um einen Zuhörer realisiert werden, über den sich – ähnlich wie bei Auro 3D (siehe Kapitel 2.2.4) – Schallquellen tatsächlich dreidimensional abbilden lassen. Der Vorteil gegenüber Auro 3D ist die variable Positionierung der Lautsprecher.

VBAP bringt dennoch Restriktionen mit sich: die Lokalisation einer Schallquelle verschiebt sich ebenso wie bei der Stereophonie entlang der Hörposition zum Lautsprecher. Wird der (ideale) Hörbereich verlassen, so verschiebt sich die Lokalisation der Phantomschallquelle wie in Kapitel 2.2.2.2 erläutert. Da das System eine variable Anzahl an Lautsprechern vorsieht, kann dieser Effekt durch das Hinzufügen weiterer Lautsprecher verringert werden.

## 6 Objektorientierte Wiedergabeverfahren (OOWV)

Weiterhin sind Lokalisationsungenauigkeiten in der Medianebene zu erwarten, oder auch in Situationen, in denen sich der Kopf des Zuhörers  $90^\circ$  gedreht zur Phantomschallquelle befindet (siehe 2.1.2 und 2.2.2.2). Hörtests für zwei- und dreidimensionale Lautsprecher-situationen wie in [46] bestätigen diese Erwartung, zeigen jedoch auf, dass eine Lokalisation hinreichend genau möglich, diese aber dennoch frequenzabhängig ist (siehe Kapitel 2.1.2).

### 6.4 Ambisonics

Ambisonics ist ebenfalls ein Verfahren zur (zwei- und) dreidimensionalen Aufzeichnung und Reproduktion von Schallfeldern. Es verfolgt dabei keinen objektorientierten Ansatz und soll daher nicht im Detail aufgeführt, dennoch aber kurz angerissen werden. Die Grundlagen dazu wurden bereits in den 1970ern von Michael Gerzon aufgeführt [11]. Es sollte zunächst als Alternative zu quadrofonen Aufnahme- und Wiedergabesystemen dienen, dennoch blieb der kommerzielle Erfolg des Systems aus – trotz der Vorteile des Systems, die im Folgenden erläutert werden sollen.

Ambisonics verfolgt ebenfalls wie die OOAU den Ansatz Schallquellensignale von Lautsprechersignalen zu entkoppeln und die Lautsprechersignale dynamisch bei der Wiedergabe zu berechnen. Dies wird unter anderem durch die Aufnahme der Schallereignisse mit einem speziellen Ambisonics-Mikrofon realisiert:

Ein konkretes Beispiel hierfür ist das sog. SoundField-Mikrofon. Es besitzt vier separate Mikrofonkapseln und kann somit Vektoren in drei Richtungen inklusive zusätzlicher Rauminformation über die vierte ungerichtete Kapsel abbilden. Das SoundField-Mikrofon zeichnet in dem von Ambisonics vorgegebenen B-Format auf, dieses setzt sich aus den folgenden Bestandteilen entlang den Vektoren zusammen [47]:

W (Druck - ungerichtet)

X (Druck-gradientenanteil vorne/hinten)

Y (Druck-gradienten-anteil links/rechts)

Z (Druckgradientenanteil oben/unten)

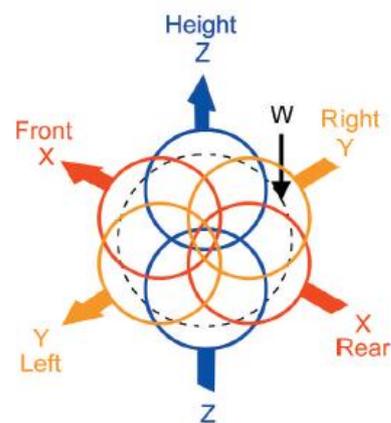


Abb. 6.6 Dreidimensionale Abbildung der Rauminformationen des Soundfield-Mikrofons  
(Quelle: [47])

Wird ein Schallereignis mit dem SoundField-Mikrofon aufgezeichnet, können die Signalanteile der Lautsprecher – in Abhängigkeit ihrer Raumposition – über Matrizenmultiplikation bestimmt werden [11]. Die rechnerische Matrizierung erfolgt dabei ähnlich wie bei VBAP, soll in diesem Zusammenhang aber nicht näher erläutert werden. Die Grundidee für die Berechnung lautet:

$$\text{Formel 6.7: } B = C * S$$

Der Vektor  $B$  trägt in diesem Zusammenhang die Ambisonics-Komponenten, die Matrix  $C$  repräsentiert die Standorte der Wiedergabelautsprecher im Raum. Der Vektor  $S$  wiederum repräsentiert die Lautsprechersignale; aufgelöst nach  $S$  ergibt sich:

$$\text{Formel 6.8: } S = C^{-1} * B$$

Dies gilt, wenn die Lautsprecheranzahl gleich der Komponentenzahl ist [11].

Dennoch ist die Nutzung eines Ambisonics-Mikrofons für die OOAP fraglich: Auch wenn es damit möglich ist Rauminformationen und Schallquellen einzufangen und mit ihren Raumanteilen zu reproduzieren, ist eine solche Aufnahme für die OOAP nicht notwendig. Die Schallquellen können – wie in Kapitel 4 aufgeführt – separat im Raum positioniert werden, die Berechnung der Lautsprecheranteile übernimmt letztlich das Rendering. Weiterhin wird eine Aufnahme von Schallquellen in direkter Verbindung mit Rauminformation am ehesten in der Aufnahme von klassischer Musik Anwendung finden.

Die Decodierungsprozesse von Ambisonics können aber auch unabhängig von der Aufnahme mit einem Ambisonics-Mikrofon durchgeführt werden. Durch die Entwicklung verschiedener Software-Plugins wie bspw. Harpex-B kann Ambisonics direkt in der Produktionsumgebung als Renderer verwendet werden. Wie mit Harpex-B produziert und anschließend Daten exportiert werden, wurde im Rahmen der Ausarbeitung jedoch nicht untersucht.

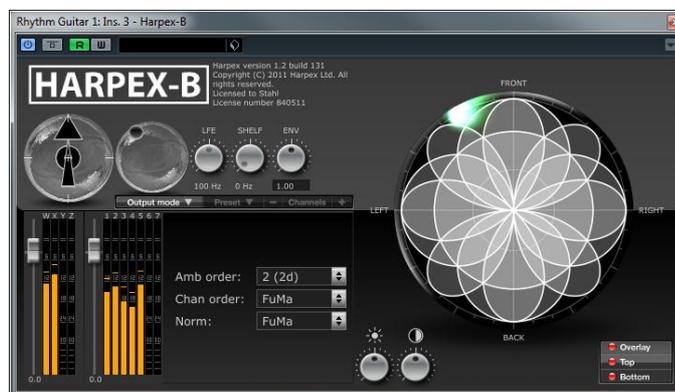


Abb. 6.7 Screenshot des Plugins „Harpex-B“

### 6.5 Wellenfeldsynthese

Die Wellenfeldsynthese (WFS) wurde mit dem Ansatz entwickelt das Schallfeld eines Raumes in einem anderen Raum realistisch nachzubilden [48]. Die WFS macht sich dabei das Huygenssche Prinzip zu Nutze: Dieses besagt, dass sich eine Wellenfront auch durch die Summe von vielen Einzelwellen – sogenannten Elementarwellen – zusammensetzen lässt. Genauer bedeutet dies, dass sich die Wellenfront einer Schallquelle durch die Verwendung von vielen einzelnen (kleinen) Schallquellen, bzw. im konkreten Falle von mehreren Lautsprechern, erzeugen lässt.

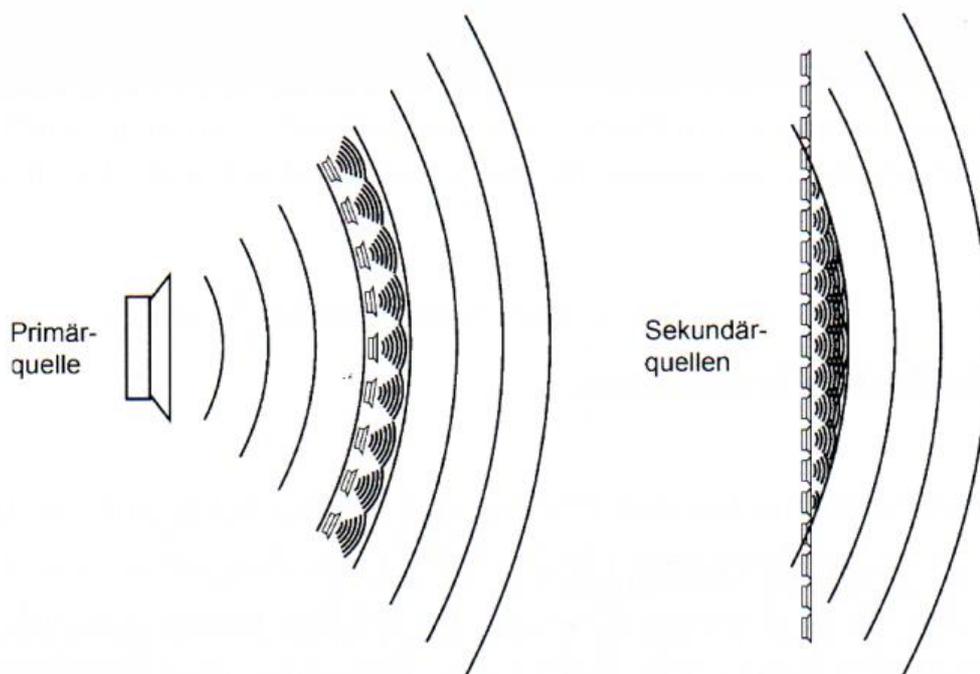


Abb. 6.8 Erzeugung von Elementarwellen mit einer großen Anzahl Lautsprecher (Quelle: [11] S. 666)

Das Huygenssche Prinzip setzt dabei eine unendliche Anzahl Elementarwellen voraus. Bei der WFS muss dahingehend ein Kompromiss gefunden werden, bedingt dadurch, dass eine endliche Anzahl Lautsprecher gewählt werden muss. Bisherige Implementationen der WFS (wie bspw. der Hörsaal H104 der TU Berlin oder die Bregenzer Festspiele) verzichten bislang auf ein dreidimensionales Modell und erzeugen die Wellenfronten lediglich in der horizontalen Ebene. Dies lässt sich u.a. dadurch begründen, dass das menschliche Gehör empfindlicher auf die Lokalisation in der horizontalen Ebene als in der vertikalen Ebene reagiert (siehe dazu Kapitel 2.1 Räumliches Hören). Ein theoretischer dreidimensionaler Ansatz ist zwar denkbar, würde allerdings zu einer Vervielfachung der Lautsprecheranzahl (entlang den räumlichen Dimensionen) führen. Ein einfaches Rechenbeispiel soll dies verdeutlichen:

Werden entlang einer 2x2 Meter Wand 8 Lautsprecher (mit je 25x25 cm) in der Horizontalen aufgestellt, so wären für einen dreidimensionalen Aufbau  $8 \times 8 = 64$  Lautsprecher notwendig. Ein solches Rechenbeispiel lässt sich auf alle Raumwände und im Zweifelsfall sogar Decke und Boden übertragen: Bei einem  $2 \times 2 \times 2 \text{ m}^3 = 8 \text{ m}^3$  Raum wären – für die komplette „Auskleidung“ des Raums – daher ca.  $6 \times 64 = 384$  Lautsprecher nötig. Dies ist bereits die 12-fache Menge im Vergleich zum zweidimensionalen Aufbau mit ca. 32 Lautsprechern.

Nichtdestotrotz ergeben sich bereits beim horizontalen Aufbau erhebliche Vorteile gegenüber der stereofonen Darstellung von Schallquellen: So können virtuelle Punktschallquellen – unabhängig der Hörposition – immer am selben Ort repräsentiert werden, auch wenn sich die Position des Zuhörers ändert. Die Position der Schallquelle bleibt stets dieselbe, und verschiebt sich nicht wie beim Verlassen des Sweetspots bei der stereofonen Wiedergabe. Weiterhin können durch die WFS ebene Wellen dargestellt werden: Eine Schallquelle kann somit immer aus derselben Richtung lokalisiert werden, auch wenn sich der Zuhörer im Raum bewegt. Eine Schallquelle, die frontal auf den Zuhörer gerichtet ist, wird somit von diesem immer frontal lokalisiert, egal, wo er sich im Raum aufhält. Schallquellen können zudem nicht nur hinter der Lautsprecherebene, sondern durch die Überlagerung der Wellenfronten auch vor der Lautsprecherebene dargestellt werden [17] (siehe Abb. 6.9).

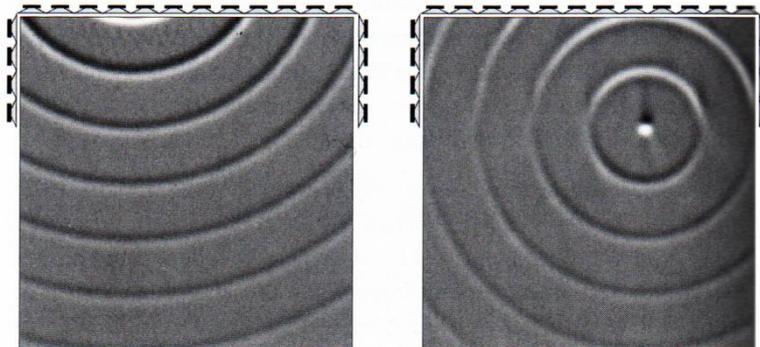


Abb. 6.9 Darstellung einer Punktschallquelle innerhalb und außerhalb der Lautsprecherebene (Quelle: [6] S. 300)

WFS als Wiedergabeverfahren bringt aber auch Einschränkungen mit sich [48]: Durch die diskreten Abstände der Lautsprecher zueinander entstehen jeweils oberhalb einer Grenzfrequenz spektrale Fehler, das sogenannte „Spatial Aliasing“. Das menschliche Gehör kann nur die Frequenzanteile unterhalb dieser Grenzfrequenz für die Lokalisierung nutzen – die Frequenzanteile oberhalb der Grenzfrequenz verursachen Lokalisationsfehler und wirken sich negativ auf die Klangfarbe der Schallquellen aus. Auch wenn bereits Ansätze zur Behebung dieser Artefakte existieren – beispielsweise durch spezielle Filterung der Frequenzen wie in [49] – ist dieser Punkt gegenüber der verwendeten Lautsprecheranzahl und des damit verbundenen

## 6 Objektorientierte Wiedergabeverfahren (OOWV)

---

Aufwands kritisch zu betrachten. Durch den Verzicht auf einen vertikalen Lautsprecheraufbau kann es zu Klangverfärbungen kommen, da sich eine Punktschallquelle nur dreidimensional exakt reproduzieren lässt. Dies kann sich nachteilig auf den zu erwartenden Raum- und Klangeindruck der Schallquelle auswirken. Zudem lässt sich feststellen, dass nicht alle Klangerzeuger reine Punktschallquellen sind: So ist bspw. bereits ein Klavier kein reiner Punktstrahler und benötigt für die korrekte klangliche Darstellung Rauminformationen, die nicht immer berechnet werden können. Ebenso ist es schwer aus einem Schallquellenverbund – wie bspw. einem Chor oder gar Applaus in einer Konzertsituation – einzelne Schallquellen zu extrahieren. Bereits die Verwendung einzelner Stützmikrofone für jeden Musiker eines Chors scheint für die zu erwartende Klangqualität fraglich; für Applaus scheint dies gar unmöglich: Hier müssten für jeden Zuhörer bzw. jede Zuhörergruppe eigene Stützmikrofone verwendet werden. Hier scheint es zielführender, auf Stereo- und Surround-Mikrofonierungsverfahren zurückzugreifen und die aufgezeichneten Quellen als Schallquellenverbund über die WFS darzustellen.

Eine weitere Schwierigkeit besteht in der Messung der Qualität eines WFS-Systems, da Lokalisationsschärfe nicht als einziger Parameter für die Verbesserung der Klangqualität genutzt werden kann: Bereits Stereofone- und Surround-Verfahren ermöglichen – zumindest für einen begrenzten Hörbereich – eine annähernd genaue Lokalisation. Lokalisationsunterschiede von bspw. Surround auf WFS müssen daher im gestalterischen Sinne nicht zwangsweise nachteilig auf die Klangqualität sein. Es scheint daher sinnig, ein WFS-System wie in [22] auf die Parameter Sinn- und Sprachverständlichkeit zu untersuchen. Der in Kapitel 2.3 beschriebene Cocktailparty-Effekt zeigt die darin inbegriffene Problemstellung auf. Unter der Berücksichtigung der Parameter Sinn- und Sprachverständlichkeit und unter Voraussetzung mit einem System Kugel- und ebene Wellen zu erzeugen, lässt sich die These aufstellen, dass sich ein System mit reduzierter Lautsprecheranzahl realisieren lässt, dass dennoch die Vorteile der WFS abbilden kann. Ein solches System muss im Zusammenhang zwar genug Lautsprecher aufweisen, um annähernd genau Kugel- und ebene Wellen zu erzeugen, könnte jedoch im Hinblick auf Sinn- und Sprachverständlichkeit eine wesentlich geringere Lautsprecherdichte aufweisen und damit einen geringeren Aufwand und weniger Kosten als die WFS erfordern.

Im Folgenden werden solche optimierten Verfahren genauer betrachtet. Da die objektorientierte Ansteuerung bzw. Wiedergabe in diesen Systemen prinzipiell auf dieselbe Art und Weise erfolgt wie bei der WFS, wird der objektorientierte Ansatz – die Ansteuerung der WFS über die Produktionsumgebung und das Auslesen der Speicherformate – an dieser Stelle nicht aufgeführt, sondern am Beispiel von konkreten Implementierungen nachfolgend erläutert.

## 6.6 IOSONO

Bisher wurden nur theoretische Ansätze zur Schallreproduktion untersucht. Im Rahmen der Ausarbeitung konnte mit dem IOSONO-Wellenfeldrenderer der IOSONO GmbH aus Erfurt ein WFS-System untersucht werden, das den oben angerissenen Ansatz der optimierten Lautsprecheranzahl verfolgt: Durch die Reduzierung der Lautsprecherdichte werden Anschaffungs- und Betriebskosten minimiert, dennoch ist das System in der Lage Kugel- und ebene Wellen zu erzeugen. Die verringerte Lautsprecherdichte äußert sich in einem kleiner werdenden Hörbereich, dieser hängt wiederum von der Raumgröße und der verwendeten Lautsprecheranzahl ab, und soll im Zusammenhang nicht weiter beachtet werden. Vielmehr soll die Funktion des IOSONO-Renderers dargelegt und untersucht werden.



*Abb. 6.10 exemplarischer Aufbau des IOSONO-Systems im Soundlabor der Klangerfinder GmbH Stuttgart*

Der IOSONO-Renderer verfolgt einen komplett objektorientierten Ansatz. Er kann einerseits direkt an eine Produktionsumgebung angeschlossen werden oder auch objektorientierte Speicherformate wiedergeben und interpretieren. Der IOSONO Renderer verwendet dabei dabei MADi als Schnittstelle (siehe Kapitel 4.4.1.1), somit ist eine zeitgleiche Übertragung von bis zu 64 Kanälen zum Renderer möglich.

Schallquellen bzw. Schallquellenkanäle werden bei der Wiedergabe als eigene Objekte in einer Szenendarstellung visualisiert: Hier werden aktuelle Positionen und Automationsverläufe dargestellt. Weiterhin wird visuell zwischen Punktschallquelle und ebener Welle unterschieden.

## 6 Objektorientierte Wiedergabeverfahren (OOWV)

Die Panoramadaten des Prototyps aus der OOAP können über UDP unmittelbar an den Renderer weitergereicht werden, dieser positioniert die Schallquellen entsprechend ihrer Panoramadaten in der Szene. Bei der Wiedergabe von Speicherformaten werden dabei die aufgezeichneten Positionsdaten decodiert und die Schallquellen entsprechend positioniert.

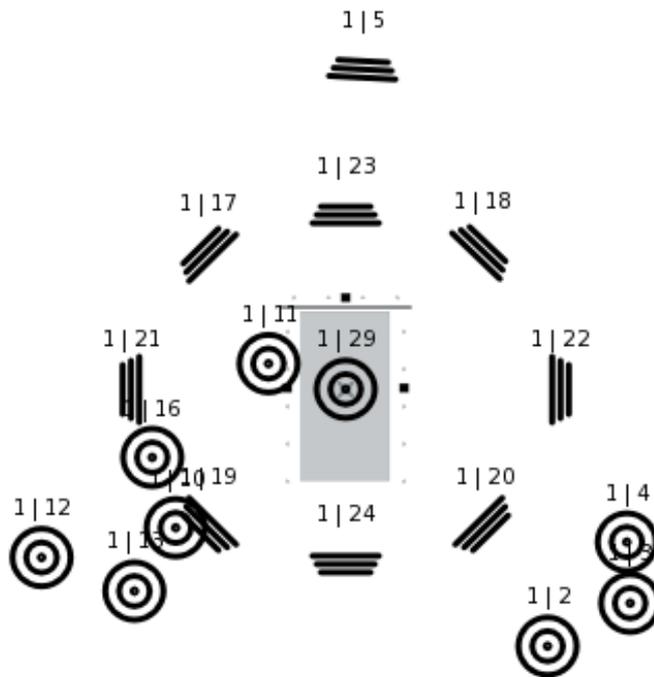


Abb. 6.11 Szenendarstellung samt Schallquellenposition bei der Wiedergabe

Innerhalb des Renderers können verschiedene Lautsprechersituationen abgebildet und gespeichert werden (siehe Abb. 6.12). Die Lautsprechersignale werden anhand der Panoramadaten und Objekteigenschaften (wie Punktschallquelle oder ebene Welle) berechnet. Die Berechnung selbst erfolgt ähnlich wie bei der konventionellen Wellenfeldsynthese: Entlang des Huygensschen Prinzips werden die einzelnen Lautsprechersignale berechnet, sodass diese in Kombination – je nach Schallquellentyp – entweder eine Kugelwelle oder eine ebene Welle darstellen.

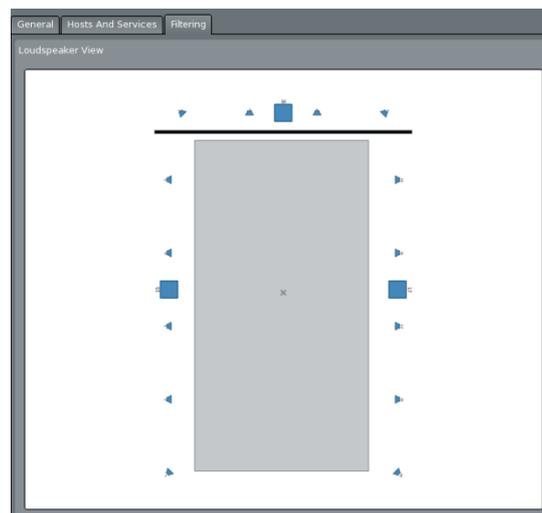


Abb. 6.12 Lautsprecheransicht und Abbildung des realen Raumes

Der im Renderer gespeicherte Lautsprecheraufbau wird dabei zur Berechnung der Signale verwendet. Die verwendeten Lautsprecher werden anhand einer Routing-Matrix auf die Ausgänge der MADI-Soundkarte übertragen, und die Lautsprechersignale somit entlang der gewählten Ausgangskanäle des Renderers über MADI an einen DA-Wandler weitergereicht. Dieser wiederum verteilt die analogen Signale auf Endstufen und Lautsprecher.

Dem System sind durch die vorhandenen Lautsprecherabstände dennoch Grenzen gesetzt. Prinzipbedingt kommt es auch beim IOSONO-Renderer zu Spatial Aliasing, was jedoch entlang des im Renderer definierten Lautsprecheraufbaus durch Filtering minimiert wird. Wird eine Schallquelle weiterhin zwischen zwei Lautsprechern positioniert, wird die Schallquelle als Phantomschallquelle abgebildet, und nicht mehr entlang des Prinzips der WFS. Die möglichen Probleme bei der Lokalisation einer Phantomschallquelle wurden schon eingehend erläutert, diese lassen sich analog auf dieses System übertragen. Dennoch genügt dieses Verfahren um Schallquellen hinreichend genau abzubilden: Die Lokalisationsfehler sind dabei als gering einzustufen, da der Hörbereich vom Renderer entlang der Lautsprecheraufstellung festgelegt werden kann. Explizite Lokalisationsfehler der Phantomschallquelle sind daher nur festzustellen, wenn der Hörbereich verlassen wird.

IOSONO ist zum Zeitpunkt der Ausarbeitung ein zweidimensionales Verfahren zur Schallreproduktion, eine dreidimensionale Abbildung wäre mit diesem System zukünftig denkbar. Die entsprechenden Voraussetzungen wie die Z-Position der Lautsprecher sind bereits im Renderer implementiert. Durch die Ansprechmöglichkeit des Systems über MADI und UDP konnte es im Zusammenhang mit dem in Kapitel 4.3 entwickelten Prototyp für einen exemplarischen Aufbau genutzt werden, um die Verbindung zwischen OOAP und OOWV und die dabei aufgeführten Arbeitsabläufe zu verifizieren.

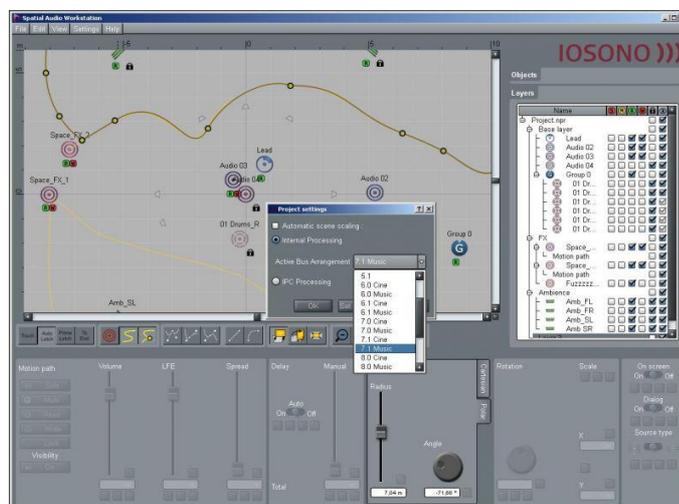


Abb. 6.13 Spatial Audio Workstation (Quelle: [50])

## 6 Objektorientierte Wiedergabeverfahren (OOWV)

IOSONO stellt mit „Spatial Audio Workstation“ zudem ein eigenes Plugin für die Produktionsumgebung Nuendo (Steinberg) bereit, mit dem unmittelbar auf die Funktionen des Renderers zurückgegriffen werden kann. Unter Verwendung der Spatial Audio Workstation kann Nuendo somit ebenfalls zu einer OOWV-Umgebung aufgerüstet werden. Dabei ist es sowohl möglich, die produzierten Inhalte für den IOSONO-Renderer zu speichern und zu exportieren, als auch explizite Downmixes auf 5.1-Surround (und weiter Surround-Varianten) sowie Stereo zu erstellen.

Mit dem Renderer selbst sind weiterhin Upmixes von bestehenden Stereo- oder 5.1-Surround-Aufnahmen möglich. Dabei werden die Lautsprechersignale für die im Renderer hinterlegten Lautsprecherpositionen berechnet.

### 6.7 Shure Atmosphaea

Shure Atmosphaea ist – ähnlich wie IOSONOs-Wellenfeldrenderer – ein optimiertes WFS-System. Dabei kommen neben Wandlautsprechern zusätzliche Deckenlautsprecher zum Einsatz, die entlang einer Gitternetzstruktur über die Raumdecke verteilt sind. Shure selbst bewirbt das Atmosphaea-System mit dem Ansatz Schauräume, Konzertsäle und Kongresszentren zu beschallen.

Durch die zusätzlichen Deckenlautsprecher (wie in Abb. 6.14) können demnach dreidimensionale Schallereignisse abgebildet werden. Atmosphaea selbst verwendet einen gemischten Rendering-Ansatz: Ein eigener (Metadaten-) Renderer decodiert die Szenendarstellung (die bspw. aus einer gespeicherten Datei gewonnen werden kann) und berechnet Meta-Informationen der Schallquellen (konkret: X-, Y- und Z-Position im Raum samt Wellenformausprägung und -ausbreitung).

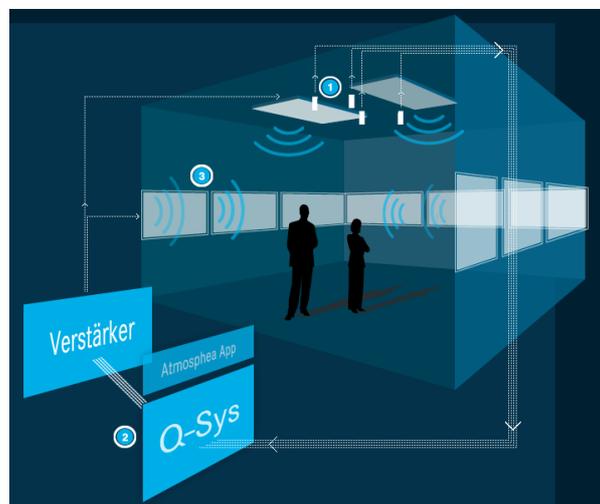


Abb. 6.14 Exemplarische Darstellung des Atmosphaea-Systems mit Wand- und Deckenlautsprechern, Deckenmikrofonen und Rendering-Engine (Quelle: [51])

Anschließend werden diese Meta-Informationen an ein (Audio-) Renderer weitergeleitet, der diese mit den Audioinhalten verrechnet und entsprechende Lautsprecher-Signale generiert (siehe Abb. 6.15).

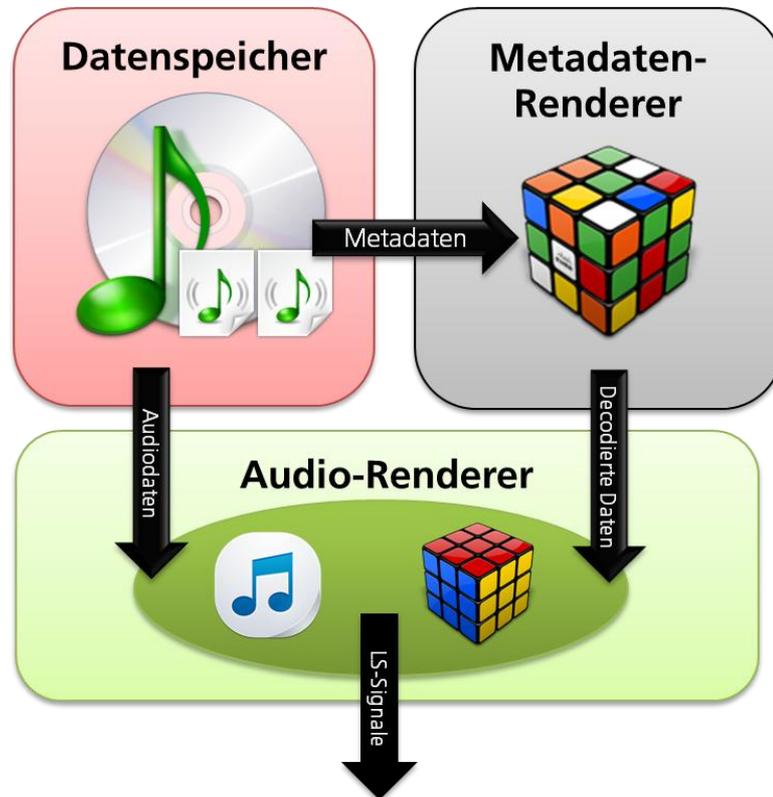


Abb. 6.15 Schematischer Aufbau des Rendering-Prozesses

Durch in die Decke eingebauten Mikrofone können mit Atmospha zudem Schallquellen im Raum aufgenommen und aufbereitet werden: Die Schallereignisse werden dabei mit Impulsantworten anderer Räume kombiniert. Der künstliche Hallanteil wird unmittelbar wiedergegeben, sodass ein (anderer) räumlicher Klangeindruck entsteht. Durch die Verwendung verschiedener Raumimpulsantworten kann somit jeder erdenkliche Raum simuliert werden.

Atmospha verwendet mit Q-Sys zudem ein eigenes System zur Übermittlung der Audioinhalte vom Renderer zu den Endstufen bzw. den Lautsprechern. Q-Sys ermöglicht die Übermittlung von bis zu 1.024 Kanälen (davon 512 Ein- und 512 Ausgangskanäle) über Gigabit-Ethernet und kann dahingehend in die Kategorie des gemischten Hardware-Ansatzes (siehe Kapitel 4.4.1.3) eingeordnet werden.

### 6.8 Dolby Atmos

Dolby nimmt im Kontext der Surround-Wiedergabe eine Sonderstellung ein. Dolbys Entwicklungen der letzten Jahrzehnte haben sowohl die Audiowiedergabe in Kinosälen als auch beim Konsumenten maßgeblich beeinflusst, unter anderem entstanden Verfahren wie Dolby Digital, Dolby Digital Plus und Dolby True HD und eine Vielzahl weiterer Encodier- und Decodier-Methoden für mehrkanalige Audioinhalte [11].

Im Zusammenhang der mehrkanaligen Audiowiedergabe wurde auch Dolby Atmos entwickelt, das wie IOSONO und Shure Atmosphaea einen objektorientierten Ansatz verfolgt. Es wurde vornehmlich für die Produktion bzw. die Wiedergabe von Filmtönen entwickelt, andere Anwendungsmöglichkeiten wären allerdings denkbar.

Der Aufbau von Dolby Atmos ähnelt dem Aufbau von Shures Atmosphaea: Beide verwenden Seitenlautsprecher und ein zusätzliches Gitternetz an Deckenlautsprechern. Beide Verfahren besitzen zudem einen (vermutlich ungewollt) ähnlichen Namen, Dolby verfolgt allerdings nicht den Ansatz der optimierten Wellenfeldsynthese, sondern steuert die Lautsprecher über ein striktes Panning-Verfahren an.

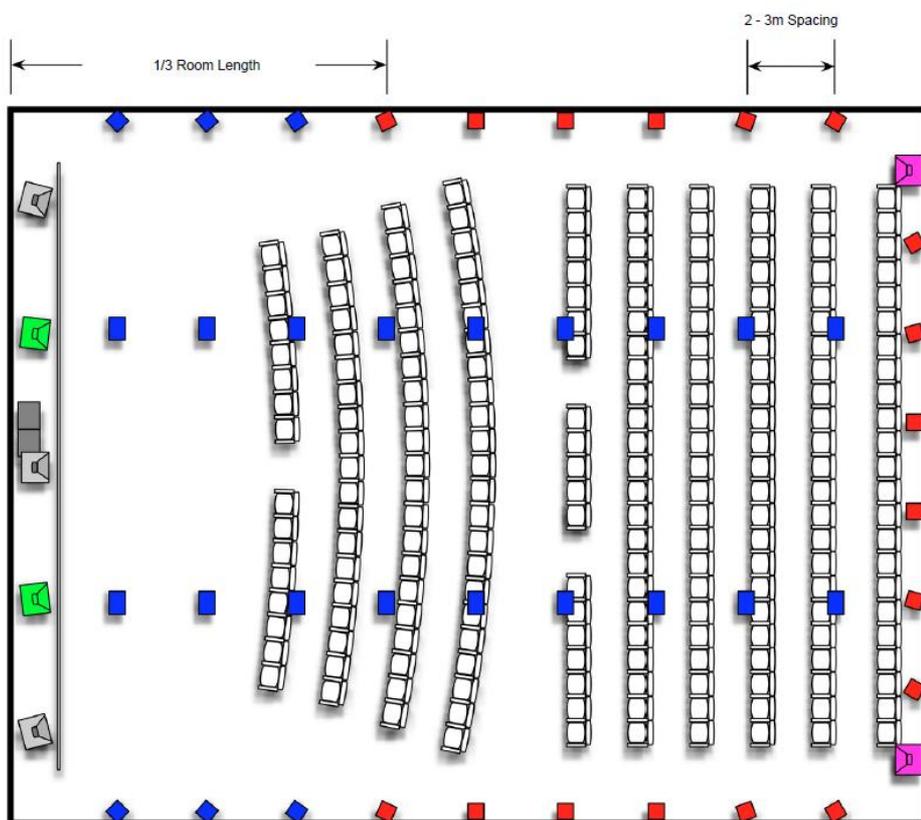


Abb. 6.16 Empfohlener Lautsprecheraufbau von Dolby Atmos (Quelle: [52] S.18)

Blau: zusätzliche Surround-Lautsprecher, pink: optionale Subwoofer, grün: optionale Leinwandlautsprecher

Durch die in Abb. 6.16 beschriebene Anordnung der Lautsprecher ist eine dreidimensionale Positionierung und Lokalisation der Schallereignisse möglich. Durch die ausreichend hohe Lautsprecherdichte kann das Verfahren zudem eine annähernd genaue Lokalisation ermöglichen (siehe dazu Kapitel 0), geringe Lokalisationsfehler sind im Zusammenhang der Phantomschallquellen – zumindest bezüglich der Filmtone wiedergabe und der zu erwartenden Nutzung der Decken- und Seitenlautsprecher (siehe Kapitel 2.3) – aber als nebensächlich zu betrachten.

Wie das Panning zwischen den Lautsprechern im Detail durchgeführt wird, konnte im Rahmen der Abschlussarbeit nicht ermittelt werden. Die Berechnung der Lautsprechersignale könnte allerdings wie in Kapitel 0 VBAP erfolgen, genauere Informationen über das tatsächliche Verfahren und die damit verbundene technische Funktionsweise von Dolby Atmos konnten jedoch nicht gewonnen werden.

Dennoch soll der entlang Dolby Atmos hervorgebrachte Produktionsansatz genauer beleuchtet werden: Ein eigens von Dolby entwickeltes Panorama-Tool (ähnlich dem Prototyp in Kapitel 4.3) ermöglicht das Positionieren von Objekten im Raum. Das System kann aktuell zwei MAD-Signalströme zusammenfassen und somit maximal 128 Objekte auditiv darstellen. Eine solche Kanalanzahl scheint im Hinblick auf Filmtoneproduktion kritisch, Dolby hat für diesen Zweck einen eigenen Ansatz für die Objektdarstellung und Lautsprecherzuweisung entwickelt:

Die 128 Kanäle werden auf 118 frei im Raum platzierbare Objekt-Kanäle und 10 statisch positionierte Bett-Kanäle aufgeteilt. Demnach können statische Elemente der Tonproduktion – wie beispielsweise Dialog, Filmmusik oder auch eine Grundatmosphäre – auf die statischen Kanäle entsprechend ihrer Position gemischt werden. Sämtliche bewegte Objekte (wie exemplarisch über die Zuhörer fliegende Hubschrauber, vorbeifahrende Autos etc.) können dabei als eigene Objekte dargestellt werden. Dies stellt durchaus einen hybriden Ansatz der objektorientierten Audioproduktion dar, ist jedoch im Hinblick auf die zu erwartenden statischen und dynamischen Objekte gänzlich zweckerfüllend umgesetzt.

Weiterhin erfolgt die Dateispeicherung in einem hybriden Format (siehe Abb. 6.17): So wird zusätzlich zum Atmos-Mix aus Objekt- und Bettkanälen ein weiterer Surround Mix (bspw. 5.1- oder 7.1-Surround) in das sogenannte „Material Exchange Format“ (MXF) für die Distribution an Kinos codiert. Dies hat den Vorteil, dass auch ältere Dolby-Wiedergabesysteme verwendet werden können, und ein nahtloser Umstieg auf Dolby Atmos möglich ist.

## 6 Objektorientierte Wiedergabeverfahren (OOWV)

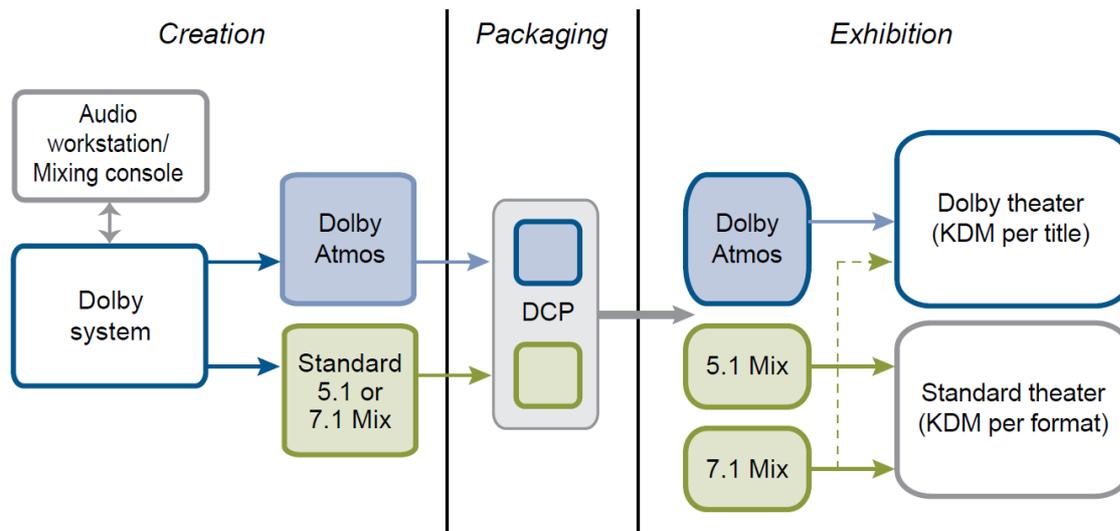


Abb. 6.17 Produktions-, Speicher- und Wiedergabeprozess von Dolby Atmos (Quelle: [52] S.11)

Über die Dolby Atmos-Infrastruktur können zeitgleich 64 Ausgangskanäle angesteuert werden. Somit sind exemplarische Aufbausituationen wie 61.3 realisierbar, die letztliche Anzahl der zu verwendenden Lautsprecher muss bzw. kann durch die objektorientierte Produktion, Speicherung und Wiedergabe allerdings dynamisch auf die Kinosaalgröße angepasst werden.

Fraglich ist, ob Dolby Atmos Chancen hat sich auf dem Konsumentenmarkt zu etablieren – das System wurde jedoch nicht unter diesen Gesichtspunkten entwickelt. Dennoch bietet Dolby Atmos durch die hybride Produktionsweise über statische und dynamische Kanäle, die hybride Speicherung eines kanalbasierten und objektorientierten Formats und letztlich der Möglichkeit eines variabel wählbaren Wiedergabesystems genügend technisches Potenzial, um auch auf Seite der Konsumenten auf Akzeptanz zu stoßen.

## **7 Objektorientierte Audioumgebungen – Epilog**

In den vorhergehenden Kapiteln wurden objektorientierte Audioumgebungen samt ihren Bestandteilen Produktion, Datenspeicherung und Wiedergabe untersucht und ausführlich dargelegt. Weiterhin wurde verdeutlicht, wie durch die Anwendung eines objektorientierten Ansatzes auf variable und sich ändernde Wiedergabesituationen eingegangen werden kann und somit eine Kompatibilität zu zukünftigen Wiedergabesystemen ermöglicht werden kann. In diesem Kapitel werden die Bestandteile der OOAU in einen gemeinsamen Kontext gesetzt um die unmittelbaren Chancen und Risiken aufzuzeigen. Zudem werden die Stakeholder und somit die unmittelbar mit den OOAU in Verbindung stehenden Nutzer- und Konsumentengruppen aufgeführt, um die Barrieren für eine erfolgreiche Integration eines solchen Ansatzes zu veranschaulichen.

### **7.1 Risiken und Probleme der OOAU**

In Kapitel 3.3 wurde die Zielgruppe der OOAU aufgeführt und den entsprechenden Bestandteilen zugeordnet. Dennoch können nicht nur Audio-Schaffende und Audio-Konsumierende in unmittelbare Verbindung mit OOAU gebracht werden:

#### **7.1.1 Produktion & Stakeholder der OOAP**

Ausgehend von der Produktion müssen zunächst neue Soft- und Hardwarekonzepte erstellt werden, um einen objektorientierten Ansatz durchdacht zu implementieren. Der aufgeführte Prototyp kann zwar als Plugin in eine bestehende Produktionsumgebung integriert werden, jedoch kann dies lediglich als Kompromisslösung betrachtet werden: Um den objektorientierten Ansatz unmittelbar in die Produktionsumgebungen einzubinden und um eine Kompatibilität zu bestehender Studio-Hardware zu ermöglichen, müssen Hersteller objektorientierte Panoramaverfahren integrieren und sich von kanalorientierten Mix-Prozessen lösen. Ebenso müssen Hardwarehersteller neue Konzepte für die dreidimensionale Eingabe der Panoramadaten entwickeln, und im Zweifelsfall Hardware (wie bspw. Mischpulte) auf einen objektbasierten Ansatz anpassen. Ein solcher Prozess ist jedoch nicht einfach durchzuführen, da auch die Nutzer der Soft- und Hardware (wie bspw. Produzenten, Toningenieure und Musiker) von diesen Entwicklungen abhängen. Auch wenn der Prozess der Positionierung von Schallquellen im Raum weitestgehend an die bisherigen Methoden adaptiert werden kann, so erfordert der Umstieg auf eine OOAP die Akzeptanz für das Verfahren und ein Vertrauen in die nachfolgenden Beteiligten (siehe 7.1.2 und 7.1.3).

### 7.1.1.1 Produktionsdetails – große Kanalzahl

Über den Verlauf einer Audioproduktion hinweg lässt sich in der Regel eine Tendenz zur letztendlichen Kanalanzahl feststellen. Lassen sich kleine (Musik-)Projekte bereits mit einer relativ geringen Anzahl zwischen 2 und 32 Kanäle realisieren, kann eine Filmttonmischung eine drei- bis vierstellige Kanalanzahl erfordern. Da jeder einzelne Kanal die Panoramainformation speichert und für das Rendering bereitstellt, kann es bei einem solchen Projekt zu Software- bzw. Hardware-seitigen Engpässen kommen: Ist es in der Produktion noch möglich eine quasi unendliche Anzahl an Kanälen zu verarbeiten, müssen die objektorientierte Daten bzw. Audioszenen für die Wiedergabe zwingend an einen Renderer weitergereicht werden.

Kapitel 4.4 zeigt, dass die aktuelle Grenze der Übermittlung zwischen Produktionsumgebung und Renderer über Hardware zwischen 192 (MADI) und 512 (Netzwerkschnittstelle) Kanälen liegt. Für einen erfolgreichen Ansatz der OOAP ist es daher zwingend notwendig, dass die Hersteller gängiger Produktionsumgebungen sowohl die Aufzeichnungen der Panoramadaten (wie im Prototyp dargelegt) als auch explizite Rendering-Prozesse (siehe Kapitel 6) in ihre Software integrieren. Dies würde zumindest das Problem der Kanalübermittlung aushebeln.

Sofern die Produktionsumgebungen dahingehend nicht erweitert werden, ist das System auf eine fixe Kanalzahl begrenzt und widerspricht prinzipiell dem Ansatz der objektorientierten Produktionsweise. Dolby Atmos führt an dieser Stelle einen hybriden Ansatz auf, in dem statische und dynamische Positionen kombiniert werden. Dies schränkt zwar einerseits die OOAP ein, ist aber im Sinne der Filmttonwiedergabe durchaus zweckdienlich: Nicht jede Produktion basiert ausschließlich auf bewegten Schallquellen, sondern kann ebenso ein gewisses Kontingent an statischen Schallquellen aufweisen.

Im Bereich des Filmtons gibt es exemplarisch viele statische Schallquellen wie Filmmusik oder Atmosphäre. Es scheint daher redundant die Panoramadaten einer solchen (statischen) Schallquelle aufzuzeichnen und diese für das Rendering bereitzustellen und dort auszuwerten, da sich die Position über den Verlauf des Stücks nicht ändert. Eine Mischung auf 10 fest im Raum positionierte Kanäle (und somit 10 Objekte) entspricht weitgehend dem Vorgehen bei 5.1-Surround (bei dem jedoch auf 6 Lautsprecherkanäle gemischt wird). Die Interpretation der Position dieser 10 Objekte im Raum wird schlussendlich aber durch den Renderer durchgeführt:

Bei der Produktion werden also keine Lautsprechersignale generiert, sondern lediglich verschiedene Kanäle zu einem Objekt zusammengefasst. In der Produktion könnten Schallquellen exemplarisch auf verschiedene Aux-Kanäle gemischt werden (siehe Kapitel 2.2.3), jeder Aux-Kanal stellt dabei ein eigenes Objekt in der Audioszene dar. Dieses Verfahren lässt sich analog auf die komplette OOAP und wiedergabeseitig auf VBAP, die Wellenfeldsynthese, den ISONO Wellenfeldrederer und Shure Atmospea übertragen.

### **7.1.2 Dateispeicherung & Stakeholder der OODS**

Kapitel 5 listet Dateiformate auf, die für die OODS geeignet sind. Dennoch gibt es kaum Software, die über einfache Mittel eine objektorientierte Speicherung in eines dieser Formate ermöglicht. An sich müssen diese Speicherformate lediglich von den Herstellern in die Produktionsumgebungen integriert werden, um eine objektorientierte Datenspeicherung für die OOAP bereitzustellen. Dennoch sind auch die Entwickler der Dateispeicherformate unmittelbar in diesen Prozess involviert, da sie letztlich über Lizenzgebühren und genau Softwarespezifikation der Speicherformate entscheiden.

Die Speicherformate bilden die Schnittstelle zum Konsumenten, die Distribution liegt allerdings in den Händen der Content-Aggregatoren, Labels und Verlage. Diese müssen für eine erfolgreiche Verteilung neben den bereits etablierten Formaten auch die objektorientierte Variante bewerben und vertreiben. Über digitale Distributionswege stellt dies tendenziell kein Problem dar, lediglich die zu erwartende Dateigröße einer objektorientierten (Audio-) Datenspeicherung könnte – im Vergleich zu etablierten Formaten wie mp3 – eine Hürde darstellen.

#### **7.1.2.1 Separation der Kanäle und Urheberrechtsverletzungen**

Wie in Kapitel 2.3 aufgeführt kann die Separation der Audioinhalte in einzelne Objekte durchaus Vorteile bieten, dennoch entstehen dadurch auch gewisse Risiken: Die objektbasierte Speicherung jedes einzelnen Kanals ermöglicht im Zweifelsfall den unmittelbaren Zugriff auf Audioinhalte. Dies muss nicht immer im Sinne der Musikschaffenden sein, denn die einzelnen Inhalte könnten extrahiert, gesondert betrachtet und weiterverarbeitet werden. Ein Beispiel ist das Wiederverwenden einer Gesangspur einer Tonproduktion. Dies kann für sogenannte Remixes – eine neue Interpretation eines bestehenden Klangmaterials – von Vorteil sein, jedoch muss ein solcher Remix nicht immer im Sinne des Urhebers des Ursprungswerks sein. Eine objektorientierte Speicherung bietet daher ein erhöhtes Risiko für Urheberrechtsverletzungen bzw. das Verwenden künstlerischen Materials ohne Einwilligung der Urheber.

### 7.1.3 Wiedergabe & Stakeholder der OOWV

Kapitel 6 listet verschiedene objektorientierte Wiedergabeverfahren auf, ob sich eines der aufgeführten Verfahren auch auf dem Konsumentenmarkt etablieren kann, bleibt abzuwarten. Systeme, die VBAP zur dynamischen Berechnung der Lautsprecheranteile nutzen, haben jedoch durchaus das Potenzial sich auch auf dem Konsumentenmarkt durchzusetzen (siehe dazu Kapitel 7.2 Chancen und Potenziale der OOAU). Entsprechende Renderer müssen allerdings zunächst softwareseitig programmiert, hardwareseitig implementiert und auf entsprechende Endgeräte angepasst werden. Exemplarisch wurden Smartphones als Wiedergabehardware aufgeführt: diese könnten zwar durchaus Lautsprechersignale berechnen und somit zum Renderer umfunktioniert werden, allerdings besteht bisher nicht die Möglichkeit ein solches System mehrkanalig mit Endstufen zu verbinden. Letztendlich existieren bis zum heutigen Tag nur Sonderlösungen, um OODS zu interpretieren und auf einen beliebigen Lautsprecheraufbau zu übertragen. Für den Konsumentenmarkt wäre es aber auch möglich bestehende Media-Center-Systeme mit mehrkanaligen Ausgängen zu einem Renderer umzufunktionieren, um bspw. eine mehrkanalige Endstufe (bspw. eine bestehende 5.1-Surround-Anlage) mit den entsprechenden Lautsprechersignalen anzusteuern. An einem solchen Ansatz sind sowohl Hardwarehersteller als auch Konsumenten beteiligt. Es ist unwahrscheinlich, dass sich ein solcher Hardwareansatz ohne bereits bestehende objektorientierte Produktionen und somit objektorientierten Medien- und Surround-Inhalten durchsetzen kann. Tendenziell könnte hier ein hybrider Ansatz wie Dolby Atmos greifen: Durch den zeitgleichen Rollout mehrerer Formate (bspw. Stereo, 5.1 und OODS) könnte der Renderer entscheiden, was als Wiedergabeverfahren verwendet wird. Aktuell gibt es aber – wie bereits angemerkt – keine Hardware-Renderer, die einen solchen Ansatz ermöglichen. Daher ist zunächst die Entwicklung von Hardware-Endgeräten notwendig, die ein solches mehrkanaliges und mehrformatiges Speichermedium interpretieren und decodieren können.

## 7.2 Chancen und Potenziale der OOAU

Die Risiken und Probleme der OOAU stellen den derzeitigen Stand der Technik und die damit verbundenen Herausforderungen an die Stakeholder dar. Werden die technischen Faktoren ausgeklammert, und wird von einer technisch umsetzbaren OOAU ausgegangen, so lassen sich folgende Vorteile und mögliche Potenziale beschreiben:

### 7.2.1 Raum-unabhängiger Lautsprecheraufbau

Wie die Kapitel 4, 5 und 6 aufführen, können Panoramadaten und Audioinhalte getrennt aufgezeichnet werden und Lautsprechersignale erst bei der Wiedergabe berechnet werden. Somit können bestehende Lautsprechersysteme (wie bspw. 5.1-Surround) optimiert und neue, variable Lautsprechersituationen (bspw. 11.1, 22.2) realisiert werden, ohne, dass das Trägermedium gewechselt werden muss. Der Vorteil der OOAU liegt dabei auf der Hand: Ein System kann zunächst mit wenigen Lautsprechern (bspw. angefangen mit einem einzigen Lautsprecherpaar) zu einem großen Aufbau erweitert werden. Das Hinzufügen neuer Lautsprecher erfordert dabei keine Änderung der Produktionsweise, der Dateispeicherung oder der Wiedergabe, sondern lediglich eine Erweiterung der Lautsprecheranzahl. (Die Lautsprecheranzahl samt Lautsprecherposition muss dabei allerdings in den Renderer übertragen werden, damit dieser eine korrekte Berechnung der Lautsprechersignale vornehmen kann.) Mit diesem Verfahren werden ebenfalls die Lokalisationsprobleme behoben, die sich bei einer unvorteilhaften Aufstellung der Lautsprecher abweichend von ITU-R BS.775 ergeben (siehe Kapitel 2.4.4).

### 7.2.2 Eingriff in das Klanggeschehen

Eine Audioszene repräsentiert die Position der Schallquellen im Raum – da diese innerhalb des Speicherformats aufgelistet sind, kann der Zuhörer auch unmittelbaren Einfluss auf die einzelnen Objekte der Audioszene nehmen. Bei einer Musikwiedergabe könnte der Zuhörer beispielsweise einzelne Instrumente gesondert betrachtet – bspw. durch Lautstärke-Absenkung oder – Anhebung – oder sie sogar anders im Raum positionieren, und somit eine alternative Darstellungsmethode des Gesamtwerks schaffen.

Es wäre zudem möglich alle Schallquellen – bzw. im Falle eines Musikstücks das gesamte Ensemble – zu verschieben, und damit die eigene Position im Raum zu verändern – obwohl die Hörposition dabei dieselbe bleibt.

Das bisher passive Musikhören könnte durch solche Möglichkeiten durchaus zu einer aktiven, fast spielerischen Herangehensweise an Musik bzw. Audioinhalte werden. Als Möglichkeit zur Ansteuerung der Panoramadaten in der Produktion (siehe Kapitel 4.5) wurde bereits die Verwendung von Smartphones (bzw. im Allgemeinen touchfähige Endgeräte) in Verbindung mit „Control“ aufgeführt. Ein solcher Ansatz ließe sich zur Einflussnahme auf die Position der Schallquellen bei der Wiedergabe übertragen.

### **7.3 Zukünftige Forschungsarbeiten**

Um den Rahmen der Abschlussarbeit nicht zu überschreiten, konnten einige Faktoren der OOAU nicht berücksichtigt und untersucht werden: Eine Audioproduktion beinhaltet viele Arbeitsschritte, die alle im Zuge der OOAP berücksichtigt werden müssen. Dabei vorgenommene Effektprozesse (wie bspw. das Verwenden von Effekten wie Equalizer, Compressor, Limiter etc.) ließen sich eventuell ebenfalls in ein Speicherformat übertragen, um anschließend wiederum Einfluss auf die Effektparameter zu nehmen. Weiterhin muss für die Verwendung von Hall (bzw. der Faltung mit Impulsantworten) ein neuer Ansatz gewählt werden, da die Berechnung der Hallanteile unmittelbar von der Raumposition der Schallquelle abhängt und die Hallanteile für die Lautsprecher signale (die letztlich von der Aufstellung der Lautsprecher im Raum abhängen) dynamisch generiert werden müssen. Ein Ansatz wäre die Berechnung des Halls durch den Renderer: Bei der OODS wird neben den Audioinhalten eine zusätzliche Impulsantwort abgespeichert. Diese könnte vom Renderer interpretiert und so die Hallanteile der einzelnen Schallquellen generiert werden. Somit können verschiedene Hallräumen bei der Produktion publiziert werden. Der Konsument kann bei der Wiedergabe zwischen verschiedenen Klangräumen wählen, ohne das Speichermedium zu tauschen. (Exemplarisch: Durch das Auswechseln der Impulsantwort kann ein kleiner Konzertsaal, eine große Bühne etc. simuliert werden – die Objekte und Audioinhalte bleiben davon unberührt, die entsprechenden Hallanteile werden bei der Wiedergabe addiert.)

Die oben aufgeführten Ansätze müssten jedoch zunächst auf Durchführbarkeit, mögliche Probleme bei der Wiedergabe und Bediener- und Nutzerakzeptanz untersucht werden.

## 7.4 Perspektiven der OOAU

Die vorliegende Arbeit zeigt, dass objektorientierte Audioumgebungen sowohl theoretisch als auch praktisch realisierbar sind. Die dazu notwendige Theorie und die ebenfalls dazu benötigte Technik steht bereits seit Jahren zur Verfügung; die Aufteilung der Audioumgebungen in verschiedene Bestandteile und den damit verbundenen Personengruppen verdeutlicht allerdings die Problematik eines solchen Systems: Zunächst müssen (objektorientierte) Audioinhalte geschaffen werden, damit der Konsumentenmarkt objektorientierte Wiedergabesysteme für sich entdecken kann. Es ist allerdings riskant ein solches Format zu produzieren, wenn die Marktrelevanz ungewiss bleibt. Weiterhin ist es schwierig objektorientierte Wiedergabesysteme zu (entwickeln und) vertreiben, wenn dafür noch keine Inhalte bereitstehen. Dieses Phänomen lässt sich mit der Entwicklung und Distribution von HD-Videoinhalten und den dafür benötigten Wiedergabedispays vergleichen.

Hier ist gewissermaßen eine Parallelentwicklung der Inhalte und Wiedergabesysteme notwendig, um den Markt zeitgleich durchdringen zu können. Ob sich ein solches System auf dem Konsumentenmarkt durchsetzen kann und sich somit eine Relevanz bzw. Nachfrage seitens der Produktion bildet, bleibt abzuwarten. Auch wenn die OOAU technische und klangliche Vorteile darlegt, müssen diese nicht über den Erfolg des Systems entscheiden.



---

## 8 Danksagung

Die vorliegende Arbeit wäre ohne die Hilfe folgender Personen nicht möglich gewesen, ich möchte mich im Zuge dessen daher persönlich bei Ihnen bedanken:

Bei Herrn Professor Jens-Helge Hergesell und Herrn Professor Oliver Curdt für die Betreuung über meine Abschlussarbeit hinweg, für ausführliche fachliche und thematisch relevante Gespräche und dafür, dass ich dieses Thema im Zuge meines Masterstudiums bearbeiten konnte. Insbesondere möchte ich mich bei Prof. Jens-Helge Hergesell für das zur Verfügung stellen seines kompletten Archivs des „Journals of the Audio Engineering Society“ bedanken, das mir als wertvolle Literaturquelle und zu meiner Wissenserweiterung diene.

Bei der Klangerfinder GmbH Stuttgart, die mir im Rahmen meiner Abschlussarbeit den Zugang zu ihrem Klanglabor samt IOSONO-Wellenfeldrenderer ermöglichte und insbesondere bei Stefan Kolbe, der mir stets mit fachlichen Tipps zu Max und insbesondere der Ansteuerungsmöglichkeiten der Renderer zur Seite stand.

Bei Heiko Schulz für die Möglichkeit die Räumlichkeiten des HDM-Tonstudios über den Zeitraum meiner Abschlussarbeit zu nutzen sowie für die vielen Gespräche, die mein tontechnisches Wissen fachlich erweitert und mich somit thematisch stets vorangebracht haben.

Insbesondere bei Herrn Dr. Markus Mehnert der IOSONO GmbH für fachliche und inhaltliche Anstöße zu meiner Abschlussarbeit; Herrn Cristoph Neumeyer und Herrn Clemens Clausen der Shure Distribution GmbH für die netten Gespräche, die ausführliche Auskunft über Shure Atmosphaea und für die Möglichkeit Atmosphaea in Aktion zu erleben, und bei Christian Lerch der Dolby Germany GmbH für weitergehende Erläuterungen zu Dolby Atmos.

Abschließend möchte ich mich bei meiner Lektorin Carolina Werli bedanken, die mir stets wertvolle Tipps zur inhaltlichen Gestaltung meiner Abschlussarbeit lieferte.



# 9 Quellenverzeichnis

## 9.1 Abbildungsverzeichnis

Abb. 1.1 Traditionelle Audioumgebung: Produktion, Speicherung und Wiedergabe .....	14
Abb. 2.1 Kopfbezogenes Koordinatensystem (Quelle: [5] S. 109).....	17
Abb. 2.2 Seitlich eintreffende Schallereignisse in der Horizontalebene (Quelle: [6] S. 11) .....	18
Abb. 2.3 Richtungsbestimmende Frequenzbänder bei Schalleinfall aus der Vertikalebene (Quelle: [7])... ..	18
Abb. 2.4 Links: 1:1-Monophonie, rechts: n:1 Monophonie durch die Verwendung eines Mischpults .....	20
Abb. 2.5 Hörfläche bei der Stereophonie (Quelle: [11] S. 611).....	21
Abb. 2.6 Phantomschallquelle durch Kombination der Lautsprechersignale (Quelle: [11] S. 658).....	23
Abb. 2.7 Mathematische Approximation Panning Laws Im konkreten Beispiel: Panoramareglers bei -3 dB in Mittelposition (Quelle: [15]).....	24
Abb. 2.8 Aufstellungsempfehlung für 5.1-Surround nach ITU-R BS.775 (Quelle: [19] S. 127).....	27
Abb. 2.9 Auro 3D 10.1, exemplarischer Aufbau (Quelle: [20]).....	29
Abb. 2.10 Kosten und zu erwartende Klangqualität bei Verdopplung der Lautsprecheranzahl.....	34
Abb. 2.11 Exemplarischer Lautsprecheraufbau entlang räumlichen Gegebenheiten.....	36
Abb. 3.1 Vereinfachtes UML-Diagramm für die Objekte „Mensch“ (Auszug) und „Schreibtischlampe“ (Auszug) .....	41
Abb. 3.2 Links: Schallquellen mit Ihrer Position im abstrahierten virtuellen Raum, rechts: „realer“ Raum mit Lautsprechern.....	42
Abb. 3.3 Projektion des virtuellen Raums auf den realen Raum links: 5 Lautsprecher, rechts: exemplarisch 13 Lautsprecher .....	43
Abb. 4.1 Visuelle kanalbasierte Darstellung von Schallquellen samt Lautstärkemixer der Produktionsumgebung „Cubase“ .....	47
Abb. 4.2 Cubase-Surround-Panning-Tool links im Bild: Eingangssignal rechts im Bild: Berechnung der Aux-Signale.....	48
Abb. 4.3 Vereinfachtes UML-Diagramm eines Kanals mit Attributen.....	50
Abb. 4.4 Visuelle Darstellung der Raumposition über das Objekt Nodes.....	53
Abb. 4.5 Routing der Produktion zum Rendering, im Beispiel 1:1-Routing von 64 Kanälen.....	55
Abb. 4.6 1:1 Routing der (Objekt-)Kanäle von Ableton Live auf die physikalischen Ausgänge .....	55
Abb. 4.7 Die Routing Matrix bei MADI-Soundkarten des Hersteller RME Audio (davon 32 Kanäle im 1:1 Routing).....	56
Abb. 4.8 Verknüpfung von Applikationen durch JACK.....	57
Abb. 4.9 Verknüpfung von Audiogeräten mit Audinates "Dante Virtual Soundcard" .....	58
Abb. 4.10 Ansteuerungsmöglichkeiten von PortAudio auf Betriebssystemebene (Quelle: [29]).....	59
Abb. 4.11 Darstellung aller Schallquellen im Prototyp .....	62
Abb. 4.12 Hardware-Surround-Panner für Pro Tools (Quelle: [34]) .....	63
Abb. 4.13 Schallquellenpositionierung mit der Android-App Control.....	64

## 9 Quellenverzeichnis

---

Abb. 4.14 Vitruvianischer Mensch (Quelle: [35]).....	65
Abb. 4.15 Nintendo Wii-Mote mit Achsendarstellung (Quelle: [36]) .....	65
Abb. 4.16 Kinect mit Kameraerfassung und Erkennung der Körperachsen (Quelle: [37]) .....	66
Abb. 5.1 Evolution der Audio-Dateiformate (Quelle: [39] S. 691).....	70
Abb. 5.2 Speicherung einer multichannel-Interleaved-Datei mit Audacity.....	71
Abb. 5.3 Exemplarische Aneinanderreihung der Panoramawerte X-, Y- und Z in einer WAVE-Datei mit willkürlichen Werten. Bei einer Quantisierung von 16 Bit werden dabei lediglich 128 Abstufungen (=7 Bit) beansprucht.....	72
Abb. 5.4 Encoding- und Decoding-Prozess von MPEG SOAC (Quelle: [44] S. 657).....	78
Abb. 5.5 MPEG SOAC als Transcoder für MPEG Surround (Quelle: [44] S. 659).....	79
Abb. 6.1 n:m Beziehung zwischen Schallquelle und Lautsprecher (blau) und Schallquellen (rot).....	81
Abb. 6.2 Berechnung der Position der Phantomschallquelle anhand von Vektoren (Quelle: [45] S. 458)..	83
Abb. 6.3 Links: Zweidimensionaler Surround-Aufbau mit VBAP (Quelle: [45] S. 459).....	84
Abb. 6.4 Beispielkonfiguration 3d-Panning mit VBAP (Quelle: [45]S. 480).....	84
Abb. 6.5 Dreidimensionales VBAP mit fünf Lautsprechern (Quelle: [45] S. 461 .....	85
Abb. 6.6 Dreidimensionale Abbildung der Raum-informationen des Soundfiled-Mikrofons (Quelle: [47])	86
Abb. 6.7 Screenshot des Plugins „Harpex-B“ .....	87
Abb. 6.8 Erzeugung von Elementarwellen mit einer großen Anzahl Lautsprecher (Quelle: [11] S. 666)....	88
Abb. 6.9 Darstellung einer Punktschallquelle innerhalb und außerhalb der Lautsprecherebene (Quelle: [6] S. 300) .....	89
Abb. 6.10 exemplarischer Aufbau des IOSONO-Systems im Soundlabor der Klangerfinder GmbH Stuttgart .....	91
Abb. 6.11 Szenendarstellung samt Schallquellenposition bei der Wiedergabe .....	92
Abb. 6.12 Lautsprecheransicht und Abbildung des realen Raumes.....	92
Abb. 6.13 Spatial Audio Workstation (Quelle: [50]).....	93
Abb. 6.14 Exemplarische Darstellung des Atmospha-Systems mit Wand- und Deckenlautsprechern, Deckenmikrofonen und Rendering-Engine (Quelle: [51]) .....	94
Abb. 6.15 Schematischer Aufbau des Rendering-Prozesses .....	95
Abb. 6.16 Empfohlener Lautsprecheraufbau von Dolby Atmos (Quelle: [52] S.18) Blau: zusätzliche Surround-Lautsprecher, pink: optionale Subwoofer, grün: optionale Leinwandlautsprecher .....	96
Abb. 6.17 Produktions-, Speicher- und Wiedergabeprozess von Dolby Atmos (Quelle: [52] S.11) .....	98

### 9.2 Tabellenverzeichnis

Tabelle 4.1: Überblick über den Transfer von Audioinhalten über Netzwerk, angelehnt an [33].....	60
Tabelle 5.1: Übersicht Container-Formate (Quelle: [39] S. 705).....	74

## 9.3 Literaturverzeichnis

- [1] W. Weaver, „Recent Contributions to the Mathematical Theory of Communication,“ in *The Mathematical Theory of Communication. 4. Paperback Edition Urbana*, Chicago, The University of Illinois Press, 1969.
- [2] C. Shannon, *The Mathematical Theory of Communication. 4. Paperback Edition Urbana*, Chicago: The University of Illinois Press, 1948.
- [3] T. Görne, „Phantomschallquellen und Stereophonie,“ in *Tontechnik*, Fachbuchverlag Leipzig, 2006, pp. 121-122.
- [4] J. Blauert und J. Braasch, „Räumliches Hören,“ in *Handbuch der Audiotechnik*, 2008, pp. 87-119.
- [5] M. Dickreiter, V. Dittel, W. Hoeg und M. Wöhr, „Schallwahrnehmung,“ in *Handbuch der Tonstudioteknik*, München, K.G. Sauer Verlag, 2008, pp. 95-112.
- [6] T. Görne, „3.3 Binaurales Hören,“ in *Tontechnik*, Fachbuchverlag Leipzig, 2006, p. 118.
- [7] Wikipedia, „Wikipedia - Blauertsche Bänder,“ 2013. [Online]. Available: [http://de.wikipedia.org/wiki/Blauertsche\\_B%C3%A4nder](http://de.wikipedia.org/wiki/Blauertsche_B%C3%A4nder). [Zugriff am 2013].
- [8] J. Blauert, *Räumliches Hören*, S. Hirzel-Verlag, 1974.
- [9] E. Torick, „Highlights in the History of Multichannel Sound,“ in *Journal of the AES*, 1998, pp. 27-31.
- [10] M. F. Davis, „History of Spatial Coding,“ in *Journal of the AES*, 2003, pp. 554-569.
- [11] S. Weinzierl und K. M. Slavik, „Wiedergabeverfahren,“ in *Handbuch der Audiotechnik*, 2008, pp. 609-682.
- [12] V. Pulkki, M. Karjalainen und J. Huopaniemi, „Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model,“ in *Journal of the AES*, April 1999, pp. 203-217.
- [13] M. Dickreiter, V. Dittel, W. Hoeg und M. Wöhr, „Stereophonie,“ in *Handbuch der Tontechnik*, K. G. Saur, 2008, pp. 183-210.
- [14] T. Görne, „10.4 Mischpulte,“ in *Tontechnik*, Fachbuchverlag Leipzig, 2006, p. 337.
- [15] M. Dickreiter, V. Dittel, W. Hoeg und M. Wöhr, „Klanggestaltung,“ in *Handbuch der Tonstudioteknik*, Nürnberg, K. G. Saur, 2008, p. 321 ff.

## 9 Quellenverzeichnis

---

- [16] M. Dickreiter, V. Dittel, W. Hoeg und M. Wöhr, „Zweikanal-Stereofonie,“ in *Handbuch der Tontechnik*, K.G. Saur, 2008, pp. 210-240.
- [17] T. Görne, „9.6 Mehrkanaltechnik,“ in *Tontechnik*, Fachbuchverlag Leipzig, 2006, pp. 294-301.
- [18] T. Holman, „A Brief History,“ in *Surround Sound*, Focal Press, 2008, pp. 1-21.
- [19] T. Görne und S. Bergweiler, „Monitoring,“ in *Monitoring*, PPVMEDIEN GmbH, 2004, pp. 7-18.
- [20] Auro Technologies, „Auro Technologies,“ [Online]. Available: <http://www.auro-technologies.com/>. [Zugriff am Januar 2013].
- [21] B. Reichert, *3D-Sound im Kinofilm - Eine wirtschaftliche Analyse*, Stuttgart: Bachelor Arbeit, vorgelegt an der Hochschule der Medien Stuttgart, SS 2012.
- [22] Beracochea, „Intelligibility in virtual acoustic opening Environments,“ *AES Journal*, pp. 339-353, Mai 2008.
- [23] Cherry, „Some experiments on the recognition of speech, with one and with two ears,“ *Journal of the Acoustical Society of America*, p. 975–979, 1953.
- [24] AES, „Surround Sound - A Chance for Enhanced Creativity,“ *AES Journal*, pp. 540-541, Juny 2006.
- [25] J. Reuter, *Wahrnehmung und Wirkung von Musik im Film*, Stuttgart: Diplomarbeit vorgelegt an der Hochschule der Medien Stuttgart, WS 2006/2007.
- [26] Y. Geutskens, „sa-cd.net,“ [Online]. Available: <http://www.sa-cd.net/faq#playback11>. [Zugriff am 5 Februar 2013].
- [27] AES, „AES MADI,“ [Online]. Available: <http://www.iis.ee.ethz.ch/~felber/DataSheets/AES-EBU/aes10-2003.pdf>. [Zugriff am Januar 2013].
- [28] RME Audio, „HDSPe MADI FX,“ [Online]. Available: [http://www.rme-audio.de/products\\_hdspe\\_madi\\_fx.php](http://www.rme-audio.de/products_hdspe_madi_fx.php). [Zugriff am Januar 2013].
- [29] PortAudio, „PortAudio API Overview,“ [Online]. Available: [http://portaudio.com/docs/v19-doxydocs/api\\_overview.html](http://portaudio.com/docs/v19-doxydocs/api_overview.html). [Zugriff am Januar 2013].
- [30] IEEE, „IEEE AVB,“ [Online]. Available: <http://www.ieee802.org/1/pages/avbridges.html>. [Zugriff am Januar 2013].
- [31] R. Schreiner, *Computer Netzwerke*, Hanser, 2007.

- [32] Audinate, „Audinate Dante QA,“ [Online]. Available: [http://www.audinate.com/index.php?option=com\\_content&view=article&id=99#How%20is%20audio%20transmitted%20over%20the%20network?](http://www.audinate.com/index.php?option=com_content&view=article&id=99#How%20is%20audio%20transmitted%20over%20the%20network?). [Zugriff am Januar 2013].
- [33] AES, „AES White Paper: Best Practices in Network Audio,“ *AES Journal*, pp. 729-741, September 2009.
- [34] Avid, „Pro Tools Surround Panner,“ [Online]. Available: Technology Transport Network capacity Max available sampling rate. [Zugriff am Januar 2013].
- [35] Vitruvianischer Mensch. [Online]. Available: [http://www.bezregdetmold.nrw.de/200\\_Aufgaben/060\\_Arbeitsschutz/010\\_Dezernat\\_56/Bilder/human.JPG](http://www.bezregdetmold.nrw.de/200_Aufgaben/060_Arbeitsschutz/010_Dezernat_56/Bilder/human.JPG).
- [36] Wii Axis. [Online]. Available: [http://www.temperies.com/blog/wp-content/uploads/2011/07/wiimote\\_axis.gif](http://www.temperies.com/blog/wp-content/uploads/2011/07/wiimote_axis.gif).
- [37] Kinect SDK. [Online]. Available: <http://www.karthikk.net/wp-content/uploads/2011/11/kinect-for-windows-sdk.jpg>.
- [38] F. Rumsey, „Blu-ray or downloads for HD audio delivery?,“ *AES Journal*, pp. 149-154, März 2011.
- [39] K. Petermichel, „Dateiformate für Audio,“ in *Handbuch der Audiotechnik*, 2008, pp. 687-718.
- [40] P. Lindner, „ZIP,“ [Online]. Available: <http://www.iana.org/assignments/media-types/application/zip>. [Zugriff am Januar 2013].
- [41] J. Coalson, „FLAC,“ [Online]. Available: <http://flac.sourceforge.net/format.html>. [Zugriff am Januar 2013].
- [42] The Moving Picture Expert Group, „MPEG,“ Januar 2013. [Online]. Available: <http://mpeg.chiariglione.org/standards/mpeg-4/>.
- [43] TU Berlin, „MPEG-4 Audio Lossless Coding,“ Januar 2013. [Online]. Available: [http://www.nue.tu-berlin.de/menue/forschung/projekte/beendete\\_projekte/mpeg-4\\_audio\\_lossless\\_coding\\_als/](http://www.nue.tu-berlin.de/menue/forschung/projekte/beendete_projekte/mpeg-4_audio_lossless_coding_als/). [Zugriff am 2013].
- [44] J. Herre und et al, „Spatial Audio Object Coding (SAOC),“ *AES Journal*, pp. 655-673, September 2012.
- [45] V. Pulkki, „Virtual Sound Source Positioning Using Vector Base Amplitude Panning,“ *AES Journal*, pp. 456-466, June 1997.

## 9 Quellenverzeichnis

---

- [46] V. Pulkki und M. Karjalainen, „Localization of Amplitude-Panned Virtual Sources I & II,“ *AES Journal*, pp. 739-767, September 2001.
- [47] Soundfield, „Soundfield B-Format,“ Januar 2013. [Online]. Available: [http://www.soundfield.com/downloads/b\\_format.pdf](http://www.soundfield.com/downloads/b_format.pdf).
- [48] M. Dickreiter, V. Dittel, W. Hoeg und M. Wöhr, „Wellenfeldsynthese,“ in *Handbuch der Tonstudiotechnik*, K.G. Saur, 2008, pp. 304-319.
- [49] E. Corteel, „Equalization in an Extended Area Using Multichannel Inversion and Wave Field Synthesis,“ *AES Journal*, pp. 1140-1161, Dezember 2006.
- [50] IOSONO, „IOSONO - Spatial Audio Workstation,“ [Online]. Available: <http://www.iosono-sound.com/spatial-audio-workstation.html>. [Zugriff am Januar 2013].
- [51] Atmospha, „Die Atmospha Technologie,“ [Online]. Available: <http://www.atmospha.de/technologie>. [Zugriff am Januar 2013].
- [52] Dolby, „Dolby Atmos Whitepaper,“ [Online]. Available: <http://www.dolby.com/uploadedFiles/Assets/US/Doc/Professional/Dolby-Atmos-Next-Generation-Audio-for-Cinema.pdf>. [Zugriff am Februar 2013].
- [53] H. Robjohns, „Sound on Sound,“ [Online]. Available: <http://www.soundonsound.com/sos/nov10/articles/stereoprocessing.htm>. [Zugriff am Februar 2013].
- [54] M. Dickreiter, V. Dittel, W. Hoeg und M. Wöhr, „Prinzipien der räumlichen Tonübertragung,“ in *Handbuch der Tontechnik*, K.G. Saur, 2008, pp. 183-196.