

# **Aufnahme – und Wiedergabesystem zur Berechnung eines dynamisch modifizierbaren binauralen Signalpaares**

## **Diplomarbeit**

durchgeführt von

**Stefan Leitner**

im Rahmen des interuniversitären Studiums Elektrotechnik – Toningenieur

Institut für Nachrichtentechnik und Wellenausbreitung  
an der Technischen Universität Graz  
Vorstand: o.Univ.-Prof.Dipl.-Ing.DDr.Willibald Riedler

und

Institut für Elektronische Musik  
an der Universität für Musik und darstellende Kunst in Graz  
Vorstand: o.Univ.-Prof.Mag.Dipl.-Ing.Dr.Robert Höldrich

Begutachter: o.Univ.-Prof.Dipl.-Ing.DDr.Willibald Riedler  
Betreuer: o.Univ.-Prof.Mag.Dipl.-Ing.Dr.Robert Höldrich

Graz, im März 2000

Für meine Eltern

## **Danksagung**

Herzlichen Dank Herrn Dr. Robert Höldrich für die Betreuung dieser Arbeit.

Herrn Thomas Musil vielen Dank für die Hilfe und Erklärungen bei der Implementation an der SGI Workstation.

## **Zusammenfassung**

In dieser Arbeit werden verschiedene Methoden der 3 – dimensionalen Schallreproduktion über Kopfhörer in Hinblick auf ein Aufnahme- und Wiedergabesystem betrachtet. Die grundlegenden Mechanismen des räumlichen Hörens bilden den Ausgangspunkt der Überlegungen. Von Wichtigkeit sind dabei die kleinen Kopfbewegungen des Hörers bei der Lokalisation eines Schallereignisses im natürlichen Schallfeld.

Die Methoden der Richtstrahlbildung, Ambisonic und Holophonie werden vorgestellt. Der Vorteil von Ambisonic liegt in der Möglichkeit ein binaurales Kopfhörersignal in Abhängigkeit von der Kopfposition recheneffizient zu erhalten.

## **Abstract**

This thesis deals with various methods of 3D – sound reproduction via headphones with the aim of a recording and reproduction system. Psychoacoustic basics and fundamental directional hearing mechanisms of the human auditory system are considered at first. A central point in this context is the importance of small head movements during localisation of a small sound source in natural hearing.

Firstly computing binaural signals with beamforming, secondly with Ambisonic and Holophony is examined.

The advantage of Ambisonic is the possibility to update the headphone signals according to head position in a computational efficient way. The holophonic sound pick up is a possibility to obtain the Ambisonic signals from a practical point of view.

# Inhaltsverzeichnis

1	Psychoakustische Grundlagen des räumlichen Hörens	1
1.1	Akustische Faktoren in der Beurteilung der Position einer Schallquelle	1
1.1.1	Monaurale Faktoren	2
1.1.2	Binaurale Faktoren	4
1.2	Wovon werden die Faktoren zur Schalllokalisierung beeinflusst?	7
1.2.1	Reliabilität oder Konsistenz	7
1.2.2	Der Einfluss der a priori Kenntnis der Quellcharakteristik	8
1.2.3	Frequenzinhalt des Quellspektrums	9
1.2.4	Einfluss der Plausibilität der Lokalisationsfaktoren	10
1.3	Zusätzliche Faktoren – Auflösung der Vorne/Hinten Ambiguität	11
1.4	Auswirkung von spontanen Kopfbewegungen auf das räumliche Hören	12
1.4.1	Rückblick	13
1.4.2	Zusammenhang von Kopfbewegungen und Hören	14
1.4.3	Beschreibung der Hörversuche von Mackensen, Reichenauer, Theile	16
2	Raumsimulation	20
2.1	Berechnung der Raumimpulsantwort mit Modellmethoden	21
2.2	Binaurale Raumsimulation	21
2.3	Schallfeldberechnung mittels geometrischer Raum-Akustik	22
3	Richtstrahlbildung – Räumliche Filterung mittels einer Sensorgruppe	24
3.1	Einführung	24
3.2	Grundlegende Begriffsdefinitionen	25
3.2.1	Beamforming (Richtstrahlbildung) und räumliche Filterung	25
3.2.2	Statistik Zweiter Ordnung	30
3.3	Klassifikation von Richtstrahlern	31
3.3.1	Klassische Richtstrahlbildung	32
3.3.2	Entwurf einer allgemeinen Frequenzantwort eines datenunabhängigen Richtstrahlers	33
3.4	Modellierung der Aussenohrübertragungsfunktion unter Verwendung des Prinzips der Richtstrahlbildung (Beamforming)	35
3.4.1	Einführung	35
3.4.2	Beamformerdesign durch Minimierung des mittleren quadratischen Fehlers (LMS Beamformerdesign)	36
3.4.3	Beschreibung der Experimente von Chen et al. zur HRTF – Modellierung durch Richtstrahlbildung	40
3.4.4	Implementation – Zerlegung in Teilbänder	47
4	Berechnung eines binauralen Stereosignals durch Anwendung der Ambisonic-Technik	49
4.1	Ambisonic – Grundlagen und Begriffsdefinition	49
4.2	Ambisonic-Signale für den 3 – dimensionalen Fall	52
4.3	Berechnung eines binauralen Stereosignals	53
4.4	Berechnung der Ambisonic-Signale W,X,Y,U,V	55

4.5	System zur Berechnung binauraler Stereosignale unter Berücksichtigung von Kopfbewegungen des Hörers	57
4.6	Berechnung der Ambisonic-Signale über den Ansatz der Holophonie	58
5	Verwendete Aussenohrübertragungsfunktionen	65
6	Literaturverzeichnis	70
7	Audio CD mit Beispielen verschieden berechneter Binauralsignale	72

# 1 Psychoakustische Grundlagen des räumlichen Hörens<sup>1</sup>

Die wahrgenommene Position einer Schallquelle im Raum wird von verschiedenen Faktoren beeinflusst. Dazu gehören Pegel- und Zeitunterschiede (ILD-Interaural Level Difference, ITD-Interaural Time Difference) sowie spektrale Verfärbung durch die Filterwirkung der Pinna (Ohrmuschel). Die relative Bedeutung dieser Komponenten hängt vom ‚a priori‘-Wissen des Hörers über die Quellcharakteristik, Frequenzinhalt der Quelle, Reliabilität und Plausibilität der Komponenten und Beschaffenheit der Komponenten über das Frequenzspektrum ab.

## 1.1 Akustische Faktoren in der Beurteilung der Position einer Schallquelle

Jeder physikalische Aspekt einer Schallwelle, der durch die Veränderung der Position im Verhältnis zum Hörer verändert wird, kommt als Faktor zur Bestimmung der Position in Frage.

Mögliche akustische Lokalisationsfaktoren:

	Temporal	Spectral
Monaural	Monaurale Phase	1.Gesamtpegel 2.Monaural Spektrale Faktoren
Binaural	Interaurale Zeitunterschiede (ITD)	1.Interaurale Pegelunterschiede (ILD) 2.Binaurale Spektrale Unterschiede

Die Trennung in zeitliche und spektrale Faktoren ist willkürlich, da es sich um verschiedene Betrachtungsweisen des selben Vorganges handelt. Da die Mechanismen der Signalverarbeitung im Ohr im Zeit und Frequenzbereich verschieden sind, ist es aber nützlich zeitliche und spektrale Faktoren getrennt zu betrachten.

Die Unterscheidung in monaurale und binaurale Faktoren betont die Tatsache, dass die Veränderungen der Schallquellenposition sowohl zu Änderungen der Wellenform individuell für jedes Ohr, als auch zur Änderung der Relation zwischen den Wellenformen an beiden Ohren führt.

---

<sup>1</sup> Dieser Abschnitt folgt [1].

### 1.1.1 Monaurale Faktoren:

Der *monaural zeitliche Faktor* ist die Beeinflussung der Schallwelle durch die positionsabhängige Änderung der Impulsantwort des akustischen Systems, bestehend aus Kopf und Pinna, an einem Ohr. Die Übertragungsfunktion dieses Systems wird als Außenohrübertragungsfunktion (head related transfer function = HRTF) bezeichnet. Aus Abbildung 1.1 sieht man die unterschiedliche Gestalt der Impulsantwort für zwei verschiedene Richtungen. Man kann deutliche Unterschiede in der Feinstruktur der beiden Impulsantworten erkennen.

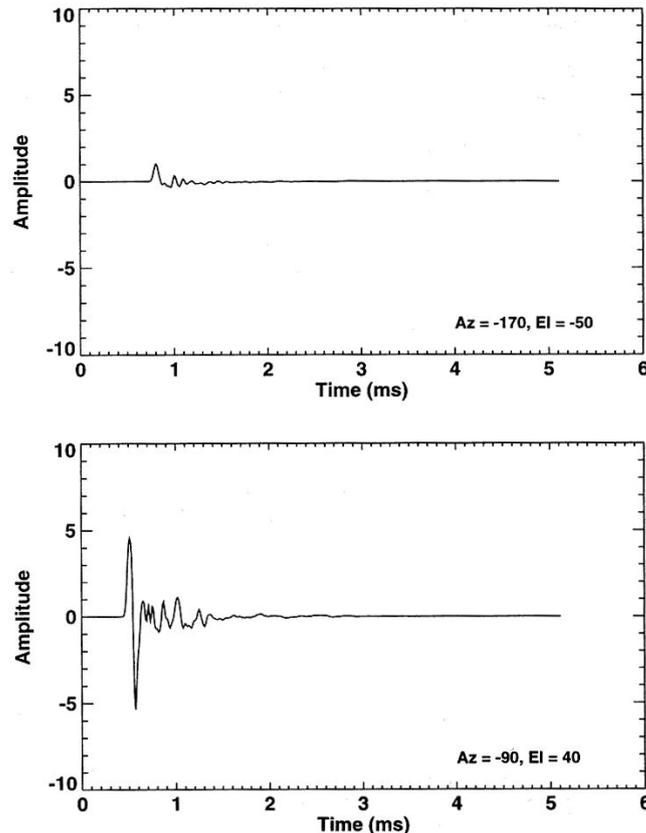


Abbildung 1.1: Beispiel für HRTF Impulsantworten, aufgenommen bei einem linkem Ohr, für Schallquellenpositionen auf der linken Seite des Hörers.

Zwei Gründe sprechen aber dafür, dass dieser monaural zeitliche Faktor nicht zu menschlichen Schallquellenlokalisierung beiträgt.

Erstens ist die Länge der Impulsantwort (ca. 2ms) im Bereich des zeitlichen Auflösungsvermögen des menschlichen Gehörs, das ebenfalls bei 2 ms liegt. Daher ist die zeitliche Feinstruktur für das Ohr nicht auflösbar.

Zweitens folgt aus psychoakustischen Experimenten von Wightman und Kistler [18], dass die Änderung der zeitlichen Feinstruktur in der Impulsantwort nicht zu dementsprechenden Änderungen in der Wahrnehmung der Schallquellenposition führt. Bei diesen Experimenten beurteilten Testpersonen über Kopfhörer dargebotene, mit HRTF- Kurven gefilterte Signale. Einerseits wurden die HRTF's originalgetreu synthetisiert (Betrag und Phase), andererseits erfolgte eine minimalphasige Rekonstruktion (gleiches Amplitudenspektrum aber andere

Feinstruktur in der Impulsantwort). Die beiden verschieden dargebotenen Signale führten zu keiner unterschiedlichen Beurteilung in der Schallquellenposition.

Der *monaural spektrale Faktor* ist die richtungsabhängige Änderung des Amplitudenspektrums der HRTF-Kurven. Abbildung 1.2 zeigt die unterschiedlichen HRTF's für zwei Hörer. Die DTF (Directional Transfer Function) ergibt sich aus der Division einer HRTF durch den Mittelwert der HRTF's aus allen Richtungen.

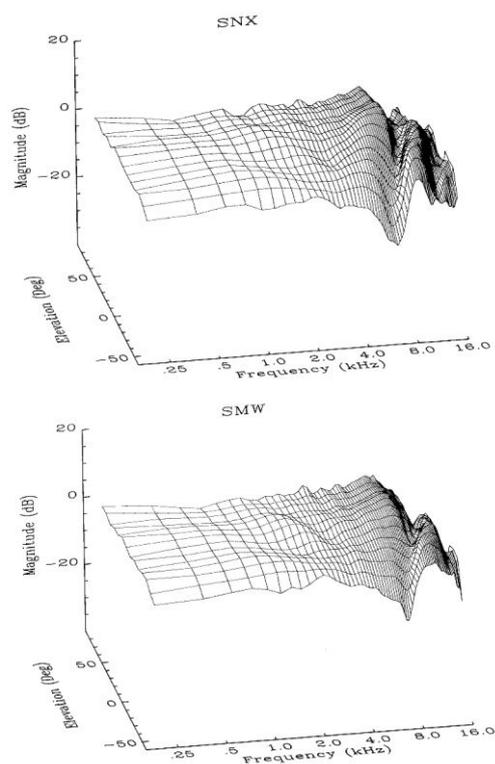


Abbildung 1.2: DTF's zweier verschiedener Hörer. Azimuth der Quellen: 90°

Damit der monaurale spektrale Faktor genutzt werden kann, muss der Hörer nicht nur Kenntnis über die Eigenschaften der HRTF und wie sie sich mit der Position verändern, sondern auch Kenntnis über die spektralen Eigenschaften der Schallquelle selbst haben. Das Spektrum des empfangenen Signals an einem Ohr ergibt sich aus dem Produkt des Quellspektrums mit der HRTF. Um aus dem empfangenen Signal auf eine bestimmte HRTF und damit auf die Richtung schließen zu können, ist zuerst eine Kenntnis des Quellspektrums notwendig.

Weiters sieht man aus Abbildung 1.3, die starken individuellen Unterschiede zwischen den DTF's verschiedener Hörer.

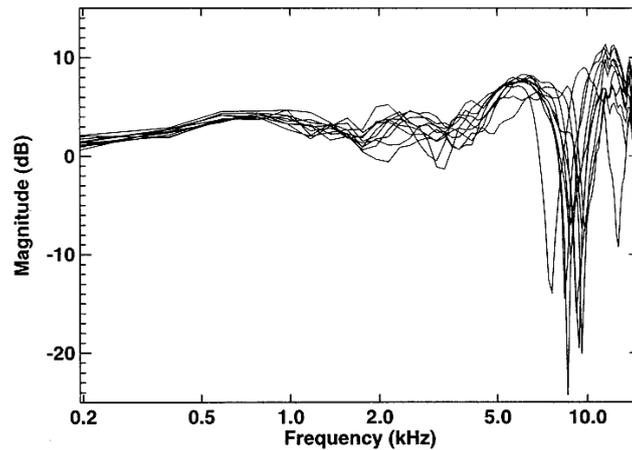


Abbildung 1.3: DTF's für eine Quelle bei 90° Azimuth und 0° Elevation für das rechte Ohr von zehn verschiedenen Hörern.

Außerdem gibt es diese spektralen Unterschiede nur für hohe Frequenzen. Eine Hauptkomponentenanalyse aus einer großen Anzahl von Richtungen führt zu Basisfunktionen die bis 5kHz im Wesentlichen flach verlaufen. Da jede DTF als gewichtete Summe solcher Basisfunktionen dargestellt werden kann, folgt daraus dass die HRTF's bis 5kHz auch flach verlaufen.

Daher hängt der Nutzen monaural-spektraler Unterschiede von einem ausreichenden Frequenzinhalt des Quellspektrums und einer ausreichenden Empfindlichkeit des Hörers für hohe Frequenzen ab.

### **1.1.2 Binaurale Faktoren:**

Man nimmt an, dass die binauralen Faktoren aus einem Verhältnis der HRTF's beider Ohren abgeleitet werden. Weil das Spektrum des Quellspektrums sowohl im Zähler als auch im Nenner erscheint, fällt es nicht ins Gewicht. Daher ist der Nutzen binauraler Faktoren weniger abhängig vom Quellspektrum und a priori Wissen des Hörers.

### **Interaurale Zeitdifferenzen (interaural time difference = ITD):**

Die interaurale Zeitdifferenz ist mit der Phase der HRTF beider Ohren verknüpft und wird als einer der wichtigsten Faktoren zur Schalllokalisierung betrachtet. In erster Näherung ist die ITD für alle Frequenzen gleich. Obwohl in Messungen größere ITD's für tiefere Frequenzen (<1,5kHz) beobachtet wurden, kann man aus psychoakustischen Experimenten (Kistler und Wightman [25]) schließen, dass diese Tatsache keinen Einfluss auf die Lokalisation einer Schallquelle hat.

In Abbildung 1.4 sieht man Messungen von ITD's; deutlich sind auch die annähernd kreisförmigen Linien konstanter ITD zu erkennen.

(Betrachtet man den Kopf als Kugel ergeben sich Kreise gleicher ITD's -Cone of Confusion)

Daraus folgt, dass ein Ort nicht eindeutig durch eine Zeitdifferenz festgelegt ist.

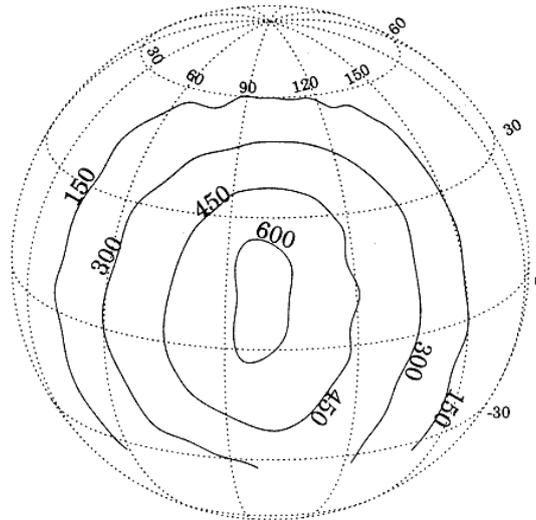


Abbildung 1.4: Kurven gleicher interauraler Zeitdifferenzen (in  $\mu\text{s}$ ).

### **Interaurale Pegeldifferenz (interaural level difference = ILD):**

Die interaurale Pegeldifferenz wird von der Amplitude des HRTF Verhältnisses beider Ohren abgeleitet. Die ILD ist eine komplizierte Funktion der Frequenz, weil für jede Position die Spitzen und Täler in den HRTF's an anderen Stellen liegen (Abbildung 1.5).

Weiters ist die ILD klein für tiefe Frequenzen, da die Größenordnung des Kopfes und der Pinna klein im Verhältnis zur Wellenlänge ( $<1,5\text{kHz}$ ) ist, und der Schall gebeugt wird. Mehrdeutigkeiten kommen wie bei den ITD's dadurch zu Stande, dass es mehrere Positionen gibt, die die gleichen Pegeldifferenzen hervorrufen (Cone of Confusion).

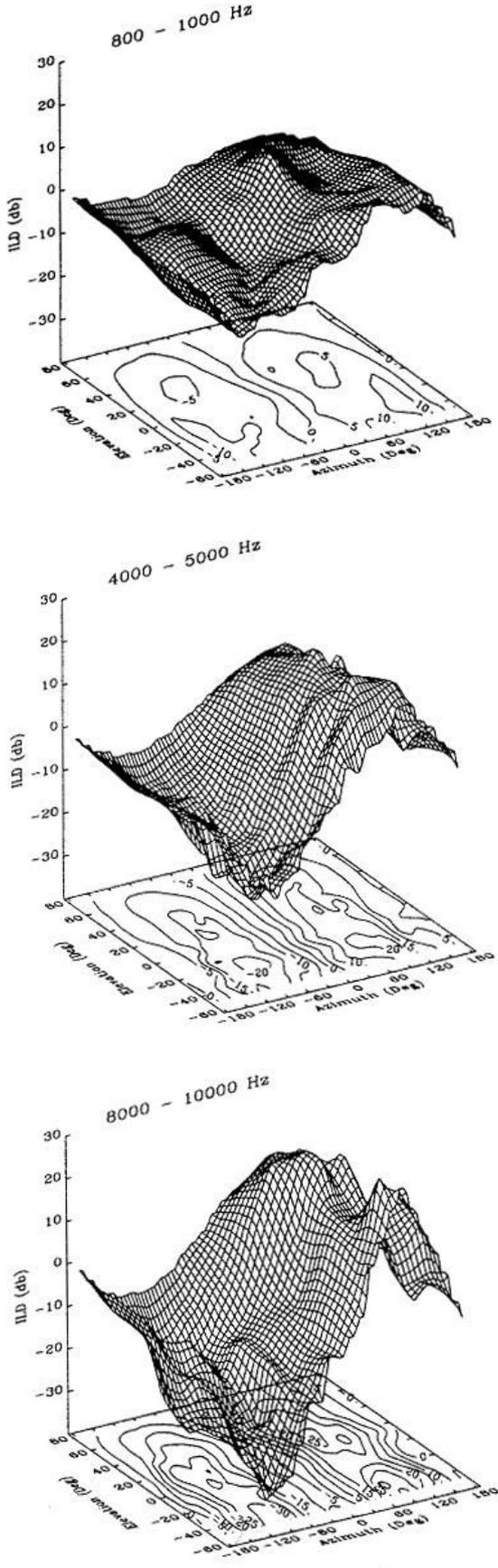


Abbildung 1.5: ILD's in drei verschiedenen Frequenzbändern aus der HRTF Messung eines einzelnen Hörers.

## 1.2 Wovon werden die Faktoren zur Schalllokalisierung beeinflusst?

Vier Dinge beeinflussen die relative Wichtigkeit und Bedeutung der einzelnen Lokalisationsfaktoren:

- die Reliabilität oder Konsistenz des Faktors über Anrengungsbedingung, Hörer und Frequenz
- a priori Kenntnis über die Quellcharakteristik
- Quellspektrum
- Plausibilität des Faktors

### 1.2.1 Reliabilität oder Konsistenz:

Es gibt mehrere Möglichkeiten die Zuverlässigkeit eines Lokalisationsfaktors zu betrachten. Dazu zählt das Ausmaß inwiefern der Faktor von der Quellcharakteristik abhängt, ob er die gleiche Information über das ganze Frequenzband liefert, ob er stark von Hörer zu Hörer variiert, oder ob er eindeutig ist. Hier wird im Folgenden ein Faktor als zuverlässig betrachtet, wenn er mehr zur Bestimmung einer Position beiträgt als ein wenig zuverlässiger Faktor, und sich in Konfliktsituationen mit einem anderen Faktor durchsetzt.

Wenn man die vorher beschriebenen Lokalisationsfaktoren mit der Definition der Reliabilität vergleicht, ergibt sich für die ITD die beste Zuverlässigkeit.

Die ITD hängt nicht von der Quellcharakteristik ab, liefert ungefähr die gleiche Information in jedem Frequenzband, und die Beziehung von ITD und Position der Schallquelle variiert wenig zwischen unterschiedlichen Hörern. Es ergibt sich jedoch eine Mehrdeutigkeit, weil eine ITD eine ganze Gruppe von möglichen Schallquellenpositionen beschreibt.

Psychoakustische Experimente von Wightman und Kistler [18] belegen die Dominanz von ITD gegenüber anderen Lokalisationsfaktoren.

In diesen Experimenten mit realen und virtuellen Quellen kam es häufig zu einer vorne/hinten Verwechslung. Diese Tatsache untermauert nicht nur die Mehrdeutigkeit der ITD sondern auch die Dominanz dieses Lokalisationsfaktors.

Obwohl auch die ILD mehrdeutig sind, ändern sich die Positionen gleicher ILD in jedem Frequenzband. Daher scheint es unwahrscheinlich, dass die Ursache der vorne/hinten Verwechslung in der Mehrdeutigkeit der ILD liegt. Im Gegenteil kann die ILD, da sie über die Frequenz nicht mehrdeutig ist, als unterstützender Faktor zur Beseitigung der vorne/hinten Verwechslung betrachtet werden. Das Muster der ILD über die Frequenz kann auch als zuverlässiger Lokalisationsfaktor betrachtet werden, weil er wenig von der Quellcharakteristik abhängt. Die Tatsache, dass die ILD's nur für hohe Frequenzen ausschlaggebend sind, und sehr stark von Hörer zu Hörer variieren, spricht jedoch gegen deren Nutzen als Lokalisationsfaktoren. Aus Experimenten von Wightman und Kistler folgt, dass die ILD hauptsächlich zur Beseitigung der vorne/hinten Verwechslung dienen.(Abbildung 1.6).

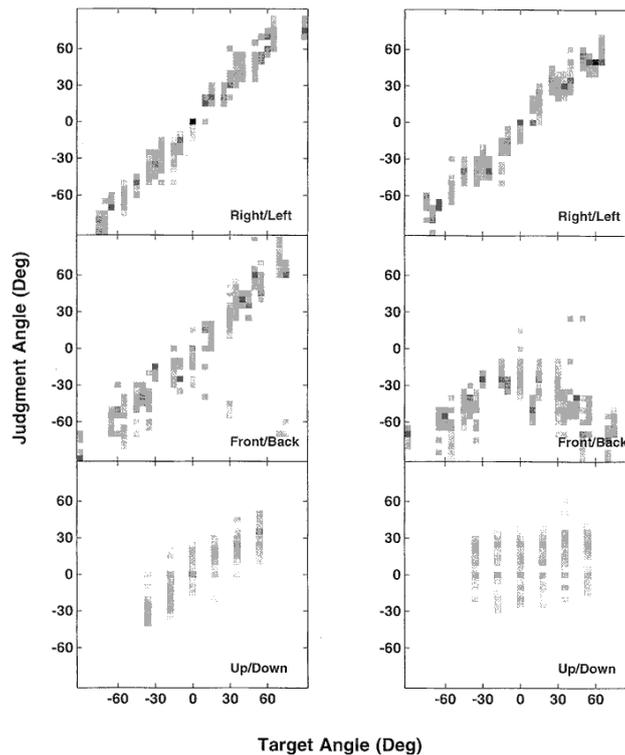


Abbildung 1.6: Richtungsbeurteilung in einem Experiment, bei dem ITD und ILD unterschiedliche Richtungen beschreiben. Die linke Tafel zeigt die Ergebnisse der Versuche für gewöhnliche Bedingungen. In der rechten Tafel ist die ILD für alle Frequenzen 0dB. Man beachte die vorne/hinten Verwechslung auf der rechten Seite.

Die monaural spektralen Faktoren sind in diesem Zusammenhang die am Wenigsten zuverlässigen. Sie unterscheiden sich sehr stark von Hörer zu Hörer, und hängen stark von der a priori Kenntnis der Quellcharakteristik ab.

## 1.2.2 Der Einfluss der a priori Kenntnis der Quellcharakteristik:

Binaurale Faktoren (ITD und ILD) werden nicht von der Quellcharakteristik beeinflusst, und sind auch am bedeutensten bei der Lokalisation von Schallquellen. Es besteht aber kein Zweifel, dass auch monaurale Faktoren, die von der Quellcharakteristik abhängen, zur Lokalisation beitragen.

Bei Experimenten mussten Versuchspersonen Quellen mit flachem und beschränktem Spektrum lokalisieren. Bei Quellen mit zerhacktem Spektrum kam es zu einer Zunahme der vorne/hinten Verwechslungen und Verzerrungen in der Elevationswahrnehmung. Wären nur binaurale Faktoren ausschlaggebend, dürfte das Zerhacken des Spektrums keinen Einfluss haben.

Ein zweites Experiment testete Hörer in monauraler Hörumgebung (ein Ohr zugestopft im Freifeld bzw. ein Kopfhörersignal ausgeschaltet).

Bei flachen Quellspektren konnte noch eine geringe Fähigkeit zur richtigen Lokalisation beobachtet werden, bei zerhacktem Spektrum war den Testpersonen jedoch keine Lokalisation mehr möglich.

Bei beiden Experimenten wurden die Testpersonen im Freifeld und mit virtuellen Quellen (Kopfhörerwiedergabe) getestet.

Um vergleichbare Ergebnisse zu erhalten, musste das Quellspektrum im Freifeld stärker zerhackt werden. Ein Grund dafür können die zusätzlichen Kopfbewegungen sein, die bei der Lokalisation im Freifeld genutzt werden.

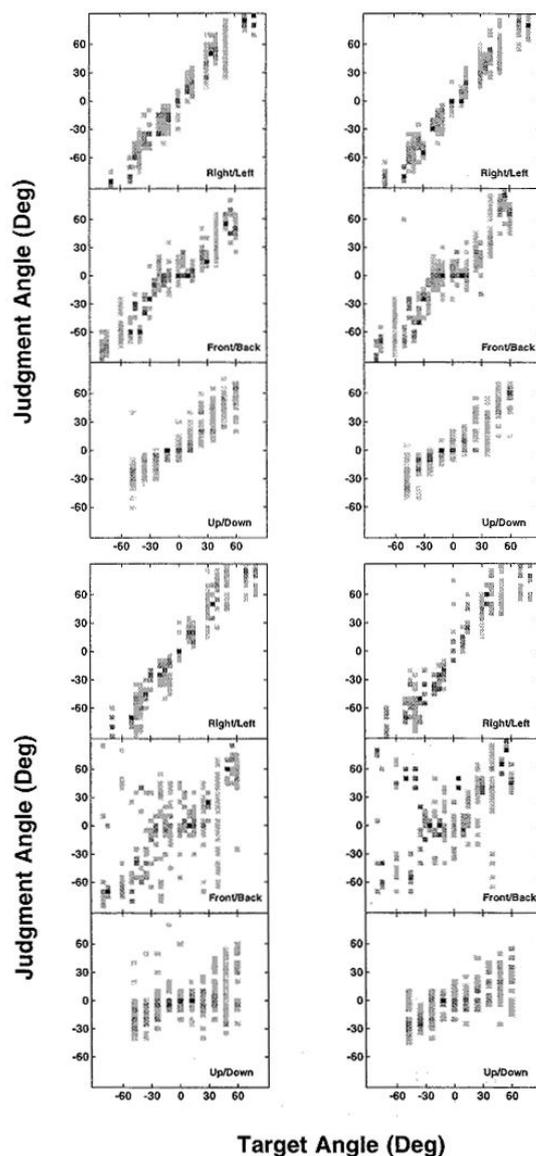


Abbildung 1.7: Richtungsbeurteilungen mit flachen Frequenzspektrum (obere Tafel) und beschränkten Spektren (untere Tafeln). Getestet wurde im Freifeld (links) und virtuellen Freifeld (rechts).

### 1.2.3 Frequenzinhalt des Quellspektrums:

Genauere Schallquellenlokalisierung ist nur für breitbandige Quellspektren möglich. Besteht eine Schallquelle nur aus einer Sinusschwingung oder Schmalbandrauschen, gibt es große Unterschiede zwischen dargebotener und wahrgenommener Position. Es gibt mehrere Gründe, warum die Lokalisation bei schmalbandigen Quellen schlecht funktioniert.

Schmalbandige Quellen führen zu einem verkleinerten aber vieldeutigem Set von Lokalisationsfaktoren, weil weder die ILD-Muster über die Frequenz, noch die monauralen spektralen Faktoren zur Verfügung stehen.

Die Dominanz der ITD vermindert sich mit steigender Frequenz, dafür nimmt die Bedeutung spektraler Faktoren (ILD's in verschiedenen Frequenzbändern und monaurale spektrale Faktoren) für höhere Frequenzen zu.

Ein Experiment, bei dem Quellspektren mit flachem Frequenzgang und gefilterte Quellspektren (Bandsperrung zwischen 5-10kHz) getestet wurden, führte zu einer Zunahme der vorne/hinten Verwechslung.

### **1.2.4 Einfluss der Plausibilität der Lokalisationsfaktoren:**

Die Konstellation, wie die Lokalisationsfaktoren dem Hörer dargeboten werden und mit der Erfahrung und Erwartung des Hörers übereinstimmen, hat signifikante Auswirkungen auf die wahrgenommene Position und die relative Gewichtung der einzelnen Faktoren.

Jene Faktoren, die unnatürlich oder ungewöhnlich sind, werden weniger gewichtet zur Bestimmung der Richtung herangezogen.

Experimente mit virtuellen Quellen, die mit HRTF's anderer Hörer (nicht die Ohrkurven der Testperson) synthetisiert wurden, zeigen vor allem eine starke Zunahme der vorne/hinten Verwechslung.

Unter solchen Testbedingungen erwartet man, dass bei gleicher Kopfgröße die ITD im Wesentlichen übereinstimmen, die ILD's und spektralen Faktoren aber starke Unterschiede aufweisen. Die oben erwähnten Testresultate legen den Schluss nahe, dass ILD zur Auflösung der vorne/hinten Ambiguität dienen, hier aber durch die Ungewöhnlichkeit und Unnatürlichkeit schwach gewichtet werden.

Beim Hören im natürlichen Schallfeld gibt es nur Lokalisationsfaktoren die konsistent über den ganzen Frequenzbereich sind. Da der Schall aus einer natürlichen Quelle kommt, ist die Position die in ILD, ITD, und monauralen Faktoren kodiert ist, für alle Frequenzbänder gleich. Bei natürlichen Quellen kann es nicht vorkommen, dass die ILD in einem Frequenzband die Quelle auf der einen Seite des Kopfes, die ILD in einem anderem Frequenzband die Quelle auf die andere Seite des Kopfes positioniert.

Bei Experimenten mit virtuellen Quellen wurde die Auswirkung untersucht, wenn Lokalisationsfaktoren in verschiedenen Frequenzbändern unterschiedliche Richtungen bestimmen. In einem Fall bestimmten ILD und spektrale Faktoren eine Richtung über das gesamte Frequenzband, ITD's bestimmten aber in jedem von vier Bändern eine andere Richtung. Die ILD's und spektralen Faktoren sind daher konsistent über den Frequenzbereich, die ITD's inkonsistent. Auch der umgekehrte Fall wurde untersucht.

Das Ergebnis dieser Untersuchung zeigt, dass die wahrgenommene Richtung immer den konsistenten Faktoren folgt.

Auch wenn die ITD's nur in einem hohen Frequenzbereich (>5kHz) inkonsistent sind, werden alle ITD's schwach gewichtet und den ILD's und spektralen Faktoren die größere Bedeutung beigemessen.

Daraus sieht man, dass realistische Faktoren stärker als unrealistische Faktoren gewichtet werden, aber auch, dass ITD's im hohen Frequenzbereich gleich wichtig wie im tieffrequenten Bereich sein können.

### 1.3 Zusätzliche Faktoren – Auflösung der Vorne/Hinten Ambiguität

Zerhacken des Quellspektrums, Entfernen der hochfrequenten Anteile des Quellspektrums und Darbieten nichtvertrauter Lokalisationsfaktoren führt zu einer Zunahme der vorne/hinten Verwechslung. Aber auch bei Hörbedingungen im Freifeld, wo der ganze Satz an Lokalisationsfaktoren einschließlich ILD, ITD und spektralen Faktoren vorhanden ist, kommt es zu einer gewissen vorne/hinten Verwechslung.

Es gibt aber einige Unterschiede zwischen den Freifeldhörbedingungen bei den Experimenten und den Hörbedingungen im natürlichen Schallfeld.

Am augenscheinlichsten ist das Fehlen von Reflexionen in der Testumgebung. Aber auch das Hinzufügen von Reflexionen eines einfachen quaderförmigen Raumes führte zu keiner Verminderung der vorne/hinten Verwechslung.

Der stärkste akustische Unterschied zwischen hinten und vorne spielt sich im Frequenzbereich zwischen 3kHz und 7kHz ab (Abbildung 1.8).

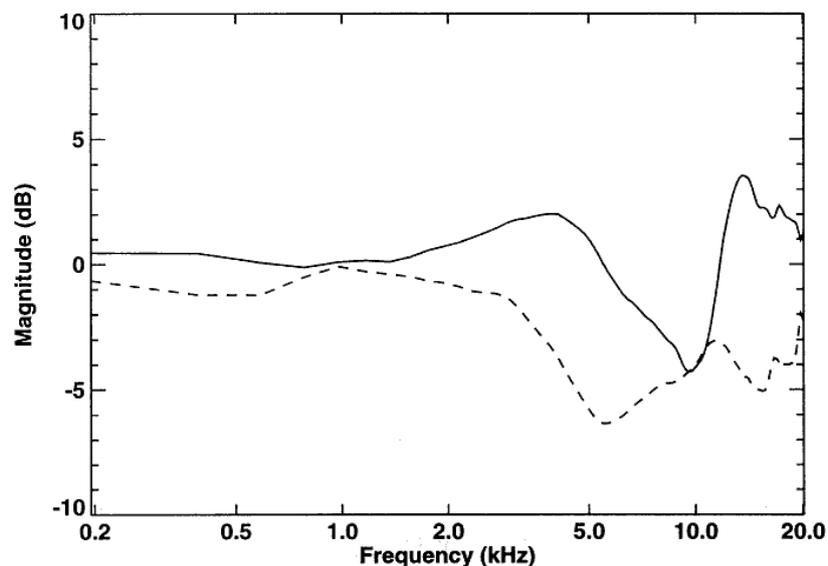


Abbildung 1.8: Durchschnittliche DTF Amplitude (12 Hörer) für Schallquellen von vorne (durchgezogene Linie, Quellen zwischen  $-30^\circ$  und  $30^\circ$  Azimuth und  $-40^\circ$  und  $40^\circ$  Elevation) und von hinten (strichlierte Linie, Quellen zwischen  $-150^\circ$  und  $150^\circ$  Azimuth und  $-40^\circ$  und  $40^\circ$  Elevation).

Eine künstliche Verstärkung dieser Unterschiede zwischen Hinten und Vorne führte zu keiner Verminderung der vorne/hinten Vertauschung.

Experimente bei denen die Kopfbewegung berücksichtigt wurden, führten zu einer Verminderung der vorne/hinten Verwechslung (Abbildung 1.9).

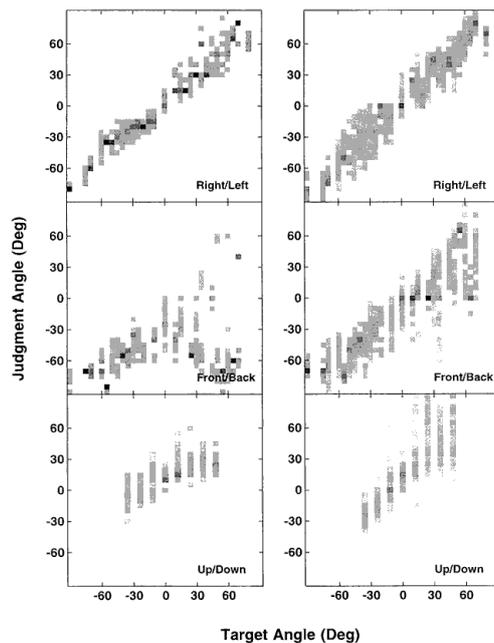


Abbildung 1.9: Richtungsbeurteilung mit virtuellen Quellen mit flachem Frequenzspektrum für zwei verschiedene Fälle. Links: Kopfbewegung unmöglich. Rechts: Kopfbewegung möglich.

Zusammenfassend kann gesagt werden, dass die wahrgenommene Position einer Schallquelle von wesentlich mehr Faktoren als den tieffrequenten ITD und hochfrequenten ILD (Duplextheorie, Lord Rayleigh [26]) abhängt. Weitere Faktoren, wie monaurale spektrale Faktoren, spielen eine Rolle. Die Wichtigkeit der einzelnen Komponenten wird von einer Reihe Quelle- und hörerabhängiger Eigenschaften beeinflusst.

## 1.4 Auswirkung von spontanen Kopfbewegungen auf das räumliche Hören

Experimente mit zwei identischen Ohreingangssignalen zeigen, dass monaurale Faktoren hauptsächlich zur Definition von Hinten-Vorne in der Medianebene, der Elevation und der Distanz des akustischen Ereignisses herangezogen werden. Interaurale Signalattribute hängen vor allem mit der lateralen Positionierung des Schallereignisses zusammen. Im freien Schallfeld, mit frei beweglichen Kopf, werden weitere Faktoren zur Bewertung der Position herangezogen, welche im Folgenden beschrieben werden [2].

Die Außenohrübertragungsfunktionen hängen von der Position der Schallquelle ab. Diese Filterung, die das Signal auf seinem Weg bis zum Trommelfeld erfährt, hängt von der Richtung und Entfernung der Schallquelle relativ zum Kopf des Hörers ab. Wenn sich der Kopf einer Person relativ zur Schallquelle bewegt, ändern sich die monauralen und interauralen Attribute des Signals am Trommelfell auf eine bestimmte Weise. Theorien des räumlichen Hörens, die die Position des akustischen Ereignisses und die Änderungen der Eingangssignale infolge der Kopfbewegungen beschreiben, werden als Bewegungs- oder Motorische Theorien bezeichnet. Motorische Theorien beschreiben auch andere Attribute des akustischen Ereignisses

durch Kopfbewegungen, wie zum Beispiel Lautheit oder Klangfarbe, aus denen die Person auch Information über die Position der Schallquelle gewinnen kann.

Zwei Arten von Kopfbewegungen können in diesem Zusammenhang unterschieden werden:

- Eine mehr oder weniger unbewusste, spontane Bewegung des Kopfes in Richtung des akustischen Ereignisses und daher zur wahrscheinlichsten Position der Schallquelle. Diese Bewegungen verlangen ein mehr oder weniger genau lokalisiertes akustisches Ereignis, bevor die Bewegung beginnt. Wird der Kopf bewegt sinkt der Winkel der maximalen Richtungsauflösung, da das Schallereignis in den Bereich der größten Auflösung gebracht wird.
- Bewegungen des Kopfes zum Suchen und Orientieren, die mehr oder weniger bewusst gemacht werden. Ihr Ziel ist, mehr Informationen über das Schallereignis zu erhalten, um eine endgültige Position des zuvor nicht klar erkennbaren akustischen Ereignisses zu erhalten. Im Allgemeinen wird die Lokalisation des akustischen Ereignisses durch die Kopfbewegungen genauer.

### 1.4.1 Rückblick:

Bewegungs- bzw. Motorische Theorien wurde das erste Mal in Zusammenhang mit Zeitdifferenztheorien von Hornbostel und Wertheimer (1920) [28] Aufmerksamkeit zuteil. Diese Autoren nahmen zwei 21cm voneinander entfernte Punkte als Modell der Schallrezeption der zwei Ohren an. In diesem Modell sind die Differenzen zwischen Distanzen zu den beiden Ohren für eine bestimmte Hyperbel in der Frontalebene gleich. Die Zeitdifferenz an den beiden Ohren ist für alle Punkte auf der Hyperbel gleich. Für größere Entfernungen kann die Hyperbel durch ihre Asymptoten angenähert werden. Daraus ergibt sich im Raum ein Kegel auf dem alle möglichen Schallquellen die gleichen Zeitdifferenzen haben. Aus diesem einfachen Modell folgen Mehrdeutigkeiten bezüglich der Zeitdifferenzen, vgl. Abbildung 1.10.

Van Soest (1929) [27] zeigte als Erster auf, dass es trotz dieser Mehrdeutigkeiten der Zeitdifferenz möglich ist, die Richtung des Schalleinfalls infolge von Kopfbewegungen richtig wahrzunehmen. Das Prinzip wird in Abbildung 1.11 dargestellt.

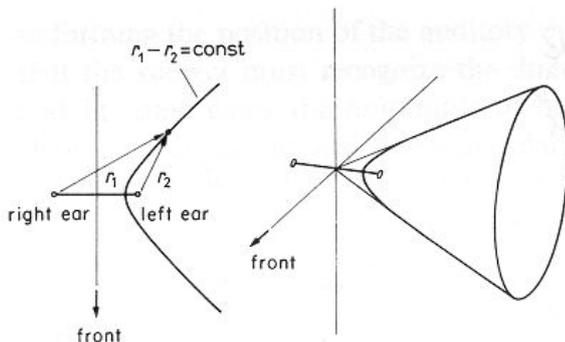


Abbildung 1.10: Orte mit gleicher Zeitdifferenz. a.) Hyperbel in der Ebene. b.) Kegel im Raum.

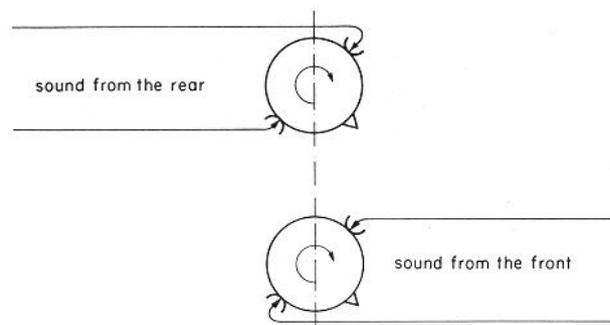


Abbildung 1.11: Die Änderung der Ohreingangssignale bei einer Kopfdrehung ist unterschiedlich für Schalle von hinten und von vorne.

Angenommen es gibt keine interauralen Signalunterschiede, so erreicht der Schall die beiden Trommelfelle zur gleichen Zeit und mit gleichem Pegel. Dann sind alle Positionen in der Medianebene mögliche Schalleinfallrichtungen. Der Schall kann von vorne oder von hinten in der Horizontalebene kommen. Wenn die Person ihren Kopf um eine vertikale Achse nach rechts dreht, kommt das linke Ohr näher zur Schallquelle, wenn die Quelle vorne liegt. Das rechte Ohr kommt näher, wenn die Quelle hinten liegt. Für die selbe Kopfbewegung kann der Wechsel der interauralen Zeitdifferenz entweder positiv oder negativ sein, in Abhängigkeit ob die Schallquelle vorne oder hinten liegt. Analoges gilt für die interauralen Pegeldifferenzen.

Van Soest nahm an, dass die Polarität des Wechsels der interauralen Zeitunterschiede vom Gehör ausgewertet und zur Bestimmung der Position des akustischen Ereignisses verwendet werden. Aber nicht nur die Wechsel der Ohreingangssignale, sondern auch die Stärke und Richtung der Kopfbewegung sind für diese Auswertung wichtig. Informationen über die Kopfbewegungen erhält das Gehirn durch das Gleichgewichtsorgan (Vestibularorgan), den Sehsinn und durch die Position und Spannung der Hals und Nackenmuskel.

Verlassen wir das System von Hornbostel, Wertheimer mit der Annahme eines symmetrischen Kopfes in Hinblick auf die Achse zwischen den Ohren, verkompliziert sich die Situation. Berücksichtigt man die Auswirkungen der Außenohrübertragungsfunktionen scheint es sehr unwahrscheinlich, dass es zu Mehrdeutigkeiten zwischen den interauralen Signaldifferenzen und der Schalleinfallrichtung in Zusammenhang mit Breitbandsignalen kommt. Anderes gilt für Schmalbandsignale, für die es sehr wohl möglich ist, dass es zu gleichen interauralen Signalunterschieden aus verschiedenen Richtungen kommt. Motorische Theorien sind für das räumliche Hören dann von Bedeutung, wenn die Einflüsse des Kopfes und der Außenohren auch schon in Betracht gezogen wurden.

### **1.4.2 Zusammenhang von Kopfbewegungen und Hören:**

Drei Fragen werden im Folgenden untersucht:

Gibt es überhaupt Kopfbewegungen in Zusammenhang mit natürlichen Hören mit zwei Ohren und wie können diese Bewegungen beschrieben und analysiert werden?

Welche Attribute zur Interpretation ergeben sich für das Gehör aus bestimmten Kopfbewegungen?

Welche Auswirkungen haben Kopfbewegungen auf die Position des akustischen Ereignisses?

Umfangreiche Untersuchungen zu Punkt eins wurden von Thurlow, Mangels und Runge (1967) [29] durchgeführt, die Lokalisationsversuche mit über 50 Personen in einem schalltoten Raum durchführten. Den Versuchspersonen wurde bei verbundenen Augen Schmalbandrauschen (500-700 Hz oder 7.5-8kHz) an 10 im Raum verteilten Positionen dargeboten. Die Kopfbewegungen wurden mit einer Filmkamera aufgezeichnet und anschließend ausgewertet. Die Bewegungen wurden in die Klassen 'rotieren', 'nicken', und 'neigen', eingeteilt. (Abbildung 1.12).

Am Häufigsten kommen Rotationsbewegungen vor. Als Kombination von Kopfbewegungen treten rotieren und nicken am häufigsten auf. Die Experimente zeigten auch, dass die Personen die Kopfbewegungen fortführten oder wiederholten um die Position der Schallquelle exakt zu erkennen.

Die Frage wie die Ohreingangssignale infolge der Kopfbewegungen verändert werden, wurde ausführlich von Wallach (1938) untersucht.

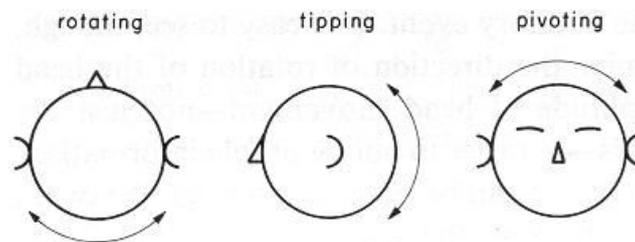


Abbildung 1.12: Klassifikation von Kopfbewegungen in rotieren, nicken und neigen. (nach Thurlow et al. 1967)

Es wurde schon beschrieben, dass Rotationsbewegungen die Vorne/Hinten - Zuordnung einer Schallquelle ermöglichen. Für eine bestimmte Richtung der Rotation ergibt sich eine bestimmte Änderung der interauralen Signaleigenschaften. Zusätzlich kann ein Neigen des Kopfes (Drehbewegung um die Schnittachse von Horizontal und Medianebene) zur Unterscheidung von Quellen aus der oberen und unteren Hemisphäre herangezogen werden. Die vier Sektoren vorne+oben, vorne+unten, hinten+oben und hinten+unten können also aus der Änderung der interauralen Signaleigenschaften infolge der Kopfbewegungen 'rotieren' und 'neigen' abgeleitet werden. Wird auch der Betrag der Änderung der interauralen Signaleigenschaften als Funktion der Kopfbewegungen in Betracht gezogen, kann daraus Information über die Elevation der Schallquelle und auch über den Winkel bezüglich der Frontalebene abgeleitet werden.

Dazu zwei Beispiele:

- Die Schallquelle liegt vorne in der Horizontalebene. Der Kopf wird bewegt bis die Schallquelle genau links liegt. Während der Bewegung (Rotation) steigt die interaurale Differenz des Signals von Null bis zu ihrem Maximalwert.
- Die Schallquelle liegt genau über dem Kopf. Bei erneuter Rotation des Kopfes um  $90^\circ$  bleibt die interaurale Signaldifferenz Null, da die Schallquelle in der Medianebene bleibt.

Zwischen diesen beiden Fällen von maximaler bzw. keiner Änderung der interauralen Signaleigenschaften besteht ein kontinuierlicher Übergang, der durch den Winkel  $\gamma$  (Winkel zwischen Schalleinfallrichtung und Medianebene) beschrieben wird. Die geometrische Beziehung zwischen Azimuth  $\varphi$ , Elevation  $\delta$  und  $\gamma$  lautet  $\sin \gamma = \cos \delta \sin \varphi$ . Abbildung 1.13 zeigt diesen Zusammenhang.

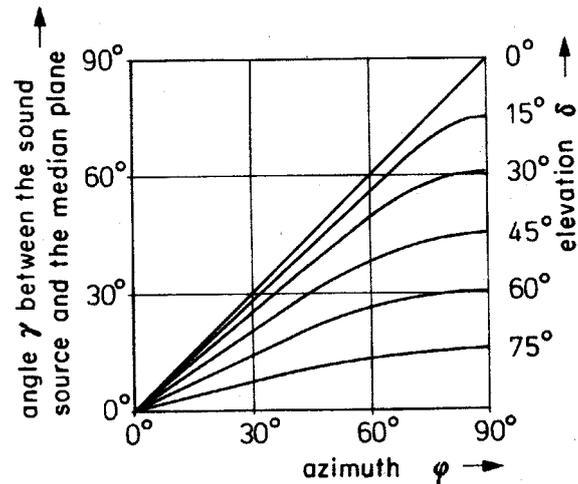


Abbildung 1.13: Der Winkel  $\gamma$  zwischen Schalleinfallrichtung und Medianebene als Funktion des Azimuth und der Elevation. (nach Wallach 1938).

Für Neigungsbewegungen können ähnliche Überlegungen wie für die Rotation angewendet werden. Der Fall der Nickbewegung (Drehbewegung um die Schnittachse von Frontal und Horizontalebene) verlangt eine gesonderte Betrachtung. Wird der Kopf als einfache Kugel betrachtet, bleiben bei einer solchen Bewegung die interauralen Differenzen konstant. Bei wirklichen Köpfen tritt aber eine Änderung der interauralen und auch monauralen Attribute als Funktion der Richtung und auch des Betrages einer Nickbewegung auf.

Untermauert wird die Behauptung, dass diese Information aus den Kopfbewegungen vom Gehör für die Lokalisation verarbeitet wird durch Experimente von Klensch (1948), Jongkees und Van de Veer (1958). Neuere psychoakustische Testreihen stammen von Mackensen, Reichenauer und Theile (1998).

#### **1.4.3 Beschreibung der Versuche von Mackensen, Reichenauer, Theile (1998):**

Untersucht wurde die Lokalisation von Schallquellen, wobei man sich auf Quellen in der Horizontalebene beschränkte. Die Auswirkung von Drehbewegungen auf die Lokalisationsgenauigkeit war Gegenstand der Untersuchung. Weiters wurde die Auswirkung von Reflexionen auf die Richtungslokalisierung beobachtet.[3].

Der Versuchsaufbau bestand aus einem nachführbaren diffusfeldentzerrten Kunstkopf (Neumann KU 100), einem elektronischen Headtrackingsystem (Polhemus Fastrak) und diffusfeldentzerrten Kopfhörer (Stax SR-Lambda Professional). Die Latenzzeit des Gesamtsystems, bestehend aus Headtrackingsystem, Computer und Dreheinheit betrug nur 50 ms, die Winkelgenauigkeit der Motordreheinheit  $1^\circ$ . Dadurch wurden auf akustisch und mechanischer Seite Störungen auf ein nicht wahrnehmbares Maß reduziert.

Untersucht wurden die Personen unter folgenden Hörbedingungen: starrer Kunstkopf, nachführbarer Kunstkopf, eigene Ohren. Außerdem erfolgten die Hörtests zum einen in einem Abhörraum mit Reflexionen (Studio), zum anderen in einem

reflexionsarmen Raum (RAR), um die Auswirkungen der Reflexionen auf die Lokalisation zu prüfen.

In den Diagrammen Abbildung 1.14 bis 1.19 kann man die Ergebnisse dieser Versuche sehen.

### **Ergebnisse der Hörversuche in natürlicher akustischer Umgebung (Reflexionen treten auf):**

- **Starrer Kunstkopf:** Es sind deutlich die vorne-hinten Vertauschungen als Abweichung der Punkte von der Diagonale (Idealfall) zu sehen. Besonders stark tritt dieser Fehler im Bereich  $330^{\circ}$ - $0^{\circ}$  und  $0^{\circ}$ - $30^{\circ}$  auf. In allen Bereichen sind die Streuungen groß, von guter Lokalisation kann keinesfalls gesprochen werden.
- **Bewegter Kunstkopf:** Im Vergleich zum starren Kunstkopf verschwinden die vorne-hinten Vertauschungen fast völlig. (Wenig Abweichung von der Idealdiagonalen). Die Streuung zwischen den einzelnen Messwerten ist geringer, man kann von guter Richtungslokalisierung sprechen.
- **Person hört mit eigenen Ohren:** Die Abweichungen von der Ideallinie sind klein, aber die Streuungen sind vergleichbar mit der Situation des mitbewegten Kunstkopfes. Was also die reine Richtungslokalisierung angeht kann behauptet werden, dass das Hören mit drehbaren Kunstkopf zu denselben Ergebnissen führt, als wenn die Versuchsperson mit eigenen Ohren hört und den Kopf dabei bewegen darf.

### **Hörversuche im reflexionsarmen Raum:**

- **Starrer Kunstkopf:** Die vorne-hinten Vertauschung kommt deutlich zum Vorschein, wobei die Wahrscheinlichkeit zur Inversion im Vergleich zum normalen Raum noch größer ist. Auch die einzelnen Streuungen sind hier sehr groß. Im schalltoten Raum ist bei starrem Kunstkopf keine gute Richtungslokalisierung möglich.
- **Bewegter Kunstkopf:** Trotz Fehlen von Reflexionen sind hier die Ergebnisse denen des natürlichen Raumes sehr ähnlich. Die Streubreite ist geringfügig größer und es treten noch vereinzelt vorne-hinten Vertauschungen auf.
- **Person hört mit eigenen Ohren:** Diese Ergebnisse sind praktisch mit denen unter Normalbedingungen identisch. Verglichen mit dem Hörversuch vorher (bewegter Kunstkopf im reflexionsarmen Raum) scheint die Richtungslokalisierung geringfügig besser zu sein.

Es herrscht Übereinstimmung in der Literatur, dass im Fall des natürlichen räumlichen Hörens, Kopfbewegungen die Fähigkeit die Schalleinfallrichtung zu erkennen, verbessern. In Hinblick auf Lokalisation bedeutet das, dass die Unterschiede zwischen der Richtung der Schallquelle und der Richtung des wahrgenommenen akustischen Ereignisses nach spontanen Kopfbewegungen kleiner sind. Im Speziellen werden Richtungsvertauschung (vorne-hinten) zwischen Quelle und Hörereignis fast vollständig vermieden.

Kopfbewegungen, die bewusst zum Suchen und Orientieren eingesetzt werden, können zu einem Wechsel der Position des wahrgenommen Schallereignisses

führen, wenn das Signal lang genug dauert. Wenn Information zur Lokalisation aus den Kopfbewegungen abgeleitet wird, wird Information von monauralen Signalattributen geringer bewertet bzw. ersetzt. Das folgt aus Experimenten von Wallach und Blauert (1969).

Zusammenfassend kann gesagt werden, dass Kopfbewegungen die genaue Lokalisation von Schallereignissen mit großer Sicherheit ermöglichen. Nach den Kopfbewegungen ist das wahrgenommene Schallereignis praktisch immer am Ort der tatsächlichen Schallquelle.

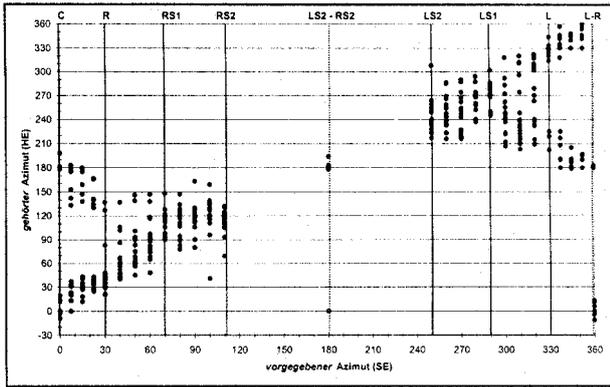


Abbildung 1.14: Richtungslokalisierung. Starrer Kunstkopf im Studio

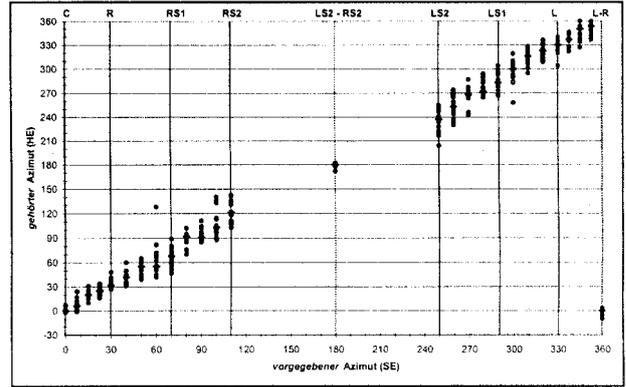


Abbildung 1.15: Richtungslokalisierung. Drehbarer Kunstkopf im Studio.

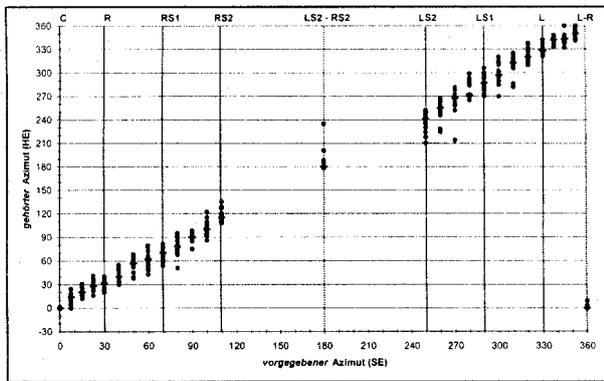


Abbildung 1.16: Richtungslokalisierung. 'eigene Ohren' im Studio.

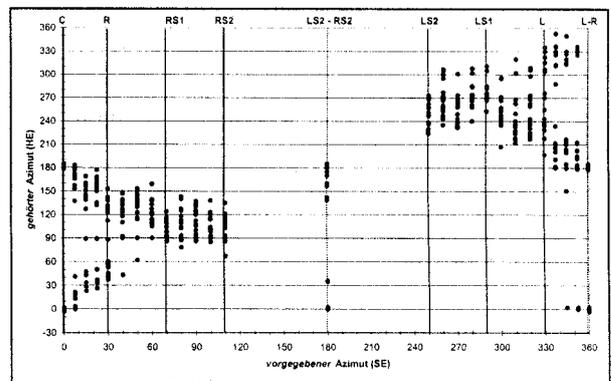


Abbildung 1.17: Richtungslokalisierung. Starrer Kunstkopf im Reflexionsarmen Raum (RAR).

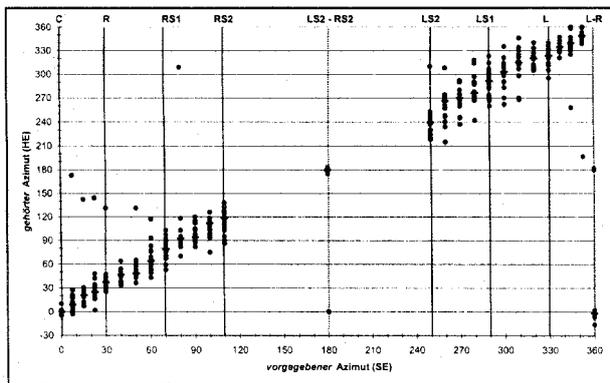


Abbildung 1.18: Richtungslokalisierung. Drehbarer Kunstkopf im RAR.

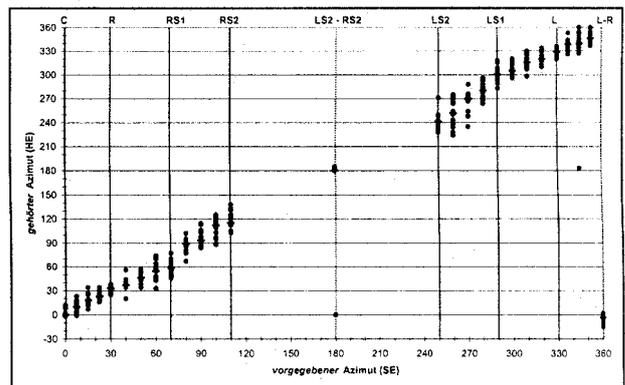


Abbildung 1.19: Richtungslokalisierung. 'eigene Ohren' im RAR.

## 2 Raumsimulation

Unter dem Begriff Raumsimulation versteht man die künstliche Erzeugung einer Raumakustik. Das zunächst direkte Signal ohne zusätzlichen Raumeindruck soll in einen bestimmten akustischen Raum abgebildet werden.

Im systemtheoretischen Sinn entspricht die Nachbearbeitung eines Audiosignals mittels Raumsimulation der Faltung des Audiosignals mit einer entsprechenden Raumimpulsantwort.

Die Raumimpulsantwort zwischen zwei Punkten innerhalb eines Raumes lässt sich in das Direktsignal, erste Reflexionen von den Wänden und den diffusen Nachhall aufspalten (Abbildung 2.1).

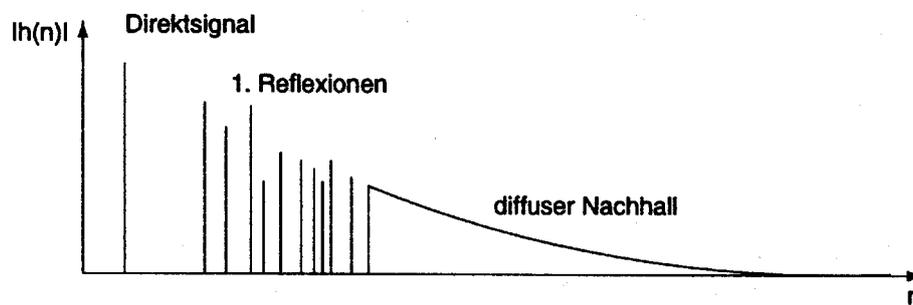


Abbildung 2.1: Raumimpulsantwort

Die ersten Reflexionen nehmen mit fortschreitender Zeit in ihrer zeitlichen Dichte ständig zu und gehen dann in ein exponentiell abfallendes Zufallssignal über, welches als Nachhall bezeichnet wird.

Die Nachhallzeit (Abnahme der Schallenergie um 60 dB) lässt sich aus der Geometrie des Raumes und den absorbierenden Teilflächen innerhalb des Raumes angeben:

$$T_{60} = 0.163 \frac{V}{\alpha S} = \frac{0.163}{m/sec} \frac{V}{\sum_n \alpha_n S_n}$$

$T_{60}$ .....Nachhallzeit in sec.

$V$ .....Volumen in  $m^3$

$S_n$ .....Teilfläche in  $m^2$

$\alpha_n$ .....Schallabsorptionsgrad der Teilfläche  $S_n$

Die Geometrie des Raumes bestimmt ebenfalls die Eigenfrequenzen eines dreidimensionalen quaderförmigen Raumes:

$$f_e = \frac{c}{2} \sqrt{\left(\frac{n_x}{l_x}\right)^2 + \left(\frac{n_y}{l_y}\right)^2 + \left(\frac{n_z}{l_z}\right)^2}$$

$n_x, n_y, n_z$ .....Anzahl der Halbwellen (0,1,2,...)  
 $l_x, l_y, l_z$ .....Kantenlängen des Raumes  
 $c$ .....Schallgeschwindigkeit

## 2.1 Berechnung der Raumimpulsantwort mit Modell-Methoden:

Die Verfahren zur analytischen Bestimmung von Raumimpulsantworten basieren auf dem Strahlenmodell oder dem Spiegelquellenmodell.

Beim Strahlenmodell wird von einer punktförmigen und radial abstrahlenden Schallquelle ausgegangen. Es werden die Laufzeiten und die Absorptionsfaktoren der Wände, Decken und Böden zur Bestimmung der Raumimpulsantwort herangezogen. Für das Spiegelquellenmodell werden zusätzliche Spiegelräume mit sekundären Spiegelquellen gebildet, die wiederum neue Spiegelräume und Spiegelquellen erhalten. Die Summation über alle Spiegelquellen mit den entsprechenden Laufzeiten und Dämpfungen liefert die geschätzte Raumimpulsantwort. (Abbildung 2.2).

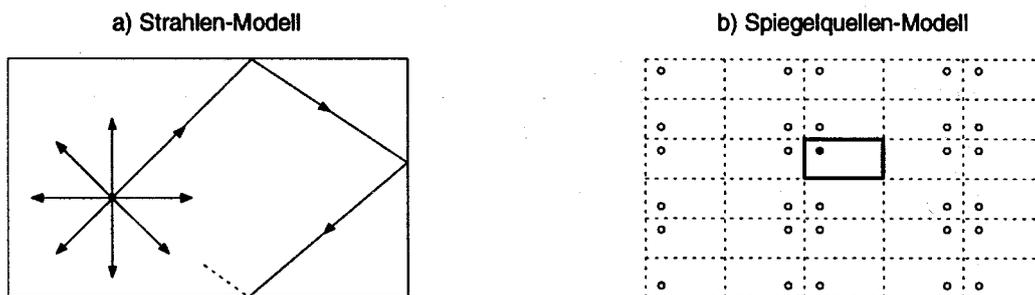


Abbildung 2.2: Modellmethoden zur Berechnung der Raumimpulsantwort

## 2.2 Binaurale Raumsimulation:

Parameter wie Geometrie und akustische Eigenschaften der reflektierenden Wände, Position, Orientierung und Richtcharakteristik von Sender und Empfänger bestimmen das wahrgenommene Schallfeld. Aufgabe der binauralen Raumsimulation ist es akustische Reize zu schaffen, welche die natürlichen Umgebung möglichst exakt nachbilden.

Die binaurale Raumsimulation besteht aus zwei Teilen. Erstens aus der Berechnung des Schallfeldes und zweitens aus der Auralisation (Hörbarmachen des berechneten Schallfeldes).

In einer Umgebung mit reflektierenden Wänden treffen auf den Hörer außer dem Direktschall auch eine Vielzahl von reflektierten Schallwellen mit unterschiedlichen Laufzeiten und unterschiedlicher spektraler Veränderung ein. Diese Situation kann mit so genannten ‚Virtuellen Quellen‘ beschrieben werden.

Dabei wird das Schallfeld innerhalb des Raumes durch eine Wolke von virtuellen Schallquellen ohne die reflektierenden Wände dargestellt.

Ein lineares Filter nach jeder virtuellen Quelle beschreibt den Einfluss des Übertragungsmedium, die Filterwirkung durch die Wände und die Richtcharakteristik der Primärquelle.

Bei der Schallfeldberechnung wird basierend auf den geometrischen Daten die räumliche Verteilung der Spiegelquellen berechnet, bei der Auralisation wird durch Faltung eines echofreien Monosignals mit der binauralen Impulsantwort und anschließende Wiedergabe über Kopfhörer das Hineinhören in einen virtuellen Raum ermöglicht.

### **2.3 Schallfeldberechnung mittels geometrischer Raum-Akustik:**

Die geometrische Raum-Akustik basiert auf der speziellen Lösung der Wellengleichung, wenn die Wellenlänge klein im Verhältnis zu den Raumabmessungen und groß im Verhältnis zu deren Oberflächenbeschaffenheit ist. Es gibt zwei grundsätzlich verschiedene Algorithmen zur Berechnung des Schallfeldes mittels geometrischer Akustik: die Spiegelquellenmethode und das Strahlenmodell.

#### **Spiegelquellenmodell:**

Die Basis der Spiegelquellenmethode ist, dass der Weg einer Reflexion durch eine gerade Linie vom Empfänger zu einer virtuellen Quelle beschrieben werden kann. Bei Anwendung dieses Prinzips auf einen quaderförmigen Raum ergibt sich ein regelmäßiges Netz von Spiegelquellen. (Abbildung 2.2b).

Für unregelmäßige Raumformen ist es schwieriger die virtuellen Quellen zu berechnen [7]. Zusätzlich muss die ‚Sichtbarkeit‘ einer Spiegelquelle im Raum berücksichtigt werden.

Finden der Spiegelquellenkoordinaten:

Als Ausgangspunkt dient eine allgemeine Ebene und die Koordinaten eines Quellpunktes. Die Ebene (reflektierende Fläche) wird festgelegt durch ihren Normalvektor, und den Normalabstand der Ebene zum Ursprung.

Nach Abbildung 2.3 folgt

$$d = p - \mathbf{P}n_e \quad \mathbf{R} = \mathbf{P} + 2dn_e \quad (2.1)$$

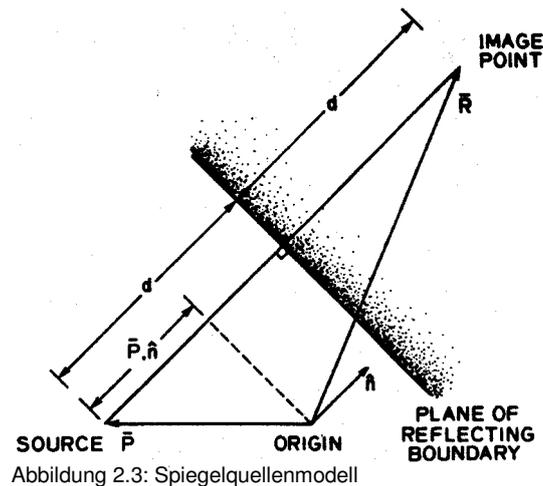


Abbildung 2.3: Spiegelquellenmodell

### Strahlenmodell:

Bei dieser Methode werden Schallstrahlen in mehreren Richtungen verfolgt, bis sie nach mehreren Reflexionen entweder auf den Hörer treffen, oder der Schallweg ein vorher definiertes Maximum überschreitet.

Der Vorteil dieser Methode ist die Einfachheit und Recheneffizienz bei der Implementation.

Der Hauptnachteil liegt in der Ungenauigkeit, da nur in endlich viele diskrete Richtungen Schallstrahlen ausgesandt werden können.

Daraus folgt, dass entweder ausschlaggebende Reflexionen, die den Hörer treffen fehlen, oder höhere Reflexionen durch Reflexionen niedrigerer Ordnung abgeschattet werden.

Solche Fehler können nur vermindert werden, indem das räumliche Inkrement verkleinert wird und auch die Verfolgung des Schallstrahls, nachdem er den Hörer getroffen hat, fortgesetzt wird. Solche Erweiterungen des Grundalgorithmus führen aber zu einer beträchtlichen Verlängerung der Rechenzeit.

Eine Verbesserung dieser Methode führte zur Schallteilchenmethode:

Der einzige Unterschied zum oben beschriebenen Verfahren liegt darin, dass Schallteilchen im Gegensatz zu Schallstrahlen emittiert werden.

Die Teilchen werden gleichmäßig verteilt verschickt und am Ort des Hörers detektiert. Da die Wahrscheinlichkeit, dass ein Teilchen einen ‚punktförmigen‘ Hörer trifft sehr gering ist wird dem Hörer ein Volumen zugeordnet. Nach jeder Reflexion wird kontrolliert, ob der Weg des Teilchens das kugelförmige Volumen des Hörers schneidet. Wenn ein Teilchen dieses Volumen trifft, wird aus der Richtung, der zurückgelegten Weglänge, und der Energie des erfolgreichen Teilchens die Position und Energie einer äquivalenten virtuellen Schallquelle berechnet. Wenn das Teilchen einen bestimmten Weg, eine bestimmte Anzahl an Reflexionen, oder eine bestimmte Energie erreicht hat, wird die Berechnung abgebrochen. Wenn alle Teilchen - ihre Anzahl hängt von der gewünschten räumlichen Auflösung ab - auf diese Art verarbeitet wurden, wird aus den resultierenden virtuellen Quellen die Raumimpulsantwort berechnet.

## 3 Richtstrahlbildung – Räumliche Filterung mittels einer Sensorgruppe<sup>1</sup>

### 3.1 Einführung

Das Wort ‚Beamforming‘ (Richtstrahlbildung) stammt daher, dass frühe räumliche Filter so dimensioniert wurden, dass Signale aus einer bestimmten Richtung empfangen und der Rest abgeschwächt werden sollte ( to form Pencil Beams). Das Wort Richtstrahler weist auf die Abstrahlung von Energie hin. Richtstrahlbildung ist aber sowohl auf die Abstrahlung als auch auf den Empfang von Energie anwendbar. Im Folgenden geht es um die Richtstrahlbildung für den Empfang von Energie.

Systeme, die dazu bestimmt sind räumlich verteilte Signale zu empfangen, berücksichtigen häufig die Gegenwart von Stör- oder Interferenzsignalen. Wenn das gewünschte Signal und die Störung das selbe Frequenzband belegen, führt zeitliche Filterung nicht zum Ziel Signal und Störer zu trennen. Störsignal und gewünschtes Signal kommen jedoch aus verschiedenen Richtungen. Diese räumliche Trennung kann jetzt verwendet werden, um mit Hilfe eines räumlichen Filters Interferenz und Signal zu trennen. Bei der Implementation eines zeitlichen Filters ist es notwendig das Signal über einen bestimmten Zeitraum zu beobachten. Analog dazu verlangt die Implementation eines räumlichen Filters das Verarbeiten von Daten, die von einer räumlichen Sensoranordnung stammen.

Wenn die räumliche Abtastung diskret erfolgt (z.B. durch eine Gruppe von Mikrofonen) bezeichnet man den Prozessor, der die räumliche Filterung durchführt, einen Beamformer.

Ein Beamformer berechnet eine Linearkombination von räumlich und zeitlich abgetasteten Sensorsignalen in der gleichen Weise wie ein FIR Filter eine Linearkombination von zeitlich abgetasteten Signalen liefert. Die räumliche Filterung mit einer Sensorgruppe hat zwei prinzipielle Vorteile, die im Folgenden beschrieben werden.

Die räumliche Auflösung hängt von der räumlichen Ausdehnung der Sensorgruppe ab; wird die Ausdehnung größer, verbessert sich die räumliche Auflösung. Die absolute Größe der Ausdehnung der Sensorgruppe ist nicht entscheidend, vielmehr ihr Verhältnis zur Wellenlänge ist der kritische Parameter.

Eine einzelne physikalische Antenne (kontinuierliche räumliche Ausdehnung), die die gewünschte räumliche Auflösung liefert, ist oft praktisch für hohe Frequenzen. Wenn hingegen tiefe Frequenzen von Interesse sind, kann mit einer Gruppe von Sensoren leichter eine große Ausdehnung erreicht werden als dies mit einer einzelnen Antenne möglich wäre.

Ein zweiter entscheidender Vorteile bei Gebrauch einer Sensorgruppe, der für jede Wellenlänge gilt, ist die Vielseitigkeit des räumlichen Filters welche sich durch die

---

<sup>1</sup> [30], [31], [34], [9 bis 17]

diskrete Abtastung ergibt. In vielen Anwendungen ist es erforderlich die Eigenschaften des Filters in Echtzeit zu verändern, um eine effiziente Unterdrückung von Störsignalen zu gewährleisten. Diese Änderung kann in einem diskreten System durchgeführt werden, indem die Art und Weise wie die Sensordaten linear kombiniert werden, beeinflusst wird. Ein Wechsel der Filterfunktion bei einer kontinuierlichen Antenne ist hingegen praktisch schwierig durchzuführen.

### 3.2 Grundlegende Begriffsdefinitionen

#### 3.2.1 Beamforming (Richtstrahlbildung) und räumliche Filterung:

Abbildung 3.1 zeigt zwei Beamformer. Der Erste, der die fortschreitende Welle räumlich abtastet, wird für die Verarbeitung von schmalbandigen Signalen eingesetzt. Der Ausgang zum Zeitpunkt  $k$ ,  $y(k)$ , wird durch eine Linearkombination der Daten an den  $J$  Sensoren zum Zeitpunkt  $k$  festgelegt:

$$y(k) = \sum_{l=1}^J w_l^* x_l(k) \quad (3.1)$$

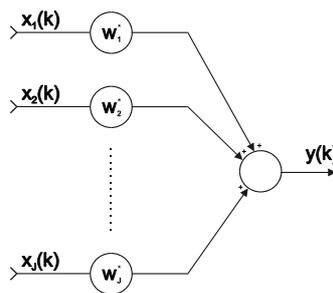


Abbildung 3.1a) Beamformer für Schmalbandsignale. Jeder Sensorausgang wird mit einem komplexen Gewicht multipliziert.

Es ist üblich die Daten mit den konjugierten Gewichten zu multiplizieren, um die Schreibweise zu vereinfachen. Wir nehmen weiters an, dass die Daten und Gewichte komplex sind, da in vielen Anwendungen ein Empfänger verwendet wird, der ‚in phase‘ und ‚quadrature‘ Daten erzeugt (I und Q Daten). Es wird weiters angenommen, dass jeder Sensor die notwendige Empfängerelektronik und einen A/D Konverter beinhaltet.

Der zweite Beamformer in Abbildung 3.1 tastet das fortschreitende Wellenfeld zeitlich und räumlich ab, und wird benutzt, wenn breitbandige Signale von Interesse sind. Der Ausgang kann in diesem Fall durch folgende Gleichung beschrieben werden:

$$y(k) = \sum_{l=1}^J \sum_{p=0}^{K-1} w_{l,p}^* x_l(k-p). \quad (3.2)$$

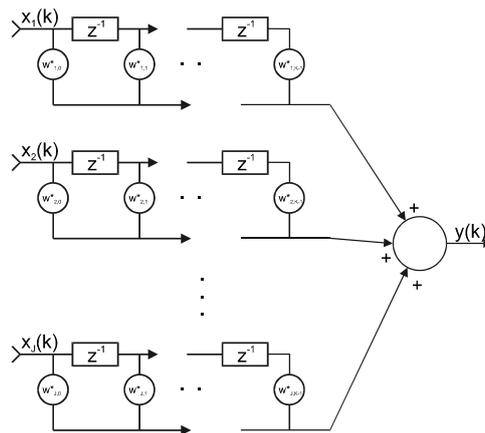


Abbildung 3.1b) Beamformer für Breitbandsignale. Jedem Sensorausgang folgt ein FIR Filter.

K-1 beschreibt hier die Anzahl der Delays pro Sensorkanal. Wenn das Signal an jedem Sensor als Eingang betrachtet wird, stellt ein Beamformer ein Mehrfacheingang und Einzelausgang System da.

Folgende Schreibweise umfasst sowohl den Fall für einen Schmalband und einen Breitbandbeamformer. Gleichung 3.1 und 3.2 können geschrieben werden als:

$$y(k) = \mathbf{w}^H \mathbf{x}(k). \quad (3.3)$$

Fettgeschriebene klein und großgeschriebene Buchstaben bezeichnen im Folgenden Vektoren bzw. Matrizen. Das hochgestellte H steht für Hermitesch (= komplex konjugiert transponiert). Nehmen wir an  $\mathbf{w}$  und  $\mathbf{x}$  sind N dimensional. Daraus folgt, dass  $N=KJ$  in Hinblick auf Gleichung 3.2 und  $N=J$  in Hinblick auf Gleichung 3.1.

Die Frequenzantwort eines FIR Filters mit den Gewichten  $w_p^*$ ,  $0 \leq p \leq J$ , und einer Zeitverzögerung von T Sekunden pro Verzögerungsglied ergibt sich zu:

$$r(\omega) = \sum_{p=1}^J w_p^* e^{-j\omega T(p-1)}$$

oder

$$r(\omega) = \mathbf{w}^H \mathbf{d}(\omega).$$

mit

$$\mathbf{w}^H = [w_1^* w_2^* \dots w_J^*]$$

und

$$\mathbf{d}(\omega) = [1, e^{j\omega T}, e^{j2\omega T}, \dots, e^{j(J-1)\omega T}]^H \quad (3.4)$$

$r(\omega)$  stellt hier die Antwort des Filters auf einen komplexen Sinus der Frequenz  $\omega$  dar, und  $\mathbf{d}(\omega)$  ist ein Vektor, der die Phase des komplexen Sinus an jeder Stufe im FIR Filter relativ zur Stufe  $w_1$  beschreibt.

Die Antwort des Beamformer ist definiert als die Amplituden und Phasenänderung, die auf eine komplexe ebene Welle als Funktion des Ortes und der Frequenz wirkt.

Abbildung 3.2 zeigt wie eine Sensorgruppe ein ebene Welle räumlich abtastet. Hier nehmen wir an, dass das einfallende Signal eine ebene komplexe Welle mit dem Einfallswinkel  $\theta$  und der Frequenz  $\omega$  ist. Die Phase am ersten Sensor wird mit Null angenommen. Daraus folgt  $x_1(k) = e^{j\omega k}$  und  $x_l(k) = e^{j\omega[k-\Delta_l(\theta)]}$  mit  $2 \leq l \leq J$ .  $\Delta_l(\theta)$  ist hier die Zeitverzögerung infolge der Ausbreitung vom ersten bis zum  $l$ -ten Sensor. Für den Ausgang des Beamformers kann man jetzt schreiben:

$$y(k) = e^{j\omega k} \sum_{l=1}^J \sum_{p=0}^{K-1} w_{l,p}^* e^{-j\omega[\Delta_l(\theta)+p]} = e^{j\omega k} r(\theta, \omega) \quad (3.5)$$

Hier gilt  $\Delta_1(\theta)=0$ .  $r(\theta, \omega)$  ist die Antwort des Beamformers und kann in Vektorform in der Form

$$r(\theta, \omega) = \mathbf{w}^H \mathbf{d}(\theta, \omega) \quad (3.6)$$

ausgedrückt werden.

Die Elemente von  $\mathbf{d}(\theta, \omega)$  sind die komplexen Exponentialfunktionen  $e^{-j\omega[\Delta_l(\theta)+p]}$ . Im Allgemeinen kann man schreiben:

$$\mathbf{d}(\theta, \omega) = [1 \quad e^{j\omega\tau_2(\theta)} \quad e^{j\omega\tau_3(\theta)} \quad \dots \quad e^{j\omega\tau_N(\theta)}]^H \quad (3.7)$$

Hier sind  $\tau_i(\theta), 2 \leq i \leq N$ , die Zeitverzögerungen infolge der Ausbreitung in Bezug auf den Referenzpunkt mit der Phase Null.

Der Vektor  $\mathbf{d}(\theta, \omega)$  wird in der Folge als Arrayresponse-Vektor bezeichnet. In der Literatur findet man auch die Namen ‚steering vector‘ und ‚direction vector‘.

Als ‚beampattern‘ definiert man das Amplitudenquadrat von  $r(\theta, \omega)$ . Man beachte, dass jedes Gewicht in  $\mathbf{w}$  sowohl die zeitliche als auch räumliche Antwort des Beamformers beeinflusst.

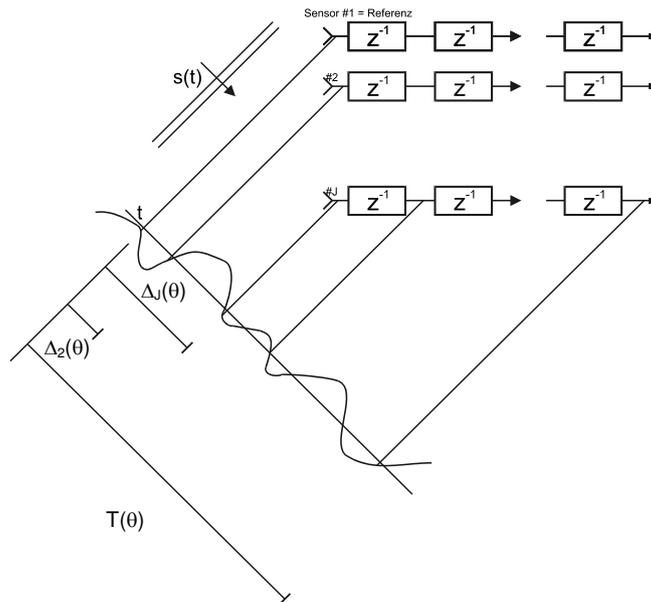


Abbildung 3.2: Diese Abbildung illustriert die Abtastung einer ebenen Welle aus der Richtung  $\theta$ . Mit  $J$  Sensoren und  $K$  Verzögerungsstufen pro Sensor wird das Signal an  $N=JK$  ungleichmäßig verteilten Punkten abgetastet.  $T(\theta)$  ist die Dauer vom ersten Sample des ersten Sensors bis zum letzten Sample des letzten Sensors und wird als zeitliche Öffnung bezeichnet. Die zeitliche Öffnung ist eine Funktion des Einfallswinkel  $\theta$ . Ebene Wellenausbreitung heisst, daß das Signal zu jedem Zeitpunkt  $k$  auf einer Wellenfront normal zur Ausbreitungsrichtung die gleiche Intensität hat. Die Ausbreitung des Signals im Raum wird daher durch eine reine Zeitverzögerung charakterisiert. In dieser Abbildung stellt  $\Delta_i(\theta)$  die Verzögerung der Welle vom Referenzsensor #1 bis zum Sensor #i dar.

Der Zusammenhang zwischen einer FIR Filterung und Beamforming ist am ersichtlichsten, wenn der Beamformer auf eine einzelne zeitliche Frequenz  $\omega_0$  wirkt, und die Sensorgeometrie linear und äquidistant ist (Abbildung 3.3).

Der Abstand der Sensoren sei  $d$ , die Ausbreitungsgeschwindigkeit sei  $c$ ,  $\theta$  ist der Einfallswinkel relativ zur Breitseite der Gruppe (‚broadside‘ im Gegensatz zu ‚end fire‘). Bei  $0^\circ$  blickt man normal auf die lineare Gruppe. Daraus ergibt sich  $\tau_i=(i-1)(d/c)\sin\theta$ . In diesem Fall ergibt sich der Zusammenhang zwischen der zeitlichen Frequenz  $\omega$  in  $\mathbf{d}(\omega)$  (FIR-Filter) und Richtung  $\theta$  in  $\mathbf{d}(\theta, \omega_0)$  (beamformer) zu  $\omega=\omega_0(d/c)\sin\theta$ . Daraus erkennt man, dass die zeitliche Frequenz in einem FIR Filter dem Sinus der Einfallsrichtung in einem schmalband, linear-äquidistanten Beamformer entspricht.

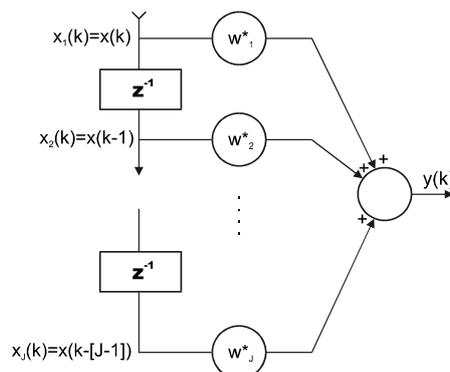


Abbildung 3.3a: Analogie einer äquidistanten, omnidirektionalen und linearen Sensorgruppe (hinten), und einem FIR Filter (oben).

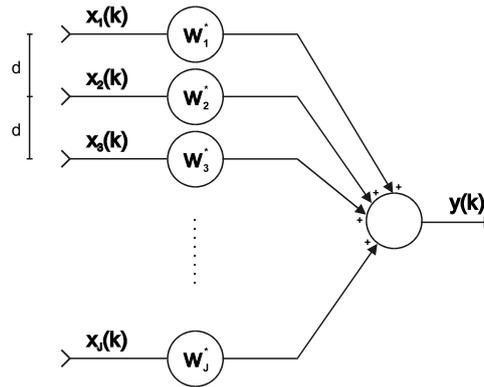


Abbildung 3.3b: Analogie einer äquidistanten, omnidirektionalen und linearen Sensorgruppe (oben), und einem FIR Filter (vorne).

Aus der Notation in Gleichung (3.6) lässt sich eine Vektorrauminterpretation für die Richtstrahlbildung ableiten. Diese Sichtweise ist sowohl beim Beamformertwurf als auch bei der Analyse nützlich. Der Gewichtsvektor  $\mathbf{w}$  und die Arrayresponsevektoren  $\mathbf{d}(\theta, \omega)$  sind Vektoren in einem  $N$ -dimensionalen Vektorraum. Die Winkel zwischen  $\mathbf{w}$  und  $\mathbf{d}(\theta, \omega)$  bestimmen die Antwort  $r(\theta, \omega)$ . Wenn zum Beispiel der Winkel zwischen  $\mathbf{w}$  und  $\mathbf{d}(\theta, \omega)$   $90^\circ$  beträgt ( $\mathbf{w}$  ist orthogonal zu  $\mathbf{d}(\theta, \omega)$ ), ist die Antwort gleich Null. Ist der Winkel in der Nähe von  $0^\circ$ , wird die Amplitude der Antwort relativ groß sein. Die Fähigkeit des Beamformers zwischen einzelnen Richtungen und/oder Frequenzen  $(\theta_1, \omega_1)$  und  $(\theta_2, \omega_2)$  zu unterscheiden wird durch den Winkel Ihrer Arrayresponsevektoren  $\mathbf{d}_1(\theta, \omega)$  und  $\mathbf{d}_2(\theta, \omega)$  bestimmt.

Die allgemeinen Effekte der räumlichen Abtastung sind ähnlich zur Abtastung im Zeitbereich. Räumliches Aliasing entspricht einer Mehrdeutigkeit der Orte der Schallquellen. Dieser Fall tritt auf, wenn Quellen an verschiedenen Orten den selben Arrayresponsevektor haben. (für Schmalbandquellen:  $\mathbf{d}(\theta_1, \omega_0) = \mathbf{d}(\theta_2, \omega_0)$ ). Das kann vorkommen wenn die Sensoren zu weit von einander positioniert sind. Sind die Sensoren einander zu Nahe, leidet die räumliche Auflösung infolge der zu kleinen Ausdehnung der Sensorgruppe. ( $\rightarrow$  die Arrayresponsevektoren sind im  $N$ -dimensionalen Vektorraum schlecht verteilt).

Ein weiterer Fall von Mehrdeutigkeit tritt bei Breitbandsignalen auf, wenn eine Quelle an einem Ort mit einer bestimmten Frequenz, nicht von einer Quelle an einem anderem Ort mit unterschiedlicher Frequenz unterschieden werden kann:  $\mathbf{d}(\theta_1, \omega_1) = \mathbf{d}(\theta_2, \omega_2)$ . Dieser Fall tritt bei einem linearen äquidistanten Beamformer immer dann auf wenn  $\omega_1 \sin \theta_1 = \omega_2 \sin \theta_2$ .

Das Hauptziel in der Folge wird der Entwurf einer Beamformerantwort durch Wahl der Gewichte sein. Gleichung (3.6) zeigt aber auch, dass die Antwort von der Geometrie der Sensoranordnung abhängt. (Und von der Sensorcharakteristik, falls das Prinzip der idealen kugelförmigen Richtcharakteristik verletzt wird).

Im Gegensatz zur Einkanalfilterung, wo A/D Umsetzer eine gleichmäßige zeitliche Abtastung vorgeben, besteht kein zwingender Grund die Sensoren im Raum gleichmäßig zu positionieren. Die Sensorpositionen liefern beim Entwurf der Antwort zusätzliche Freiheitsgrade und können so gewählt werden, dass über einen

gewünschten Bereich  $(\theta, \omega)$  die Arrayresponsevektoren eindeutig und gut im N-dimensionalen Vektorraum verteilt sind.

### 3.2.2 Statistik Zweiter Ordnung:

Die Beurteilung der Leistungsfähigkeit eines Beamformer nimmt Bezug auf die Leistung oder Varianz. In dieser Hinsicht spielt die Statistik (zweiter Ordnung) der Sensordaten eine wichtige Rolle. Wir nehmen an, dass die Daten, die an den Sensoren empfangen werden den Mittelwert Null haben. Die Varianz oder erwartete Leistung am Ausgang des Beamformer ist gegeben durch

$$E\{|y|^2\} = \mathbf{w}^H E\{\mathbf{x}\mathbf{x}^H\} \mathbf{w} \quad (3.8)$$

Sind die Daten stationär, so ist die Kovarianzmatrix  $\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\}$  unabhängig von der Zeit. Obwohl die Daten oft nicht stationär sind, wird die Annahme der Stationarität dazu verwendet statistisch optimale Beamformer zu entwerfen, oder den eingeschwungenen Zustand zu beurteilen.

Nehmen wir an  $\mathbf{x}$  repräsentiert äquidistante zeitliche Abtastwerte mit einem Leistungsdichtespektrum  $S(\omega)$  und hat keine Energie außerhalb des spektralen Bandes  $[\omega_a, \omega_b]$ .  $\mathbf{R}_x$  kann in Termen des Leistungsdichtespektrums ausgedrückt werden, indem man die Beziehungen der Fouriertransformation anwendet:

$$\mathbf{R}_x = \frac{1}{2\pi} \int_{\omega_a}^{\omega_b} S(\omega) \mathbf{d}(\omega) \mathbf{d}(\omega)^H d\omega. \quad (3.9)$$

Die Daten an den Sensoren sollen von einer Quelle in der Richtung  $\theta$  stammen. Für die Kovarianzmatrix gilt jetzt:

$$\mathbf{R}_x = \frac{1}{2\pi} \int_{\omega_a}^{\omega_b} S(\omega) \mathbf{d}(\theta, \omega) \mathbf{d}(\theta, \omega)^H d\omega. \quad (3.10)$$

Eine Quelle kann als schmalbandig mit der Frequenz  $\omega_0$  bezeichnet werden, wenn für  $\mathbf{R}_x$  gilt:

$$\mathbf{R}_x = \sigma_s^2 \mathbf{d}(\theta, \omega_0) \mathbf{d}^H(\theta, \omega_0) \quad (3.11)$$

$\sigma_s^2$  ist hier die Varianz der Quelle bzw. die Leistung.

Die Bedingungen, unter denen eine Quelle als schmalbandig betrachtet werden kann, hängen sowohl von der Bandbreite der Quelle, als auch von der Zeit, in der die Quelle beobachtet wird ab. Als Parameter zur Entscheidung, ob eine Quelle als schmalbandig betrachtet werden kann, dient das Zeit-Bandbreite Produkt. (TBWP-Time Bandwidth Product).

Eine Sensorgruppe bietet eine bestimmte zeitliche Öffnung für eine Quelle aus einer bestimmten Richtung. Die zeitliche Öffnung ist abhängig vom Einfallswinkel  $\theta$ . Daraus folgt, dass auch das TBWP vom Einfallswinkel (=DOA-Direction of Arrival) abhängig ist. Die Annahme eines Schmalbandbeamformer gelten, wenn das TBWP für die beobachtete Richtung viel kleiner ist als eines für alle anderen möglichen Richtungen.

Eine Sensorgruppe mit einer nachfolgenden Gruppe von gewichteten Verzögerungsgliedern ermöglicht eine räumliche und zeitliche Filterung einer fortschreitenden Welle. Die Abbildung 3.2 zeigt das Abtasten einer fortschreitenden ebenen Welle aus der Richtung  $\theta$ . Mit  $J$  Sensoren und  $K$  Verzögerungsgliedern pro Sensor wird das Signal an  $N=JK$  ungleichmäßig verteilten Punkten abgetastet.  $T(\theta)$  ist die Zeitdauer vom ersten Abtastwert des ersten Sensors bis zum letzten Abtastwert des letzten Sensors und wird als zeitliche Öffnung oder Ausdehnung für die Richtung  $\theta$  bezeichnet. Wie aus der Notation ersichtlich, ist diese Zeit abhängig vom Einfallswinkel  $\theta$ . Für die Ausbreitung einer Ebenen Welle gilt, dass die Intensität in Abhängigkeit von der Entfernung zur Quelle konstant bleibt. Daraus folgt, dass die Ausbreitung der Welle im Raum durch eine reine Zeitverzögerung beschrieben werden kann. In dieser Abbildung ist  $\Delta_l(\theta)$  die Zeitdauer infolge der Ausbreitung der ebenen Welle vom ersten Sensor(Referenzpunkt) bis zum  $l$ -ten Sensor.

### 3.3 Klassifikation von Richtstrahlern

Beamformer können unterteilt werden in datenunabhängige oder statistisch optimale, abhängig davon, in welcher Weise die Gewichte für die anschließende FIR Filterbank berechnet werden.

Die Gewichte eines datenunabhängigen Richtstrahlers sind nicht abhängig von den Eingangsdaten und werden gewählt, um eine spezifizierte Antwort für alle Signal/Stör Szenarien zu erreichen.

Die Gewichte eines statistisch optimalen Beamformers werden auf Basis der statistischen Eigenschaften der Eingangsdaten berechnet und so gewählt, dass die Antwort der Sensorgruppe optimiert wird. Im Allgemeinen wird die Antwort der Sensorgruppe für Richtungen, aus denen Störsignale erwartet werden auf Null gesetzt, in der Absicht das Signal/Rausch Verhältnis am Ausgang des Beamformers zu verbessern. Weiters sind die statistischen Eigenschaften der Sensordaten a priori nicht bekannt, und sie ändern sich auch mit der Zeit. In diesem Fall werden adaptive Algorithmen angewandt, um die Gewichte zu bestimmen. Ein solcher Algorithmus wird so bestimmt, dass die Antwort der Sensorgruppe nach einer bestimmten Zeit zu einer statistisch optimalen Lösung konvergiert.

Datenunabhängige Richtstrahler:

Die Gewichte eines datenunabhängigen Beamformers werden so bestimmt, dass die Antwort der Sensorgruppe, unabhängig von den Eingangsdaten oder der Datenstatistik eine bestimmte gewünschte Antwort approximiert.

Der Ansatz - Approximation einer bestimmten Frequenzantwort - ist vergleichbar mit klassischen FIR Filterdesign. Man kann daher auch Analogien zum FIR Filterdesign herstellen

Zunächst wird der klassische Fall betrachtet, dass die Antwort aus einer bestimmten Richtung 1, und aus anderen Richtungen 0 sein soll (forming pencil beams).

### **3.3.1 Klassische Richtstrahlbildung:**

Bei einem einfachen FIR Filter ist es beispielsweise erwünscht eine einzelne komplexe Frequenzkomponente von anderen Frequenzen mittels einer gewichteten Verzögerungskette (J Stufen) , vgl. Abb. 3.3, zu trennen. Wenn die Frequenz  $\omega_0$  von Interesse ist, ist die gewünschte Frequenzantwort bei  $\omega_0$  gleich 1, und anderswo gleich 0. Eine mögliche Lösung für dieses Problem findet man bei der Wahl  $\mathbf{w}=\mathbf{d}(\omega_0)$ . Dieser Ansatz liefert den kleinsten quadratischen Fehler zwischen der tatsächlichen und der gewünschten Frequenzantwort. Die tatsächliche Antwort wird durch eine Hauptkeule und eine Reihe von Nebenkeulen charakterisiert. Da  $\mathbf{w}=\mathbf{d}(\omega_0)$  ist, hat jedes Element von  $\mathbf{w}$  den Betrag 1. Wird über diese Gewichte ein Fenster gelegt, kann die Form der Hauptkeule und ihr Verhältnis zu den Nebenkeulen beeinflusst werden.

Bei räumlicher Filterung ist es erwünscht ein Signal aus einer bestimmten Richtung  $\theta_0$  zu empfangen. Wenn das Signal schmalbandig ist (Frequenz  $\omega_0$ ) kann für den Gewichtsvektor der Arrayresponsevektor  $\mathbf{d}(\theta_0, \omega_0)$  gewählt werden. Diese Sensorgruppe wird als 'phased array' bezeichnet, da jeder Sensoreingang vor der Summation phasenverschoben wird.

Wie beim zuvor beschriebenen FIR Filter sind die Breite der Keule und die Pegel der Nebenkeulen wichtige Charakteristiken des Beamformer. Gewichtung der Amplituden der Gewichte kann die Form der Antwort beeinflussen (-form the beam).

Die Äquivalenz der linear-äquidistanten Sensorgruppe für Schmalbandsignale mit einem gewöhnlichen FIR Filter führt dazu, dass die selben Techniken zur Wahl der Fensterfunktionen auf beide Probleme angewandt werden können.

Hat man es mit einem Breitbandsignal zu tun, kann die Gewichtung mit einer Fensterfunktion unabhängig voneinander auf den Ausgang der Sensoren und die FIR Filter angewandt werden.(Abbildung 3.4).

Die Gewichte  $t_j$  formen die räumliche Antwort, die FIR Koeffizienten bestimmen ein gewünschtes zeitliches Verhalten.

Wie schon erwähnt, stehen die zeitliche und räumliche Antwort in einer Wechselwirkung, sodass sie nicht vollständig unabhängig voneinander synthetisiert werden können.

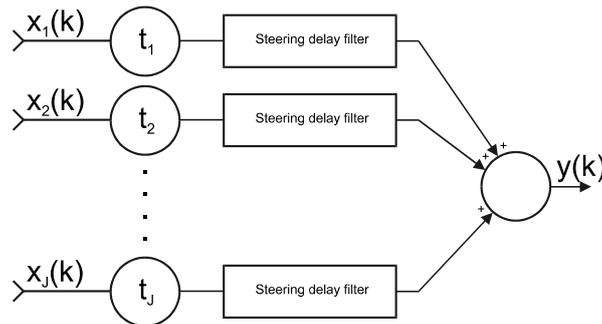


Abbildung 3.4: Darstellung eines klassischen Richtstrahlers für Breitbandssignale. Beim klassischen Beamformer (im speziellen für lineare Gruppen) bestimmen die Gewichte  $\{t_1, t_2, \dots, t_J\}$  die Form des Beampattern. Filter steuern den Richtstrahler in eine bestimmte Richtung indem Verzögerungszeiten für eine Quelle aus einer bestimmten Richtung kompensiert werden.

### 3.3.2 Entwurf einer allgemeinen Frequenzantwort eines datenunabhängigen Richtstrahlers:

Im folgenden Abschnitt werden Beamformer betrachtet, die eine beliebige gewünschte Frequenzantwort approximieren. Es ist zum Beispiel gewünscht, Signale aus einem Bereich von Richtungen zu empfangen, und aus anderen Bereichen zu unterdrücken. Dieses Beispiel ist mit einem Bandpass oder einer Bandsperre bei einem FIR Filter zu vergleichen. Obwohl in diesem Zusammenhang keine Richtstrahlen (beams) im eigentlichen Sinn mehr geformt werden, spricht man bei dieser Art von räumlicher Filterung auch von Richtstrahlbildung oder Beamforming.

Man wähle die Gewichte  $\mathbf{w}$  des Beamformers so, dass die tatsächliche Frequenzantwort  $r(\theta, \omega) = \mathbf{w}^H \mathbf{d}(\theta, \omega)$ , eine gewünschte Frequenzantwort  $r_d(\theta, \omega)$  approximiert. Im Folgenden erfolgt die Berechnung der Gewichte durch Minimierung der  $L_2$  Norm, dh. die Minimierung des mittleren quadratischen Fehlers zwischen tatsächlicher und gewünschter Frequenzantwort an  $P$  Punkten  $(\theta_i, \omega_i)$  mit  $1 \leq i \leq P$ . Ist  $P > N$ , dann erhalten wir ein überbestimmtes Gleichungssystem

$$\min_{\mathbf{w}} \|\mathbf{A}^H \mathbf{w} - \mathbf{r}_d\|^2 \quad (3.12)$$

hier ist

$$\mathbf{A} = [\mathbf{d}(\theta_1, \omega_1) \quad \mathbf{d}(\theta_2, \omega_2) \quad \dots \quad \mathbf{d}(\theta_P, \omega_P)]$$

und

$$\mathbf{r}_d = [r(\theta_1, \omega_1) \quad r(\theta_2, \omega_2) \quad \dots \quad r(\theta_P, \omega_P)]^H$$

Wenn  $\mathbf{A}\mathbf{A}^H$  invertierbar ist, ergibt sich die Lösung zu

$$\mathbf{w} = \mathbf{A}^+ \mathbf{r}_d \quad (3.13)$$

In dieser Gleichung ist  $\mathbf{A}^+ = (\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{A}$  die Pseudoinverse von  $\mathbf{A}$ .

Ein Augenmerk muss man auf die Verstärkung des Beamformer bei weißem Rauschen legen (white noise gain). Diese ist definiert als die Ausgangsleistung des Beamformer infolge von weißen Rauschen mit der Varianz eins.

Das Betragsquadrat des Gewichtsvektors,  $\mathbf{w}\mathbf{w}^H$ , repräsentiert diese Verstärkung. Wenn diese Verstärkung groß ist, ist die Genauigkeit, mit der die gewünschte Frequenzantwort approximiert wird, ein nicht entscheidender Punkt, da der Ausgang infolge der starken Verstärkung des weißen Rauschen ein sehr kleines Signal/Rausch Verhältnis hat.

Bestimmte Richtungen und Frequenzen können durch die Auswahl der Abtastwerte  $(\theta_i, \omega_i)$ , und/oder durch unterschiedliche Gewichtung des Fehlers an jedem Punkt  $(\theta_i, \omega_i)$  betont werden.

### **3.4 Modellierung der Aussenohrübertragungsfunktion unter Verwendung des Prinzips der Richtstrahlbildung (Beamforming)**

#### **3.4.1 Einführung:**

Bei diesem Modell wird im Gegensatz zu physiologischen Modellen nicht versucht den Mechanismus des Ohres nachzubilden, sondern man will direkt die Übertragungseigenschaften des Außenohrs modellieren - ein funktioneller Ansatz.

Im Folgenden wird das Konzept der Richtstrahlbildung beschrieben. Ein Beamformer ist ein räumlicher und zeitlicher Filter, der aus Eingangsdaten, die von im Raum verteilten Sensoren stammen, eine gewichtete Kombination berechnet. Die Gewichte der einzelnen Filterstufen bestimmen das räumliche und zeitliche Verhalten vollständig. Das Ohr wird als akustische Antenne betrachtet, deren Übertragungseigenschaften wir mittels des Beamformers nachzubilden versuchen. Es wird für eine bestimmte Sensorgeometrie versucht unter Verwendung eines Algorithmus, der den mittleren quadratischen Fehler zwischen der tatsächlichen und gewünschten Frequenzantwort minimiert, die Gewichte zu berechnen. Die Antwort des Modells ergibt sich aus dem In-Produkt von Gewichtsvektor und Arrayresponse-Vektor. Im Arrayresponse-Vektor ist die Sensorgeometrie und die Position der Schallquelle kodiert. Im Folgenden wird die Effektivität des Modells anhand von Versuchen beschrieben, die von Chen, VanVeen, Hecox [20] durchgeführt wurden. Dieses Modell beschreibt die Übertragungsfunktion des Außenohres in Form einer Eingangs-Ausgangsbeziehung. Die physikalischen Eigenschaften werden nur allgemein durch die Sensorgeometrie und FIR Filter in jedem Kanal beschrieben.

Der funktionelle Ansatz dieses Modells hat folgende Vorteile:

1. Es stellt einen expliziten mathematischen Ausdruck dar, der die Richtungsabhängigkeit der Schallwelle am Trommelfell beschreibt.
2. Es repräsentiert die Charakteristik der Außenohren als kontinuierliche Funktion der Richtung, obwohl nur diskrete Messwerte zur Bestimmung der Parameter des Modells verwendet werden.
3. Es können durch Änderung der Gewichte verschiedene Sets von HRTFs modelliert werden (zB.: HRTFs verschiedener Personen)
4. Es führt zu einer substantiellen Datenreduktion, wenn das System dazu benutzt wird HRTFs an vielen Richtungen und Frequenzen darzustellen.

Anmerkungen zur Notation:

Dickgeschriebene kleine und große Buchstaben beschreiben Vektoren bzw. Matrizen. Hochgestellte T und H stehen für transponierte Matrix bzw. hermitesche Matrix (= konj. komplex transponiert). Das Symbol \* steht für konjugiert komplex.

### 3.4.2 Beamformerdesign durch Minimierung des mittleren quadratischen Fehlers (LMS Beamformerdesign):

Ein Beamformer generiert ein einzelnes Ausgangssignal als gewichtete Summe von Signalen, die an mehreren Sensoren empfangen werden. Historisch gesehen stellt ein Beamformer eine Alternative zu einer Antenne mit einer kontinuierlichen Öffnung dar (Schüssel). Der Vorteil eines Beamformers ist, dass die räumliche Antwort durch die Wahl der Gewichte im Gegensatz zur Form einer kontinuierlichen Öffnung gebildet wird. Weiters wird die räumliche Antwort nicht durch die physikalischen Grenzen, die mit der Konstruktion einer kontinuierlichen Form verbunden sind, beeinträchtigt. Die Gewichte der ersten Beamformer wurden so gewählt, dass eine Empfangscharakteristik einer normalen kontinuierlichen Antenne nachgebildet wurde. Ähnlich wird hier versucht die Frequenzantwort des Aussenohres, unser Äquivalent zu einer kontinuierlichen Antenne, nachzubilden.

Die Struktur eines Beamformers, wenn breitbandige Signale von Interesse sind, schaut folgendermaßen aus:

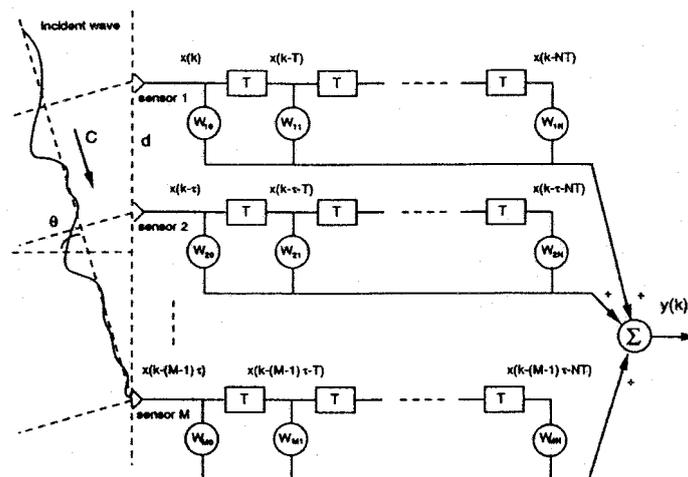


Abbildung 3.5: Struktur eines Richtstrahlers mit FIR Filterkette in jedem Sensorkanal. C steht für die Ausbreitungsgeschwindigkeit der Welle,  $\theta$  ist der Winkel zwischen Ausbreitungsrichtung und der Normalen zur Achse der Sensorgruppe, d ist der Abstand zwischen den Sensoren und T steht für das Abtastintervall.

Die Antwort ist

$$h(\omega, \theta) = \mathbf{w}^H \mathbf{d}(\omega, \theta) \quad (3.14)$$

Den Arrayresponsevector kann man auch schreiben als

$$\begin{aligned} \mathbf{d}^H(\omega, \theta) = & [1, \exp(j\omega T), \dots, \exp(j\omega NT), \exp(j\omega \tau_1), \\ & \exp(j\omega(\tau_1 + T)), \dots, \\ & \exp(j\omega(\tau_1 + NT)), \dots, \exp(j\omega(\tau_{M-1})), \\ & \exp(j\omega(\tau_{M-1} + T)), \dots, \exp(j\omega(\tau_{M-1} + NT))] \end{aligned} \quad (3.15)$$

Der Vektor  $\mathbf{w}^H$  ist

$$\mathbf{w}^H = [\omega_{0,0}^*, \dots, \omega_{0,N}^*, \omega_{1,0}^*, \dots, \omega_{1,N}^*, \omega_{M-1,1}^*, \dots, \omega_{M-1,N}^*] \quad (3.16)$$

In diesen Gleichungen ist N die Ordnung der Filter und T das Abtastintervall. In der folgenden Betrachtung wird  $T=1$  gesetzt. Das Symbol  $\tau_i$  beschreibt die Zeitverzögerung infolge der fortschreitenden Welle vom ersten bis zum (i+1)-ten Sensor. Der erste Index bei der Definition von  $\mathbf{w}$  bezieht sich auf den Sensorkanal, der Zweite auf den Index der Verzögerungsstufe.

Die Frequenzantwort  $h(\omega, \theta)$  ist eine Funktion des Gewichtsvektors  $\mathbf{w}$  und der Art und Weise, wie das fortschreitende Signal durch die Sensorgruppe abgetastet wird. Diese Charakteristik ist abhängig von der Geometrie der Sensorgruppe und spiegelt sich in der Definition des Arrayresponsevektor  $\mathbf{d}(\omega, \theta)$  wider. Im Folgenden wird  $\mathbf{w}$  für eine bestimmte Geometrie berechnet, sodass eine gewünschte Frequenzantwort approximiert wird.

Die gewünschte Frequenzantwort wird als  $h_d(\omega, \theta)$  bezeichnet. In diesem Fall stellt  $h_d(\omega, \theta)$  die Außenohrübertragungsfunktion dar.

Der Fehler zwischen tatsächlicher und gewünschter Frequenzantwort beträgt

$$e(\omega, \theta) = h_d(\omega, \theta) - \mathbf{w}^H \mathbf{d}(\omega, \theta). \quad (3.17)$$

Die Minimierung des mittleren quadratischen Fehlers über einen Bereich  $\omega$  und  $\theta$  wird folgendermaßen beschrieben:

$$\min_{\mathbf{w}} \int_{\theta \in \Theta} \int_{\omega \in \Omega} \gamma(\omega, \theta) |e(\omega, \theta)|^2 d\omega d\theta \quad (3.18)$$

In dieser Gleichung ist  $\gamma(\omega, \theta)$  eine nicht negative Gewichtsfunktionen, die es ermöglicht bestimmte Richtungen und Frequenzen gegenüber anderen zu bevorzugen. In der Praxis sind die Außenohrübertragungsfunktionen nicht als kontinuierliche Funktionen der Richtung und Frequenz gegeben, sondern ergeben sich aus Messwerten für bestimmte diskrete Richtungen  $\theta_i$  mit  $i=1,2,\dots,Q$ , und Frequenzen  $\omega_j$ , mit  $j=1,2,\dots,P$ .

Nun lässt sich das Problem schreiben als

$$\min_{\mathbf{w}} \sum_{i=1}^Q \sum_{j=1}^P \gamma(\omega_j, \theta_i) |e(\omega_j, \theta_i)|^2. \quad (3.19)$$

Diese Gleichung entspricht dem Problem ein System linearer Gleichungen zu lösen.

Man definiert nun einen Fehlervektor:

$$\mathbf{e}^H = [e(\omega_1, \theta_1), e(\omega_2, \theta_1), \dots, e(\omega_p, \theta_1), e(\omega_1, \theta_2), \\ e(\omega_2, \theta_2), \dots, e(\omega_p, \theta_2), \dots, e(\omega_1, \theta_Q), \\ e(\omega_2, \theta_Q), \dots, e(\omega_p, \theta_Q)] \quad (3.20)$$

Es gilt

$$\mathbf{e}^H = \mathbf{h}_d^H - \mathbf{w}^H \mathbf{D}^H \quad (3.21)$$

worin

$$\mathbf{h}_d^H = [h_d(\omega_1, \theta_1), h_d(\omega_2, \theta_1), \dots, h_d(\omega_p, \theta_1), h_d(\omega_1, \theta_2), \\ h_d(\omega_2, \theta_2), \dots, h_d(\omega_p, \theta_2), \dots, h_d(\omega_1, \theta_Q), h_d(\omega_2, \theta_Q), \dots, \\ h_d(\omega_p, \theta_Q),] \quad (3.22)$$

und

$$\mathbf{D}^H = [\mathbf{d}(\omega_1, \theta_1), \mathbf{d}(\omega_2, \theta_1), \dots, \mathbf{d}(\omega_p, \theta_1), \mathbf{d}(\omega_1, \theta_2), \\ \mathbf{d}(\omega_2, \theta_2), \dots, \mathbf{d}(\omega_p, \theta_2), \dots, \mathbf{d}(\omega_1, \theta_Q), \mathbf{d}(\omega_2, \theta_Q), \dots, \\ \mathbf{d}(\omega_p, \theta_Q),]. \quad (3.23)$$

Definition der Gewichtungsmatrix  $\Gamma$ :

Man schreibt die Gewichtungsfaktoren für jede Richtung und Frequenz in eine Diagonalmatrix der Dimension  $PQ \times PQ$ .

$$\Gamma = \text{diag}\{\gamma^{1/2}(\omega_1, \theta_1), \gamma^{1/2}(\omega_2, \theta_1), \dots, \gamma^{1/2}(\omega_p, \theta_Q)\} \quad (3.24)$$

Gleichung (3.18) kann jetzt geschrieben werden als

$$\min_{\mathbf{w}} \|\Gamma(\mathbf{h}_d - \mathbf{D}\mathbf{w})\|_2^2 \quad (3.25)$$

Das entspricht folgendem linearen Gleichungssystem mit  $S=PQ$  linearen Gleichungen in  $L=M(N+1)$  Unbekannten, wobei der mittlere quadratische Fehler minimiert wird:

$$\Gamma \mathbf{D}\mathbf{w} = \Gamma \mathbf{h}_d \quad (3.26)$$

Im Allgemeinen ist  $S > L$ .

Bis jetzt wurde  $\mathbf{w}$  komplexwertig angenommen. Die Frequenzantwort des physikalischen Systems (HRTF-Kurven), ist bezüglich der Frequenzantwort konjugiert komplex symmetrisch:  $h_d(\omega, \theta) = h_d^*(-\omega, \theta)$ .

Daraus folgt, dass wir die Lösung der modellierten Frequenzantwort auch auf konjugiert komplex symmetrische Werte beschränken.  $h^*(\omega, \theta) - h(-\omega, \theta) = 0$ .

Jetzt ist

$$\begin{aligned}
 h^*(\omega, \theta) &= \mathbf{w}^T \mathbf{d}^*(\omega, \theta) \\
 \text{und} & \\
 h(-\omega, \theta) &= \mathbf{w}^H \mathbf{d}(-\omega, \theta).
 \end{aligned}
 \tag{3.27 a,b}$$

Man beachte, dass  $\mathbf{d}^*(\omega, \theta) = \mathbf{d}(-\omega, \theta)$ , sodass folgende Gleichung gilt:

$$h_d(\omega, \theta) - h_d^*(-\omega, \theta) = (\mathbf{w}^T - \mathbf{w}^H) \mathbf{d}(-\omega, \theta) = 0. \tag{3.28}$$

Daraus folgt, dass  $\mathbf{w}^T = \mathbf{w}^H$  und somit  $\mathbf{w}$  reellwertig ist.

Man spaltet nun den Fehlervektor in einen reellen und einen komplexwertigen Teil auf:

$\mathbf{e} = \mathbf{e}_R + j\mathbf{e}_I$ . Wenn der Vektor  $\mathbf{w}$  reellwertig ist, kann man schreiben:

$$\begin{aligned}
 \mathbf{e}_R &= \mathbf{h}_{dR} - \mathbf{D}_R \mathbf{w} \\
 \mathbf{e}_I &= \mathbf{h}_{dI} - \mathbf{D}_I \mathbf{w}
 \end{aligned}
 \tag{3.29 a,b}$$

In dieser Gleichung stellen  $\mathbf{h}_{dR}$ ,  $\mathbf{h}_{dI}$ ,  $\mathbf{D}_R$  und  $\mathbf{D}_I$  die Real bzw. Imaginärteile von  $\mathbf{h}_d$  und  $\mathbf{D}$  dar.

Da  $\|\Gamma \mathbf{e}\|_2^2 = \|\Gamma \mathbf{e}_R\|_2^2 + \|\Gamma \mathbf{e}_I\|_2^2$ , ist Gleichung (3.25) folgendem reellwertigen Problem äquivalent:

$$\min_{\mathbf{w}} \|\mathbf{g} - \mathbf{C}\mathbf{w}\|_2^2 \tag{3.30}$$

In dieser Gleichung ist

$$\begin{aligned}
 \mathbf{g}^T &= [\mathbf{h}_{dR}^T \Gamma \quad \mathbf{h}_{dI}^T \Gamma] \\
 \text{und} & \\
 \mathbf{C}^T &= [\mathbf{D}_R^T \Gamma \quad \mathbf{D}_I^T \Gamma].
 \end{aligned}
 \tag{3.31),(3.32}$$

Zur Lösung dieses Gleichungssystems wird das Prinzip der Singulärwertzerlegung angewandt. (SVD-singular value decomposition).

$\mathbf{C}$  kann geschrieben werden als

$$\mathbf{C} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \tag{3.33}$$

Darin ist  $r$  der Rang von  $\mathbf{C}$ ,  $\mathbf{u}_i$  und  $\mathbf{v}_i$  sind die links bzw. rechtsseitigen Singulärvektoren, und  $\sigma_i$  sind die Singulärwerte. Es sei

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_r \geq \sigma_{r+1} = \dots = \sigma_L = 0$$

Die Lösung für  $\mathbf{w}$  ergibt sich zu

$$\mathbf{w} = \sum_{i=1}^r \sigma_i^{-1} \mathbf{v}_i \mathbf{u}_i^T \mathbf{g}. \quad (3.34)$$

$\mathbf{w}$  hängt also von den inversen Singulärwerten ab.

Die sehr kleinen Singulärwerte beeinflussen das Lösungsergebnis daher sehr stark und führen zu einer großen Norm von  $\mathbf{w}$ . Das wiederum führt zu einer großen 'array gain'. Die Approximation der Frequenzantwort genau an den Stützstellen wird zwar besser, die Interpolation zwischen den Stützstellen sinkt aber stark. Daher werden nur Singulärwerte in die Berechnung einbezogen die einen gewissen zu definierenden Schwellwert nicht unterschreiten. Man berücksichtigt also nur  $\rho$  Singulärwerte.

$$\mathbf{w} = \sum_{i=1}^{\rho} \sigma_i^{-1} \mathbf{v}_i \mathbf{u}_i^T \mathbf{g} \quad (3.35)$$

Der mittlere quadratische Fehler ergibt sich zu:

$$\|\mathbf{g} - \mathbf{C}\mathbf{w}\|_2^2 = \mathbf{g}^T \sum_{i=\rho+1}^{2S} \mathbf{u}_i \mathbf{u}_i^T \mathbf{g}. \quad (3.36)$$

Die Norm von  $\mathbf{w}$  ergibt sich zu:

$$\|\mathbf{w}\|_2^2 = \mathbf{g}^T \sum_{i=1}^{\rho} \sigma_i^{-2} \mathbf{u}_i \mathbf{u}_i^T \mathbf{g}. \quad (3.37)$$

Aus den letzten beiden Gleichungen sieht man, dass der Fehler steigt und die Norm sinkt, wenn man  $\rho$  verkleinert.

### **3.4.3 Beschreibung der Experimente von Chen et al. zur HRTF-Modellierung durch Richtstrahlbildung [20]:**

Die Effektivität des Modells der Richtstrahlbildung wurde anhand der HRTF-Modellierung einer Katze nachgewiesen. Die Außenohrübertragungsfunktionen einer Katze erstrecken sich über einen größeren Frequenzbereich als beim Menschen, weisen aber eine ähnliche Form auf.

Beim Messvorgang wurden die nicht eindeutig definierten Zeitverzögerungen bis zum Beginn der Impulsantwort entfernt. Diese haben keinen Einfluss auf das Amplitudenspektrum sehr wohl aber auf die Phase. Aus diesem Grund ergibt sich eine Mehrdeutigkeit der Phaseninformation. Die Eindeutigkeit der Darstellung der HRTF ist aber von großer Bedeutung. Bei der Modellierung durch Richtstrahlbildung sind eindeutig definierte HRTFs  $\mathbf{h}_d$  der Ausgangspunkt der Berechnung. Es wurden drei verschiedene Methoden zur Beseitigung der Phasenambiguität getestet.

1. Nachbilden einer linearen Phase. Die ursprüngliche Phase wird durch eine Phase, deren Gruppenlaufzeit proportional zum Zentrum des Beamformers ist, ersetzt.
2. Ersetzen der ursprünglichen Phase durch ihre minimalphasige Komponente. Das garantiert eine eindeutige Darstellung der ursprünglichen Übertragungsfunktion mit minimaler Zeitverzögerung.
3. Aufspalten der originalen Phase in eine minimalphasige und eine Allpasskomponente. Die Allpasskomponente wird geglättet und ihr linearphasiger Anteil, der eine reine Zeitverzögerung darstellt, wird entfernt. Die restliche nichtlineare Phasenkomponente wird zur minimalphasigen Komponente addiert.

Für eine gegebene Struktur des Array erhält man die beste Approximation des gewünschten Amplitudenspektrum durch Anwenden der Methode 1. Die gemessene Phase wird jedoch vollkommen ignoriert, und daraus folgt ein großer Unterschied zur Gestalt der gemessenen Impulsantwort. Methode 3 führt zu geringster Verzerrung der Gestalt der Impulsantwort, aber zum größten Fehler zwischen gemessener und modellierter Impulsantwort. Die Anwendung von Methode 2 führt zu einem guten Kompromiss.

### Resultate der Berechnungen von Chen et al.:

Das Modell besteht aus drei Gruppen von Parametern. Einen Teil bildet der Gewichtsvektor, der durch die Lösung eines Fehlerminimierungsproblems berechnet wird. (Minimierung des mittleren quadratischen Fehlers). Der zweite Teil bezieht sich auf die physikalische Struktur der Sensorgruppe. Damit meint man die Geometrie der Anordnung, die Anzahl der Sensoren, die Anzahl der Taps (bzw. Verzögerungsglieder) pro Sensorkanal, den Abstand zwischen den Sensoren (gleiche Abstände angenommen) und die Ausdehnung der Sensorgruppe. Für eine lineare Sensorgruppe ist die Ausdehnung (aperture) definiert als Produkt von (Sensoranzahl-1) und Abstand der Sensoren. Der dritte Teil besteht aus der Gewichtsfunktion  $\Gamma$  und dem Schwellwert  $\delta$ .

Die Leistungsfähigkeit des Modells wird qualitativ und quantitativ ausgewertet. Zur qualitativen Beurteilung werden Grafiken der gemessenen und modellierten Amplituden und Phasengänge beurteilt. Die quantitative Beurteilung erfolgt durch die Berechnung des mittleren Fehlers (average approximation error).

$$\|e\|_2^2 = \frac{1}{Q} \sum_{i=1}^Q \frac{\sum_{j=1}^P |h_d(\omega_j, \theta_i) - \mathbf{w}^H \mathbf{d}(\omega_j, \theta_i)|^2}{\sum_{j=1}^P |h_d(\omega_j, \theta_i)|^2} \quad (3.38)$$

$$\|e_i\|_2^2 = \frac{\sum_{j=1}^P |h_d(\omega_j, \theta_i) - \mathbf{w}^H \mathbf{d}(\omega_j, \theta_i)|^2}{\sum_{j=1}^P |h_d(\omega_j, \theta_i)|^2} \quad (3.39)$$

Gleichung (3.39) beschreibt den mittleren quadratischen Fehler über alle Frequenzen, normiert auf die Gesamtenergie der gemessenen Antwort, für eine Richtung. Gleichung (3.38) ist der mittlere Fehler über alle Richtungen und Frequenzen. Der mittlere Fehler in Prozent wird in der Folge zur Beurteilung des Modells herangezogen.

Die Gewichtsfunktion  $\Gamma$  wird in der Folge für alle Frequenzen auf 1 gesetzt. Der Schwellwert wird so gewählt, dass alle möglichen Singulärwerte zur Berechnung der Gewichte herangezogen werden. Dies führt zum kleinstmöglichen quadratischen Fehler für eine gegebene Sensoranordnung.

### Auswirkungen verschiedener Strukturen der Sensorgruppe:

#### 1. Fixe Anzahl von Gewichten bei Variation der Sensoranzahl und Tapanzahl

Das Außenohr stellt einen räumlichen und zeitlichen Filter dar. Diese Tatsache wird durch die Abbildungen 3.6 und 3.7 veranschaulicht. Die Abbildungen zeigen Sensorgruppen mit 400 Gewichten in verschiedener Sensoranzahl - Tapanzahl Konfiguration. Die Anordnung ist eine lineare Gruppe mit äquidistanten Sensorabständen. ( $d=0.0043m=const.$ ).

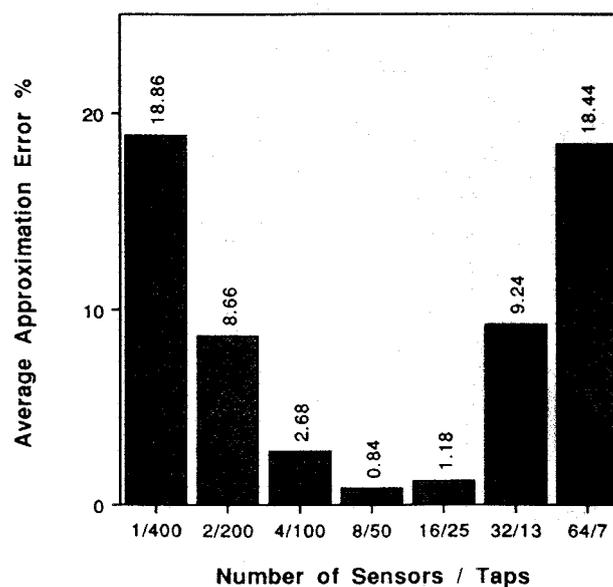


Abbildung 3.6: Mittlerer Fehler als Funktion von Sensoranzahl und Verzögerungsstufen. Die Gesamtanzahl der Gewichte ist konstant 400. Verwendet wurde eine lineare Sensoranordnung.

Die 1 Sensor 400 Tap Konfiguration stellt einen einfachen FIR Filter dar und hat keine Möglichkeit zwischen verschiedenen Richtungen zu unterscheiden. Das Modell liefert nur eine mittlere Übertragungsfunktion und hat mit 19% einen großen Fehler. Die Anordnung mit 2 Sensoren 200 Taps hat einen wesentlich geringeren Fehler mit 8,66%, weil die Unterscheidung zwischen verschiedenen Richtungen jetzt möglich ist. Im Gegensatz dazu hat die Anordnung mit 64 Sensoren und 7 Taps/Sensor wieder einen großen Fehler, weil die zeitliche Öffnung zu gering ist. Die Unterschiede zwischen den einzelnen Frequenzen können nicht mehr erfasst werden. Mit der Anordnung 8 Sensoren 50 Taps erreicht man die beste Approximation der gewünschten Übertragungsfunktion. Dieses Modell hat eine ausreichende zeitliche Öffnung (50 Taps bei einem Abtastintervall von  $12,5\mu\text{s}$  führt zu einem Zeitfenster von  $625\mu\text{s}$ ) und eine ausreichende räumliche Ausdehnung ( $0,0043 \text{ mal } 7$  führt zu einer räumlichen Ausdehnung von  $30\text{mm}$ , die der Größenordnung der Pinna entspricht.)

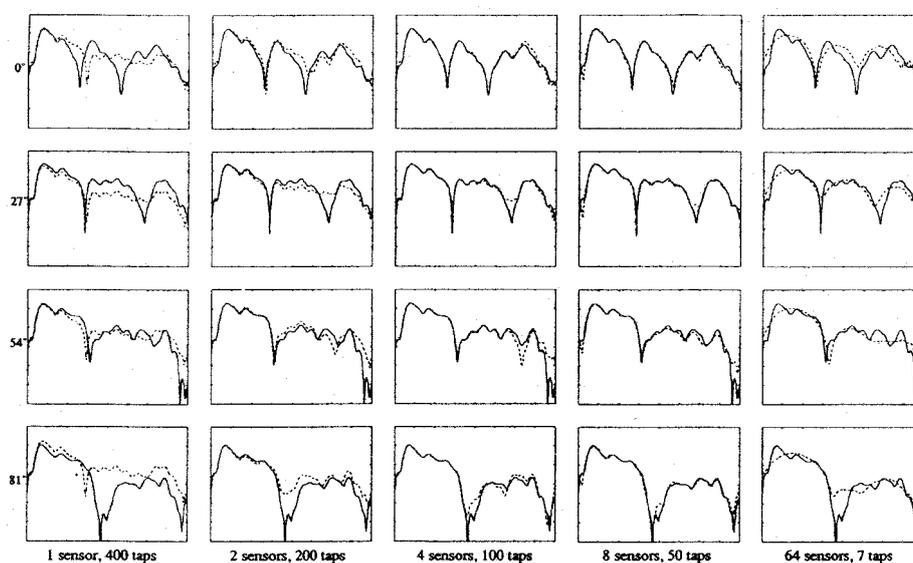


Abbildung 3.7: Gemessene (durchgezogen) und modellierte (strichliert) Frequenzantwort für vier Richtungen mit unterschiedlichen Sensorkonfigurationen. Y-Achse von  $-30\text{dB}$  bis  $30\text{dB}$ . X-Achse von  $0-40\text{kHz}$ .

## 2. Auswirkungen verschiedener Sensoranordnungen:

Da die Geometrie der Anordnung in die Gleichung der Frequenzantwort eingeht, ergibt sich ein komplexer Zusammenhang. Im Folgenden werden nur einige Anordnungen untersucht, die aufgrund der Geometrie der Pinna nahe liegen. Die lineare Gruppe hat Schwierigkeiten mit der Abbildung der gewünschten Frequenzantwort für große Azimutwinkel  $\theta$ . In einer linearen Gruppe hängt die Antwort von der Zeitverzögerung zwischen den Sensoren ab. Diese ist eine Funktion von  $\sin(\theta)$ . Für kleine  $\theta$  ist der Beamformer sensitiv für Richtungsunterschiede, weil  $\sin(\theta) \approx \theta$ . Für  $\theta \text{ ca. } 90^\circ$  ist der Beamformer insensitiv für Richtungsunterschiede, weil  $\sin(\theta) \text{ ca. } 1$ .

Die Pinna selbst ist zweidimensional (bei Betrachtung in der Azimuthalebene) und lässt daher darauf schließen, dass eine L-förmige Anordnung zu besseren Ergebnissen führt.

Ein L-förmiger Sensor hat einen Sensor immer im Schnittpunkt zwischen x und y Achse. Die anderen Sensoren sind auf der x und auf der y Achse verteilt. Die y Achse entspricht dem linearen Array der ersten Versuche. Die x Achse steht normal dazu, dh. sie korrespondiert mit einem Winkel von  $\theta = 0^\circ$ .

Abbildung 3.8 zeigt den durchschnittlichen Fehler als Funktion der Anzahl der Sensoren entlang der beiden Achsen. Die Gesamtanzahl der Sensoren ist hier 11. Jeder Sensorkanal enthält 46 Verzögerungsstufen. Die Größe der Ausdehnung im Falle einer L-förmigen Gruppe ist hier definiert als Distanz zwischen den beiden Enden der L-förmigen Gruppe. Für den Fall einer 5-1-5 Konfiguration ist die Ausdehnung 1,414 mal die Länge eines Arms. Bei gleicher Anzahl von Sensoren in jedem Arm (5 entlang der x Achse, 5 entlang der y Achse, 1 im Schnittpunkt) ist der durchschnittliche Fehler 0.155% im Vergleich zu 0.6% und 1,06% bei den Linearen Gruppen auf der x beziehungsweise auf der y Achse.

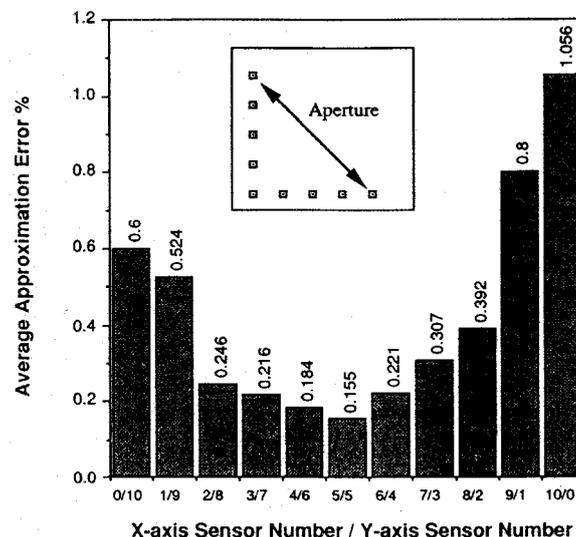


Abbildung 3.8: Mittlerer Fehler als Funktion der Sensoranzahl entlang einer Achse einer L-förmigen Gruppe. Der Abstand zwischen den Sensoren beträgt in allen Fällen 8mm.

### 3. Fixe Anzahl von Sensoren bei Variation der Tapanzahl:

Aus Abbildung 3.9 erkennt man, dass der Fehler bei steigender Anzahl der Verzögerungsstufen sinkt. Auffallend ist, dass es zu keiner weiteren Reduktion des Fehlers für eine Tapanzahl  $> 50$  kommt. Das ist der Fall, weil die gemessenen Impulsantworten im Zeitbereich mit einem  $625\mu\text{s}$  Fenster gefenstert wurden. Das entspricht einer Zeitauflösung von 50 Taps ( $1/80\text{kHz} \times 50 = 625\mu\text{s}$ ). Im Allgemeinen führt eine zusätzliche Anzahl von Taps nur zu einer Erhöhung des Rechenaufwandes.

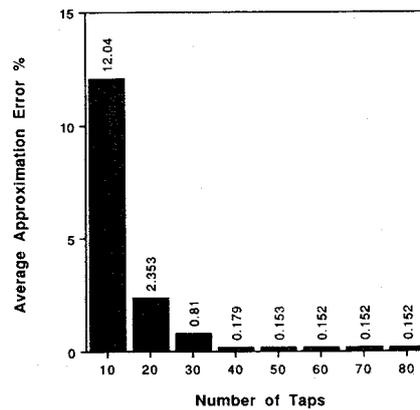


Abbildung 3.9: Mittlerer Fehler als Funktion der Taps in jedem Sensorkanal bei einer 5-1-5 L-förmigen Sensorgruppe. Der Abstand der Sensoren beträgt 8mm.

Änderung der Ausdehnung der Gruppe und der Distanz zwischen den Sensoren:

Die Wahl der ISD (Intersensor Distance) ist einfach, wenn es sich um eine lineare Gruppe und ein Schmalbandsignal handelt. Der optimale Abstand ergibt sich zu  $\lambda_0/2$ , wobei  $\lambda_0$  die Wellenlänge ist. Diese Wahl liefert die größtmögliche Ausdehnung bei Vermeidung von räumlichen Aliasing. Es gibt keine einfachen Regeln zur Auswahl der ISD bei L-förmiger Geometrie und Breitbandsignalen.

Für eine gegebene ISD sinkt der durchschnittliche Fehler bei steigender Anzahl von Sensoren wegen der Anzahl der Freiheitsgrade und der größeren Ausdehnung der Gruppe. Die 5-1-5 Konfiguration mit einer ISD von 8mm führt hier zu einem guten Kompromiss zwischen erreichtem Fehler und Komplexität der Rechnung. Abbildung 3.10.

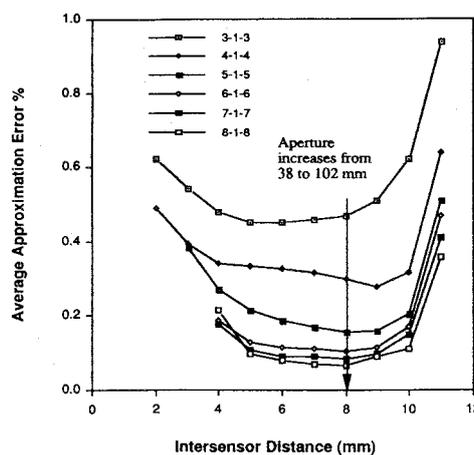


Abbildung 3.10: Mittlerer Fehler als Funktion des Sensorabstandes für 6 verschiedene Sensoranordnungen. Alle Anordnungen haben 46 Taps pro Kanal.

## 4. Interpolationsfehler:

Eine gewünschte Charakteristik dieses Modells ist, dass auch Außenohrübertragungsfunktionen an Richtungen dargestellt werden können, an denen keine Messungen vorliegen. Das Modell interpoliert die Antwort zwischen den

gemessenen Richtungen. Abbildung 3.11 zeigt gemessene und interpolierte Antworten für einige Richtungen. Ein Vergleich zeigt, dass der Fehler der interpolierten Antworten im gleichen Bereich liegt wie der Fehler bei Berücksichtigung dieser Richtungen bei der Bestimmung der Filterkoeffizienten.

Die Eigenschaften der Interpolation sind abhängig vom Schwellwert  $\delta$ . Dieser Wert bestimmt die Anzahl der verwendeten Singulärwerte bei Berechnung der Koeffizienten.

Wenn  $\delta$  steigt, sinkt der mittlere quadratische Fehler. Die Norm des Gewichtsvektors  $\mathbf{w}$  steigt dadurch aber dramatisch an, weil die Inversen sehr kleiner Singulärwerte zur Berechnung von  $\mathbf{w}$  verwendet werden. Der Fehler an den gemessenen Punkten ist zwar geringfügig besser, zwischen den Punkten mit Messwerten ist der Fehler jetzt aber sehr groß.

Wenn  $\delta$  sinkt, steigt der mittlere quadratische Fehler, aber die gewünschten Kurven zwischen den Messwerten werden gut angenähert.

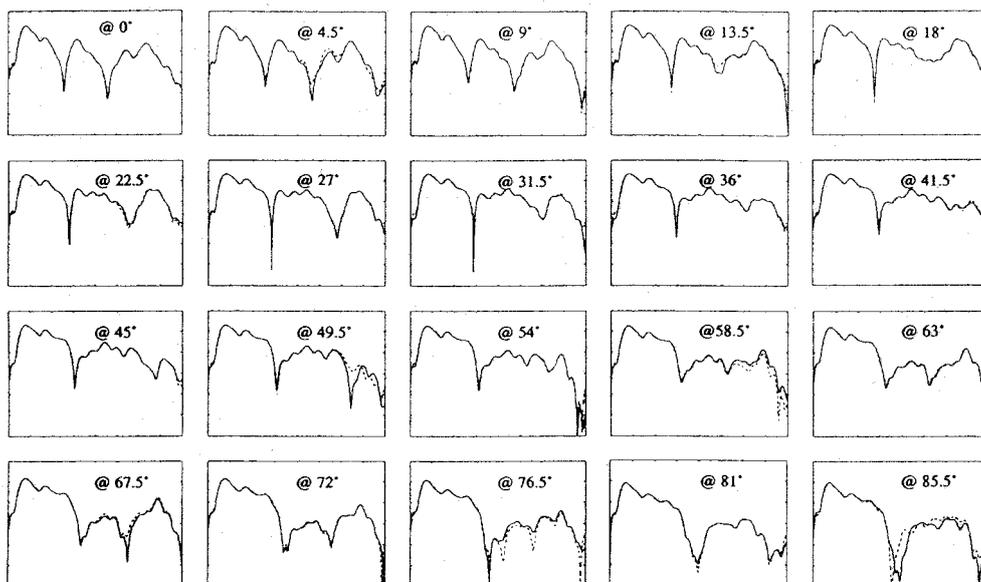


Abbildung 3.11: Gemessene (durchgezogen) und modellierte (strichliert) Frequenzantwort zur Demonstration der Interpolation. Die Gewichte wurden nach Meßdaten bei 0,9,18...90° bestimmt. Die Frequenzantworten bei 4,5, 13,5,...,85,5° sind interpoliert. Y-Achse von -30dB bis 30dB, X-Achse von 0 bis 40kHz.

### 3.4.4 Implementation - Zerlegung in Teilbänder:

Den Ausgangspunkt des Filterdesign bilden die 710 vermessenen KEMAR Außenohrübertragungsfunktionen (Impulsantworten zu je 128 Punkten) [24].

Jeder Impulsantwort wird die für eine bestimmte Richtung charakteristische Zeitverzögerung (Onsetdelay) hinzugefügt, indem der Frequenzgang mit einem linearphasigen Filter mit Betrag 1 multipliziert wird.

Es sind Richtungen und Frequenzpunkte für jede Richtung zu definieren, welche die Eingangswerte für den Beamformingalgorithmus bilden.

Gewählt werden Punkte auf einer linearen oder logarithmischen Frequenzachse zwischen einer unteren und oberen Grenzfrequenz. Durch Interpolation erhält man die Frequenzgangswerte an den neuen Stützstellen.

Bei der Berechnung des Beamformingalgorithmus - im Speziellen bei der Singulärwertzerlegung - gibt es bezüglich der Größe der Matrizen Grenzen, dh. es kann nur eine begrenzte Anzahl an Punkten festgelegt werden. Man kann daher für viele Richtungen Frequenzgänge mit einer geringen, oder wenige Richtungen mit einer höheren Frequenzauflösung definieren.

Umgehen kann man dieses Problem, indem man den Frequenzbereich in Frequenzbänder unterteilt, für jedes Frequenzband die Filterkoeffizienten bestimmt, und anschließend die Gesamtkoeffizienten für den gesamten Frequenzbereich durch Überlagerung der Teilkoeffizienten berechnet. Folgende Abbildungen illustrieren diese Vorgangsweise.

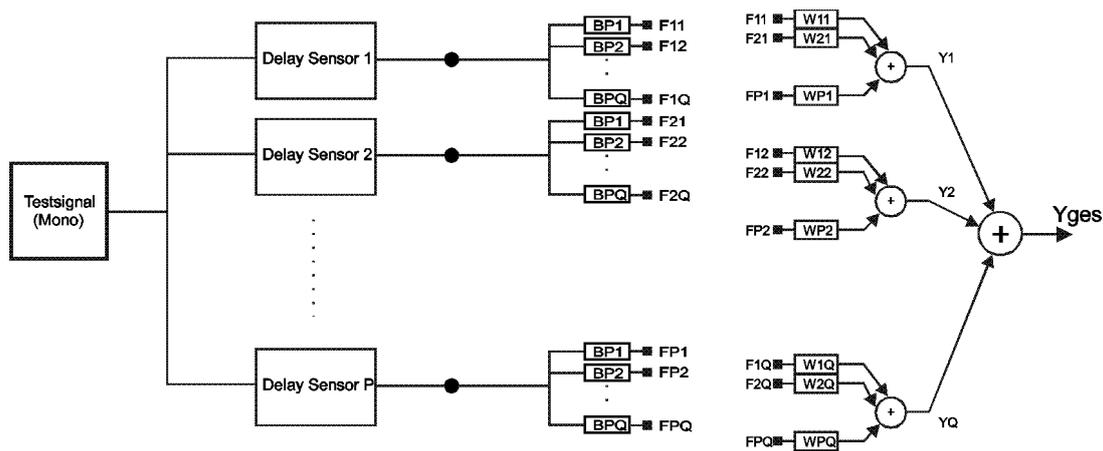


Abbildung 3.12: Zerlegung des Frequenzbereiches in einzelne Teilbänder. Die Beamformerkoeffizienten werden für die einzelnen Teilbänder berechnet. Durch Überlagerung der Teilbeamformer ergeben sich die Gesamtbeamformerkoeffizienten.

P: Anzahl der Sensoren

Q: Anzahl der Frequenzbänder

$$Y_{ges} = Y_1 + Y_2 + \dots + Y_Q \quad (3.40)$$

$$\begin{aligned}
 Y_{ges} = & \\
 & F_{11}W_{11} + F_{21}W_{21} + \dots + F_{P1}W_{P1} + \\
 & F_{12}W_{12} + F_{22}W_{22} + \dots + F_{P2}W_{P2} + \\
 & \dots \\
 & \dots \\
 & F_{1Q}W_{1Q} + F_{2Q}W_{2Q} + \dots + F_{PQ}W_{PQ}
 \end{aligned} \quad (3.41)$$

Die Signale F ergeben sich durch Bandpassfilterung der Mikrofonsignale M.

Aus Gleichung (3.41) folgt

$$\begin{aligned}
 Y_{ges} = & \\
 & M_1[(W_{11}BP_1) + (W_{12}BP_2) + \dots + (W_{1Q}BP_Q)] + \\
 & M_2[(W_{21}BP_1) + (W_{22}BP_2) + \dots + (W_{2Q}BP_Q)] + \\
 & \dots \\
 & \dots \\
 & M_P[(W_{P1}BP_1) + (W_{P2}BP_2) + \dots + (W_{PQ}BP_Q)]
 \end{aligned} \quad (3.42)$$

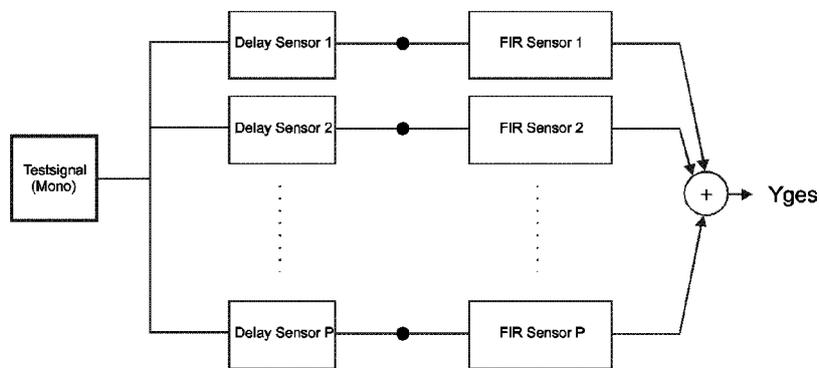


Abbildung 3.13: Resultierender Richtstrahler für alle Frequenzbänder.

Die Koeffizienten  $W_g$  (= FIR Sensor1 bis FIR SensorP) des Beamformers für alle Frequenzbänder sind jetzt:

$$\begin{aligned}
 W_{g1} &= (W_{11}BP_1) + (W_{12}BP_2) + \dots + (W_{1Q}BP_Q) \\
 W_{g2} &= (W_{21}BP_1) + (W_{22}BP_2) + \dots + (W_{2Q}BP_Q) \\
 &\dots \\
 &\dots \\
 W_{gP} &= (W_{P1}BP_1) + (W_{P2}BP_2) + \dots + (W_{PQ}BP_Q)
 \end{aligned} \quad (3.43)$$

## 4 Berechnung eines binauralen Stereosignals durch Anwendung der Ambisonic – Technik

### 4.1 Ambisonic - Grundlagen und Begriffsdefinition

Ambisonic bietet die Möglichkeit ein Schallfeld dreidimensional über Lautsprecher zu reproduzieren. Im Folgenden werden anhand des zweidimensionalen Falls die Grundgleichungen für dieses Aufnahme- und Wiedergabesystem betrachtet [21].

Die Anzahl der benötigten Übertragungs-Kanäle für ein zweidimensionales System ist  $(2m+1)$ , für ein dreidimensionales  $(m+1)^2$ , wobei  $m$  in diesem Fall die Ordnung des Systems ist.

Das Modell basiert auf einem Vergleich einer Referenzwelle mit der Welle, die das System produziert. Eine Voraussetzung für die folgenden Überlegungen ist, dass die Referenzwelle eine ebene Welle ist. Das gilt für Schallquellen, die ausreichend weit vom Aufnahmepunkt entfernt sind. Die zweite Voraussetzung ist, dass die von den Lautsprechern abgestrahlten Wellen auch ebene Wellen sind. Dazu muss sich der Hörer in ausreichendem Abstand zu den Lautsprechern befinden.

Bei Ambisonic ist es üblich das Koordinatensystem um  $90^\circ$  gedreht darzustellen. Die  $x$ -Achse zeigt daher zur Raumvorderseite, die  $y$ -Achse zeigt nach links. Die  $z$ -Achse für den dreidimensionalen Fall zeigt nach oben. Der Ursprung des Koordinatensystems bildet das Zentrum der Aufnahmeanordnung.

Wir nehmen an, dass die ebene Welle aus einer Richtung  $\psi$  in Bezug auf die  $x$ -Achse eintrifft. Der Aufnahmepunkt befindet sich in einer Entfernung  $r$  und Winkel  $\phi$ .

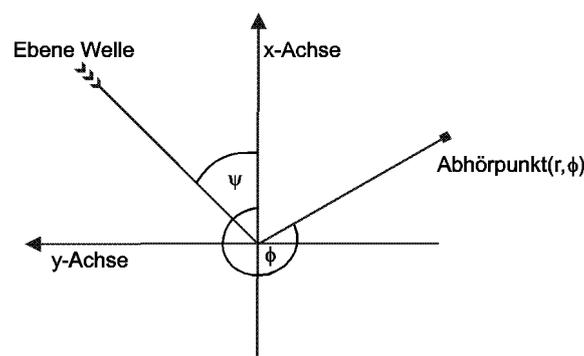


Abbildung 4.1: Achsen und Winkeldefinition bei Ambisonic.

Eine ebene Welle im Aufpunkt  $\vec{r}$  lässt sich ausdrücken durch:

$$P_\psi e^{i\vec{k}\vec{r}}. \quad (4.1)$$

$P$  ist der Druck der ebenen Welle,  $k$  ist die Wellenzahl  $2\pi/\lambda$  und  $\lambda$  ist die Wellenlänge. Der Vektor  $\vec{k}$  repräsentiert eine Welle mit der Wellenzahl  $k$ , die sich in eine bestimmte Richtung bewegt. Die Gleichung kann auch geschrieben werden als

$$S_\psi = P_\psi e^{ikr \cos(\phi - \psi)}. \quad (4.2)$$

$S_\psi$  ist hier die originale ebene Welle. Das Ziel des Ambisonic-Systems ist, diese ebene Welle im Zentrum der Hörzone zu reproduzieren.

Gleichung (4.3) beschreibt den Zusammenhang zwischen einer ebenen Welle und einer Reihenentwicklung aus Cosinus und Zylindrischen Besselfunktionen.

$$S_\psi = P_\psi J_0(kr) + 2P_\psi \sum_{m=1}^{\infty} i^m J_m(kr) \cos m(\phi - \psi). \quad (4.3)$$

Der Cosinus Term kann weiters umgeschrieben werden und es folgt:

$$S_\psi = P_\psi \left( J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) [\cos(m\psi) \cos(m\phi) + \sin(m\psi) \sin(m\phi)] \right). \quad (4.4)$$

Aus dieser Gleichung kann man sehen, dass die ursprüngliche ebene Welle nur dann hundertprozentig wiederhergestellt werden kann, wenn eine unendliche Anzahl von Kanälen zur Verfügung steht. Bei der Implementierung beschränkt man sich auf ein System erster oder zweiter Ordnung.

Eine typische Ambisonic Anordnung besteht aus  $N$  kreisförmig angeordneten Lautsprechern, (Siehe Abbildung 4.2 für  $N=8$ ).

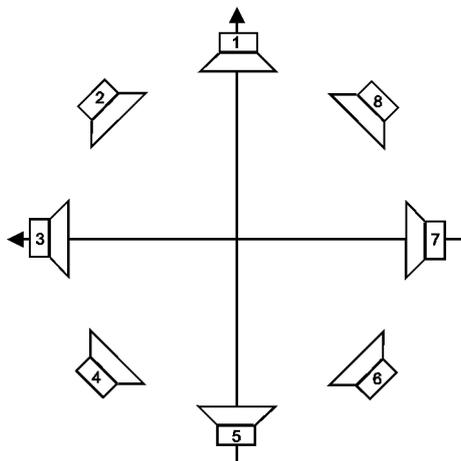


Abbildung 4.2: Typische Lautsprecheranordnung für ein Ambisonicmodell in der Ebene.

Wenn jeder Lautsprecher eine ebene Welle abstrahlt, ist der Output des  $n$ -ten Lautsprechers:

$$S_n = P_n \left( J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) [\cos(m\phi) \cos(m\phi_n) + \sin(m\phi) \sin(m\phi_n)] \right). \quad (4.5)$$

$\phi_n$  ist hier der Winkel des n-ten Lautsprechers. Jeder Lautsprecher trägt zur resultierenden Welle bei, daher ergibt sich die resultierende Welle als Summe der einzelnen Lautsprecherwellen. Die Gesamtwelle ist

$$S = \sum_{n=1}^N P_n J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) \left( \sum_{n=1}^N P_n \cos(m\phi_n) \cos(m\phi) + \sum_{n=1}^N P_n \sin(m\phi_n) \sin(m\phi) \right). \quad (4.6)$$

Damit die Lautsprecher die originale Welle nachbilden, müssen folgende Bedingungen eingehalten werden. Der Koeffizientenvergleich mit Gleichung (4.4) ergibt:

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

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

$$P_\psi \sin(m\psi) = \sum_{n=1}^N P_n \sin(m\phi_n) \quad (4.9)$$

Jede Ordnung m entspricht der sphärischen Harmonischen m-ter Ordnung. Wenn eine unendliche Anzahl an Kanälen zur Verfügung stehen würde, könnte die originale Welle exakt reproduziert werden. Eine unendliche Anzahl an Kanälen heisst aber auch, dass eine unendliche Anzahl an Lautsprechern zur Verfügung stehen müsste. Eine begrenzte Anzahl an Lautsprechern bei der Implementation bedeutet, dass die Reihe bei einer bestimmten Ordnung abgebrochen wird. Ein System erster, bzw. zweiter Ordnung bedeutet, dass die sphärischen Harmonischen bis zur ersten bzw. zweiten Ordnung in Übereinstimmung gebracht werden.

Die Bedingungen (4.7) bis (4.9) beschreiben die Gleichungen, aus denen die Informationen für das Ambisonic-Signal bezogen werden. Bei einem System nullter Ordnung muss nur (4.7) übereinstimmen. Wenn (4.7) eingehalten wird, stimmt der Druck der beiden Wellen überein. Das heisst, dass die originale und reproduzierte Welle die gleiche Amplitude haben. Gleichung (4.7) beschreibt also den Druck der ebenen Welle der bei den Ambisonic-Signalen das W-Signal darstellt. Für Systeme erster Ordnung wird zusätzlich die Schnelle im Ursprung in Übereinstimmung gebracht, das heisst, die originale und reproduzierte Welle bewegen sich in die gleiche Richtung. Für m=1 ergibt sich aus den Gleichungen (4.8) und (4.9) das X und Y Ambisonic-Signal.

$$\begin{aligned} X &= P_\psi \cos(\psi) \\ Y &= P_\psi \sin(\psi) \end{aligned} \quad (4.10)$$

Diese drei Komponenten bilden ein Ambisonic-System erster Ordnung. Für  $m=2$  in Gleichung (4.8) und (4.9) folgen die Ambisonic-Signale U und V.

$$\begin{aligned} U &= P_\psi \cos(2\psi) \\ V &= P_\psi \sin(2\psi) \end{aligned} \quad (4.11)$$

Ein System zweiter Ordnung besteht aus folgenden fünf Signalen: W;X;Y;U;V. Bei Berücksichtigung dieser fünf Signale wird die ebene Welle im Zentrum der Hörregion bis zur Ordnung 2 korrekt reproduziert. Im Folgenden werden die Eingangssignale für die einzelnen Lautsprecher berechnet. Die Frage lautet: Welche Signale  $P_n$  müssen die Lautsprecher bekommen damit die Gleichungen 4.7 bis 4.9 eingehalten werden. Ein System erster Ordnung berücksichtigt die Signale W,X,Y. Die Lautsprechersignale ergeben sich zu

$$P_n = \frac{1}{N} (W + 2X \cos \phi_n + 2Y \sin \phi_n). \quad (4.12)$$

Für ein System zweiter Ordnung ergeben sich zwei zusätzliche Terme:

$$P_n = \frac{1}{N} (W + 2X \cos \phi_n + 2Y \sin \phi_n + 2U \cos 2\phi_n + 2V \sin 2\phi_n). \quad (4.13)$$

Bis jetzt wurde vorausgesetzt, dass sich die Lautsprecher in einer regelmäßigen Anordnung kreisförmig ums Zentrum der Hörregion verteilen müssen. Um nicht an bestimmte Lautsprecherlayouts gebunden zu sein, werden die Parameter  $\alpha$  bis  $\varepsilon$  in Gleichung 4.13 eingefügt. Die Bedingung, dass der Winkel zwischen den Lautsprechern gleich sein muss, kann fallen gelassen werden, und man erhält die Gleichung:

$$P_n = \frac{1}{N} (\alpha W + 2\beta X \cos \phi_n + 2\gamma Y \sin \phi_n + 2\delta U \cos 2\phi_n + 2\varepsilon V \sin 2\phi_n). \quad (4.14)$$

## 4.2 Ambisonic-Signale für den 3 - dimensionalen Fall

Im Folgenden werden die Ambisonic-Signale 1.Ordnung für den 3 dimensionalen Fall berechnet. Es ergeben sich 4 Signale W, X, Y und Z.

Im Vergleich zu Gleichung (4.10) ergibt sich ein zusätzlicher Term für die z-Komponente.

$$\begin{aligned} X &= P_\psi \cos(\psi) \cdot \cos(\varphi) \\ Y &= P_\psi \sin(\psi) \cdot \cos(\varphi) \\ Z &= P_\psi \cos(\varphi) \end{aligned} \quad (4.15)$$

mit  $\psi$  = Azimuth und  $\varphi$  = Elevation.

Die Lautsprechersignale für die virtuellen Lautsprecher an der Position  $(\phi_n, \gamma_n)$  berechnen sich zu

$$P_n = \frac{1}{N} (W + 2X \cos \phi_n \cos \gamma_n + 2Y \sin \phi_n \cos \gamma_n + 2Z \sin \gamma_n). \quad (4.16)$$

mit  $(\phi_n, \gamma_n)$  = Azimuth und Elevation des n-ten Lautsprechers.

Die Wiedergabe wird in Folge nicht mit realen Lautsprechern durchgeführt. Durch eine Filterung mit Außenohrübertragungsfunktionen wird ein binaurales Kopfhörersignal berechnet.

### 4.3 Berechnung eines binauralen Stereosignals

Um ein binaurales Stereosignal zu erhalten ist es notwendig, dass eine einfallende Schallwelle aus einer bestimmten Raumrichtung mit einer entsprechenden Außenohrübertragungsfunktion gefiltert wird. Als einfallende Schallwellen werden in diesem Fall die Signale der Lautsprecher aus dem Ambisonic-System betrachtet. Da bei den KEMAR Kunstkopfmessungen 710 HRTF Impulsantworten vermessen wurden, können die Lautsprecherfeeds dieser virtuellen Lautsprecher mit den entsprechenden Außenohrübertragungsfunktionen gefiltert werden, und man erhält durch Superposition ein Kopfhörersignal.

Das Stereosignal L und R ergibt sich durch Filterung der Lautsprechersignale  $P_1$  bis  $P_N$  mit den HRTF Impulsantworten:

$$\begin{pmatrix} L \\ R \end{pmatrix} = \begin{pmatrix} H_{1,L} & H_{2,L} & \cdots & H_{N,L} \\ H_{1,R} & H_{2,R} & \cdots & H_{N,R} \end{pmatrix} * \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} \quad (4.17)$$

Die Lautsprechersignale berechnen sich aus den Ambisonic-Signalen W,X,Y für ein System 1.Ordnung und W,X,Y,U und V für ein System 2.Ordnung.

$$\begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} = \frac{1}{N} \begin{pmatrix} \alpha_1 & 2\beta \cos \phi_1 & 2\gamma \sin \phi_1 & 2\delta \cos 2\phi_1 & 2\varepsilon \sin 2\phi_1 \\ \alpha_2 & 2\beta \cos \phi_2 & 2\gamma \sin \phi_2 & 2\delta \cos 2\phi_2 & 2\varepsilon \sin 2\phi_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_N & 2\beta \cos \phi_N & 2\gamma \sin \phi_N & 2\delta \cos 2\phi_N & 2\varepsilon \sin 2\phi_N \end{pmatrix} \cdot \begin{pmatrix} W \\ X \\ Y \\ U \\ V \end{pmatrix} \quad (4.18)$$

Die Faktoren  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  und  $\varepsilon$  sind für regelmäßige Lautsprecheranordnungen gleich eins.

Das Schallfeld, dass in den Ambisonic-Signalen W,X,Y,U und V abgebildet ist, kann in beliebige Richtungen gedreht werden, das heisst, der Schall wird bei der

Wiedergabe aus einer anderen Richtung wahrgenommen als das originale Schallfeld. Man erhält bei einer Drehung die neuen Ambisonic-Signale  $W', X', Y', U'$  und  $V'$ . Die unrotierten Signale lauten:

$$W = \frac{P_\psi}{\sqrt{2}} \quad X = P_\psi \cos(\psi) \quad U = P_\psi \cos(2\psi) \\ Y = P_\psi \sin(\psi) \quad V = P_\psi \sin(2\psi) \quad (4.19)$$

Bei einer Rotation gilt:  $\cos(\psi) \longrightarrow \cos(\psi + \rho)$ ,  $\sin(\psi) \longrightarrow \sin(\psi + \rho)$ ,  $\cos(2\psi) \longrightarrow \cos(2(\psi + \rho))$  und  $\sin(2\psi) \longrightarrow \sin(2(\psi + \rho))$ . Mit den Sumsensätzen für die Sinus und Cosinusfunktionen folgt für rotierte Signale um einen Winkel  $\rho$ :

$$\begin{pmatrix} W' \\ X' \\ Y' \\ U' \\ V' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \cos \rho & -\sin \rho & 0 & 0 \\ 0 & \sin \rho & \cos \rho & 0 & 0 \\ 0 & 0 & 0 & \cos 2\rho & -\sin 2\rho \\ 0 & 0 & 0 & \sin 2\rho & \cos 2\rho \end{pmatrix} \cdot \begin{pmatrix} W \\ X \\ Y \\ U \\ V \end{pmatrix} \quad (4.20)$$

Aus den obigen Gleichungen berechnet sich jetzt ein binaurales Stereosignal für ein rotiertes Schallfeld um den Winkel  $\rho$  folgendermaßen aus den Ambisonic-Signalen  $W, X, Y, U$  und  $V$ .

Gleichung (4.21):

$$\begin{pmatrix} L \\ R \end{pmatrix} = \begin{pmatrix} H_{1,L} & H_{2,L} & \dots & H_{N,L} \\ H_{1,R} & H_{2,R} & \dots & H_{N,R} \end{pmatrix} \cdot \frac{1}{N} \cdot \begin{pmatrix} \alpha_1 & 2\beta \cos \phi_1 & 2\gamma \sin \phi_1 & 2\delta \cos 2\phi_1 & 2\varepsilon \sin 2\phi_1 \\ \alpha_2 & 2\beta \cos \phi_2 & 2\gamma \sin \phi_2 & 2\delta \cos 2\phi_2 & 2\varepsilon \sin 2\phi_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_N & 2\beta \cos \phi_N & 2\gamma \sin \phi_N & 2\delta \cos 2\phi_N & 2\varepsilon \sin 2\phi_N \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \cos \rho & -\sin \rho & 0 & 0 \\ 0 & \sin \rho & \cos \rho & 0 & 0 \\ 0 & 0 & 0 & \cos 2\rho & -\sin 2\rho \\ 0 & 0 & 0 & \sin 2\rho & \cos 2\rho \end{pmatrix} \cdot \begin{pmatrix} W \\ X \\ Y \\ U \\ V \end{pmatrix}$$

### 4.3.1 Rotationsmatrizen für den 3 dimensionalen Fall erster Ordnung:

Rotationsmatrix bei einer Drehung des Kopfes um die z-Achse (Drehwinkel  $\rho$ ):

$$\begin{pmatrix} W' \\ X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \rho & -\sin \rho & 0 \\ 0 & \sin \rho & \cos \rho & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} W \\ X \\ Y \\ Z \end{pmatrix} \quad (4.22)$$

Zusätzlich zur Rotation (Drehung des Kopfes um die z-Achse) können auch Bewegungen des Kopfes um die x-Achse (Neigen), und Bewegungen um die y-Achse berücksichtigt werden. Die Matrix bei einer Drehung um die x-Achse (Drehwinkel  $\xi$ ) ist

$$\begin{pmatrix} W' \\ X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \xi & -\sin \xi \\ 0 & 0 & \sin \xi & \cos \xi \end{pmatrix} \begin{pmatrix} W \\ X \\ Y \\ Z \end{pmatrix}. \quad (4.23)$$

Bei einer Drehung um die y-Achse (Nicken) lautet die Matrix (Drehwinkel  $\zeta$ )

$$\begin{pmatrix} W' \\ X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \zeta & 0 & -\sin \zeta \\ 0 & 0 & 1 & 0 \\ 0 & \sin \zeta & 0 & \cos \zeta \end{pmatrix} \begin{pmatrix} W \\ X \\ Y \\ Z \end{pmatrix}. \quad (4.24)$$

Zusammengesetzte Bewegungen lassen sich durch Multiplikation der einzelnen Matrizen berücksichtigen.

#### 4.4 Berechnung der Ambisonic-Signale W,X,Y,U,V

Ziel ist es mittels einer geeigneten Mikrophonanordnung und Signalverarbeitung die Ambisonic-Signale zu erhalten. W entspricht dem reinen Drucksignal unabhängig von der Schalleinfallrichtung, X und Y ist ein Schnellesignal und hängt vom Cosinus bzw. Sinus des Schalleinfallswinkels ab. U und V sind abhängig vom Cosinus und Sinus des doppelten Winkels. Die folgenden Diagramme zeigen die Abhängigkeit der Ambisonic-Signale von der Schalleinfallrichtung. Betrachtet wird hier nur die Horizontalebene (Elevation = 0°).

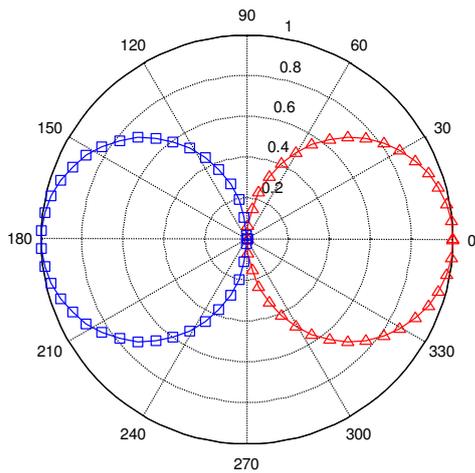


Abbildung 4.3: Richtcharakteristik des X Signals.  $\Delta$ =positiv,  $\square$ =negativ.

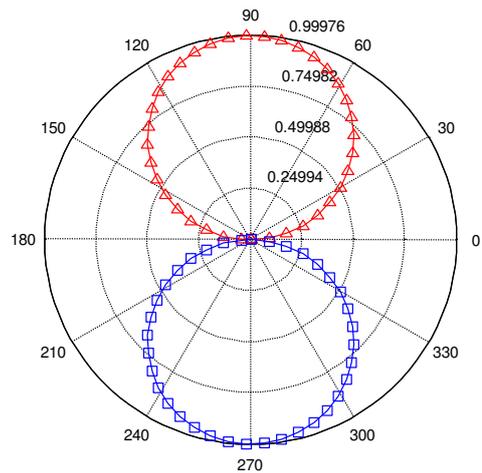


Abbildung 4.4: Richtcharakteristik des Y Signals.  $\Delta$ =positiv,  $\square$ =negativ.

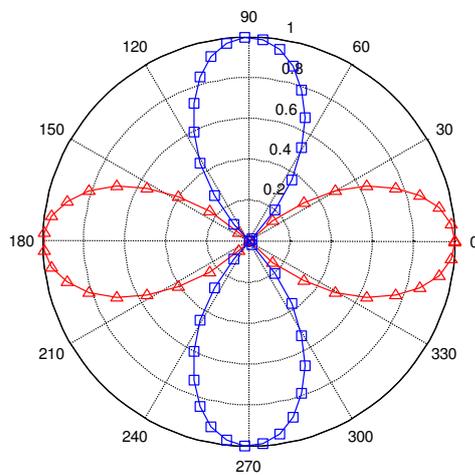


Abbildung 4.5: Richtcharakteristik des U Signals.  $\Delta$ =positiv,  $\square$ =negativ.

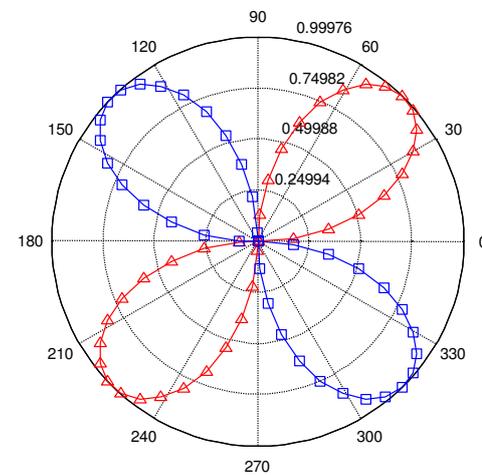


Abbildung 4.6: Richtcharakteristik des V Signals.  $\Delta$ =positiv,  $\square$ =negativ.

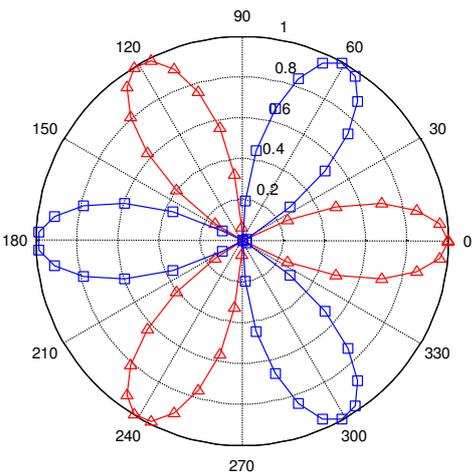


Abbildung 4.7: Richtcharakteristik des cos-Signals 3.Ordnung.  $\Delta$ =positiv,  $\square$ =negativ.

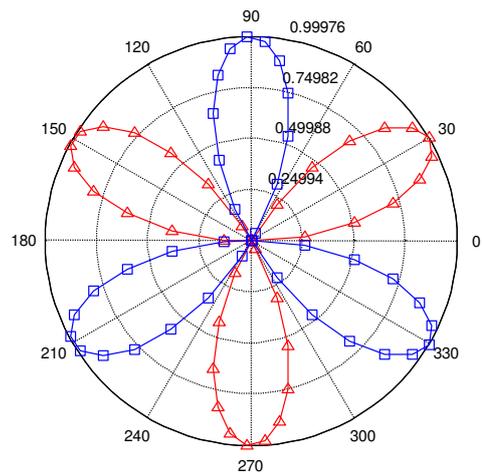


Abbildung 4.8: Richtcharakteristik des sin-Signals 3.Ordnung.  $\Delta$ =positiv,  $\square$ =negativ.

Diese Kurven dienen jetzt als Eingangsparameter für die Koeffizientenberechnung einer Beamformer FIR Filterbank. Den Ausgang der einzelnen Beamformer bilden dann die Ambisonic-Signale W,X,Y,U und V.

#### 4.5 System zur Berechnung binauraler Stereosignale unter Berücksichtigung von Kopfbewegungen des Hörers:

Um die Vorne-Hinten Lokalisation zu ermöglichen, werden beim natürlichen Hören auch die Kopfbewegungen (vor allem die Rotation nach links und rechts) und die damit verbundenen Signalveränderungen ausgewertet. Es gilt nun die Kopfhörersignale L und R entsprechend der Kopfposition so zu verändern, dass diese Signalveränderungen bei der Lokalisation zu Verfügung stehen.

Aus Gleichung (4.21) sieht man, dass die Stereosignale direkt aus den gefilterten Ambisonic-Signalen berechnet werden können. Der einzige Teil der sich zeitlich verändert ist die Rotationsmatrix. Den Rotationswinkel gewinnt man durch detektieren der Kopfdrehung mittels Headtracker, und daraus ergeben sich für jeden Zeitpunkt neue Filterkoeffizienten zur Berechnung der Stereosignale.

Die folgende Abbildung zeigt die Änderungen bei einer Kopfbewegung.

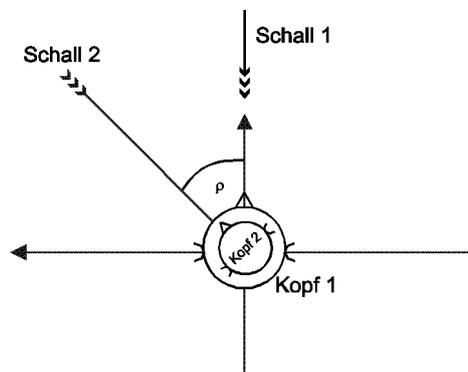


Abbildung 4.9: Auswirkungen einer Kopfdrehung auf die Filtercharakteristik des Ohres.

Betrachtet werden die benötigten HRTF Filter für Schallereignisse aus zwei verschiedenen Richtungen ( $0^\circ, 45^\circ$ ) und für zwei verschiedene Kopfpositionen. In der Stellung Kopf 1 (nach vorne blickend) muss der Schall1 die HRTF für  $0^\circ$ , der Schall2 die HRTF für  $45^\circ$  erhalten. Bei einer Kopfdrehung um  $\rho^\circ$  nach links (Kopf 2  $\rho=45^\circ$ ) muss der Schall2 die HRTF für  $0^\circ$  und der Schall1 die HRTF für  $-45^\circ$  erhalten.

## 4.6 Berechnung der Ambisonic-Signale über den Ansatz der Holophonie

Durch Kombination der beiden Methoden sollen die Vorteile von Ambisonic (sehr effiziente räumliche Kodierung) und Holophonie (Vorteile in der Aufzeichnung) genutzt werden. Dazu wird vorher das Prinzip der Holophonie erläutert [32, 33].

### 4.6.1 Holophonie

Abgeleitet wird die Holophonie vom Hygens'schen Prinzip. Demnach kann ein Schallfeld innerhalb eines quellenfreien Volumens  $\Omega$  und der Begrenzung  $\partial\Omega$  durch

$$p(\vec{r}) = \iint_{\partial\Omega} \left[ \vec{\nabla} p_0 \cdot \vec{n} - \frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4\pi R} dS_0 \quad \forall \vec{r} \in \Omega \quad (4.25)$$

beschrieben werden.

$p$  entspricht dem Druck innerhalb von  $\Omega$  und  $p_0$  dem Druck auf der Oberfläche des Volumens.

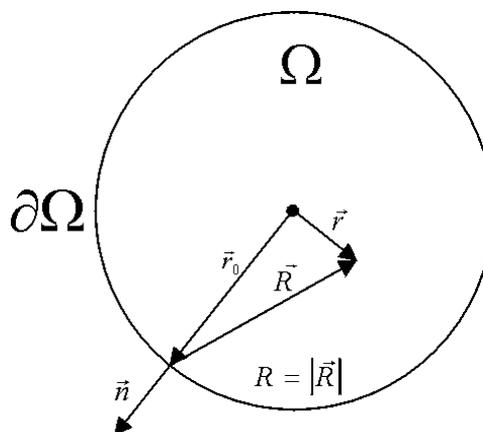


Abbildung 4.10: Bezeichnungen des Kirchhoff – Helmholtz Integrals.

Dieses Integral (Kirchhoff Helmholtz Integral) kann als Verteilung von Sekundärquellen auf einer Oberfläche verstanden werden, und bietet eine Möglichkeit das Schallfeld (Druck) innerhalb eines begrenzten Gebietes eindeutig zu bestimmen.

Zuerst muss das Schallfeld durch Druck ( $p_0$ ) und Schnellmikrophone ( $\vec{\nabla} p_0 =$  Druckgradient) aufgezeichnet werden. Im zweiten Schritt werden die Mikrofone durch Lautsprecher ersetzt, sodass das ursprüngliche Schallfeld exakt reproduziert wird.

Ganz genaue Reproduktion könnte nur mit kontinuierlichen Schallwandlern (Mikrofone und Lautsprecher) erreicht werden. Da in der Praxis Mikrofone das Schallfeld nur diskret abtasten, tritt hier das Phänomen des räumlichen Aliasing auf. Probleme bekommt man daher bei höheren Frequenzen.

Wie bei der Abtastung im Zeitbereich muss hier im Vergleich zum Zeitabstand die räumliche Entfernung der Sensoren dem Shannon'schen Theorem genügen, um in diesem Fall räumliches Aliasing zu vermeiden.

Der minimale Abstand der Sensoren wird durch die maximale Frequenz des räumlichen Spektrums [30] bestimmt.

$$\Delta_{\min} = \frac{\pi}{k \cdot \sin \theta} \quad (4.26)$$

Man beachte, dass der erforderliche Abstand der Sensoren nicht nur von der zeitlichen Frequenz sondern auch vom Einfallswinkel  $\theta$  abhängt. Untersuchungen zeigen, dass räumliches Aliasing die Lokalisationsgenauigkeit unwesentlich beeinflusst, solange die Aliasingfrequenz über 1.5kHz liegt [32].

#### Ansatz für kreisförmiges Lautsprecherlayout und ebene Welle

Das Kirchhoff Helmholtz Integral wird für den Fall einer ebenen Welle und für ein kreisförmiges Lautsprecherlayout umgeschrieben.

Das 2 - dimensionale Integral aus Gleichung (4.25) wird eindimensional, sodass  $\partial\Omega$  zu einem Kreis mit Radius  $r_0$  wird. Das Druckfeld  $p_0$  sei eine ebene Welle.

Mathematisch kommt das Prinzip der Stationary Phase Approximation zum Tragen [32].

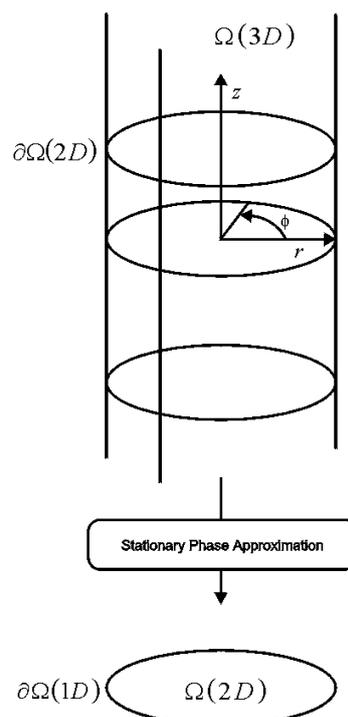


Abbildung 4.11: Stationary Phase Approximation.

Der Druck innerhalb eines Kreises mit dem Radius  $r_0$  lässt sich daher folgendermaßen beschreiben.

$$p(r, \phi) = \frac{r_0}{2\sqrt{2\pi jk}} \int_0^{2\pi} \left[ \frac{\partial p_0}{\partial r_0} - \cos \alpha (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{\sqrt{R}} d\phi_0 \quad (4.27)$$

In dieser Gleichung ist

$$R = \sqrt{r^2 + r_0^2 - 2rr_0 \cos(\phi - \phi_0)} \quad (4.28)$$

$$\cos \alpha = \frac{r \cos(\phi - \phi_0) - r_0}{R}. \quad (4.29)$$

$\vec{R}$  beschreibt den Ausbreitungspfad von der Sekundärquelle bei  $\vec{r}_0$  und dem Abhörpunkt bei  $\vec{r}$ .

Zur Rekonstruktion des Schallfeldes nach Gleichung (4.27) werden sowohl 2 Typen von Mikrofonen (Druck und Schnellempfänger) als auch 2 Typen von Lautsprechern (Monopol und Dipollautsprecher) benötigt.

Laut Gleichung (4.25) ist der Ausbreitungsterm der Monopol Quellen

$$\frac{e^{-jkR}}{4\pi R} \quad (4.30)$$

und jener der Dipolquellen

$$\frac{\vec{R}}{R} \cdot \vec{n} (1 + jkR) \frac{e^{-jkR}}{4\pi R}. \quad (4.31)$$

Durch Einführung einer räumlichen Gewichtung die mit der Position der Primärquelle verbunden ist, ist es möglich das Schallfeld nur mit Monopol Quellen zu rekonstruieren [32].

Das rekonstruierte Schallfeld ergibt sich jetzt zu

$$p(r, \phi) = \sqrt{\frac{2\pi}{jk}} r_0 \int_0^{2\pi} w(\phi_0) \frac{\partial p_0}{\partial r_0} \frac{e^{-jkR}}{2\pi\sqrt{R}} d\phi_0 \quad \forall \vec{r} \in \Omega \quad (4.32)$$

Die Gewichtungsfunktion in dieser Gleichung ist

$$w(\phi_0) = \text{Re ct} \left( \frac{\phi - \phi_0}{\pi} \right) = \begin{cases} 1 & \forall \phi_0 \in [\theta - \frac{\pi}{2}, \theta + \frac{\pi}{2}] \\ 0 & \text{sonst} \end{cases} \quad (4.33)$$

### Verbinden der Ansätze Ambisonic und Holophonie

Für die Berechnung des Schallfeldes mittels Ambisonic gilt

$$p(r, \phi) = \frac{1}{N} \sum_{i=0}^{N-1} \left[ 1 + 2 \sum_{n=1}^M \cos[n(\phi_i - \theta)] \right] e^{jkr \cos(\phi - \phi_i)}. \quad (4.34)$$

N = Anzahl der Lautsprecher

M = Ordnung

$\theta$  = Schalleinfallrichtung,  $\phi_i$  = Position der Lautsprecher

Für das Schallfeld mittels Berechnung über die Holophonie gilt für einer diskrete Anordnung von N Lautsprechern nach Gleichung (4.32).

$$p(r, \phi) = \frac{1}{N} \sqrt{\frac{2\pi}{jk}} r_0 \times \sum_{i=0}^{N-1} w(\phi_i) \frac{\partial p_0(r_0, \phi_i)}{\partial r_0} \frac{e^{-jkR_i}}{\sqrt{R_i}}. \quad (4.35)$$

In dieser Gleichung ist

$$R_i = \sqrt{r^2 + r_0^2 - 2rr_0 \cos(\phi - \phi_i)} \quad (4.36)$$

Vergleicht man (4.34) und (4.35) kann man eine gemeinsame Struktur erkennen. Man kann den Druck als Faltungssumme einer gewichteten Funktion  $f$ , welche die Aufzeichnung des Schallfeldes beschreibt und einer Funktion  $g$ , welche die Ausbreitung des Schalles von der Sekundärquelle  $n$  bis zum Aufpunkt  $(r, \phi)$  beschreibt darstellen.

$$\begin{aligned} p(r, \phi) &= \sum_{i=0}^{N-1} w(\phi_i) f(\phi_i) g(r, \phi - \phi_i) \\ &= [(f \times w) * g_r](\phi) \end{aligned} \quad (4.37)$$

$$g_r(\phi) = g(r, \phi) \quad (4.38)$$

Die Funktion  $f$  beschreibt also das räumliche Kodieren des Schallfeldes. Die Funktion  $g$  ist für das Dekodieren des aufgezeichneten Schallfeldes verantwortlich.

Wünschenswert ist es das Schallfeld mit einer kreisförmigen Anordnung von Mikrofonen aufzuzeichnen (Kodieren = Holophonie), die aufgezeichneten Signale in Ambisonicsignale zu transformieren und als solche zu dekodieren. Dadurch umgeht man das Problem der Aufzeichnung von Ambisonicsignalen  $W, X, Y$  die nur mit dem Soundfieldmikrofon von Gerzon, und außerdem nur bis zur Ordnung 1 möglich ist. Im Folgenden werden die Funktionen  $f, w$  und  $g$  getrennt für Holophonie und Ambisonic betrachtet.

$$f(\phi) = \sum_{i=-\infty}^{\infty} f(\phi_i) \delta(\phi - i \frac{2\pi}{N})$$

Für die Holophonie ist

$$fh(\phi) = \cos(\phi - \theta) e^{jkr_0 \cos(\phi - \theta)} \times \sum_{i=-\infty}^{+\infty} \delta(\phi - i \frac{2\pi}{N}) \quad (4.39)$$

$$wh(\phi) = \text{Re ct}(\frac{\phi - \theta}{\pi}) \quad (4.40)$$

$$gh_r(\phi) = \frac{1}{N} gh_0 \frac{e^{-jk\rho}}{\sqrt{\rho}} \quad (4.41)$$

wobei

$$gh_0(\phi) = -\sqrt{2\pi j k r_0} \quad (4.42)$$

$$\rho = \sqrt{r^2 + r_0^2 - 2rr_0 \cos(\phi)} \quad (4.43)$$

gilt.

Für Ambisonic ist

$$fa(\phi) = \left[ 1 + 2 \sum_{n=1}^M \cos[n(\phi - \theta)] \right] \times \sum_{i=-\infty}^{+\infty} \delta(\phi - i \frac{2\pi}{N}) \quad (4.44)$$

$$wa(\phi) = 1 \quad (4.45)$$

$$ga_r = \frac{1}{N} e^{jkr \cos \phi} \quad (4.46)$$

Jede der Funktionen  $f$ ,  $w$  und  $g_r$  wiederholt sich mit der Periode  $2\pi$ . Aus diesem Grund kann jede der Funktionen in eine komplexe Fourier Reihe entwickelt werden. Man erhält dann das räumliche Spektrum in Abhängigkeit der Variable  $\nu$  (räumliche Frequenz, engl. spatial frequency, angular frequency).

$$g_r(\phi) = \sum_{n=-\infty}^{+\infty} \gamma_n e^{jn\phi} \quad (4.47)$$

$$G_r(\nu) = \sum_{n=-\infty}^{+\infty} \gamma_n \delta(\nu - \frac{n}{2\pi}) \quad (4.48)$$

Im dualen Bereich  $\nu$  wird die räumliche Faltung aus Gleichung (4.37) zu einer Multiplikation der Spektren:

$$P(r, \nu) = [F \times W](\nu) \times G_r(\nu) . \quad (4.49)$$

Es wird jetzt das räumliche Spektrum der Funktion  $f$  für Holophonie und Ambisonic betrachtet.

$$Fh(\nu) = \frac{1}{2\pi} \left[ \sum_{n=-\infty}^{+\infty} \mathcal{H}_n \delta\left(\nu - \frac{n}{2\pi}\right) \right] * \left[ \sum_{i=-\infty}^{+\infty} \delta\left(\nu - i \frac{N}{2\pi}\right) \right] \quad (4.50)$$

mit

$$\mathcal{H}_n = j^{|n|-1} J_{|n|}'(kr_0) e^{-jn\theta} \quad (4.51)$$

$$Fa(\nu) = \frac{1}{2\pi} \left[ \sum_{n=-M}^{+M} \gamma_n \delta\left(\nu - \frac{n}{2\pi}\right) \right] * \left[ \sum_{i=-\infty}^{+\infty} \delta\left(\nu - i \frac{N}{2\pi}\right) \right] \quad (4.52)$$

mit

$$\gamma_n = e^{-jn\theta}. \quad (4.53)$$

Der Zustand einer ebenen Welle wird durch den Einfallswinkel  $\theta$  vollständig beschrieben. Aus den vorhergehenden Gleichungen kann man sehen, dass diese Information in den räumlichen Spektren für Ambisonic und Holophonie im Term  $e^{-jn\theta}$  enthalten ist. Deswegen bietet sowohl die Holophonie (Aufnahme des Schallfeldes entlang eines Kreises mit Radius  $r_0$ ) und Ambisonic (Aufnahme des Schallfeldes in einem Punkt) eine äquivalente Beschreibung der räumlichen Eigenschaften.

Da es sich um eine äquivalente Beschreibung handelt können die Fourierkoeffizienten  $\gamma_a$  und  $\gamma_h$  voneinander abgeleitet werden. Die Ambisonicsignale können daher aus den aufgezeichneten Holophoniesignalen abgeleitet werden. Es gilt die Beziehung:

$$\gamma_n = \frac{\mathcal{H}_n}{j^{|n|-1} J_{|n|}'(kr_0)} \quad (4.54)$$

Gleichung (4.54) beschreibt eine räumliche Filterung und kann sowohl im räumlichen Frequenzbereich  $\nu$  (Multiplikation der Spektren), oder im Bereich  $\phi$  (räumliche Faltung) berechnet werden.

Weiters können bei einer Holophonieaufzeichnung sowohl Druck als auch Schnellemikrofone zum Einsatz kommen. Die Gleichung (4.39) beschreibt eine Holophonieaufzeichnung mit Schnellemikrofonen. Werden diese Druckgradientenempfänger durch Druckempfänger ersetzt ergibt sich die Gleichung für  $fh$  zu:

$$fh_o(\phi) = e^{jkr_0 \cos(\phi-\theta)} \times \sum_{i=-\infty}^{+\infty} \delta\left(\phi - i \frac{2\pi}{N}\right). \quad (4.55)$$

Die Koeffizienten dieses Spektrums sind

$$\mathcal{H}_o_n = j^{|n|} J_{|n|}(kr_0) e^{-jn\theta}. \quad (4.56)$$

Die Koeffizienten  $\mathcal{H}_n$  können daher von den Koeffizienten  $\mathcal{H}_o_n$  abgeleitet werden.

$$\mathcal{H}_n = -j \frac{J'_{|n|}(kr_0)}{J_{|n|}(kr_0)} \mathcal{H}_{0n} \quad (4.57)$$

Aus dieser Beziehung sieht man, dass die Schnellemikrofonsignale von Druckempfängersignalen abgeleitet werden können.

### Gegenüberstellung der Methoden Ambisonic und Holophonie

Aus theoretischer Sicht ist Ambisonic zu bevorzugen, da die Enkodierung sehr effizient ist, und die Dekodierung eine perfekte Rekonstruktion ermöglicht. Der Fehler (Abschneiden der Besselfunktionen höherer Ordnung) ist grundsätzlich ein Problem der Aufnahme. Im Gegensatz dazu gibt es bei der Holophonie zusätzliche Fehler durch räumliches Aliasing und die ‚stationary phase approximation‘.

Bei der Implementation von Ambisonic gibt es aber das Problem der Aufzeichnung von Signalen höherer Ordnung und der Rekonstruktion von ebenen Wellen. Aus praktischer Sicht ist daher die Holophonie zu bevorzugen.

## 5 Verwendete Außenohrübertragungsfunktionen

Bei den Versuchen mit dem Beamforming Modell und bei der Berechnung des binauralen Stereosignals aus den Ambisonicsignalen wurden die folgenden Zielfilterkurven verwendet.

Die Messungen von Gardner und Martin (1994) [24] an einem KEMAR Kunstkopf bilden das Basismaterial. Dabei handelt es sich um 512 Punkte Impulsantworten aus 710, Richtungen die bei Elevationen von  $-40^\circ$  bis  $90^\circ$  an unterschiedlich vielen Azimuthstellen pro Elevation gemessen wurden. Um kürzere Filter realisieren zu können, werden diese Originalimpulsantworten gekürzt. Wichtig ist, daß die Phaseninformation, die beim Beamformingmodell ausschlaggebend ist, nicht verloren geht (Fall1) beziehungsweise wieder richtig rekonstruiert wird (Fall2).

### 5.1 Fall1 – Gekürzte Impulsantworten

Die 512 Punkte Impulsantworten werden gekürzt.

Dabei werden die ersten 25 Punkte (Laufzeit des Schalls bis zum Ohr) weggeschnitten. Die nächsten  $N$  Punkte werden verwendet ( $N = 256, 128$  und  $64$ ), und die letzten 50, 30 bzw. 20 Punkte werden mit einem halben Hanningfenster ausgeblendet. Die originale Phaseninformation ist erhalten.

### 5.2 Fall2 – Rekonstruierte Impulsantworten

Die KEMAR Messungen liegen als 128 Punkte Impulsantworten mit extrahierten Onsetdelay vor. Diese Kurven wurden bei der Implementation der Hauptkomponentenmethode [18] verwendet. Die Information der interauralen Zeitdifferenzen kann mit diesen Kurven nicht modelliert werden. Das entsprechende Onsetdelay muß wieder hinzugefügt werden.

#### 5.2.1 Berechnung der Onsetdelays

Bei der Modellierung der Außenohrübertragungsfunktionen mittels Hauptkomponentenmethode wurde folgendermaßen vorgegangen.

Als Ausgangsmaterial wurden die Meßdaten an einem KEMAR Kunstkopf verwendet, (Gardner, Martin 1994).

Bei Berechnung der Basisvektoren und Gewichte mittels Hauptkomponentenmethode wurde die ITD (bzw. das Onsetdelay) für jede gemessene Richtung extrahiert. Der Funktionsverlauf dieser Zeitverzögerung folgt auf der dem Ohr zugewandten Seite einer Sinusfunktion und verläuft auf der dem Ohr abgewandten Seite linear.

Um für jede beliebige Richtung das Onsetdelay bestimmen zu können, wird ein funktioneller Zusammenhang definiert.

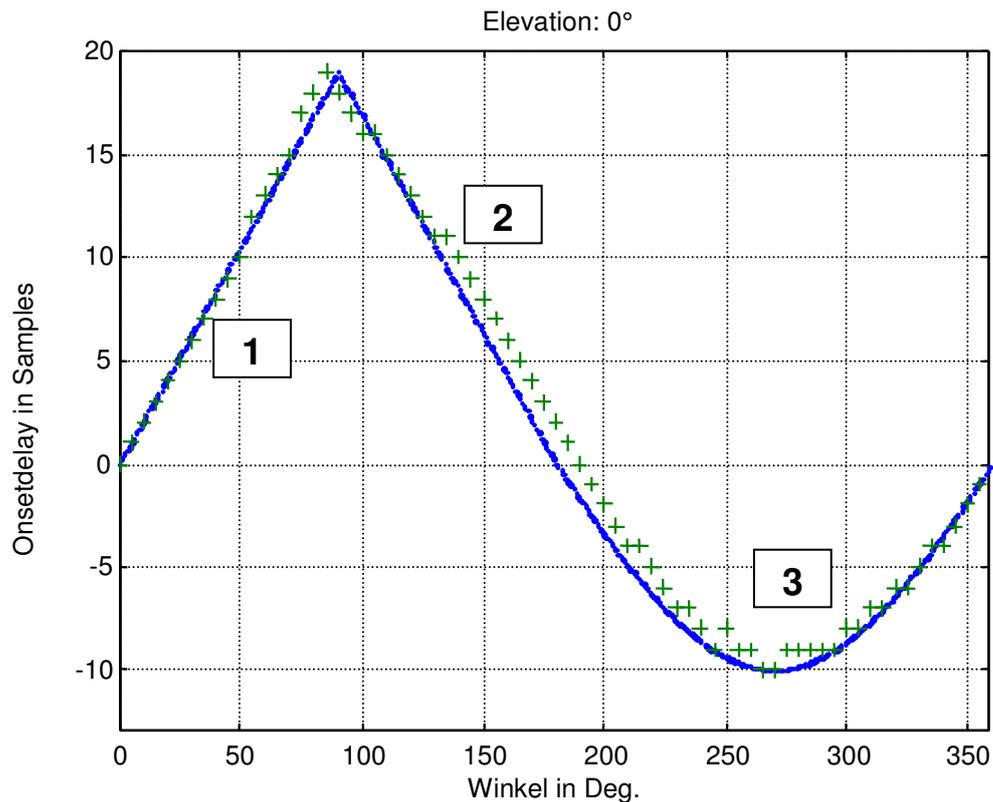


Abbildung 5.1: Funktioneller Verlauf des Onsetdelay für das linke Ohr. (KEMAR Kunstkopf).

'+'...gemessene Werte  
'.'...berechnete Werte

Zur Berechnung der Parameter wurden nur Meßdaten der linken Pinna des KEMAR Kunstkopfes verwendet. Wegen der Symmetrie erhält man die Laufzeit für das rechte Ohr, indem man den Winkel  $\theta_{\text{rechts}}=360^\circ-\theta_{\text{links}}$  in die Gleichung für das Onsetdelay einsetzt.

Anmerkung: Die Differenz ( $\text{Laufzeit}_{\text{linkes Ohr}} - \text{Laufzeit}_{\text{rechtes Ohr}}$ ) ist das ITD (interaural time delay).

Die Funktionskurve wird für die Laufzeiten mit der Elevation  $0^\circ$  aufgestellt. Die Funktion besteht aus drei Teilen.

1. Bereich ( $0^\circ \leq \text{azimuth} < 90^\circ$ ):

$$OD = \frac{19}{90} * \text{Winkel} \quad (5.1)$$

2. Bereich ( $90^\circ \leq \text{azimuth} < 180^\circ$ ):

$$OD = -\frac{19}{90} * \text{Winkel} + 38 \quad (5.2)$$

3. Bereich ( $180^\circ \leq \text{azimuth} < 360^\circ$ ):

$$OD = 10 * \sin(\text{Winkel}) \quad (5.3)$$

OD...Onsetdelay in Samples

Winkel...Azimuth in  $^\circ$

Bei Elevationen  $\neq 0^\circ$  reduzieren sich die Onsetdelays nach der Funktion

$$y = \cos(x) \quad (5.4)$$

y...Reduktionsfaktor

x...Elevation

Onsetdelay\_korrekt = (Onsetdelay für  $0^\circ$  Elevation)\*Reduktionsfaktor

Die Periodendauer bei einer Abtastrate von 44,1kHz beträgt  $22,7\mu\text{s}$ . Das maximale Onsetdelay ergibt sich daher mit  $19*22,7\mu\text{s}$  zu  $430\mu\text{s}$ .

Folgende Grafiken zeigen die gemessenen Laufzeiten (Onsetdelays) und den festgelegten funktionellen Zusammenhang für das linke Ohr.

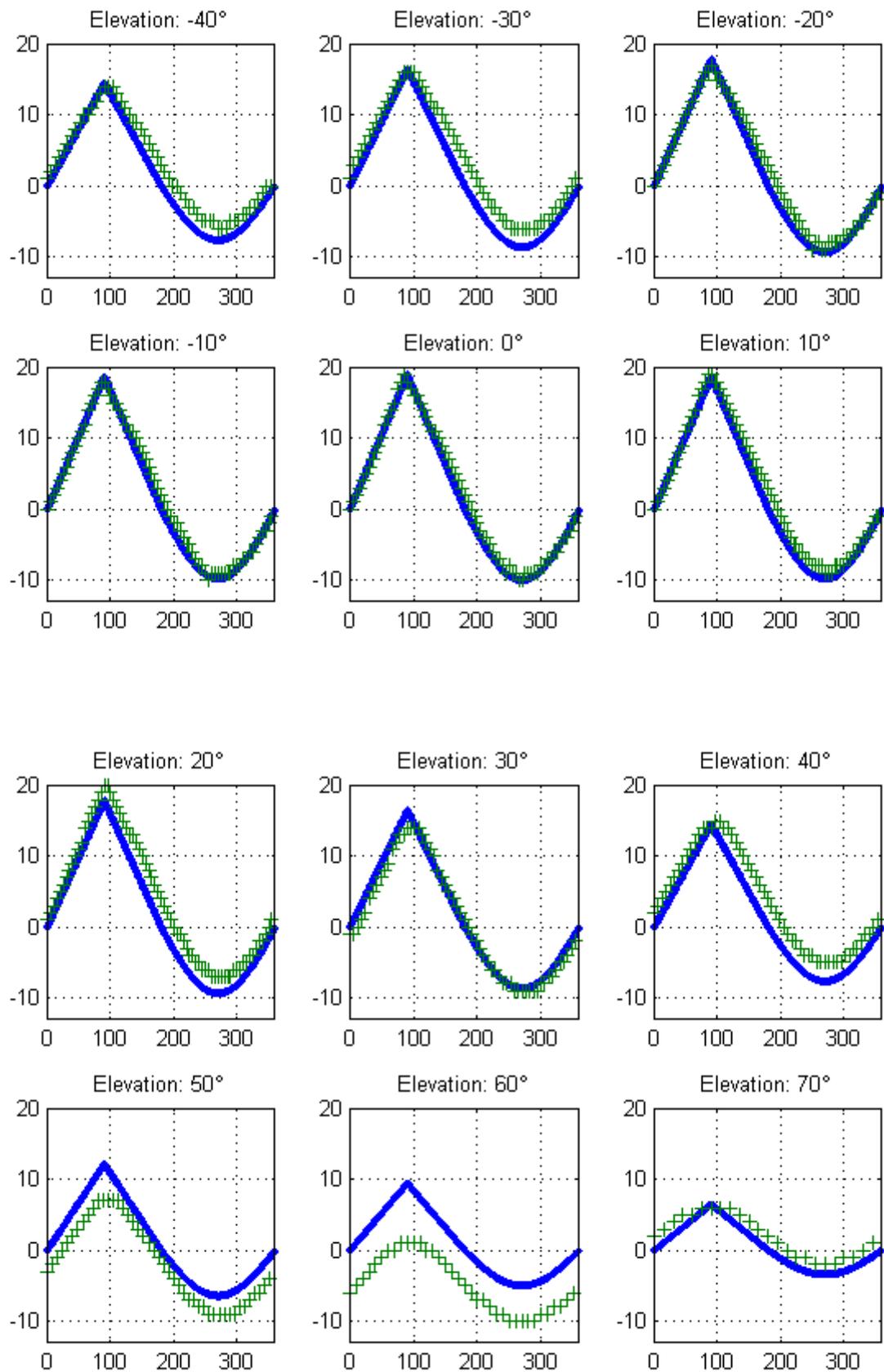


Abbildung 5.2: Funktioneller Zusammenhang des Onsetdelay für verschiedene Elevationen.

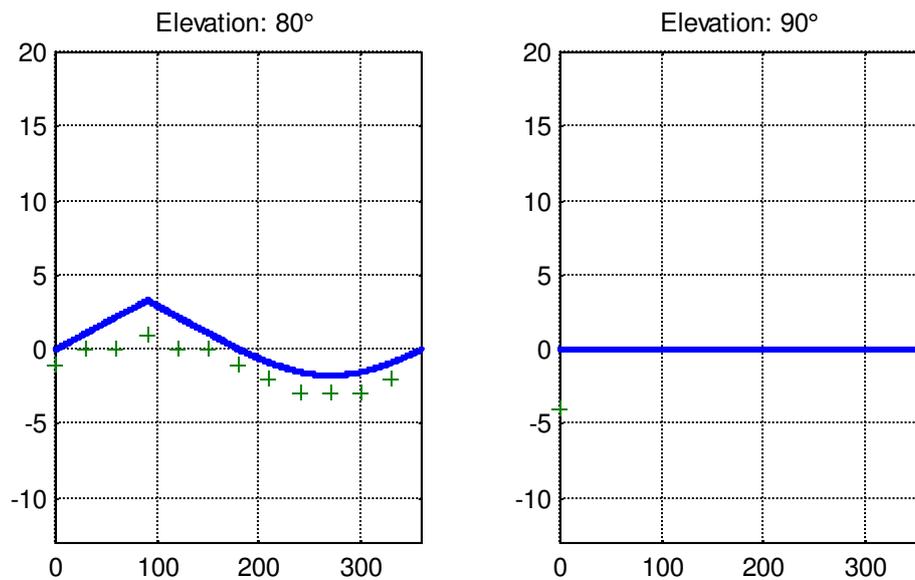


Abbildung 5.2f: Funktioneller Zusammenhang des Onsetdelay für verschiedene Elevationen.

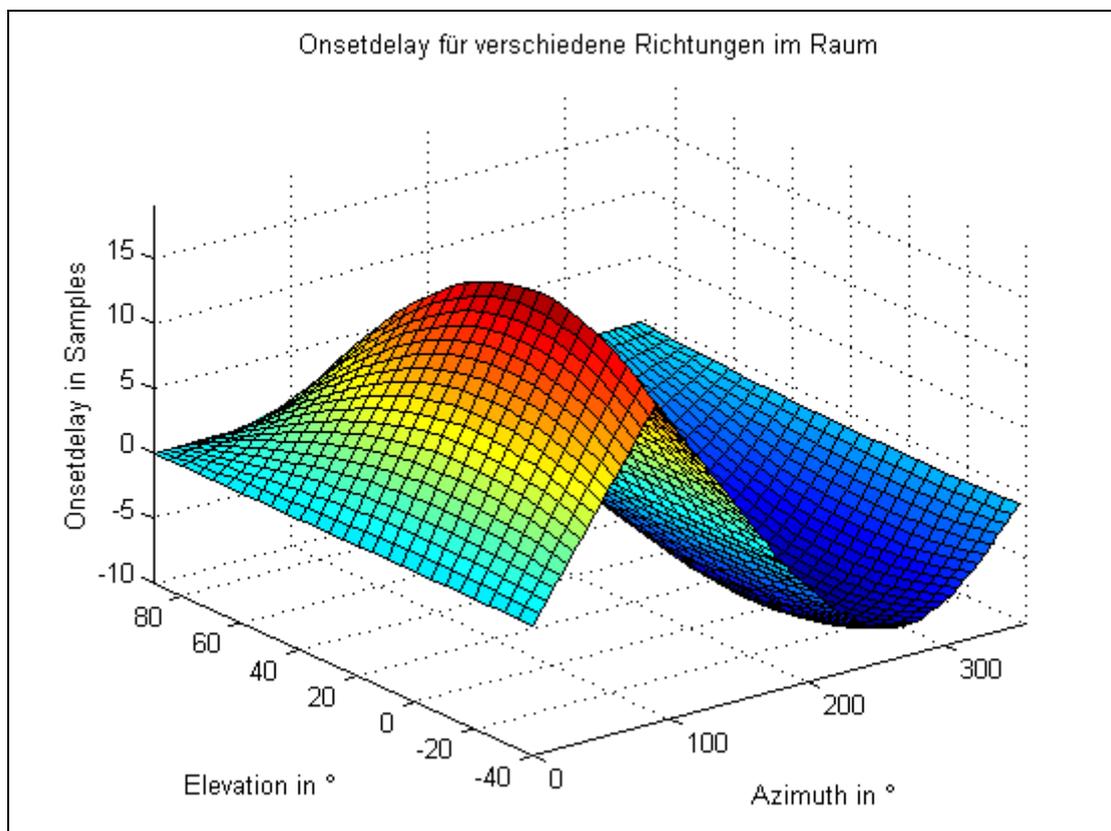


Abbildung 5.3: Verwendete Zeitverzögerungen bei der Rekonstruktion der HRTF Impulsantworten.

Das Hinzufügen des Onsetdelays erfolgt durch multiplizieren der Frequenzantworten mit einem entsprechenden linearphasigen Filter. (Die Multiplikation eines linearphasigen Filters mit der Amplitude 1 entspricht einer reinen Zeitverzögerung.)

## 6 Literaturverzeichnis

- [1] Robert H. Gilkey, Timothy R. Anderson: "Binaural and Spatial Hearing in Real and Virtual Environments"; Lawrence Erlbaum Associates, Publishers; Mahwah, New Jersey; 1997.
- [2] Jens Blauert: "The Psychophysics of Human Sound Localization"; The MIT Press; Revised Edition 1997.
- [3] Philip Mackensen, Klaus Reichenauer, Günther Theile: "Einfluß der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören"; Tagungsband, 20.Tonmeistertagung 1998.
- [4] Gary S. Kendall, William L. Martens, Shawn L. Decker: "Spatial Reverberation: Discussion and Demonstration"; Current Directions in Computer Music Research ed Max Mathews and John Pierre Cambridge: MIT Press, 1991.
- [5] Udo Zölzer: "Digitale Audiosignalverarbeitung"; Teubner Verlag Stuttgart; 1996.
- [6] Hilmar Lehnert, Jens Blauert: "Principles of Binaural Room Simulation"; Applied Acoustics 36 (p.259-291), 1992.
- [7] R.MacLaverty, D.J. Furlong: "Room Acoustical Simulation Algorithms Compared"; presented at the 93<sup>rd</sup> AES Convention, San Francisco ;Oct.1992.
- [8] Chris Kyriakis, Panagiotis Tsakalides, Tomlinson Holman: "Surrounded by Sound, Acquisition and Rendering Methods for Immersive Audio"; IEEE Signal Processing Magazine; January 1999.
- [9] Hamid Krim, Mats Viberg: "Two Decades of Array Signal Processing Research"; IEEE Signal Processing Magazine; July 1996.
- [10] Barry D. Van Veen, Kevin M. Buckley: "Beamforming: A Versatile Approach to Spatial Filtering"; IEEE ASSP Magazine; April 1988.
- [11] Wolfgang H. Kummer: "Basic Array Theory"; Proceedings of the IEEE, Vol. 80, No.1; January 1992.
- [12] William F. Gabriel: "Adaptive Arrays – An Introduction"; Proceedings of the IEEE, Vol.64, No.2; February 1976.
- [13] R. C. Hansen: "Array Pattern Control and Synthesis"; Proceedings of the IEEE, Vol.80, No.1; January 1992.
- [14] Saeed Gazor, Yves Grenier: "Criteria for Positioning of Sensors for a Microphone Array"; IEEE Transactions on Speech and Audio Processing, Vol.3, No.4; July 1995.
- [15] Kevin M. Buckley: "Spatial/Spectral Filtering with Linearly Constrained Minimum Variance Beamformers"; IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.ASSP-35, No.3; March 1987.
- [16] S.S. Shanan, C.A. Pomalaza-Raez: "The use of nonuniform element spacing in array processing algorithms"; J.Acoust.Soc.Am., Vol.86(4); October 1989.
- [17] J.L. Flanagan, J.D. Johnston, R.Zahn, G.W. Elko: "Computer-steered microphone arrays for sound transduction in large rooms"; J.Acoust.Soc.Am.,Vol.78(5); November 1985.

- [18] Doris J.Kistler, Frederic L. Wightman: "A model of head-related transfer functions based on principal component analysis and minimum-phase reconstruction"; J.Acoust.Soc.Am., Vol.91(3); March 1992.
- [19] Michael J. Evans, James A.S. Angus, Anthony I. Tew: "Analyzing head related transfer function measurements using surface spherical harmonics"; J.Acoust.Soc.Am., Vol.104(4); October 1998.
- [20] Jiashu Chen, Barry D. Van Veen, Kurt E. Hecox: "External ear transfer function modeling: A beamforming approach"; J.Acoust.Soc.Am., Vol.92(4), Pt.1, October 1992.
- [21] J.S. Bamford: "An Analysis of Ambisonic Sound Systems of First and Second Order", Thesis presented to the University of Waterloo, Waterloo, Ontario, Canada; 1995.
- [22] David G. Malham, Anthony Myatt: "3-D Sound Spatialization using Ambisonic Techniques"; Computer Music Journal, 19:4, pp.58-70, 1995 Massachusetts Institute of Technology; Winter 1995.
- [23] M. Poletti: "The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems"; J.Acoust.Soc.Am., Vol.44, No.11; November 1996.
- [24] B. Gardner, K. Martin: "HRTF Measurements of a KEMAR Dummy-Head Microphone"; <http://sound.media.mit.edu/KEMAR.html>; 1994.
- [25] F.L.Wightman, D.J.Kistler: "Headphone Simulation of free-field listening I: Stimulus Synthesis"; J.Acoust.Soc.Am., Vol.85, pp.858-867; 1989.
- [26] Lord Rayleigh (J.W.Strutt, 3<sup>rd</sup> Baron of Rayleigh): "On our perception of sound direction"; Philos.Mag.13, pp.214-232; 1907.
- [27] J.L. Van Soest: "Richtungshooren bij sinusvormige geluidstrillingen [Directional hearing of sinusoidal sound waves]"; Physica, Vol.9, pp.271-282; 1929.
- [28] E.M. von Hornbostel, M.Wertheimer: "Über die Wahrnehmung der Schallrichtung [On the perception of the direction of sound]"; Sitzungsber. Akad. Wiss. Berlin, pp.388-396; 1920.
- [29] W.R. Thurlow, J.W. Mangels, P.S. Runge: "Head Movements during sound localization"; J.Acoust.Soc.Am., Vol.42, pp.489-493, 1347; (1967).
- [30] Lawrence J. Ziomek: "Acoustic Field Theory and Space-Time Signal Processing"; CRC Press, Inc.; 1995.
- [31] I.Haykin, S.Simon: "Advances in spectrum analysis and array processing"; Prentice-Hall, Inc.; 1991.
- [32] Rozenn Nicol, Marc Emerit: "3D-Sound Reproduction Over An Extensive Listening Area, A Hybrid Method Derived From Holophony And Ambisonic"; AES 16<sup>th</sup> International Conference, pp.436-453; 1999.
- [33] Rozenn Nicol, Marc Emerit: "Reproducing 3D-Sound for Videoconferencing: a Comparison between Holophony and Ambisonic"; DAFX98 Proceedings pp.17-20; 1998.
- [34] Carsten Sydow: "Selbstausrichtende Mikrofonarraysysteme"; Fortschr.-Ber. VDI Reihe 10 Nr.429; Düsseldorf; 1996.

## 7. Audio CDs mit Beispielen verschieden berechneter Binauralsignale

### CD 1

Alle Beispiele auf CD 1 wurden mit MATLAB 5.1 generiert.

Track	Inhalt
1	Monosignal. 10 mal 100 ms Rosa Rauschen. Ausgangsmaterial für die folgenden Audiobeispiele.
	<b>Rosa Rauschen Testsignal (aus 1) gefaltet mit KEMAR Impulsantworten verschiedener Länge.</b>
2	512 Punkte
3	256 Punkte
4	128 Punkte
5	64 Punkte
6	Rosa Rauschen Testsignal (aus 1) gefaltet mit minimalphasig rekonstruierten KEMAR Impulsantworten. Länge = 128 Punkte.
	<b>Binaurale Stereosignale berechnet mittels Ambisonic verschiedener Ordnung. 72 virtuelle Lautsprecher. Gewicht des Gliedes höchster Ordnung = 2.</b>
7	1.Ordnung
8	2.Ordnung
9	3.Ordnung
10	4.Ordnung
11	5.Ordnung
12	6.Ordnung
	<b>Binaurale Stereosignale berechnet mittels Ambisonic 3. Ordnung. 72 virtuelle Lautsprecher. Verschiedene Gewichte des Gliedes höchster Ordnung.</b>
13	Gewicht = 2.0
14	Gewicht = 1.5
15	Gewicht = 1.0
16	Gewicht = 0.5
17	Gewicht = 0.1
	<b>Binaurale Stereosignale berechnet mittels Ambisonic 3. Ordnung. Verschiedene Anzahl virtueller Lautsprecher (L). Verschiedene Gewichte (G) des Gliedes höchster Ordnung. (2.0 oder 1.0).</b>
18	L = 72, G = 2.0
19	L = 72, G = 1.0
20	L = 36, G = 2.0
21	L = 36, G = 1.0
22	L = 18, G = 2.0
23	L = 18, G = 1.0
24	L = 8, G = 2.0
25	L = 8, G = 1.0
26	L = 7, G = 2.0
27	L = 7, G = 1.0
	<b>Binaurale Stereosignale berechnet mittels Ambisonic verschiedener Ordnung (O). Verschiedene Anzahl virtueller Lautsprecher (L). Verschiedene Gewichte (G) des Gliedes höchster Ordnung.</b>
28	O = 1, L = 3, G = 2.0
29	O = 2, L = 5, G = 2.0
30	O = 2, L = 5, G = 1.0

31	O = 2, L = 5, G = 0.8
32	O = 3, L = 7, G = 2.0
33	O = 3, L = 7, G = 1.0
34	O = 3, L = 7, G = 0.8
	<b>Binaurales Stereosignal berechnet mittels Faltung der originalen 512 Punkte KEMAR Impulsantworten. Verschiedene Elevationen.</b>
35	Elevation = -40°
36	Elevation = -20°
37	Elevation = 0°
38	Elevation = 20°
39	Elevation = 40°
40	Elevation = 60°
41	Elevation = 80°
42	Binaurales Stereosignal berechnet mittels Ambisonic 2. Ordnung. 5 virtuelle Lautsprecher. Gewicht des höchsten Gliedes = 0.8. Berücksichtigung von Hall aus 4 verschiedenen Richtungen.
43	Thema Vivaldi mono.
44	Binaurales Stereosignal mittels Ambisonic. Thema Vivaldi.
45	Thema Emerson mono.
46	Binaurales Stereosignal mittels Ambisonic. Thema Emerson.

## CD 2

Alle Beispiele auf CD 2 wurden mit dem auf der SGI O2 implementierten max-patch generiert.

Track	Inhalt
1	Introduktion aus Don Giovanni. Rotation der Stereolautsprecher (+30°, -30°). Rotation um 360°. 1 Umlauf dauert 1 Minute. 3 Umläufe (links –rechts –links)
2	Arie des Don Giovanni. Stereolautsprecher auf +30° und -30°. Rotationen mit 90° Inkrementen. Dauer der Bewegung = 10 Sekunden.
3	Letzte Szene aus Don Giovanni. Rotationen um unterschiedliche Winkel. Kurze Dauer der Bewegung (1 Sekunde – 10 Sekunden).
4	Clara Rockmore, Theremin: Saint Saens – The Swan. Rotationen um 360°. Dauer der Bewegung = 1 Minute.
5	Clara Rockmore, Theremin: Rachmaninoff – Vocalise. Stereolautsprecher auf +45° und -45°. Schnelle Rotationen.
6	Clara Rockmore, Theremin: Song of Grusia. Rotation = 0°. Unterschiedliche Position der beiden Lautsprecher.
7	Michel Chion: Requiem, Domine Deus. Stereolautsprecher auf +30° und -30°. Steuerung der Rotation mit MIDI.