

Lehrstuhl für Nachrichtentechnik  
Fachrichtung 6.2 - Informatik  
Prof. Dr.-Ing Thorsten Herfet



# Studienarbeit

## Generation of point sound sources and surround sound effects

Autor: Tobias Jung (2026323)

Tutor: Dipl.Inf. Sylvia Schulz

Abgabe: 22. November 2006

**Erklärung**

Hiermit erkläre ich, dass ich die vorliegende Studienarbeit selbständig angefertigt habe. Es wurden nur die in der Arbeit ausdrücklich benannten Quellen und Hilfsmittel benutzt. Wörtlich oder sinngemäss übernommenes Gedankengut habe ich als solches kenntlich gemacht.

---

(Ort und Datum)

---

(Tobias Jung)

# Nomenklatur

$P_\psi$	Leistung der im Winkel $\psi$ einfallenden Welle
$S_\psi$	Zu reproduzierende ebene Welle
$J_m$	Besselfunktion mit Parameter $m$
$\vec{k}$	Vektor in Richtung der einfallenden Welle
$\vec{r}$	Vektor in Richtung der Zuhörerposition
$\phi_m$	Winkel des $m$ -ten Lautsprechers zur X-Achse
$S$	Erzeugtes Signal, als Summe über den einzelnen Signalen
$g_i$	Amplitude des Signals des $i$ -ten Lautsprechers
$P$	Gesamtamplitude aller Signale
$E$	Gesamtenergie aller Signale
$P_x$	Amplituden der Signalanteile aus X-Richtung
$E_x$	Energie der Signaleanteile aus X-Richtung
$P_y$	Amplituden der Signalanteile aus Y-Richtung
$E_y$	Energie der Signaleanteile aus Y-Richtung
$SPos_i$	Winkel des $i$ -ten Lautsprechers
$\lambda, k$	Korrekturfaktoren
$C_F$	Center-Lautsprecher vorne
$L_F$	Lautsprecher links vorne
$R_F$	Lautsprecher rechts vorne
$L_B$	Lautsprecher links hinten
$R_B$	Lautsprecher rechts hinten

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
<b>2</b>	<b>Die Entwicklung von Surround-Systemen - von Mono bis 7.1</b>	<b>5</b>
2.1	Aktueller Stand der Entwicklung . . . . .	7
2.1.1	Das 5.1 System . . . . .	7
2.1.2	Dolby Digital . . . . .	9
2.1.3	Sony Dynamic Digital Sound (SDDS) . . . . .	9
2.1.4	Digital Theater Sound . . . . .	10
<b>3</b>	<b>Methoden zur Generierung von virtuellen Klangquellen und Raumklangeffekten</b>	<b>11</b>
3.1	Lokalisation von Schallereignissen . . . . .	11
3.2	Binaurale Tonaufnahmen . . . . .	14
3.3	Wellenfeldsynthese . . . . .	15
3.4	Vector Based Panning (VBP) . . . . .	17
<b>4</b>	<b>Ambisonic</b>	<b>21</b>
4.1	Aufnahmetechnik und Speicherformate . . . . .	22
	UHJ . . . . .	22
	G-Format . . . . .	23
4.2	Grundlagen zu Ambisonic . . . . .	23
4.3	Kodierung . . . . .	27
4.4	Dekodierung . . . . .	28
4.5	Probleme mit asymmetrischen Anordnungen . . . . .	29
<b>5</b>	<b>Versuchsaufbau</b>	<b>32</b>
<b>6</b>	<b>Das Programm <i>ambimove</i></b>	<b>34</b>
6.1	Systemvoraussetzungen . . . . .	34
6.2	Aufbau . . . . .	34
6.2.1	Kurzer Ausflug in csound . . . . .	35
6.3	Bedienung und Einstellungen . . . . .	36
6.4	Die Kodierungsparameter . . . . .	38
6.5	Beispiele . . . . .	40
6.5.1	Punktquellen . . . . .	40
6.5.2	Geraden . . . . .	41
6.5.3	Kreise . . . . .	42
<b>7</b>	<b>Ausblick und Erweiterungen</b>	<b>45</b>

# Kapitel 1

## Einleitung

Diese Arbeit gehört zum Forschungsprojekt „Binaural Computational Source Separation“ des Lehrstuhls für Nachrichtentechnik an der Universität des Saarlandes.

Ziel dieses Projektes ist herauszufinden, ob Eigenschaften und Anatomie eines menschlichen Kopfes genutzt werden können, um computergestützte Quellen-separation zu ermöglichen bzw. zu verbessern. Ziel der folgenden Studienarbeit ist die Generierung von einer oder mehreren (virtuellen) Klangquellen in beliebigen Raumpunkten und Raumklangeffekten für ein gegebenes Lautsprecher-Setup. Raumklangeffekte sind hier Figuren, die von einer virtuellen Klangquelle im Raum beschrieben werden (z.B. Kreise, Linien etc.).

In den ersten Kapiteln wird zunächst eine Übersicht über den aktuellen Stand von Raumklangsystemen bzw. die Entwicklung bis zum heutigen Stand gegeben. Danach werden die gängigsten Methoden zur Generierung von virtuellen Klangquellen vorgestellt und Vor- und Nachteile hervorgehoben.

In den letzten Abschnitten wird das hier verwendete Verfahren „Ambisonic“ erläutert bevor dann das hier benutzte Programm *ambimove* vorgestellt und dessen Funktionsweise und Bedienung erklärt wird.

## Kapitel 2

# Die Entwicklung von Surround-Systemen - von Mono bis 7.1

### **Mono**

Die einfachste Form der Wiedergabe von Klang sind so genannte Monoaufnahmen. Mono bedeutet, dass nur ein Signal aufgenommen und später wiedergegeben wird, das aber durchaus über mehrere Lautsprecher wiedergegeben werden kann. Die erste Mono-Aufnahme wurde 1877 angefertigt. Jedes Tonwiedergabesystem erlaubt auch heute noch Mono-Wiedergabe, dabei wird über alle Kanäle bzw. Lautsprecher das gleiche Signal wiedergegeben. Die spezifischen Eigenschaften des jeweiligen Tonwiedergabesystems (z.B. Stereo) fallen dabei weg. Statt des für das jeweilige Tonwiedergabesystem spezifischen Raumklangs wird dabei das Monosignal lediglich auf einer breiteren Basis wiedergegeben und klingt meistens voller, als bei der Wiedergabe über nur einen Lautsprecher. Selbst Dolby Digital bietet eine 1.0-Kodierung.

Da diese Form der Aufnahme bzw. Wiedergabe aber keinerlei räumliche Informationen des Klangfeldes enthalten, wurde nach Methoden gesucht, die eine möglichst natürliche Wiedergabe des ursprünglichen (aufgenommenen) Klangfeldes ermöglichen.

Der nächste Schritt in der Entwicklung war daher die Einführung eines zweiten Kanals zum s.g. Stereo-System.

### **2 Kanäle**

Bei diesen Verfahren verfügen beide Kanäle über die gleiche Bandbreite. Das System der stereofonen Tonwiedergabe wurde in den 30er Jahren des 20. Jahrhunderts entwickelt, aber erst seit den 60er Jahren z.B. auch im Fernsehen und Rundfunk verwendet. Stereo verwendet beide Kanäle um räumliche Klanginformation zu übermitteln und stellt somit das einfachste System zur Erzeugung eines räumlichen Eindrucks dar. Noch heute ist es der am weitesten verbreitete Standard.

### 3 Kanäle

Die nächste Stufe der Entwicklung war die Einführung eines dritten Lautsprechers, sei es als Hilfskanal, oder als gleichwertiger Lautsprecher mit gleicher Bandbreite. Der Hilfskanal konnte z.B. bei Filmen als Center-Lautsprecher für Dialoge genutzt werden. Bei drei gleichwertigen Kanälen wurde das Klangfeld auf diese aufgeteilt, wobei alle drei Lautsprecher vor dem Zuhörer platziert waren (Links-Mitte-Rechts). Es gilt als ein Übergangsschritt von der Stereo- zur Quadrophonie.

### 4 Kanäle

Auch hier kann man wieder eine Unterscheidung treffen zwischen vier vollen Kanälen und zwei vollen mit zwei Hilfskanälen. Dolby Stereo z.B. ist ein solches. Es ist ein 1976 von der Firma Dolby eingeführtes analoges Lichttonformat für Kinofilme. Vier Kanäle (Links, Mitte, Rechts, Hinten) sind hierbei in einem zweispurigen Lichtton matriziert, die beim Abspielen vom Kinoprozessor wieder in vier Kanäle dematriziert werden. Zur benötigten Rauschunterdrückung werden von Dolby entwickelte Kompander-Verfahren verwendet.

Dolby SR ist z.B. ein solches Verfahren, weshalb das Format dann technisch dementsprechend Dolby Stereo SR heisst. Jeder heute hergestellte Kinofilm enthält, unabhängig von den digitalen Tonspuren, eine Dolby Stereo SR Tonspur, auf die die Projektionseinrichtung automatisch umschaltet, falls der digitale Ton nicht mehr in ausreichender Qualität gelesen werden kann. Dies wird durch ein Knacken und eine plötzliche Änderung des Tons für den Zuschauer hörbar.

Die entsprechende digitale Lichttonvariante nennt sich Dolby Digital oder technisch Dolby SRD.

Die Variante für Heimkinowandler ist das bekannte Dolby Surround, das die gleiche Matrizierungstechnik, nicht jedoch die Rauschunterdrückung beinhaltet.

1986 wurde das Dolby Pro Logic Verfahren erstmals vorgestellt. Es ist für den Heimkino-Bereich entwickelt worden und basiert auf dem oben genannten Dolby Stereo. Es dekodiert Raumklang-Informationen aus einer Stereoquelle in folgender Weise:

In zwei vollwertige Kanäle sind zwei weitere Hilfskanäle mithilfe einer Matrix kodiert. Diese Kanäle sind für Effekte und Akzente gedacht. Aus den beiden Tonspuren werden dafür die Anteile herausgefiltert die in beiden Kanälen gegenphasig oder phasenkohärent vorliegen.

Die Anteile, die gegenphasig sind, werden auf den so genannten Center Lautsprecher gegeben. Diejenigen, die gleiche Phase besitzen liegen dann am Effektkanal (Surround/Rear) an, welcher allerdings in seiner Bandbreite begrenzt ist. Die Anteile, die gänzlich unkorreliert sind, werden auf die beiden Hauptkanäle Links und Rechts dekodiert.

Die Quadrophonie ist ein Verfahren das mit vier gleichwertigen Kanälen arbeitet. Hierfür werden bereits bei der Aufnahme vier separate Tonspuren erzeugt.

Dann gibt es zwei verschiedene Vorgehensweisen:

- diskret 4-4-4

- matriziert 4-2-4

Dabei werden also entweder alle 4 Spuren getrennt übertragen (4-4-4) oder mithilfe einer Matrix zur Übertragung auf zwei Kanäle kodiert um dann bei der Wiedergabe wieder entsprechend auf 4 Kanäle dekodiert zu werden.

Ein Grund, warum sich dieses Verfahren nie wirklich durchsetzen konnte ist der, dass es zu viele verschiedene, konkurrierende (damit unkompatible) Matrix-Verfahren gibt (z.B. SQ, CD4, QS, UD4, EV4, QM, UMX usw.).

Ausserdem ging man von der falschen Annahme aus, dass ein Lautsprecher-system mit vier Lautsprechern, aufgestellt in jeweils  $90^\circ$  Winkeln zueinander (Vorne:Links,Rechts-Hinten:Links,Rechts), die Möglichkeit einer realitätsnahen räumlichen Wiedergabe bieten würde. Tatsächlich zeigten sich aber Lücken im Klangbild.

## 5 Kanäle

Wenn man so will ist das erste Tonsystem mit fünf Lautsprechern das oben bereits erwähnte, von DOLBY in den 1980er Jahren entwickelte Surround Pro Logic I. Im Gegensatz zur Quadrophonie wurde ein Center-Lautsprecher eingeführt und damit ein neuer Kanal, über den z.B. Dialoge in Filmen wiedergegeben werden. Allerdings geben bei DOLBY Surround Pro Logic I die beiden hinteren Lautsprecher das Selbe wieder, weshalb es kein echtes 5-Kanal-System ist. Eine echte Weiterentwicklung ist dagegen Dolby Pro Logic II welches auf obiger Technik basierend 5 gleichwertige Kanäle mit vollem Frequenzumfang kodiert bzw. dekodiert. Das Pro Logic II Verfahren überwindet also den Nachteil des monophonen Rear-Signals und ersetzt es durch ein echtes stereofones. Ausserdem kann hier jetzt die Wirkungsweise des Center-Filters eingestellt werden.

## 2.1 Aktueller Stand der Entwicklung

### 2.1.1 Das 5.1 System

Das von Dolby entwickelte Dolby Surround Verfahren ist das erste mit einem s.g. Low Frequency Effect (LFE) Kanal. Es gehört zu den 5.1 Systemen, welche 5 getrennte, volle Kanäle und einen LFE-Kanal beinhalten, was durch das .1 signalisiert wird. Abgeleitet wurde das 5.1-System für den Heimanwender aus im Kino eingesetzten Verfahren wie Dolby-Prologic und besteht aus folgenden Bestandteilen:

- Hauptlautsprecher Links und Rechts
- Centerlautsprecher
- Surroundlautsprecher Links und Rechts
- Subwoofer (LFE)

Im Gegensatz zu Dolby Surround Prologic und ähnlichen Verfahren werden bei 5.1 alle 6 Kanäle diskret, d.h. einzeln und vollständig gespeichert und wiedergegeben. Die 5 vollen Kanäle (Center, Vorne links, Vorne rechts, Hinten links und Hinten rechts) enthalten den kompletten Frequenzbereich zwischen 20Hz



und 20 kHz, der LFE-Kanal hingegen nur die tiefen Frequenzen unter 120Hz. Er soll den Klangeindruck verbessern und realistischer machen. Da herkömmliche 5.1 Anlagen häufig relativ schwache „Satelliten“ (alle LS ausser Center) verwenden werden in der Praxis Frequenzweichen eingesetzt, die tiefe Frequenzen auf den Subwoofer legen, um die anderen Boxen zu entlasten. Die Aufstellung eines solchen Systems erfolgt normalerweise nach dem ITU-Standard. Dieser kann allerdings in den Aufstellungswinkeln der Lautsprecher variieren, abhängig davon, ob sie für Kinoanwendungen (frontlastiger) oder Musikwiedergabe (symmetrische Anordnung) verwendet wird. Hier die ITU-Aufstellung für Musikwiedergabe:

- meist 5 identische Lautsprecher für Front, Center und Surround
- identischer Hörabstand aller 5 Lautsprecher zum Hörer
- Winkelanordnung der Lautsprecher in Blickrichtung des Hörers: Center 0 Grad, Front +30 Grad, Surround +110 Grad (also eher seitlich, nicht hinter dem Hörer)

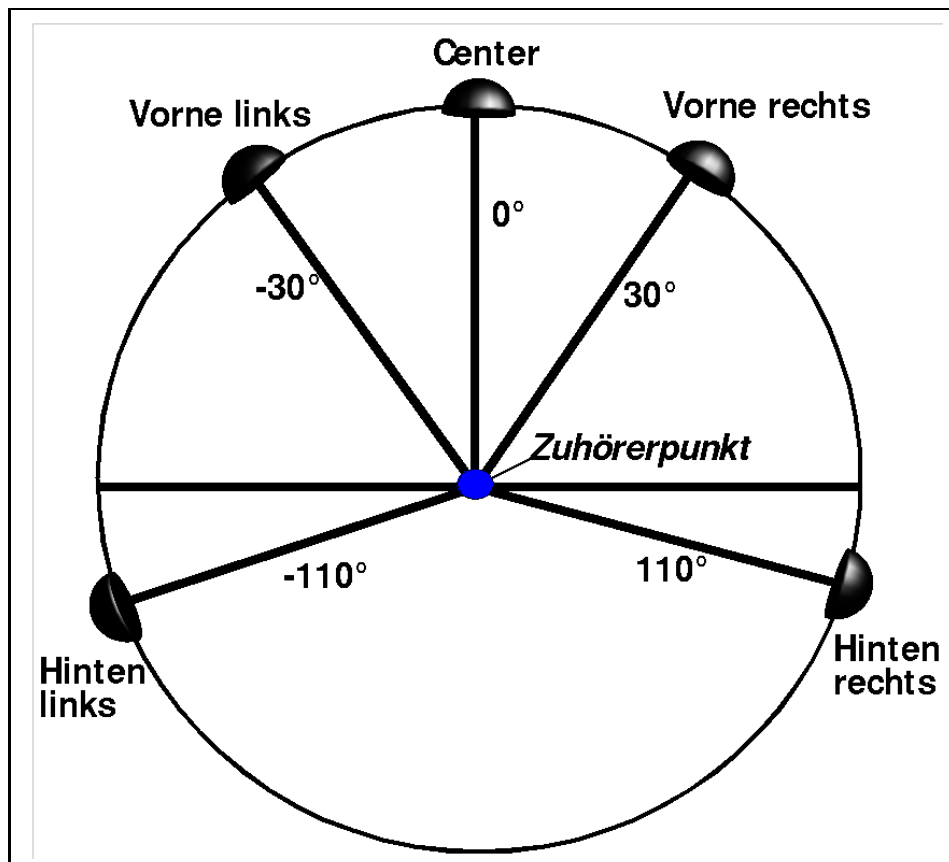


Abbildung 2.1: ITU-Aufstellung für 5.1-Systeme zur Musikwiedergabe

Eine Weiterentwicklung mit jedoch eher geringem Verbesserungspotential sind 6.1-Verfahren, bei denen ein zusätzlicher Back-Center Lautsprecher Ver-

wendung findet. Das Signal für selbigen kann entweder diskret bei entsprechenden Tonformaten vorhanden sein oder aus den Signalen für die beiden Surround-LS generiert werden<sup>1</sup>.

### 2.1.2 Dolby Digital

Das digitale „Dolby Digital“ ist ein Mehrkanal-Tonsystem, das im Kino, auf Laserdiscs, DVDs und im Fernsehen zum Einsatz kommt. Das verwendete Kodierungsverfahren heisst AC-3. Die vollfrequenten Kanäle haben wie bei Dolby Surround ein Spektrum von 20Hz bis 20kHz, der Basskanal reicht bis 120Hz. AC-3 unterstützt Bitraten zwischen 32 und 640 kbps (DVD für 5.1 Ton 384 oder 448 kbps, Stereo-Ton 192 oder 224 kbps). Die Kompression basiert wie auch bei MP3, Vorbis und AAC auf der Tatsache, dass das menschliche Ohr bestimmte Toninformationen nicht wahrnimmt, es ist also verlustbehaftet.

Zur Wiedergabe von mehr als 6 getrennten Kanälen wie z.B. beim 7.1 System muss eine Erweiterung des Dolby Digital verwendet werden, das s.g. **Dolby Digital Plus** <sup>2</sup>. Dieses verwendet eine erweiterte Form der AC-3 Kodierung, das s.g. E(nhanced)AC-3, welches bis zu 13 diskrete Kanäle übertragen kann und einige weitere Verbesserungen bereithält.

Abbildung 2.2 zeigt ein Beispiel für eine Aufstellung eines 7.1-Systems:

Die Positionierung der mittleren und hinteren Lautsprecherpaare kann zwischen 60° und 150° variieren. Für die Musikwiedergabe würde man ein symmetrischeres Layout wählen, die mittleren Lautsprecher weiter nach hinten (etwa  $\pm 90^\circ$ ) und die hinteren Lautsprecher weiter seitlich ( $\pm 120^\circ$ ) positionieren.

### 2.1.3 Sony Dynamic Digital Sound (SDDS)

Das Sony Dynamic Digital Sound System wurde von Sony für den Einsatz in Kinos entwickelt und findet auch nur dort Einsatz, ist also für Heimanwender eher uninteressant. Es ist ein digitales Verfahren zur Kodierung und Wiedergabe von bis zu acht Kanälen (eine abgespeckte Version mit sechs Kanälen existiert). Das ATRAC-Verfahren (Adaptive Transform Acoustic Coding) zur Kodierung und Wiedergabe wurde auch für DVDs spezifiziert, allerdings hat sich dieses System nie durchgesetzt, es gibt weder DVDs mit entsprechender Tonspur, noch dazugehörige Dekoder. Die Anordnung der Lautsprecher ist hier aber anders als bei den übrigen Systemen. Bei der Version mit acht Lautsprechern befinden sich ganze fünf (bei sechs Kanälen nur drei) im Frontbereich und lediglich zwei im hinteren Bereich. Ein weiterer LFE Kanal ist auch hier für Frequenzen zwischen 20 und 120 Hz zuständig und besitzt keinen festen Platz.

SDDS ist wie die Konkurrenz Dolby Digital und DTS mittlerweile bei fast allen aktuellen Hollywood-Filmen vorhanden. Es wird im Kino verhältnismässig selten eingebaut, da das System seinen wesentlichen Vorteil, die fünf Frontkanäle, nur auf sehr grossen Leinwänden ausspielen kann, wie sie meist nur in wenigen Sälen ab ca. 600 Plätzen in grossen Städten vorhanden sind <sup>3</sup>.

---

<sup>1</sup><http://de.wikipedia.org/wiki/Mehrkanal-Tonsystem>

<sup>2</sup>[http://de.wikipedia.org/wiki/Dolby\\_Digital](http://de.wikipedia.org/wiki/Dolby_Digital)

<sup>3</sup><http://de.wikipedia.org/wiki/SDDS>

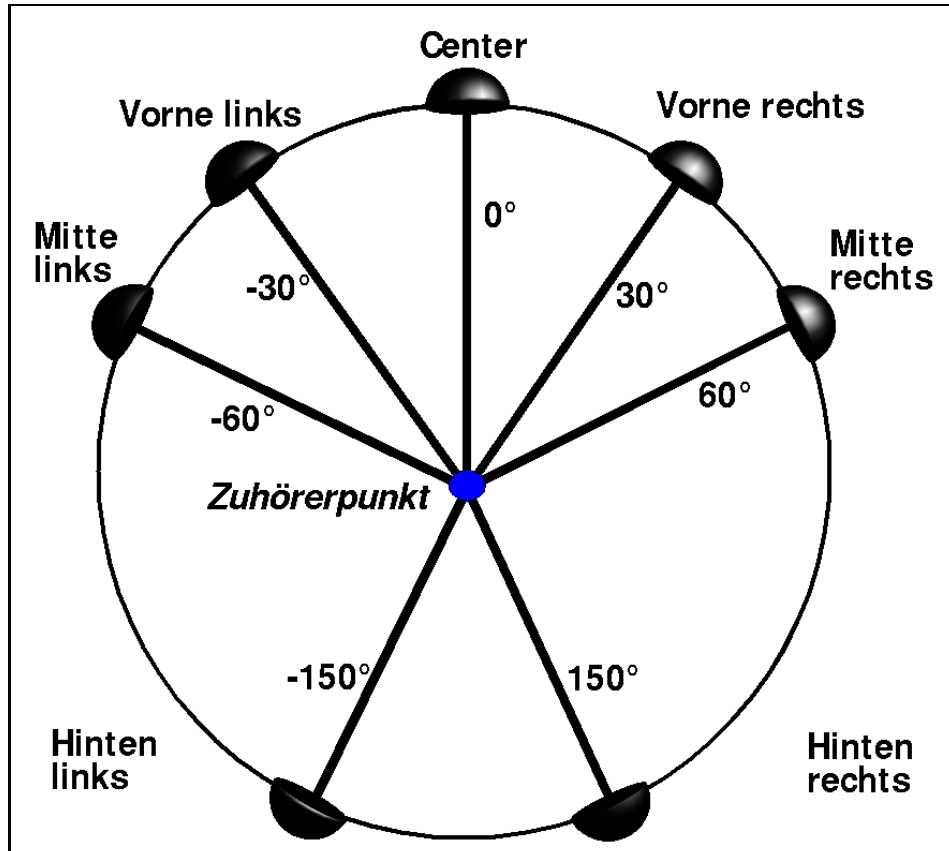


Abbildung 2.2: Die von der ITU entwickelte Aufstellung für 7.1-Systeme

#### 2.1.4 Digital Theater Sound

DTS ist die Abkürzung für „Digital Theatre Sound“ und wurde von der Firma Digital Theatre Systems Inc. entwickelt. Es ist ein Mehrkanal-Tonformat, das im Kino, bei Laserdisks und DVDs Verwendung findet. DTS verwendet ebenfalls ein verlustbehaftetes Verfahren zur Kodierung der Audiodaten, das s.g. CAC. Es kodiert bis zu acht volle Kanäle und kann in den verwendeten Datenraten variieren. DTS wird als die höherwertigere Alternative zu Dolby Digital vermarktet, die Bitrate für Video-DVDs kann hier bis zu 1509,25 kbps betragen. Ein optionaler LFE-Kanal deckt aber nur den Frequenzbereich bis 80 Hz ab, während er bei den anderen Systemen bis 120 Hz geht. Synchronisiert wird der Ton zum Film mittels eines Timecodes (besteht aus Punkten und Strichen ähnlich einem Barcode), der sich zwischen der Lichttonspur und dem eigentlichen Bild befindet und mit einem kleinen Abtaster ausgelesen wird. Da jedes einzelne Bild einen eigenen Timecode hat, wird der Ton selbst bei Filmriss bzw. fehlenden Bildern nachgeführt. Sollte der Timecode ausfallen oder nicht lesbar sein, wird der Ton ca. 6 Sekunden weiter von der CD gespielt, bevor auf den Lichtton zurückgeschaltet wird<sup>4</sup>.

<sup>4</sup>[http://de.wikipedia.org/wiki/Digital\\_Theater\\_Systems](http://de.wikipedia.org/wiki/Digital_Theater_Systems)

## Kapitel 3

# Methoden zur Generierung von virtuellen Klangquellen und Raumklangeffekten

Man kann grundsätzlich zwei Kategorien von Systemen unterscheiden. Nämlich Systeme, die ein Schallfeld in einem bestimmten Hörbereich simulieren (VBAP, Ambisonic, Wellenfrontsynthese) und solche, die Eingangsgrößen der Schallwellen an Trommelfellen reproduzieren (Binaurale- und Transaurale Verfahren). Beide werden im Folgenden kurz vorgestellt und Vor- und Nachteile analysiert. Zunächst aber eine allgemeine Einführung, wie es dem Menschen überhaupt gelingt, Schallquellen zu orten.

### 3.1 Lokalisation von Schallereignissen

Ein Mensch ist in der Lage, seine wahrgenommenen Hörereignisse bestimmten Richtungen zuzuweisen. Dies geschieht in verschiedenen Ebenen und Richtungen unterschiedlich gut, man spricht von der sogenannten „Lokalisationsschärfe“. Dies könnte daher rühren, dass bestimmte Richtungen von Menschen in der Frühzeit besonders gut lokalisierbar sein mussten, um Gefahren frühzeitig zu erkennen. Man unterscheidet Horizontalebene, Medianebene und Frontalebene, wobei diese Ebenen durch ein auf den Kopf bezogenes Koordinatensystem definiert werden.

Sobald die Schallquelle nicht mehr direkt auf der  $0^\circ$ -Achse der **Horizontalebene** liegt, kommt es zu unterschiedlichen interauralen Signalen, also interauralen Laufzeitdifferenzen (ITD) und interauralen Pegeldifferenzen (ILD). Oder einfach gesprochen, der Schall trifft auf einem Trommelfell früher auf als auf dem andern, diese Unterschiede werden vom Gehirn verarbeitet. Laufzeitunterschiede können durch das menschliche Gehör bereits ab einer Größe von  $10\mu\text{s}$  zur Richtungslokalisierung ausgewertet werden. Dieses entspricht einer Lokalisationsschärfe in der Horizontalebene von etwa einem Grad.

Bis zu einer Laufzeitdifferenz von  $0,63\text{ ms}$  erhöht sich die seitliche Lokalisation in etwa proportional zum Laufzeitunterschied. Eine Laufzeitdifferenz von  $0,63$

ms entspricht einer Wegstreckendifferenz des Schalls von 21,5 cm. Diese, auch „Hornbostel-Wertheimer Konstante“ genannte Grösse, entspricht der Wegstreckendifferenz bei Schalleinfall aus  $90^\circ$  bzw.  $270^\circ$  Einfallrichtung. Interaurale Pegeldifferenzen entstehen nicht nur durch die unterschiedlichen Weglängen und die dadurch unterschiedliche Dämpfung, sondern auch durch Abschattungen durch den Kopf.

Beide Phänomene, besonders aber das letztere, sind stark frequenzabhängig, da Frequenzen mit Wellenlängen in der Grössenordnung des Hindernisses (Kopf) kaum noch gebeugt werden, sondern am Hindernis reflektiert werden.

Tiefe Frequenzen unterhalb von etwa 300 Hz bilden dagegen keinen Schallschatten und damit keine wahrnehmbaren Pegeldifferenzen aus. Theoretisch können Signale, die genau auf der  $0^\circ$ -Achse liegen also nicht detektiert werden.

Dass in der Realität solche Schalle dennoch lokalisiert werden können, liegt möglicherweise daran, dass ein Mensch seinen Kopf nie vollkommen ruhig hält und dadurch Pegel- und Laufzeitunterschiede entstehen.

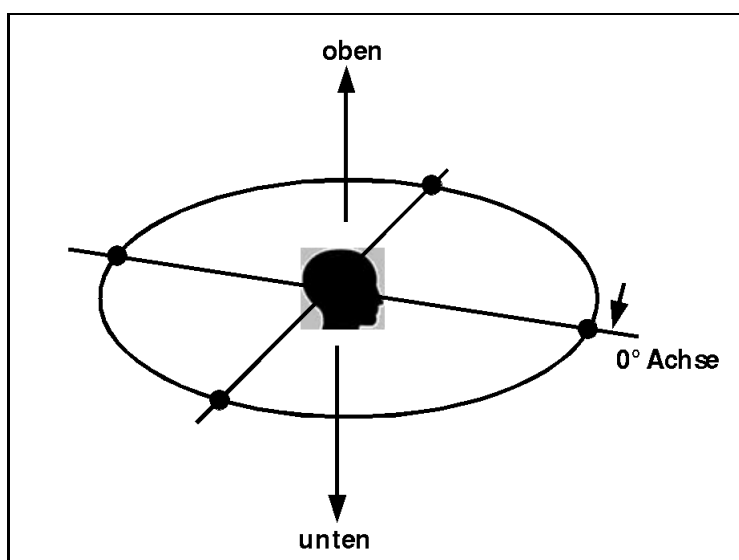


Abbildung 3.1: Darstellung der Horizontalebene mit dem Kopf im Mittelpunkt

Bei der **Medianebene** kann man davon ausgehen, dass es keine Laufzeit- und Pegeldifferenzen an den Ohren gibt. Dennoch kann ein Schallereignis in dieser Ebene lokalisiert werden (wenn auch weniger gut). Dies geschieht mit Hilfe von Klangfarbenunterschieden, den Spektraldifferenzen, die man aber nicht direkt als Klangunterschied wahrnimmt. Es sind bestimmte Frequenzbereiche, die so genannten „Blauertschen Bänder“, auch „richtungsbestimmende Frequenzbänder“ genannt. Es sind sehr komplexe Formen von Frequenzanhebungen und -absenkungen notwendig, um ein Hörereignis in dieser Ebene lokalisieren zu können. Naturgemäss dürfen Hörereignisse für diese Art der Lokalisation nicht zu schmalbandig sein.

Die Lokalisationsschärfe liegt bei unbekanntem Signalen bei rund  $17^\circ$ , bei bekannten Signalen um  $9^\circ$ . Diese Werte gelten für den Blick nach vorne.

Je weiter ein Signal aus der Vorwärtsrichtung austritt, desto schlechter wird die

Lokalisationsgenauigkeit.

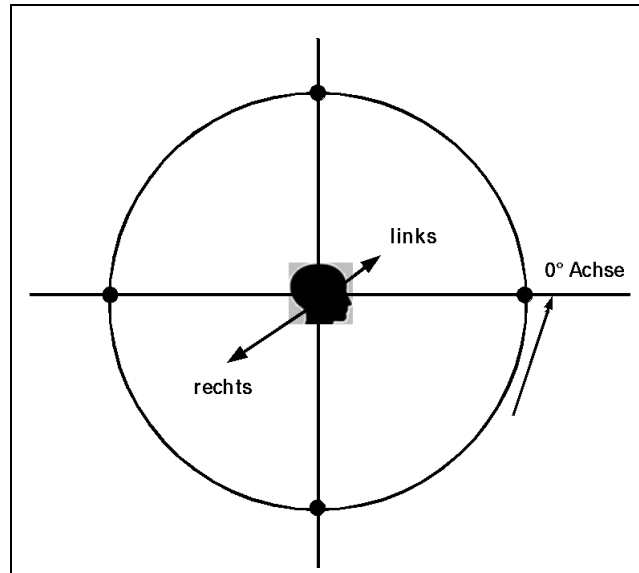


Abbildung 3.2: Darstellung der Medianebene mit dem Kopf im Mittelpunkt

Bei Schallereignissen, die auf der **Frontalebene** liegen kommen beide Mechanismen zum Tragen.

Als **Im-Kopf-Lokalisation** bezeichnet man einen als unangenehm empfundenen Effekt, der vor allem bei der Kopfhörerwiedergabe, aber auch bei Lautsprecherwiedergabe eintreten kann. Die Hörereignisse werden dann nicht mehr ausserhalb des Kopfes lokalisiert, sondern im Kopf. Das Gehirn vergleicht die eintreffenden Signale mit bekannten Signalen. Wenn die neuen Signale unbekannt sind, kann es zu einer Im-Kopf-Lokalisation kommen. Bei Lautsprechern tritt dieser Fall bei Verpolung ein und wenn man sie in einem Winkel von mehr als  $90^\circ$  aufstellt <sup>1</sup>.

<sup>1</sup>[http://de.wikipedia.org/wiki/Binaurale\\_Tonaufnahme](http://de.wikipedia.org/wiki/Binaurale_Tonaufnahme)

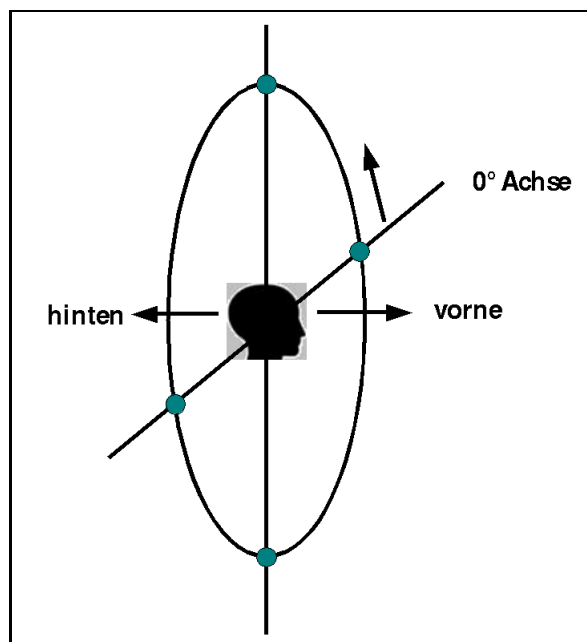


Abbildung 3.3: Darstellung der Frontalebene mit dem Kopf im Mittelpunkt

## 3.2 Binaurale Tonaufnahmen

Da binaural früher häufig mit stereophon gleichgesetzt wurde, muss man hier eine Unterscheidung treffen. Stereophone Aufnahmen sind solche, die erst bei der Wiedergabe über ein stereophones System (2 Lautsprecher) gemischt werden und so ihre Richtungsinformationen erhalten, während binaurale Tonaufnahmen speziell für die Wiedergabe über Kopfhörer entwickelt sind und die Richtungsinformationen getrennt in beiden Signalen enthalten. Deshalb wird oft auch von Kopfhörer-Stereophonie gesprochen. Die bei binauralen Aufnahmen verwendeten Signale sind so genannte „Head Related Transfer Functions“ (HRTF).

Eine Head Related Transfer Function beschreibt, wie ein gegebenes Schallereignis durch die Reflektionen und Brechungen am Oberkörper, am Kopf und an der Ohrmuschel gefiltert wird, bevor es die eigentliche Signalverarbeitung am Trommelfell und im Innenohr erreicht. Da diese Filterungseffekte abhängig vom Standort der Klangquelle sind, können diese genutzt werden, um den Eindruck von Raumklang zu erzeugen, indem die HRTFs für die entsprechend positionierte Klangquelle simuliert werden und über Kopfhörer wiedergegeben werden. Der Mensch ist also in der Lage, anhand der Filterung, die ein Schallereignis erfährt, zu erkennen, aus welcher Richtung dieses kommt. Durch eine Grosszahl von Messungen für entsprechend viele Einfallsrichtungen und Frequenzen etc. kann sozusagen ein Katalog von verschiedenen HRTFs erstellt werden, mit denen dann Richtungsinformationen den Klangereignissen mitgegeben werden können. Dies geschieht durch Filterung mit der, der Richtung entsprechenden, HRTF. Allerdings sind die HRTFs bei allen Menschen unterschiedlich, bedingt durch den individuellen Körperbau. Deshalb muss man eine entsprechend gewichtete Mittelung finden. Das einfache arithmetische Mittel eignet sich nicht, da dadurch die

Richtungsinformationen verwischt würden und der Klang verfälscht wird. Binaurale Aufnahmen sind die beste Möglichkeit, einen räumlichen Höreindruck realitätsnah zu reproduzieren. Die Aufnahmen erfolgen meist unter Verwendung eines Kunstkopfes (Kunstkopfstereophonie) [?].

Der **Vorteil** bei diesem Verfahren ist, dass die Invertierung der Übertragungswege einfach ist. Die Berücksichtigung mehrerer Zuhörer ist leicht und mit gleichbleibender Qualität möglich, durch Verwendung beliebig vieler Kopfhörerpaare. Ausserdem ist sie damit nicht an einen bestimmten Raumpunkt gebunden.

Der Vorteil der räumlichen Unabhängigkeit ist aber auch einer der **Nachteile**, da das Tragen von Kopfhörern viele Menschen als störend empfinden. Ausserdem ist es technisch eine der schwierigsten Methoden, da aufwendige HRTF-Messungen notwendig sind (ca. 700) bis die idealen Übertragungsfunktionen gefunden sind.

### 3.3 Wellenfeldsynthese

Die akustische Wellenfeldsynthese (WFS), auch Wellenfrontsynthese genannt, ist die wohl komplexeste Möglichkeit Schallereignisse über einen dreidimensionalen Raum nachzubilden. Im Gegensatz zu allen anderen Methoden, die Lautsprecheranordnungen verwenden, ist es bei der WFS möglich, auch virtuelle Quellen abzubilden, die nicht auf der Ebene der Lautsprecher liegen. Hier wird nämlich versucht, die Luftschwingungen eines Schallszenarios im kompletten Raum nachzubilden, weshalb die virtuellen Quellen sogar umgehbar werden. Basierend auf dem Huygensschen Prinzip wurde die Theorie zur WFS in den 80ern an der Technischen Universität von Delft entwickelt. Das Huygenssche Prinzip besagt, dass jeder Punkt einer Wellenfront selbst wieder als Ausgangspunkt einer neuen Welle angesehen werden kann. Durch Überlagerung dieser unendlich vielen Elementarwellen entsteht wieder die ursprüngliche Wellenfront (siehe Abb. 3.4).

Mathematische Grundlage ist das Kirchhoff-Helmholtz-Integral, das die Rekonstruktion eines beliebigen Schallfeldes mithilfe von Mono- und Dipolen innerhalb einer geschlossenen Hüllfläche beschreibt. Durch einige Vereinfachungen wurde erreicht, dass nur noch eine begrenzte Anzahl von Monoquellen (durch normale Lautsprecher näherungsweise realisiert) erforderlich ist. Diese werden auf so genannten Lautsprecherpanals rings um den Raum verteilt (die Anzahl schwankt zwischen etwa 30 bis weit über 100 Stück). Aufnahmen für solche Abspielanordnungen müssen streng genommen derart hergestellt werden, dass alle Schallereignisse mit ihren Richtungsinformationen getrennt aufgenommen werden. Bei einem Konzert z.B. müsste dann jedes Instrument mit einem eigenen Mikrofon aufgenommen werden, mit möglichst wenig Störgeräuschen. Da dies in der Praxis nicht möglich ist wurden spezielle Aufnahmetechniken und Mikrofonanordnungen entwickelt, die dieser Vorgabe Nahe kommen und gute Ergebnisse liefern. Bei der Wiedergabe werden die Klänge samt Positionsangaben computerunterstützt auf die Lautsprecherpanals gegeben. Kleine Prozessoren innerhalb dieser Panals errechnen dann auf Grundlage der Positionsangaben, welche Lautsprecher des jeweiligen Panals angesprochen werden müssen um diesen Klang zu reproduzieren.

Interessant ist, dass die grosse zu speichernde Informationsmenge (Audio-



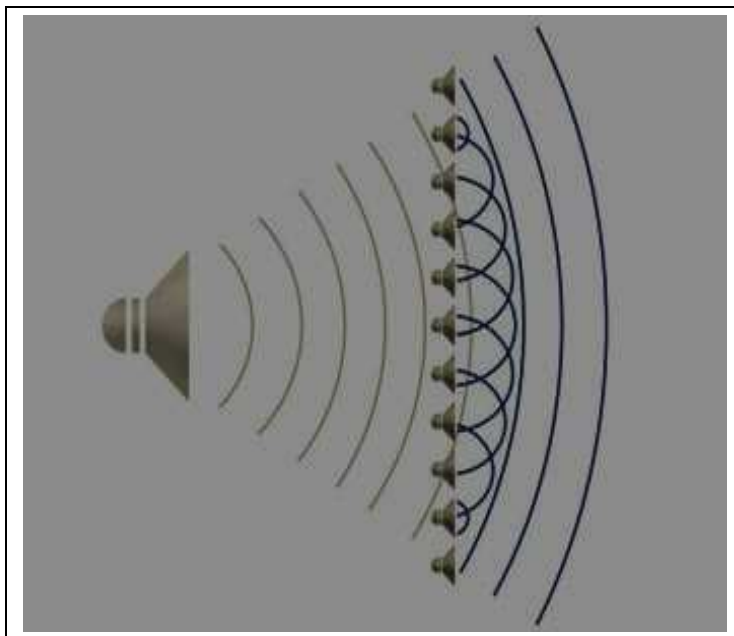


Abbildung 3.4: Das Huygenssche Prinzip der Wellenausbreitung

und Positionsdaten) auch einen Vorteil für herkömmliche Surround-Systeme (5.1 oder 7.1) bieten. So können diese Lautsprecher als virtuelle Quellen wiedergegeben werden, die scheinbar weiter ausserhalb des Raumes stehen, wodurch der optimale Zuhörerpunkt (auch „sweet spot“<sup>2</sup> genannt) erweitert wird und somit ein besseres Hörerlebnis für Zuschauer weiter am Rand der Anordnung entsteht.

Im Zusammenhang mit der WFS sollte man auch das s.g. Beamforming erwähnen. Es verwendet ähnliche Lautsprecherarrays wie die WFS. Ziel ist aber nicht die exakte Reproduktion des Schallfeldes. Vielmehr wird hier ein möglichst stark gerichtetes Signal abgestrahlt, das dann durch gezielte Reflexionen an Wänden und Decken gelenkt wird und so eine virtuelle Quelle in einem bestimmten Punkt bildet. Hierbei sind Form und Abmessungen des Abspielraums entscheidend.

Beiden Verfahren gemeinsam ist, dass die obere Grenze der wiederzugebenden Frequenzen durch den Abstand der einzelnen Lautsprecher auf den Panals und die untere Grenze durch die Länge der Panals selbst beschränkt ist<sup>3</sup>.

Vorteil der Wellenfeldsynthese ist, dass sie die wohl besten Lokalisationseigenschaften aller Systeme besitzt, da das originale Schallfeld über einen weiten Hörbereich reproduziert werden kann.

Es gibt keinen „sweet“ pot für die Zuhörerposition, d.h. es ist möglich mehrere Zuhörer zu integrieren. Ausserdem ist es relativ variabel bezüglich Raumanforderungen und Realisierungsmöglichkeiten.

Klarer Nachteil ist die aufwendige praktische Realisierung. Dadurch füllt es als mögliches Realisierungskonzept für gewöhnliche Heimkinoanlagen weg, obwohl

<sup>2</sup>sweet spot: Ort der besten Lokalisation von virtuellen Quellen. Je weiter man sich von diesem Punkt entfernt, desto mehr Verzerrungen und Fehllokalisationen treten auf.

<sup>3</sup><http://de.wikipedia.org/wiki/Wellenfeldsynthese>

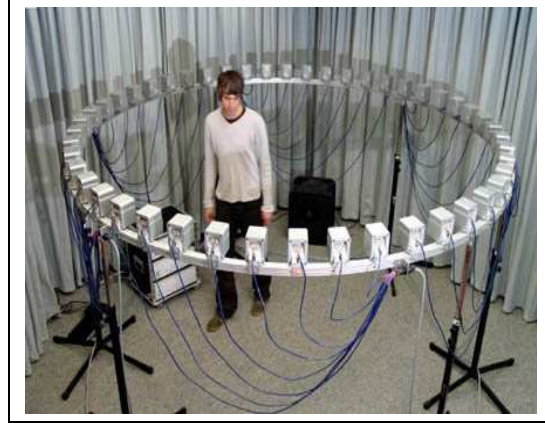


Abbildung 3.5: Versuchsaufbau einer WFS-Anlage

mittlerweile erste Lautsprecherpanal für Heimkinoanlagen angeboten werden.

### 3.4 Vector Based Panning (VBP)

Eine von V. Pulkki entwickelte Methode zur Generierung von Punktklangquellen. Im 2-d Fall wird die horizontale Ebene auf mehrere stereophone Systeme aufgeteilt. Dazu werden die Lautsprecher in einem Kreis um die vorher festgelegte Zuhörerposition aufgestellt. Damit wird die Reproduktion eines originalen Schallfeldes in diesem Punkt möglich.

Im einfachsten Fall strahlen zwei Lautsprecher ein kohärentes Signal mit unterschiedlicher Amplitude ab. Der Zuhörer nimmt ein einziges Schallereignis (virtuelle Quelle) an einem Ort, bestimmt durch die Position der Lautsprecher und des Zuhörers und die Amplituden der Signale, wahr.

Eine typische Stereoanordnung zeigt Abbildung 3.6. Zwei symmetrisch zur x-Achse im Winkel von  $30^\circ$  stehende Lautsprecher, deren Amplituden durch die Gain-Faktoren  $g_1$  und  $g_2$  kontrolliert werden.

Sinnvoll ist eine Beschränkung der virtuellen Quelle auf einen Winkel, dessen Betrag kleiner ist als der zwischen Lautsprecher und X-Achse, denn bei der Wiedergabe von Quellen ausserhalb des Bereichs zwischen den Lautsprechern erzeugt Amplituden-Panning gegenphasige Signale, welche den Klangeindruck verzerren.

Es gibt mehrere Methoden, den Winkel  $\Theta$  ( $Theta_s$  bzw.  $Theta_T$ ) der virtuellen Quelle in Beziehung mit den Gain-Faktoren der Lautsprecher zu setzen. Zunächst findet man folgende exakte Beziehung:

$$\frac{\sin \Theta_s}{\sin \Theta_0} = \frac{g_1 - g_2 \cos \Theta_0}{g_1 + g_2 \cos \Theta_s} \quad (3.1)$$

Der Faktor  $\frac{\cos \Theta_0}{\cos \Theta_s}$  liegt bei einem Winkel der Lautsprecher von  $\Theta_0 = 30^\circ$  zwischen 0,86 und 1, je nach simuliertem Winkel.

Durch das Weglassen des Terms  $\frac{\cos \Theta_0}{\cos \Theta_s}$  wird die virtuelle Quelle also in Richtung einer der Lautsprecher verschoben (vergleiche Grafik unten). Dies ent-

spricht eher der tatsächlichen Lokalisation durch das menschliche Gehör. Die Gleichungen (3.2) und (3.3) sind somit Näherungen, die die akustische Wahrnehmung eines Zuhörers berücksichtigen. Das „stereophonic law of sines“ hat dabei Gültigkeit, solange der Zuhörer starr nach vorne sieht.

$$\frac{\sin \Theta_s}{\sin \Theta_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{3.2}$$

Folgt die Kopfbewegung des Zuhörers der virtuellen Quelle, ist das „tangent law“ zutreffender, welches man durch ähnliche Korrekturen wie oben erhält.

$$\frac{\tan \Theta_T}{\tan \Theta_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{3.3}$$

Die Richtung  $\Theta$  einer virtuellen Quelle entspricht nun annähernd dem Winkel  $\Theta_s$  bzw.  $\Theta_T$ .  $\Theta_0$  ist der Winkel zwischen der X-Achse und einem Lautsprecher (gewöhnlich ca. 30°). Weil man aus diesen beiden Gleichungen jedoch lediglich

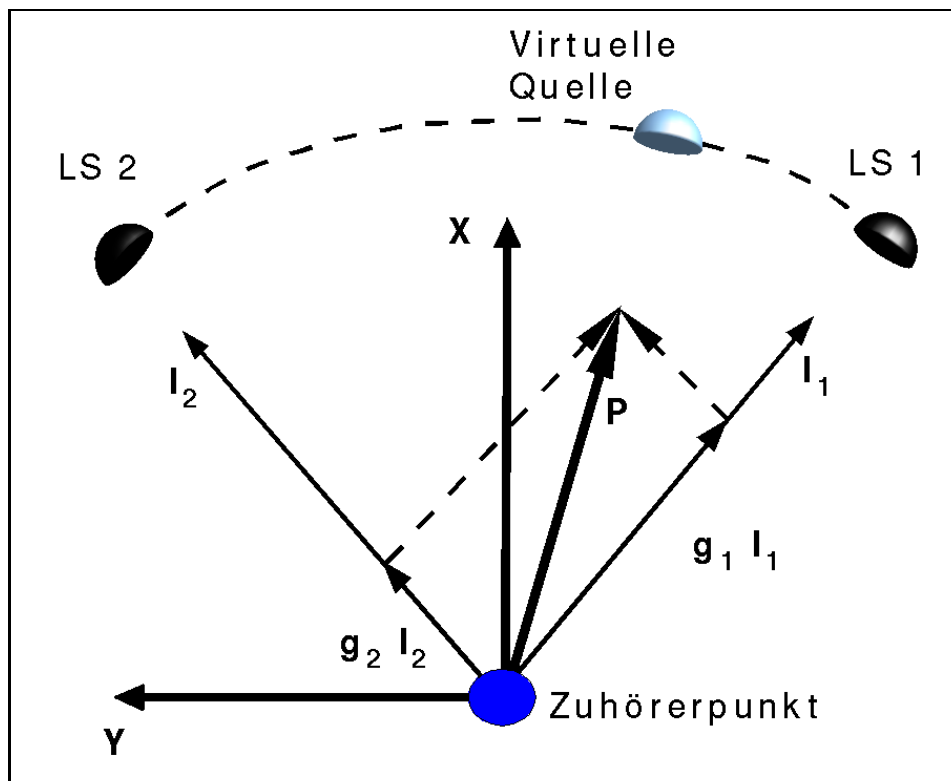


Abbildung 3.6: Schematische Darstellung des Prinzips von VBP

das Verhältnis der Werte zueinander bestimmen kann, muss man eine weitere Beziehung finden, über die man die exakten Werte bestimmen kann. Darum führt man eine Gleichung ein, die die virtuelle Quelle für den Zuhörer auf einer konstanten Entfernung hält.

$$\sqrt[p]{\sum_{n=1}^N g_n^p} = 1 \tag{3.4}$$

Der Exponent  $p$  lässt eine Anpassung an den vorliegenden Raum, bzw. dessen akkustische Eigenschaften zu:

- $p=1$ : Amplitude Panning (VBAP) Die Signalamplituden der Signale werden weitgehend konstant gehalten, was dem „Amplitudenmodell“ von Gerzon entspricht und in reflexionsarmen Räumen zu guter Lokalisation führt (Frequenzbereich  $< 700\text{Hz}$ ).
- $p=2$ : Intensity Panning (VBIP) Die Schallenergie wird auf konstantem Wert gehalten. Dies entspricht dem „Energimodell“, wird dort eingesetzt, wo mit Reflexionen zu rechnen ist. Gültigkeit im Frequenzbereich oberhalb etwa  $700\text{ Hz}$ .

Mit VBAP lassen sich Gain-Faktoren für Lautsprecher-Paare wie folgt berechnen. Dazu wird das „tangent law“ zunächst in eine Vektordarstellung überführt. Wir nehmen hier an, dass die Lautsprecher sich auf einem Einheitskreis um den Zuhörerpunkt befinden. Für abweichende Aufstellungsorte müssen Korrekturen von Laufzeit und Gainfaktoren durchgeführt werden, die die Lautsprecher wieder auf eine Kreisfläche „projizieren“. Die Einheitsvektoren  $\vec{I}_1$  und  $\vec{I}_2$ , die in Richtung des jeweiligen Lautsprechers weisen bilden die Basis. Der Vektor  $\vec{P}$ , welcher zur virtuellen Quelle zeigt, kann nun als Linearkombination der mit  $g$  gewichteten Lautsprechervektoren aufgefasst werden.

$$\vec{P} = g_1 \vec{I}_1 + g_2 \vec{I}_2 \quad (3.5)$$

oder in Matrixschreibweise mit  $\vec{g} = [g_1 g_2]$  und  $L_{12} = \begin{bmatrix} \vec{I}_1 & \vec{I}_2 \end{bmatrix}^T$

$$\vec{p}^T = \vec{g} L_{12} \quad (3.6)$$

$$\vec{g} = \vec{P}^T L_{12}^{-1} = [p_1 p_2] \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}^{-1} \quad (3.7)$$

Das Gleichungssystem (3.7) kann gelöst werden, wenn die Inverse von  $L$  existiert, was für  $\Theta_S \neq 0^\circ$  und  $\Theta_S \neq 90^\circ$  zutrifft. Beide Fälle entsprechen praxisfernen Anordnungen.

Die so berechneten Amplitudenwerte erfüllen (3.5), müssen aber, um (3.6) ebenso zu genügen, dementsprechend normiert werden, was mittels Gleichung (??) passiert.

$$\vec{g}_{normiert} = \frac{\vec{g}}{\sqrt{g_1^2 + g_2^2}} \quad (3.8)$$

In 2D-Wiedergabesystemen mit mehr als 2 Lautsprechern erfolgt paarweises Panning, d.h. jeweils nur 2 benachbarte Lautsprecher bilden eine Basis  $L_{mn}$ , die s.g. „active arcs“ (Bereiche zwischen den aktiven Lautsprechern) sollen sich nicht überlappen. Das garantiert kontinuierlich veränderliche Gainfaktoren bei der Wiedergabe bewegter Schallereignisse und somit keine Verzerrungen oder Fehllokalisationen. Wird eine virtuelle Quelle genau im Punkt eines Lautsprechers erzeugt, so ist nur dieser aktiv und die schärfste Lokalisation wird erreicht. Die Übertragung auf den 3-dimensionalen Fall ergibt sich dadurch, dass immer genau drei Lautsprecher angesprochen werden. Nämlich die, die am nächsten

zur virtuellen Quelle liegen, immer drei Lautsprecher bilden also ein Kugelsegment.

Somit ergeben sich folgende Gleichungen:

$$\vec{P} = g_1 \vec{I}_1 + g_2 \vec{I}_2 + g_3 \vec{I}_3 \quad (3.9)$$

oder in Matrixschreibweise mit  $\vec{g} = [g_1 g_2 g_3]$  und  $L_{123} = [\vec{I}_1 \vec{I}_2 \vec{I}_3]^T$

$$\vec{P}^T = \vec{g} L_{123} \quad (3.10)$$

$$\vec{g} = \vec{P}^T L_{123}^{-1} = [p_1 p_2 p_3] \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}^{-1} \quad (3.11)$$

Folgender Algorithmus kann zur Realisierung verwendet werden:

1. Positionierung der Lautsprecher
2. Berechnen und speichern der Matrizen  $L_{ij}$  (bzw.  $L_{ijk}$ )
3. Richtungsvektoren der virtuellen Schallquellen definieren
4. zuständige Lautsprecherpaare (bzw -tripel) auswählen
5. Amplituden der Lautsprechersignale berechnen

Die Schleife muss in Echtzeit durchlaufen werden.

Der **Vorteil** von Vector Based Panning ist, dass virtuelle Schallquellen an beliebigen Positionen auf der „Lautsprecherkugel“ realisierbar sind und dass dies, verglichen mit anderen Verfahren, einfach ist.

**Nachteil** ist hierbei die schwankende Lokalisationsschärfe, d.h. die Lokalisation von virtuellen Schallquellen in der Nähe eines Lautsprechers ist deutlich besser als genau zwischen den Boxen. Es hat wohl die schlechtesten Lokalisationseigenschaften der hier vorgestellten Methoden, wobei auch die Zuhörerposition genau eingehalten werden muss [1][2].

## Kapitel 4

# Ambisonic

Die Raumklangtechnologie „Ambisonic“ wurde in den 1970er Jahren vom britischen Mathematiker Michael Gerzon entwickelt. Ursprünglich handelt es sich um eine besondere Mikrophontechnik, mit der eine mehrkanalige Aufnahme erstellt wird, welche die räumliche Information des Schallfeldes trägt und schliesslich über ein Lautsprechersystem derart wiedergegeben werden kann, dass sich der Eindruck dreidimensionalen (periphonen) Hörens einstellt [3]. Man kann aber auch den umgekehrten Weg gehen, indem man dem Klang durch Verwendung bestimmter Gleichungen Rauminformation zu gibt: Eine genaue x, y und z- Koordinate wird jedem Ton zugeordnet. Die x, y und z- Komponenten eines jeden Klanges werden separat auf verschiedenen Tonspuren gespeichert. Diese Spuren enthalten die komplette Richtungs-Information, verlust- und redundanzfrei. In einem zweiten Schritt werden die so gewonnenen (kodierte) Spuren auf die gewünschte Lautsprecheraufstellung dekodiert, die im Grunde jede (symmetrische) Anordnung sein kann. Für asymmetrische Aufstellungen müssen gewisse Korrekturfaktoren eingebaut werden, dazu später mehr im Kapitel „Dekodierung“. An den Universitäten York und Derby in England wird weiter zu Ambisonic geforscht. Richard Furse und Dave Malham haben die Gleichungen zweiter Ordnung entwickelt, die noch präziser lokalisierbare Klangeindrücke aus jeder beliebigen Richtung und die Aufweitung des „sweet spot“ in der Mitte der Lautsprecheraufstellung ermöglichen. Die räumliche Wiedergabe mit Ambisonic teilt sich also in zwei grundlegende Teile auf:

- Kodieren (allgemein) und
- Dekodieren (für spezielle Anordnungen)

**Vorteil** ist, dass Ambisonic trotz des relativ geringen technischen Aufwands eine hohe Flexibilität in der Lautsprecherzahl und -anordnung bietet. Die Lokalisation ist im Zentrum gut und dieser Bereich ist grösser als bei dem VBP-Verfahren.

**Nachteil**, wie auch bei VBP, sind Verfälschungen, wenn die zentrale Zuhörerposition verlassen wird. Doch diese Nachteile können für den fest in der Mitte des Raums positionierten Kunstkopf zunächst ausser Acht gelassen werden.

## 4.1 Aufnahmetechnik und Speicherformate

Das Ziel einer Aufnahme für eine spätere Wiedergabe über Ambisonic ist, die räumliche Information des Klangfeldes zu speichern. Für eine Ambisonic-Aufnahme verwendet man deshalb meist ein sogenanntes Soundfield Mikrofon. Die Soundfield Mikrofon bestehen aus einer Anordnung von sechs cardioidischen (drei „figure-of-eight“) Mikrofonen, angeordnet in Form eines regelmässigen Tetraeders (zwei an der Unterseite aneinandergesetzte Pyramiden), d.h. jeweils orthogonal zueinander, und einem omnidirektionalen Mikrofon. Durch die Richtungscharakteristiken der cardioiden Mikrofone wird die Richtungsinformation des Klangfeldes bei der Aufnahme gespeichert. Das vom omnidirektionalen Mikrofon aufgenommene Signal beinhaltet keine Richtungsinformation, sondern die gesamte Amplitude aller im Zentrum auftreffenden Schallwellen.

Als Ausgangssignal erhält man das so genannte A-Format. Dieses Signal wird anschliessend elektronisch aufbereitet und man erhält das vierkanalige B-Format, das als Standardsignal einer Aufnahme mit dem Soundfield Mikrofon gilt. Die einzelnen Kanäle werden mit W,X,Y und Z bezeichnet. W ist ohne Richtungsinformation, X enthält die Anteile, die den Ausschlag in X-Richtung beinhalten (vorne/hinten), Y enthält die Anteile, die den Ausschlag in Y-Richtung beinhalten (links/rechts) und Z entsprechend die in Richtung der Z-Achse (oben/unten). Damit sind die Richtungsinformationen in den vier Spuren gespeichert und unabhängig von der zur Wiedergabe verwendeten Lautsprecheranordnung reproduzierbar [3][5].

Darauf aufbauend wurden weitere Formate abgeleitet, die entweder die Wiedergabe über ein stereophones System (UHJ) erlauben oder wie im Fall des G-Formats eine Dekodierung für ein bestimmtes Lautsprecherlayout (meist 5.1) beinhalteten.

### UHJ

Das nach der Aufnahme mit einem Soundfield Mikrofon erhaltene B-Format enthält zwar die Richtungsinformationen des aufgenommenen Klangfeldes, leider ist die Wiedergabe dieser vier Kanäle über ein gängiges Abspielsystem, wie LR-Stereo, aber nicht möglich. Es fehlt der entsprechende Dekoder, um die vier Signale auf oft nur zwei Wiedergabekanäle zu dekodieren. Als Ergebnis wurde in den 70ern ein Matrix-Kodierungsverfahren entwickelt, das dieses Problem löst: UHJ.

UHJ ist hierarchisch aufgebaut und nutzt in der vollen Version vier Kanäle, also genau so viele wie im B-Format. Es enthält die gleichen Informationen, allerdings ist der Inhalt der einzelnen Kanäle verschieden. Zwei dieser Kanäle sind nämlich kompatibel mit Stereo (L,R). Die beiden anderen Kanäle heissen Q und T. Sie enthalten Höheninformationen (Q) bzw. Informationen, die die Genauigkeit der horizontalen Surround-Wiedergabe verbessern. Werden nur die beiden Stereo-Kanäle mit einem Dekoder wiedergegeben, so ergibt sich ein guter horizontaler Surround Klang, ohne Dekoder ein „super-stereo“ Klang. Die Superposition beider Kanäle ergibt einen vollen Monokanal. Die Kodierungsbeziehungen für die beiden Stereokanäle von UHJ lauten <sup>1</sup>:

$$Left = (0.0928 + 0.255j)X + (0.4699 - 0.171j)W + (0.3277)Y$$

<sup>1</sup>[http://www.york.ac.uk/inst/mustech/3d\\_audio/ambis2.htm](http://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm)

$$Right = (0.0928 - 0.255j)X + (0.4699 + 0.171j)W - (0.3277)Y$$

Die imaginären Anteile der Faktoren bedeuten eine Phasenverschiebung des „normalen“ Signals über die gesamte Bandbreite um  $90^\circ$ . Die Kodierung des kompletten UHJ-Signals ist invertierbar, d.h. es ist möglich, das 4-spurige UHJ wieder in das ursprüngliche B-Format umzuwandeln.

Die Möglichkeit der Wiedergabe über ein „normales“ Stereo Lautsprecherpaar sollte die Vermarktung und Popularität von Ambisonic vorantreiben.

Dass dies trotzdem nicht in gewünschtem Mass gelang liegt an zwei Faktoren: Zum einen kann es vorkommen, dass es zu einem so genannten „beyond the speaker“ Effekt kommt, das heisst durch Phasenverschiebungen in den Signalen kommt es zu falscher Ortung des Klangs.

Andererseits ist es aber immernoch deutlich besser als herkömmliches Stereo. Zum anderen fehlte Ambisonic einfach der kommerzielle Hintergrund, den das Format benötigte um sich als Standard zu etablieren, wie ihn z.B. Dolbys „Dolby Stereo“ hatte.

### G-Format

G-Format ist die Bezeichnung für ein gerendertes B-Format, d.h. bereits für eine bestimmte Lautsprecheranordnung dekodiertes Stück. Meist meint man damit die Dekodierung für eine herkömmliche 5.1 ITU Lautsprecheraufstellung. Dieses Format sollte als Alternative zu DolbySurround-Klangspuren Einzug halten, allerdings mit geringem Erfolg.

Das G-Format hat den Vorteil gegenüber „normalem“ Ambisonic, dass die in B- oder UHJ-Format vorliegende Tonspur direkt im Studio dekodiert werden kann, s.d. die Notwendigkeit eines Decoders für den Endanwender entfällt.

## 4.2 Grundlagen zu Ambisonic

Im folgenden wird kurz die Theorie erklärt, die Ambisonic zu Grunde liegt. Ambisonic bietet die Möglichkeit, ein komplettes, drei-dimensionales Klangfeld zu erzeugen.

Wir betrachten hier aber lediglich den zwei-dimensionalen Fall, was aber völlig ausreicht, um das Prinzip zu verstehen. Ausserdem steigt die Komplexität nicht linear in der Anzahl der Dimensionen. Z.B. ist die Anzahl der Kanäle, die zur Wiedergabe des 2-dimensionalen Ambisonic nötig sind  $2m+1$ , wobei  $m$  die Ordnung des Systems ist, wohingegen im 3d Fall bereits  $(m+1)^2$  Kanäle nötig sind, was die Komplexität natürlich rasch erhöht. Das Modell basiert auf dem Vergleich einer einfallenden Referenzwelle mit der, welche vom System erzeugt wird.

Es wird die Annahme getroffen, dass das Referenzsignal eine ebene Welle ist, was zutrifft, so lange die Klangquelle genügend weit vom Hörer entfernt ist. Die Annahme, dass die ausgehende Welle des Lautsprechers ebenso eine ebene Welle ist, ist wiederum nur für Zuhörer weit genug entfernt vom Lautsprecher korrekt. Zur Begriffsklärung: Mit „einfallend“ ist hier die originale Welle gemeint, mit „ausgehend“ die Reproduzierte.



Diese Begriffe sollen verdeutlichen, dass das Originalsignal im Mittelpunkt ein- geht, dass das neu erzeugte (simulierte) Signal aber von den Lautsprechern ausgeht. In Ambisonic ist es üblich, dass das Koordinatensystem um  $90^\circ$  ge- genüber dem „normalen“ kartesischen gedreht ist. Deshalb zeigt die X-Achse zur Vorderseite des Raums und die Y-Achse nach links. (Für den drei-dimensionalen Fall: Z-Achse zeigt nach oben.)

Der Ursprung dieses Koordinatensystems liegt im Zentrum des Zuhörerraums (Raum innerhalb der Lautsprecheranordnung), auch „sweet spot“ genannt. Es wird angenommen, dass die ebene Welle in einem Winkel  $\psi$  entgegengesetzt dem Uhrzeigersinn bezüglich der X-Achse auftritt. Die Zuhörerposition ist im Punkt  $(r, \phi)$  bezüglich der X.-Achse, wie in Abbildung zu sehen. Sei  $\vec{k}$  der Vektor der

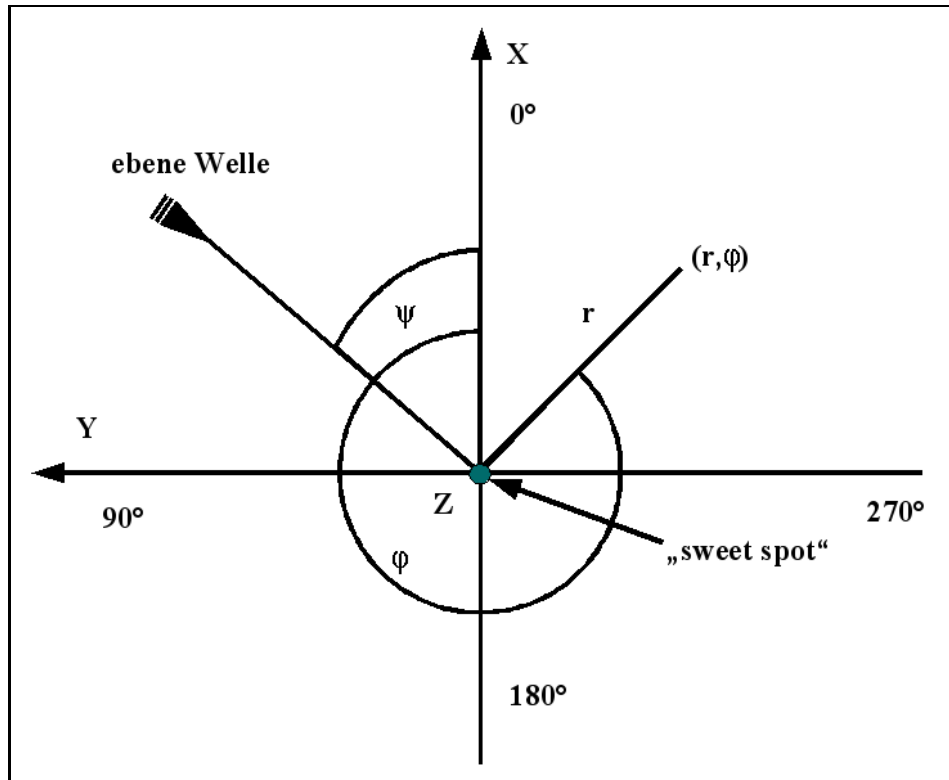


Abbildung 4.1: Schematische Darstellung einer einfallenden ebenen Welle

Welle und der Zuhörer-Vektor  $\vec{r}$ , dann ist in dieser Situation die Welle mit folgender Gleichung beschrieben:  $P_\psi e^{i\vec{k}\vec{r}}$ , wobei  $P_\psi$  der Schalldruck und  $k$  die Frequenz ist, bzw.  $\frac{2\pi}{\lambda}$ , mit der Wellenlänge  $\lambda$ .  $\vec{k}$  stellt eine Welle der Frequenz  $k$  dar, die sich in Richtung Mittelpunkt bewegt unter dem Winkel  $\psi$  bezüglich der X-Achse.

Diese Gleichung lässt sich auch folgendermassen ausdrücken:

$$S_\psi = P_\psi e^{ikr \cos(\phi - \psi)} \quad (4.1)$$

wobei  $S_\psi$  die originale, ebene Welle ist.

Das Ziel von Ambisonic ist nun, diese ebene Welle im Mittelpunkt zu reproduzieren bzw zu simulieren. Die Frage ist, wie kann das realisiert werden?

Es besteht eine Beziehung zwischen einer ebenen Welle und einer Reihenentwicklung von cosinus-Termen und Zylindrischen Bessel Funktionen (siehe Literatur). Wenn man eine ebene Welle nun als Reihenentwicklung darstellen kann, ist es vielleicht auch möglich die Terme niedriger Ordnung zu nutzen, um das Signal (in Näherung) zu rekonstruieren. Es sollte möglich sein, die ersten Terme dieser Reihe auf eine Anordnung von Lautsprechern zu verteilen.

Es gilt folgende Gleichung:

$$S_\psi = P_\psi J_0(kr) + 2P\psi \sum_{m=1}^{\infty} i^m J_m(kr) m \cos(\phi - \psi) \quad (4.2)$$

Der cosinus-Teil kann folgendermassen umgeschrieben werden:

$$S_\psi = P_\psi \left( J_0(kr) + 2P\psi \sum_{m=1}^{\infty} 2i^m J_m(kr) [\cos(m\psi) \cos(m\phi) + \sin(m\psi) \cos(m\phi)] \right) \quad (4.3)$$

Aus 4.3 kann man sehen, dass es möglich wäre eine ebene Welle exakt zu rekonstruieren, mit unendlich vielen Signalen (Lautsprechern).

Natürlich ist dies nicht realisierbar, da nur eine begrenzte Anzahl von Kanälen zur Verfügung steht. Zunächst geht es allerdings nur um die Idee, eine Angabe, bis zu welcher Ordnung Signale überlagert werden müssen ist noch nicht notwendig. Eine typische Anordnung für Ambisonic besteht aus N, äquidistant zu einem Mittelpunkt aufgebauten Lautsprechern, in einer polygonen Formation, z.B. wie in Abbildung 4.2, mit N=6 und einem Winkel von 60° zwischen den Lautsprechern.

Wenn jeder Lautsprecher eine ebene Welle produziert, dann ist das Ausgangssignal des n-ten Lautsprechers wie folgt gegeben:

$$S_n = P_n \left( J_0(kr) + 2P\psi \sum_{m=1}^{\infty} 2i^m J_m(kr) [\cos(m\psi) \cos(m\phi_n) + \sin(m\psi) \cos(m\phi_n)] \right) \quad (4.4)$$

wobei  $\phi_n$  der Winkel des n-ten Lautsprechers ist.

Da jeder der Lautsprecher zur gesamt resultierenden Welle beiträgt, wird die Summe über die N Signale die gesamte ebene Welle beinhalten, die produziert wurde.

Zu diesem Zeitpunkt müssen die Lautsprecher noch in einer regulären Anordnung stehen. Später wird die Notwendigkeit gleicher Winkel revidiert.

Aus obiger Gleichung folgt:

$$S = \sum_{n=1}^{\infty} P_n J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) \left( \sum_{n=1}^N P_n \cos(m\phi_n) \cos(m\phi) + \sum_{n=1}^N P_n \sin(m\phi_n) \sin(m\phi) \right) \quad (4.5)$$

Wenn man die Gleichungen 4.3 und 4.5 vergleicht, so kann man sehen, dass manche Terme einander entsprechen. Es ist die Bedingung gesucht, für die das

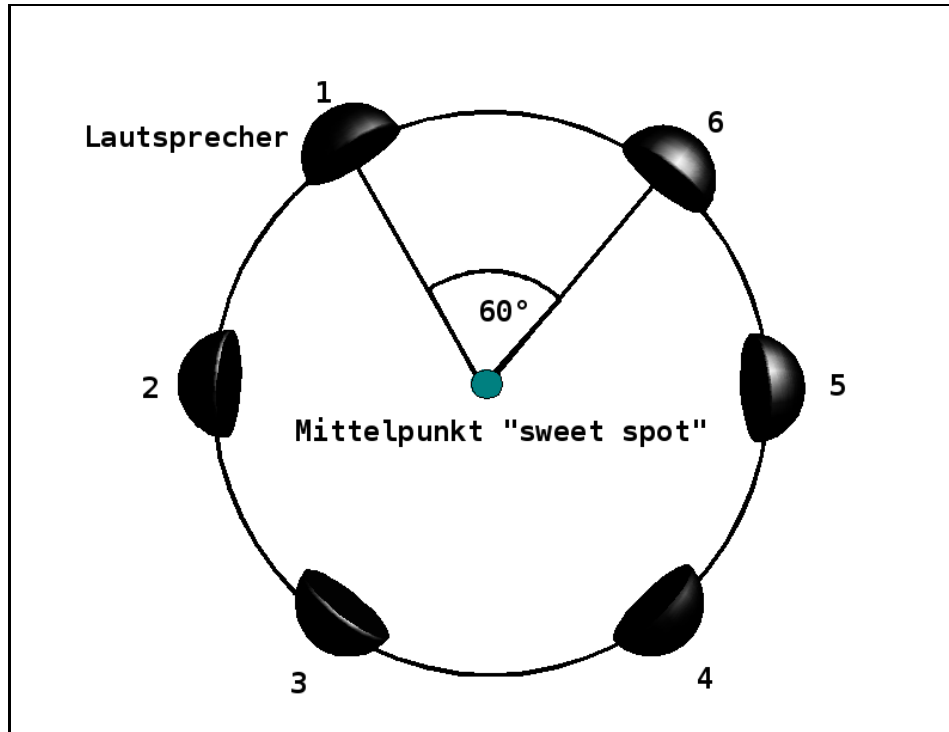


Abbildung 4.2: Ein typischer Lautsprecheraufbau für zweidimensionales Ambisonic

erzeugte Signal dem Original entspricht.  
Das ist der Fall, wenn:

$$P_\psi = \sum_{n=1}^N P_n \quad (4.6)$$

$$P_\psi \cos(m\psi) = \sum_{n=1}^N P_n \cos(m\phi_n) \quad (4.7)$$

$$P_\psi \sin(m\psi) = \sum_{n=1}^N P_n \sin(m\phi_n) \quad (4.8)$$

Die Gleichungen 4.6 bis 4.8 enthalten die Bedingungen, für die die ursprüngliche ebene Welle mit der erzeugten übereinstimmt.

Die Ordnung  $m$  meint die Kugelflächenfunktion  $m$ -ter Ordnung, die abgeglichen werden muss.

Wenn eine unendliche Anzahl von Lautsprechern (Kanälen) zur Verfügung stände, wäre die reproduzierte Welle genau die Ursprüngliche. (Die Anordnung der Lautsprecher wäre ein Kreis.) Dann würde ein Lautsprecher genau das direkte Signal wiedergeben und die übrigen die Echos (Hall) des Raums, in dem die Aufnahme gemacht wurde. Da eine unendliche Zahl von Kanälen und damit Lautsprechern aber (leider) nicht realisierbar ist, muss die Reihe nach endlicher

Zahl der Glieder abgebrochen werden. Hier betrachten wir lediglich Ambisonic Systeme erster und zweiter Ordnung. Da für eigentlich alle Anwendungen Systeme 2ter Ordnung mehr als ausreichend sind, ist diese Einschränkung akzeptabel. Die Abgleichbedingungen beschreiben nun das Gleichungssystem, das genutzt wird, um die entsprechenden Informationen für ein bestimmtes Ambisonic-Signal zu generieren. Damit ein System 0-ter Ordnung „passt“, muss lediglich die geringste Ordnung stimmen, was einer Übereinstimmung bezüglich des Schalldrucks der Wellen mit dem ursprünglichen Signal entspricht, was bedeutet, dass das Original und das Reproduzierte die gleiche Amplitude haben.

Gleichung 4.6 gibt den Schalldruck der ebenen Welle wieder, welcher in Ambisonic mit  $W$  bezeichnet wird. Die nächst höhere Ordnung ( $m=1$ ) sorgt dafür, dass beide, die reproduzierte und die originale Welle die gleiche Richtung besitzen. Dafür sorgen die Gleichungen 4.7 und 4.8, mit  $X = P_\psi \cos \psi$  und  $Y = P_\psi \sin \psi$ . Diese drei Signale umfassen ein Ambisonic System 1ter Ordnung. Die nächst höhere Ordnung ( $m=2$ ) gibt mehr Informationen über die Welle wieder. Sie enthält Informationen über die Krümmung der Welle.

Für  $m=2$  lauten diese Gleichungen, 4.7 und 4.8 wie folgt:

$$U = P_\psi \cos 2\psi \text{ und}$$

$$V = P_\psi \sin 2\psi$$

Ein System 2ter Ordnung besteht also aus folgenden fünf Signalen:  $W$ ,  $X$ ,  $Y$ ,  $U$  und  $V$ .

Enthält ein System diese fünf Signale, so kann mit der obigen Lautsprecheranordnung eine ebene Welle im Mittelpunkt des Zuhörerraums, korrekt bis zur 2-ten Ordnung, reproduziert werden.

Um in einem Ambisonic System eine ebene Welle bis zu einer gewissen Ordnung korrekt wiederzugeben, müssen die Gleichungen 4.7 und 4.8 erfüllt werden.

Die Frage bleibt, wie gross ist  $P_n$  für den  $n$ -ten Lautsprecher?

Gerzon fand folgende Relation für ein System erster Ordnung:

$$P_n = \frac{1}{N} (W + 2X \cos \phi_n + 2Y \sin \phi_n) \quad (4.9)$$

wobei  $\frac{1}{N}$  der Proportionalitätsfaktor ist. Man kann zeigen, dass die Gleichung 4.9 die Abgleichbedingungen der Gleichungen 4.6 bis 4.8 erfüllt.

Für ein System 2-ter Ordnung müssen zwei Terme ergänzt werden, damit die Bedingungen weiter erfüllt werden <sup>2</sup>:

$$P_n = \frac{1}{N} (W + 2X \cos \phi_n + 2Y \sin \phi_n + 2U \cos 2\phi_n + 2V \sin 2\phi_n) \quad (4.10)$$

Soweit zum theoretischen Teil. Um die praktische Anwendung geht es in den folgenden Abschnitten.

### 4.3 Kodierung

Beim Enkodieren werden Winkel und Höhenwinkel-Koordinaten mithilfe der Enkodiergleichungen 4.1 einer beliebigen Mono-Quelle (Tonspur/ Sprache etc.) zugeordnet. Das heisst, man legt ein bestimmtes Bewegungsprofil für die Tonspur fest und ordnet dieser damit zu jeder Zeit dementsprechende Koordinaten

<sup>2</sup><http://audiolab.uwaterloo.ca/~jefeb/thesis/node12.html>

zu. Das Ergebnis sind dann 4 Audio Spuren (W, X, Y, Z) in Ambisonic erster Ordnung oder 9 Audio Spuren (W, X, Y, Z, R, S, T, U, V) im Ambisonic Format zweiter Ordnung.

W ist omnidirektional (ungerichtet) und enthält lediglich den Schalldruckpegel des Signals, während die anderen vektorielle, d.h. gerichtete Grössen sind, die die Richtungen im Kartesischen Raum repräsentieren. X zum Beispiel enthält nur die Komponente des Klages, die in X-Richtung zeigt.

Es wird das volle Signal (Amplitude =1) für eine Tonquelle genau vor oder hinter (Amplitude =-1) dem Zuhörer liegend erhalten (also genau auf der X-Achse), aber kein Signal (Amplitude =0) von einem Signal das genau auf der Y-Achse liegt, also rechts oder links des Zuhörers ( $x = signal * \cos(A)\cos(E) = signal * \cos(90^\circ)\cos(0^\circ) = 0$ ).

Alle Töne zwischen den Hauptrichtungen der Achsen werden in Bezug auf ihre Richtungsenergie aufgeteilt, ihrer Position entsprechend. So werden die räumlichen Richtungen enkodiert.

Unten aufgeführt sind die Enkodier- Gleichungen für Ambisonic zweiter Ordnung von Richard Furse und Dave Malham (FMH -set of encoding equations). FMH -Sammlung der Enkodier-Gleichungen:

Name			Polar Koordinaten	Kartesische Koordinaten
W	=	Signal *	0.707107	0.707107
X	=	Signal *	$\cos(A)\cos(E)$	x
Y	=	Signal *	$\sin(A)\cos(E)$	y
Z	=	Signal *	$\sin(E)$	z
R	=	Signal *	$1.5\sin(E)\sin(E)-0.5$	$1.5zz-0.5$
S	=	Signal *	$\cos(A)\sin(2E)$	2zx
T	=	Signal *	$\sin(A)\sin(2E)$	2yz
U	=	Signal *	$\cos(2A)\cos(E)\cos(E)$	xx-yy
V	=	Signal *	$\sin(2A)\cos(E)\cos(E)$	2xy

Tabelle 4.1: FMH-Sammlung der Enkodier-Gleichungen

## 4.4 Dekodierung

Zur Wiedergabe wird ein Dekoder benötigt, der aus dem B-Format die Signale für die einzelnen Lautsprecher errechnet. Der grosse Vorteil wie oben erwähnt ist bei Ambisonic, dass es sich auf praktisch jede beliebige (punktsymmetrische) Lautsprecheranordnung dekodieren lässt, wobei die Qualität der räumlichen Wiedergabe besser wird, je mehr Lautsprecher zur Verfügung stehen. Dies ermöglicht die Anpassung an die gegebenen räumlichen Begrenzungen. Der Dekoder benötigt nur die Aufstellung des Lautsprechersystems. Jedem Lautsprecher wird der genaue Anteil aus jeder räumlich enkodierten Richtung zugeordnet, seiner Position im Raum entsprechend. Das omnidirektionale Signal W liegt an jedem Lautsprecher an. Die übrigen Signale werden entsprechend der Position der Lautsprecher mit Faktoren multipliziert. Diese Faktoren sind für einige Lautsprecheraufstellungen bereits hergeleitet. Sie werden derart bestimmt, dass die resultierenden Signale dem Amplitudenmodell von Gerzon entsprechen<sup>3</sup>.

<sup>3</sup>[www.sonicarchitecture.de](http://www.sonicarchitecture.de)

## 4.5 Probleme mit asymmetrischen Anordnungen

Das Problem bei asymmetrischen Anordnungen, d.h. nicht äquidistant auf einem Kreis um das Zentrum verteilten Lautsprechern ist, dass Probleme bei der Dekodierung entstehen.

Im folgenden werden die Probleme bei herkömmlichen 5.1-Anlagen betrachtet, die Ansätze können aber auch auf alle anderen asymmetrischen Anordnungen übertragen werden.

Um Decoder überhaupt vergleichen zu können, müssen zunächst Kriterien festgelegt werden, die deren Qualität beschreiben. Gerzon nahm dazu Energie- und Amplitudenvektoren des Klangfeldes (vgl. Glg.(4.11) und (4.12)).

Die Gleichungen (4.13) bis (4.16) bestimmen die jeweiligen Anteile in X- bzw. Y-Richtung. Die Vektorlänge entspricht dabei der Qualität der Lokalisation, der Winkel des Vektors entspricht dabei dem Winkel, aus dem das Signal kommt. Ein Vektor der Länge 1 entspricht dabei dem besten Lokalisationseffekt.

$$P = \sum_{j=1}^n g_j \quad (4.11)$$

$$E = \sum_{j=1}^n g_j^2 \quad (4.12)$$

$$P_x = \sum_{j=1}^n g_j \frac{\cos(\Theta_j)}{P} \quad (4.13)$$

$$E_x = \sum_{j=1}^n g_j^2 \frac{\cos(\Theta_j)}{E} \quad (4.14)$$

$$P_y = \sum_{j=1}^n g_j \frac{\sin(\Theta_j)}{P} \quad (4.15)$$

$$E_y = \sum_{j=1}^n g_j^2 \frac{\sin(\Theta_j)}{E} \quad (4.16)$$

Für reguläre Lautsprecheranordnungen entspricht der wiedergegebene Winkel immer dem kodierten Winkel (wenn das Muster des virtuellen Microphons für alle Lautsprecher identisch ist), die Qualität der Lokalisation kann lediglich durch Phasenverschiebungen beeinträchtigt werden.

Bei asymmetrischen Aufbauten müssen nicht nur die Vektorlängen in Betracht gezogen werden, sondern auch der Abspielwinkel und der gesamte Schallpegel (Amplitude) aller Signale. Anders kommt es zu Verzerrungen und unscharfen Klangbildern.

Betrachtet man z.B. eine ITU 5.1-Aufstellung, so kann man sich vorstellen, das bei gleichen Polaritäten der Lautsprecher ein Klang, aus der forderen Hälfte kommend, lauter erscheinen wird, als einer von hinten. Einfach dadurch, dass im forderen Halbkreis drei und im hinteren lediglich 2 Lautsprecher stehen. Es gibt also ein Ungleichgewicht in der X-Achse, das behoben werden muss. (In Y-Richtung ist die Anordnung symmetrisch, re-li.)

Da wir hier nur ein horizontales System betrachten, müssen wir bei Ambisonic

1-ter Ordnung nur die Signale W,X, und Y beachten:

$$W = \frac{1}{\sqrt{2}} \quad (4.17)$$

$$X = \cos(\Theta) \quad (4.18)$$

$$Y = \sin(\Theta) \quad (4.19)$$

Die Einbeziehung eines Korrekturfaktors für ein Ungleichgewicht in der X-Richtung könnte dann so aussehen:

$$W' = 0.5(\lambda + \lambda^{-1})W + 8^{-\frac{1}{2}}(\lambda - \lambda^{-1})X \quad (4.20)$$

$$X' = 0.5(\lambda + \lambda^{-1})X + 2^{-\frac{1}{2}}(\lambda - \lambda^{-1})W \quad (4.21)$$

$$Y' = Y \quad (4.22)$$

Dabei ist  $2 > \lambda > 1$  für ein System mit Front-Dominanz und  $0 < \lambda < 1$  für eines mit Dominanz hinter dem Hörzentrum. Man sieht, dass die Y-Komponente nicht beeinflusst wird.

Nach dem Einsetzen in die normalen Gleichungen erhält man dann die korrigierten Versionen für die einzelnen Kanäle:

$$C_F = (kW_C * W') + (kX_C * X') \quad (4.23)$$

$$L_F = (kW_F * W') + (kX_F * X') + (kY_F * Y') \quad (4.24)$$

$$R_F = (kW_F * W') + (kX_F * X') - (kY_F * Y') \quad (4.25)$$

$$L_B = (kW_B * W') + (kX_B * X') + (kY_B * Y') \quad (4.26)$$

$$R_B = (kW_B * W') + (kX_B * X') - (kY_B * Y') \quad (4.27)$$

$\lambda$  und  $k$  werden dabei so gewählt, dass das beste Ergebnis für die Ausgabe erreicht wird, mit  $0 < k < 1$ . Um nun die Güte eines Decoders zu berechnen, müssen folgende Bedingungen betrachtet und möglichst exakt eingehalten werden:

- Die Vektoren  $R_E$  und  $R_V$  sollten so nah wie möglich bei „1“ liegen ( $R_E \approx R_V \approx 1$ ).
- $\Theta_P = \Theta_E = \Theta$  für alle  $\Theta$ , wobei  $\Theta$  der kodierte Winkel ist,  $\Theta_P$  und  $\Theta_E$  die entsprechenden virtuellen Winkel.
- $P, E = const$  für alle Winkel  $\Theta$

$$P_x = \sum_{i=1}^N g_i * \cos(SPos_i) \frac{1}{P} P_y = \sum_{i=1}^N g_i * \sin(SPos_i) \frac{1}{P} \quad (4.28)$$

$$E_x = \sum_{i=1}^N g_i^2 * \cos(SPos_i) \frac{1}{E} E_y = \sum_{i=1}^N g_i^2 * \cos(SPos_i) \frac{1}{E} \quad (4.29)$$

$$R_E = \sqrt{E_x^2 + E_y^2} R_V = \sqrt{V_x^2 + V_y^2} \quad (4.30)$$

$$P = \sum_{i=1}^n g_i E = \sum_{i=1}^n g_i^2 \quad (4.31)$$

$$\Theta_E = \tan^{-1} \left( \frac{E_y}{E_x} \right) \quad \Theta_P = \tan^{-1} \left( \frac{E_y}{E_x} \right) \quad (4.32)$$

Für Systeme mit grosser Anzahl von Lautsprechern und beliebigen Aufstellungsformen sind alle Bedingungen gleichzeitig schwer zu erfüllen. Als Ergebnisse erhält man mehrere Koeffizientenmengen, aus denen dann mithilfe von Hörtests die Besten bestimmt werden müssen. Die Bestimmung der Koeffizientenmengen kann mit geeigneten numerischen Verfahren erfolgen.



# Kapitel 5

## Versuchsaufbau

Der Versuchsaufbau am Lehrstuhl besteht aus folgenden Einheiten:

- Verstärker: Yamaha, DSP-Z9 Av
- 7.1 Lautsprechersystem von Nubert der Serie „Nuline“
- PC mit RME HDSP9632 Soundkarte
- beweglicher (motorgesteuerter) Kunstkopf (Neumann) mit zwei Mikrofonen

Die Verbindungen zwischen den einzelnen Komponenten für die Wiedergabe der Ambisonic-Dateien ist in folgender Tabelle zu sehen (JACK Audio Connection Kit, siehe Kapitel 6.1 „Systemvoraussetzungen“):

<b>Lautsprecher</b>	<b>Eingang Verstärker</b>	<b>Jack-Ausgang</b>	<b>Ausgang Soundkarte</b>
VORNE LINKS	MCI Front L	PLAYBACK_3	2. v. links
MITTE LINKS	MCI Surround L	PLAYBACK_1	1. v. links
HINTEN LINKS	MCI SurroundBack L	PLAYBACK_5	2. v. rechts
HINTEN RECHTS	MCI SurroundBack R	PLAYBACK_6	2. v. rechts
MITTE RECHTS	MCI Surround	PLAYBACK_2	1. v. links
VORNE RECHTS	MCI Front R	PLAYBACK_4	2. v. links
CENTER	MCI Center	PLAYBACK_7	1. v. rechts
SUBWOOFER	MCI Subwoofer	PLAYBACK_8	1. v. rechts

Tabelle 5.1: Verbindung Lautsprecher – Verstärker – Jack-Ausgang – Soundkarten-ausgang

MCI (Multiple Channel Input) ist der Modus, in dem der Verstärker betrieben wird. Das heisst, dass jedes analoge Signal einzeln in den Verstärker gegeben wird. Der PC und der Verstärker befinden sich in einem separaten Raum, von dem aus man durch ein Fenster die Lautsprecheranlage und den Kunstkopf beobachten kann. Dies hat den Vorteil, dass keine Störgeräusche die Versuchsergebnisse mit den Kunstkopfmikrofonen beeinträchtigen. Man kann die Versuchsanordnung zusätzlich mit einem schallabsorbierenden Vorhang gegen Störgeräusche abschirmen (siehe Bild unten). Im Hintergrund sind die Lautsprecher zu sehen und im Bildvordergrund der Kunstkopf „Bob“. In der Metallbox darunter befinden sich die geräuscharmen Schrittmotoren, die den Kunstkopf bewegen sowie die zugehörige Elektronik.

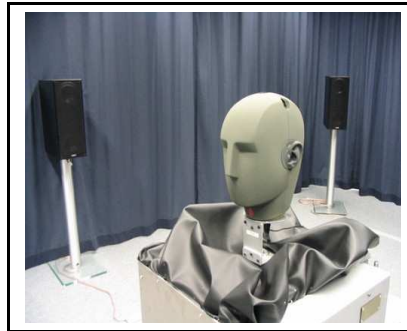


Abbildung 5.1: Versuchsaufbau am Lehrstuhl für Nachrichtentechnik

# Kapitel 6

## Das Programm *ambimove*

### 6.1 Systemvoraussetzungen

#### Hardware:

- Pc mit genügend grosser Rechenleistung (128 MB RAM) und Speicherplatz (1 GB ) zur Abspeicherung der Klangbeispiele müssen jeweils mind. 6 Wav-Dateien mit einer Grösse (abhängig von der Anzahl der Quellen und dem Bewegungsprofil) von ca. 4MB pro Minute abgespeichert werden. Für eine halbe Minute also etwa 12 MB.
- Soundkarte mit mind. 6 analogen Ausgängen zur Übertragung der einzelnen Kanäle zum Verstärker
- Verstärker mit mind. 6 analogen Ein- und Ausgängen zur Verstärkung und Ansteuerung der einzelnen Lautsprecher

#### Software:

- Linux Betriebssystem (Suse, Gentoo o.a.)
- AMBIMOVE zur Erzeugung der ambisonic-wav-Dateien
- ECASOUND zur Wiedergabe der wav-Dateien
- JACK Audio Connection Kit. Soundserver als Schnittstelle zwischen ECASOUND und der Soundkarte.
- MCOUT übernimmt die Verknüpfung der ECASOUND-Ausgangssignale mit den Ausgängen der Soundkarte in JACK (auch von Hand möglich).

### 6.2 Aufbau

Grundlage des Programms sind die csound ambisonic-Skripte von Jan Jacob Hoffmann, die dieser auf seiner Homepage [www.sonicarchitecture.de](http://www.sonicarchitecture.de) zur Verfügung stellt. Diese beinhalten die bereits erwähnten FMH-Enkodiergleichungen. Das Programm Ambimove ist ein Kommandozeilenprogramm das zur Eingabe der zur Berechnung der Ambisonic-Dateien erforderlichen Daten dient. Es ist in der

Sprache C geschrieben. Mit den abgefragten Daten werden dann insgesamt vier Textdateien erzeugt, zwei orchestra-files und zwei score-files. Diese dienen später als Eingabe für das Programm *csound* welches die eigentlichen Berechnungen durchführt und von *ambimove* aufgerufen wird. *csound* liefert als Ausgabe einmal die kodierten ambisonic Dateien (w,x,y,z,u,v,r,s,t) und im zweiten Schritt direkt die wav-files für die entsprechenden Kanäle. D.h. *ambimove* ruft *csound* insgesamt zwei mal auf.

### 6.2.1 Kurzer Ausflug in *csound*

Dieser Abschnitt ist als eine kurze Einleitung in *csound* gedacht und zum groben Überblick, was zum Verständnis der Studienarbeit beitragen soll. Für detaillierte Informationen bietet das folgende Tutorial <http://www.csounds.com/toots/index.html> einen guten Einstieg und die gesamte Homepage [www.csounds.com](http://www.csounds.com), sowie entsprechende Fachliteratur.

Die Geschichte von *csound* beginnt in den frühen 60ern. Die Bell Labs entwickeln eine Programmiersprache namens Music4, den Vorgänger von *csound*. Es ist frei und für sehr viele Plattformen erhältlich. Wahrscheinlich ist *csound* gerade deshalb mittlerweile eine der meist verbreitetsten Programmiersprachen für die Entwicklung von Musikstrukturen und Klangdesign. Obwohl ursprünglich zum Erstellen von synthetischen Klängen gedacht, lassen sich damit aber auch ganze Symphonien komponieren. Wie funktioniert *csound*?

Wie bei einer Programmiersprache gibt der „Komponist“ seine Komposition als Source Code ein, der mit Hilfe des Programms *Csound* anschliessend kompiliert wird. Am Ende dieser Kompilierung steht eine Klangdatei, z.B. im .wav oder .aif Format. Genauer gesagt besteht der Source Code aus 2 Textdateien: Einer Orchestra-File (.orc) und einer Score-File (.sco). Eine Kommandozeile zur Ausführung einer Kompilierung könnte dann z.B. so aussehen: „*csound* beispiel.orc beispiel.sco“

Hierbei stellt „*csound*“ den Programmaufruf dar. Die ausgegebene Sound-File heisst per default „test“ (ohne Endung), eine wav-Datei. Natürlich lassen sich über Kommandozeilenparameter bestimmte Kompilierungsoptionen einstellen, z.B. Skip-Zeiten, Ausgabeformate etc. Diese werden beim Eingeben von „*csound*“ ohne die .orc und .sco Files aufgelistet.

In den .orc-Dateien befinden sich die Instrumente und in den .sco-Dateien die „Noten“. Wobei man „Instrumente“ und „Noten“ nicht zu wörtlich nehmen sollte. Mit Instrumenten sind mit „i#“ (# steht für eine Zahl) gekennzeichnete Teilabschnitte gemeint, die Variablen enthalten. Diesen Variablen werden dann mithilfe der .sco-Dateien Werte zugewiesen.

Hier ein Beispiel:

```
/* .orc-file*/
sr = 44100 ; Signalrate
kr = 4410 ; Kontrollsignalrate
ksmps = 10 ; Anzahl der Samples pro Kontrollperiode ksmps = sr/kr
nchnls = 1 ; Anzahl der Ausgabekanäle
instr 1
a1 oscil 10000, 440, 1 ; der Variablen a1 wird der Wert der Funktion oscil gegeben
out a1 ; die Ausgabe ist eine mono wav-Datei (nchnls=1) mit den Werten von a1
endin ; markiert das Ende der .orc-File
```

```

/* .sco-file*/
f1 0 4096 10 1 ; nutzt die GEN10 um eine sinus-Welle zu erzeugen
;instr. start dauer
i1 0 4 ; beginnend beim Zeitpunkt 0 wird 4 Sekunden lang i1 \ "gespielt\ "
e ; markiert das Ende der .sco-File

```

Dieses Beispiel erzeugt einen Sinus-Ton von 4s mit einer Frequenz von 440Hz und einer Amplitude von 10000. Der Header (die ersten 4 Zeilen) gibt die Signalrate, die Kontrollrate, die Samples pro Kontrollrate und die Anzahl der Ausgabekanäle an. „oscil“ wie auch „out“ sind vordefinierte Funktionen. Oscil generiert periodische Kontrollsignale bestehend aus dem Wert der Amplitude (hier 10000) mal dem Wert, den die functiontable zu diesem Zeitpunkt enthält. Die Funktion out gibt ein mono-Signal aus, hier a1. „f1“ im Score bezeichnet eine s.g. „ftable“ (Functiontable) mit der Generator-Nummer 10 (GEN10). Diese Tabelle enthält Werte einer Sinus-Schwingung. Das erste Argument „0“ bezeichnet den Beginn der Gültigkeit der Wertetabelle, die Zahl „4096“ gibt die Auflösung der Tabelle an (die Zahl muss eine 2er-Potenz oder 2er-Potenz+1 sein) 10 ist die Generator-number, welche die Funktion bestimmt, die in der Tabelle gespeichert wird. Die folgende „1“ besagt, dass lediglich die erste Harmonische gespeichert wird. Damit sollte die Vorgehensweise bei der Programmierung eines csound Skriptes klar geworden sein. Natürlich gibt es fast unendlich viele Funktionen, die hier nicht alle vorgestellt werden können aber generell verfolgen alle Skripte diesen Aufbau.

### 6.3 Bedienung und Einstellungen

Die Bedienung des Programms erfolgt über Kommandozeile (Shell). Die erste Abfrage dient der Benennung des Projekts, d.h. welchen Namen das Klangstück bekommen soll:

Bitte geben sie den Namen fuer das Projekt an:

Nach der Eingabe wird überprüft, ob der Name bereits vergeben ist und gegebenenfalls eine Abfrage ausgegeben, ob die bereits vorhandene Datei überschrieben werden soll:

Projekt bereits vorhanden, ueberschreiben?(y/n)

Entsprechendes eingeben. **Achtung!** Wird die Datei überschrieben, geht sie unwiderruflich verloren. Die nächste Eingabe betrifft die Lautsprecheranordnung, die zum Abspielen der wav-Dateien verwendet wird. Man hat hier die Möglichkeit aus verschiedenen Anordnungen zu wählen, die kodiert bzw. dekodiert werden können.

Lautsprecheranordnung

2=Stereo

4=Quadrophon

5=Pentagon

51=5.1-Surround

6=Hexagon

7=7.1-Surround

81=Octagon1  
 82=Octagon2  
 8=Wuerfel(3d)  
 121=Dodecahedron  
 122=Dodecahedron2:

Der nächste Punkt betrifft das Einlesen der WAV-Dateien. Zuerst muss die Anzahl der WAV-Dateien angegeben werden, die eingelesen werden sollen. Hier muss eine Zahl zwischen 1 und 20 eingegeben werden.

Anzahl der Wavefiles, die eingelesen werden sollen:

(Die Beschränkung auf 20 wurde willkürlich festgelegt und könnte bei Bedarf erhöht werden.) Anschliessend muss für jede dieser WAV-File der Name (mit Endung) eingegeben werden.

Name der wavfile # (xxx.wav): ;# steht fuer die #-te wav-Datei

Die Dateien müssen im gleichen Verzeichnis liegen wie das Programm ambimove. Eine Fehlermeldung erscheint, wenn die Datei im entsprechenden Verzeichnis nicht gefunden wird.

Konnte Datei nicht finden!Bitte korrekte Datei angeben!

Ausserdem werden die Beginn-Zeiten jeder wavfile abgefragt, sowie deren Länge. Die Eingabe bestimmt, zu welcher Zeit Diese Angaben sind später wichtig, damit die Bewegungen nicht die Länge der WAV-Dateien überschreiten.

Start der wavfile # [s]:

Laenge der wavfile # [s]:

Nun folgt der Hauptteil des Programms. Für jede der oben angegebenen Dateien müssen nun Bewegungsprofile erstellt werden. Dies funktioniert nach folgendem Schema:

- eine Bewegungsform für die jeweilige Datei muss gewählt werden (1=Punkt,2=Gerade,3=Kreis)

Movement fuer wavfile # (1=Punkt,2=Gerade,3=Kreis):

- die genauen Parameter der Bewegung werden abgefragt  
ein Punkt:

Winkel [°]:  
 Distanz zum Mittelpunkt [Einheiten]:  
 Dauer der Bewegung [s]:

eine Gerade:

Anfangs-Winkel [°]:  
 Anfangs-Distanz [Einheiten]:  
 End-Winkel [°]:  
 End-Distanz [Einheiten]:  
 Dauer der Bewegung [s]:

ein Kreis(-segment):

```
Anfangswinkel des Kreises [°]:
Radius des Kreises [Einheiten]:
Endwinkel des Kreises [°]:
Dauer der Bewegung [s]:
```

- ist die bis dahin verwendete Zeit geringer als die Länge der WAV-File, so wird nach einer weiteren Bewegung gefragt

```
Es sind noch n sec Zeit, eine weitere Bewegung anhaengen?(y/n)
;n steht fuer eine ganze, pos. Zahl
```

Dieser Zyklus wird so lange durchschritten, bis entweder die gesamte Zeit aller Dateien mit Bewegungen gefüllt ist, oder bis man die gewünschten Bewegungen für alle wav-Dateien eingegeben hat. Eine Fehlermeldung erscheint dann, wenn die für eine Bewegung in Anspruch genommene Zeit so gross ist, dass die Länge der wav-Datei überschritten würde. Man wird gebeten, eine kürzere Angabe zu machen.

```
Die maximale Dauer betraegt n sec! ;n steht fuer eine ganze, pos. Zahl
```

Wenn dieser Teil abgeschlossen ist, beginnt das Programm mit dem Aufruf von csound entsprechend der im Programm gemachten Angaben. Eine Korrektur ist jetzt nicht mehr möglich!

## 6.4 Die Kodierungsparameter

Im folgenden Auszug aus dem Quellcode sind die einzelnen Kodierungsparameter. Da es sich um Ambisonic 1.Ordnung und einen 2-dimensionalen Versuchsaufbau handelt sind lediglich die Signale W,X und Y von Bedeutung. Diese werden entsprechend mit aw, ax und ay bezeichnet. Die Parameterwerte entsprechen denen für eine 5.1 Anordnung, wobei die beiden Mitte-Signale mithilfe von Hörtests entsprechend angepasst wurden. Da diese beiden Lautsprecher genau auf der Y-Achse liegen, muss die X-Komponente 0 sein. Entsprechend beim center-Lautsprecher die Y-Komponente. In den Kommentarzeilen stehen die Bezeichnungen der Lautsprecher sowie deren Koordinaten bezüglich des in Ambisonic verwendeten Koordinatensystems in kartesischen Koordinaten (Z-Komponente immer 0).

```
;Speaker 1 <0.8660,0.5000,0.0000>(Vorne links)
  achnl1 sum aw *0.1690,ax *0.0797,ay * 0.0891
; Speaker 2 <1.0000,0.0000,0.0000>(center)
  achnl2 sum aw *0.1635,ax * 0.0923
; Speaker 3 <0.8660,-0.5000,0.0000>(Vorne rechts)
  achnl3 sum aw *0.1690,ax *0.0797,ay * -0.0891
; Speaker 4 <0.0000,1.0000,0.0000>(Mitte links)
  achnl4 sum ay * 0.15000
; Speaker 5 <0.0000,-1.0000,0.0000>(Mitte rechts)
  achnl5 sum ay * -0.15000
```

```
; Speaker 6 <-0.5000,0.8660,0.0000>(Hinten links)
achnl6  sum aw *0.4563,ax *-0.1259,ay * 0.1543
; Speaker 7 <-0.5000,-0.8660,0.0000>(Hinten rechts)
achnl7  sum aw *0.4563,ax *-0.1259,ay * -0.1543
```

Die Verwendung dieser Parameter ergab eine gute Lokalisationsschärfe. Allerdings gibt es keinerlei theoretische Herleitung, sondern die Werte basieren alleine auf Hörtests.



## 6.5 Beispiele

Im folgenden werden jeweils ein Beispiel für die Erstellung von virtuellen Punktquellen, Raumgeraden und Kreisfiguren gegeben. Es wird eine Beispieldatei *noise.wav* eingelesen, es ist aber auch jede andere wav-Datei möglich. Es ist darauf zu achten, dass die Länge der wav-Datei mindestens der Länge des Bewegungsprofils entspricht, da sonst nicht alle Bewegungen ausgeführt werden können. Eine Entfernung von „1“ entspricht dem ursprünglichen Signal. Je grösser die Entfernung  $d$  wird, desto leiser wird der Ton (geringere Amplitudenwert) durch Multiplikation mit dem Faktor  $\frac{1}{d}$ .

### 6.5.1 Punktquellen

Beispielcode eines Programmablaufs mit einer Folge von virtuellen **Punktquellen** verteilt auf dem Kreis um den Mittelpunkt. Die Entfernung der Punktquellen wird konstant gehalten (5).

```
*****ambimove1.0*****
*****Tobias Jung 2006*****
Bitte geben sie den Namen fuer das Projekt an: test
Projekt bereits vorhanden, ueberschreiben?(y/n) y
Lautsprecheranordnung (2=Stereo
4=Quadrophon
5=Pentagon
51=5.1-Surround
6=Hexagon
7=7.1-Surround
81=Octagon1
82=Octagon2
8=Wuerfel(3d)
121=Dodecahedron
122=Dodecahedron2): 7
Anzahl der Wavefiles, die eingelesen werden sollen: 1
Name der wavfile 1 (xxx.wav): noise.wav
Start der wavfile 1 [s]: 0
Laenge der wavfile 1 [s]: 30
Movement fuer wavfile 1 (1=Punkt,2=Gerade,3=Kreis): 1
Winkel [ ]: 45
Distanz zum Mittelpunkt [Einheiten]: 5
Dauer der Bewegung [s]: 5
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y
Movement fuer wavfile 1 (1=Punkt,2=Gerade,3=Kreis): 1
Winkel [ ]: 225
Distanz zum Mittelpunkt [Einheiten]: 5
Dauer der Bewegung [s]: 5
Es sind noch 20s zeit, eine weitere Bewegung anhaengen? y
Movement fuer wavfile 1 (1=Punkt,2=Gerade,3=Kreis): 1
Winkel [ ]: 135
Distanz zum Mittelpunkt [Einheiten]: 5
Dauer der Bewegung [s]: 5
```

```

Es sind noch 15s zeit, eine weitere Bewegung anhaengen? y
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 1
Winkel [ ]: 315
Distanz zum Mittelpunkt [Einheiten]: 5
Dauer der Bewegung [s]: 5
Es sind noch 10s zeit, eine weitere Bewegung anhaengen? y
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 1
Winkel [ ]: 270
Distanz zum Mittelpunkt [Einheiten]: 5
Dauer der Bewegung [s]: 5
Es sind noch 5s zeit, eine weitere Bewegung anhaengen? y
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 1
Winkel [ ]: 90
Distanz zum Mittelpunkt [Einheiten]: 5
Dauer der Bewegung [s]: 5
****Aufruf von csound****

```

## 6.5.2 Geraden

Die Virtuelle Quelle Bewegt sich auf Geraden durch den Mittelpunkt. Es wirkt, als ob sie sich über den Zuhörer hinwegbewegen. Geraden werden durch ihren Anfangspunkt (Anfangswinkel und -entfernung) und ihren Endpunkt (Endwinkel und -entfernung) definiert.

```

*****ambimove1.0*****
*****Tobias Jung 2006*****
Bitte geben sie den Namen fuer das Projekt an: test
Projekt bereits vorhanden, ueberschreiben?(y/n) y
Lautsprecheranordnung (2=Stereo
4=Quadrophon
5=Pentagon
51=5.1-Surround
6=Hexagon
7=7.1-Surround
81=Octagon1
82=Octagon2
8=Wuerfel(3d)
121=Dodecahedron
122=Dodecahedron2): 7
Anzahl der Wavefiles, die eingelesen werden sollen: 1
Name der wavefile 1 (xxx.wav): noise.wav
Start der wavefile 1 [s]: 0
Laenge der wavefile 1 [s]: 30
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 2
Anfangswinkel [ ]: 45
Distanz zum Mittelpunkt [Einheiten]: 15
Endwinkel [ ]: 225
Distanz zum Mittelpunkt [Einheiten]: 15
Dauer der Bewegung [s]: 5
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y

```

```

Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 2
Anfangswinkel [ ]: 135
Distanz zum Mittelpunkt [Einheiten]: 15
Endwinkel [ ]: 315
Distanz zum Mittelpunkt [Einheiten]: 15
Dauer der Bewegung [s]: 5
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 2
Anfangswinkel [ ]: 270
Distanz zum Mittelpunkt [Einheiten]: 15
Endwinkel [ ]: 90
Distanz zum Mittelpunkt [Einheiten]: 15
Dauer der Bewegung [s]: 5
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 2
Anfangswinkel [ ]: 0
Distanz zum Mittelpunkt [Einheiten]: 15
Endwinkel [ ]: 180
Distanz zum Mittelpunkt [Einheiten]: 15
Dauer der Bewegung [s]: 5
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 2
Anfangswinkel [ ]: 225
Distanz zum Mittelpunkt [Einheiten]: 15
Endwinkel [ ]: 45
Distanz zum Mittelpunkt [Einheiten]: 15
Dauer der Bewegung [s]: 5
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 2
Anfangswinkel [ ]: 315
Distanz zum Mittelpunkt [Einheiten]: 15
Endwinkel [ ]: 135
Distanz zum Mittelpunkt [Einheiten]: 15
Dauer der Bewegung [s]: 5
****Aufruf von csound****

```

### 6.5.3 Kreise

Im folgenden Beispiel beschreibt die virtuelle Quelle Kreise rings um den Zuhörer. Dabei wird die Entfernung der Kreise immer grösser, die Lautstärke also immer geringer.

```

*****ambimove1.0*****
*****Tobias Jung 2006*****
Bitte geben sie den Namen fuer das Projekt an: test
Projekt bereits vorhanden, ueberschreiben?(y/n) y
Lautsprecheranordnung (2=Stereo
4=Quadrophon
5=Pentagon
51=5.1-Surround

```

6=Hexagon  
7=7.1-Surround  
81=Octagon1  
82=Octagon2  
8=Wuerfel(3d)  
121=Dodecahedron  
122=Dodecahedron2): 7  
Anzahl der Wavefiles, die eingelesen werden sollen: 1  
Name der wavefile 1 (xxx.wav): noise.wav  
Start der wavefile 1 [s]: 0  
Laenge der wavefile 1 [s]: 30  
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 3  
Anfangswinkel [ ]: 0  
Distanz zum Mittelpunkt [Einheiten]: 1  
Endwinkel [ ]: 360  
Distanz zum Mittelpunkt [Einheiten]: 3  
Dauer der Bewegung [s]: 5  
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y  
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 3  
Anfangswinkel [ ]: 0  
Distanz zum Mittelpunkt [Einheiten]: 3  
Endwinkel [ ]: 360  
Distanz zum Mittelpunkt [Einheiten]: 6  
Dauer der Bewegung [s]: 5  
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y  
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 3  
Anfangswinkel [ ]: 0  
Distanz zum Mittelpunkt [Einheiten]: 6  
Endwinkel [ ]: 360  
Distanz zum Mittelpunkt [Einheiten]: 12  
Dauer der Bewegung [s]: 5  
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y  
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 3  
Anfangswinkel [ ]: 0  
Distanz zum Mittelpunkt [Einheiten]: 12  
Endwinkel [ ]: 360  
Distanz zum Mittelpunkt [Einheiten]: 18  
Dauer der Bewegung [s]: 5  
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y  
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 3  
Anfangswinkel [ ]: 0  
Distanz zum Mittelpunkt [Einheiten]: 18  
Endwinkel [ ]: 360  
Distanz zum Mittelpunkt [Einheiten]: 25  
Dauer der Bewegung [s]: 5  
Es sind noch 25s zeit, eine weitere Bewegung anhaengen? y  
Movement fuer wavefile 1 (1=Punkt,2=Gerade,3=Kreis): 3  
Anfangswinkel [ ]: 0  
Distanz zum Mittelpunkt [Einheiten]: 25  
Endwinkel [ ]: 360

```
Distanz zum Mittelpunkt [Einheiten]: 35  
Dauer der Bewegung [s]: 5  
****Aufruf von csound****
```

## Kapitel 7

# Ausblick und Erweiterungen

Die Weiterführung des Programms *ambimove* sollte dahin gehen, dass man die Verwendung von beliebigen Lautsprecheranordnungen zulässt und sich nicht auf die beschränkt, für die bereits Dekodiergleichungen erster und zweiter Ordnung existieren.

Ziel sollte sein, die Eingabe von Koordinaten für jeden einzelnen Lautsprecher zu ermöglichen und dann auf Grundlage dieser Anordnung die Dekodierungskoeffizienten zu berechnen unter Berücksichtigung der im Kapitel 4.5 erwähnten Qualitätsmerkmale wie Energie- und Amplitudengleichgewicht.

# Literaturverzeichnis

- [1] „3D Akustik mit Lautsprechern“, Simon Fuast, VRCA Aachen
- [2] „Mehrkanalwiedergabetechniken“, Michael Strauß, HSOAT
- [3] „Dreidimensionale Klangprojektion“, Philippe Kocher, ICST (Institute for Computer Music and Sound Technology) Zürich
- [4] „The design and optimisation of surround sound decoders using heuristic methods“, Bruce Wiggins, I. Paterson-Stephens, Val Lowndes & S. Berry , University of Derby
- [5] „Stereo Microphone Techniques Explained“, SoS (SoundOn-Sound), Februar 1997