

Recheneffiziente Modellierung von Außenohrübertragungsfunktionen

Diplomarbeit

von

Maria Fellner

im Rahmen des interuniversitären Studiums Elektrotechnik-Toningenieur
durchgeführt am

Institut für Elektronische Musik

an der Universität für Musik und darstellende Kunst Graz

Vorstand: o.HProf. Mag. Gerd Kühr

in Zusammenarbeit mit dem

Institut für Nachrichtentechnik und Wellenausbreitung

an der Technischen Universität Graz

Vorstand: o.Univ.-Prof. Dipl.-Ing. DDr. Willibald Riedler

Betreuer: HAss. Mag. Dipl.-Ing. Dr. Robert Höldrich

Graz, im Oktober

1998

Meinen Eltern gewidmet

Zusammenfassung

Diese Arbeit behandelt recheneffiziente Methoden zur Modellierung der Außenohrübertragungskurven (HRTFs). Nach einer Einführung über räumliches Hören und die menschlichen HRTFs werden verschiedene mathematische Modelle genauer vorgestellt. Eines dieser Modelle, die Hauptkomponentenmethode, wird implementiert. Die vorliegenden Ergebnisse zeigen eine gute Übereinstimmung mit der Theorie und eine besonders effiziente Anwendbarkeit im Fall von mehreren zu spatialisierenden Schallquellen.

Abstract

This diploma thesis presents efficient methods for modelling the head-related transfer functions (HRTFs). After an introduction to spatial hearing and the human HRTFs, various mathematical models are discussed in detail. One approach, the principal components analysis, is implemented. The results coincide very well with theory and show a very efficient applicability especially for various sound sources to be spatialized.

Danksagung

Mein Dank gilt Dr. Robert Höldrich für die hervorragende Betreuung, Dipl.-Ing. Wolfgang Werth für die Unterstützung betreffend den Abschnitt über Balanced Model Truncation und nicht zuletzt Ferdinanda Anhofer für das geduldige Scannen unzähliger Bilder.

Inhaltsverzeichnis

| | |
|---|-----------|
| 1 Die Grundlagen räumlichen Hörens | 1 |
| 1.1 Allgemeines | 1 |
| 1.1.1 Historischer Abriß | 1 |
| 1.1.2 Umgebungskontext | 3 |
| 1.1.3 Aufbau des Ohrs | 5 |
| 1.1.4 Ein Modell des natürlichen und virtuellen räumlichen Hörens | 6 |
| 1.1.5 Die Ziele von virtuellem räumlichen Hören | 6 |
| 1.2 Wahrnehmung von Azimut und Elevation | 7 |
| 1.2.1 Interaurale Zeit- und Intensitätseffekte | 7 |
| 1.2.2 Bewegungen von Kopf und Quelle | 8 |
| 1.2.3 Spektrale Einflüsse der Pinnae | 10 |
| 1.3 Auralisation, Schallquellendistanz und Umgebungskontext | 12 |
| 1.3.1 Distanz | 12 |
| 1.3.2 Raumeindruck | 15 |
| | |
| 2 HRTF-Kurven | 17 |
| 2.1 Eigenschaften der HRTF-Kurven | 17 |
| 2.1.1 Einflußgrößen | 17 |
| 2.1.2 HRTF-Amplituden-Charakteristik | 18 |
| 2.1.3 HRTF-Phasen-Charakteristik | 21 |
| 2.2 Lokalisation mit HRTFs | 22 |
| 2.3 Messung von HRTF-Kurven | 23 |
| 2.3.1 Messung | 23 |
| 2.3.2 Equalisation von HRTFs | 23 |
| 2.3.3 Vorgangsweise im Fall der KEMAR-Kurven | 26 |
| 2.4 Vorhandenes HRTF-Material | 27 |
| 2.4.1 Meßdaten von Gardner/Martin | 27 |
| 2.4.2 Meßdaten von Wightman/Kistler | 28 |
| | |
| 3 Modellierung von HRTF-Kurven | 31 |
| 3.1 Hauptkomponentenmethoden | 31 |
| 3.1.1 Allgemeines | 31 |
| 3.1.2 Prinzip | 31 |
| 3.1.3 Hauptkomponentenanalyse und minimalphasige Rekonstruktion | 35 |

| | |
|---|------------|
| 3.1.4 Extraktion der spatialen Eigenschaften und Regularisierung | 38 |
| 3.1.5 Extraktion der spatialen Eigenschaften im Zeitbereich | 42 |
| 3.2 FIR- und IIR-Filterdesign | 44 |
| 3.2.1 FIR-Filterdesign | 44 |
| 3.2.2 IIR-Filterdesign | 45 |
| 3.2.3 Zeitvariante Implementation | 49 |
| 3.3 Modellierung durch Richtstrahlbildung | 50 |
| 3.3.1 Allgemeines | 50 |
| 3.3.2 Prinzip | 51 |
| 3.3.3 LMS-Richtstrahler-Design | 52 |
| 3.3.4 Beschreibung der experimentellen Daten | 56 |
| 3.3.5 Experimentelle Ergebnisse | 57 |
| 3.4 Balanced Model Truncation (BMT) | 61 |
| 3.4.1 Allgemeines | 61 |
| 3.4.2 Prinzip der balancierten Realisierung und Modellreduktion | 61 |
| 3.4.3 Konversionsalgorithmus FIR->IIR | 68 |
| 3.4.4 Experimentelle Ergebnisse | 70 |
| 3.4.5 Weitere mathematische Zusammenhänge | 72 |
| 4 Implementierung von HRTFs nach der Hauptkomponentenmethode | 77 |
| 4.1 Bearbeitung und Korrektur der Rohdaten | 80 |
| 4.2 Berechnung der Hauptkomponenten | 93 |
| 4.2.1 Rechnung im Zeitbereich | 93 |
| 4.2.2 Reelle Rechnung im Frequenzbereich | 107 |
| 4.2.3 Komplexe Rechnung im Frequenzbereich | 114 |
| 5 Zusammenfassung und Ausblick | 121 |
| 5.1 Zusammenfassung | 121 |
| 5.2 Ausblick | 121 |
| Literaturverzeichnis | 123 |

1 Die Grundlagen räumlichen Hörens

1 Allgemeines

Im folgenden wird eine Einführung in die wichtigsten Effekte räumlichen Hörens gegeben. Die Behandlung erfolgt dabei nur aus Sicht der physikalischen Akustik und der Psychoakustik, nicht jedoch der Neurophysiologie, da diese für die hier vorgestellten Modelle nicht von Bedeutung ist.

1 Historischer Abriß

Dieser Abschnitt folgt im wesentlichen [Wightman, Kistler, 1993] und [Duda, 1997].

Obwohl es im 19. und beginnenden 20. Jahrhundert noch keine Technologien gab, die für die Erzeugung und Kontrolle der entsprechenden Testtöne für systematische empirische Studien geeignet gewesen wären, wurden schon zu dieser Zeit einige Theorien über die Lokalisation von Schallquellen diskutiert.

Thompson veröffentlichte in seinem Essay "On the Function of the Two Ears in the Perception of Space" 1882 eine Zusammenfassung dieser Theorien. Zu dieser Zeit gab es drei grundlegende Positionen. Steinhauser und Bell berücksichtigten nur die interauralen Intensitätsdifferenzen (*interaural intensity differences*, IIDs) und gingen davon aus, daß die interauralen Zeitdifferenzen (*interaural time differences*, ITDs) irrelevant seien. A. M. Mayer ging davon aus, daß sowohl ITDs als auch IIDs relevant seien. Die dritte Theorie, Mach und Lord Rayleigh zugeschrieben, war ebenfalls eine "IID-Theorie". Mach vermutete jedoch zusätzlich, daß die Richtung, aus der der Schall wahrgenommen wird, durch die Wirkung der Pinnae als Resonatoren für hohe Frequenzen bestimmt wird.

John Strutt, der besser als Lord Rayleigh bekannt ist, formulierte 1907 die sogenannte "Duplex-Theorie", die auf IIDs und ITDs basierte und Machs Pinna-Resonanzen außer Acht ließ. Diese Theorie dominierte die Untersuchungen über menschliche räumliche Wahrnehmung die nächsten 50 Jahre. Lord Rayleigh beobachtete, daß bei Schallwellen mit kleiner Wellenlänge im Vergleich zu den Kopfabmessungen der Kopf eine Abschattung des Schalls bewirkt, sodaß der Schallintensität an dem der Schallquelle zugewandten (= ipsilateralen) Ohr höher ist als am abgewandten (= kontralateralen) Ohr. Das bedeutet, daß eine IID entsteht. Außerdem bewirken die verschiedenen Distanzen zwischen Schallquelle und ipsi- bzw. kontralateralem Ohr eine ITD. Er konnte demonstrieren, daß die ITDs besonders bei tiefen Frequenzen wirksam werden, wo die IIDs vernachlässigbar sind. Daher argumentierte er, daß die Lokalisation bei tiefen Frequenzen von ITDs und bei hohen Frequenzen von

IIDs bestimmt wird, wobei der Übergangsbereich um 1,5kHz liegt.

D. W. Batteau publizierte in den 1960ern über die Effekte der Pinnae im Zeitbereich. Er betrachtete die Pinnae als stark richtungsabhängige Reflektoren, die zu jedem einfallenden Schall ein bestimmtes Muster an Echos hinzufügen, das von Azimut und Elevation abhängt. Das menschliche Ohr würde diese Echomuster dekodieren und dadurch die Schallrichtung bestimmen. Dafür wäre aber eine viel höhere zeitliche Auflösung nötig gewesen als beim menschlichen Ohr beobachtet worden war.

J. H. Hebrank und D. Wright erkannten 1974, daß Batteaus Pinna-Reflektionen spektrale Entsprechungen haben und diese spektralen Effekte vom menschlichen Ohr sehr wohl dekodiert werden können.

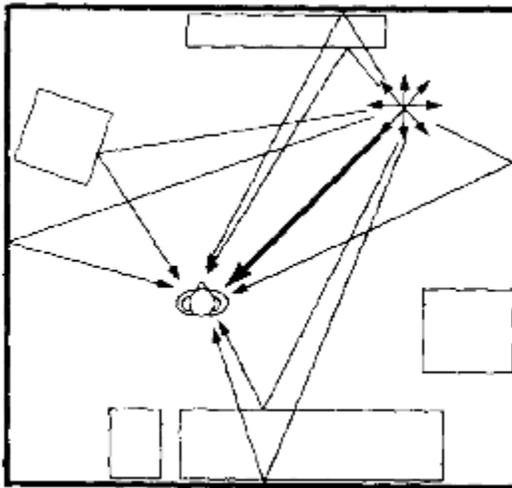
Dieses neue Augenmerk auf Pinna-Effekte führte zu einer Reihe von Untersuchungen, die den fundamentalen Einfluß der Pinna-Effekte auf die Richtungswahrnehmung zum Inhalt hatten. Es zeigte sich, daß die Pinna-Filterung besonders für die Lokalisation in der Medianebene und in der Vorne-/Hinten-Unterscheidung essentiell ist, wo die interauralen Differenzen minimal sind. Ebenso bewirkt sie die Externalisation eines Schalleindrucks.

Auch wenn sie qualitativ ähnlich sind, gibt es bei den Pinnae intersubjektive Unterschiede, und es kann ein Zusammenhang zur Lokalisationsgenauigkeit hergestellt werden.

2 Umgebungskontext

Direkter und indirekter Schall

Bei einem akustischen Ereignis in natürlicher Umgebung breiten sich die Schallwellen nach allen Richtungen aus. Dabei treffen sie auf Raumobjekte, durch die sie reflektiert oder gebeugt werden. So entsteht eine komplexe Mischung, die aus mehreren Schallquellen vorhanden sind. Eine spezielle Position ist der/die Hörende. Diese Schallwellen zu verschiedenen Richtungen trifft eine Wellenfront als erste



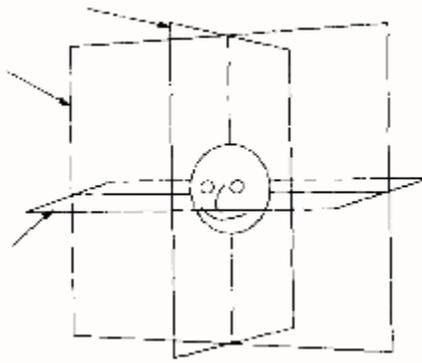
Ereignis in natürlicher Schallwellen nach allen sie auf Raumobjekte, durch werden. So entsteht eine besonders wenn mehrere Eine spezielles "Raumobjekt" Position erreichen die verschiedenen Zeiten und aus Wie aus Abb. 1.1 ersichtlich, auf geradem Weg bei dem/r

Hörenden ein, ohne zuvor an einer umgebenden Oberfläche reflektiert worden zu sein. Dieser Direktschall enthält die am wenigsten verfälschte Information über die Richtung des Schallereignisses. Danach werden die Schallwellen von der Umgebung reflektiert und kommen aus verschiedenen Richtungen mit verschiedenen Zeitverzögerungen bei dem/r Hörenden an. Dieser indirekte Schall enthält Information über die Umgebung und die relative Position des Schallereignisses, im speziellen den Abstand von dem/r Hörenden.

Am Kopf des/r Hörenden werden Schallwellen abhängig von der Wellenlänge des Schalls im Vergleich zu den Kopfabmessungen gebeugt ($<1,5$ kHz) oder reflektiert ($>1,5$ kHz). Im Übergangsbereich um 1,5 kHz sind die Wellenlängen des Schalls in der Größenordnung des Kopfdurchmessers.

Ipsilaterales und

Als ipsilaterales Ohr bezeichnet man das dem Schallereignis zugewandte Ohr, zu dem der Schall den kürzeren Weg hat. Das kontralaterale Ohr ist das dem Schallereignis abgewandte Ohr.



kontralaterales Ohr

bezeichnet man das dem Ohr, zu dem der Schall den kürzeren Weg hat. Das kontralaterale Ohr ist das dem Schallereignis abgewandte Ohr.

Medianebene, Horizontalebene und Frontalebene

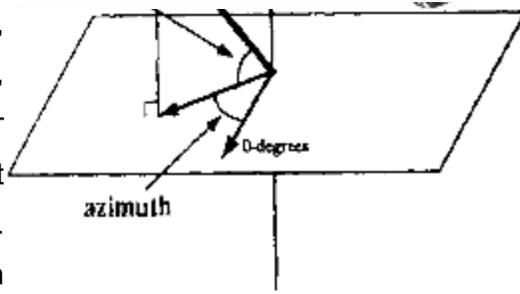
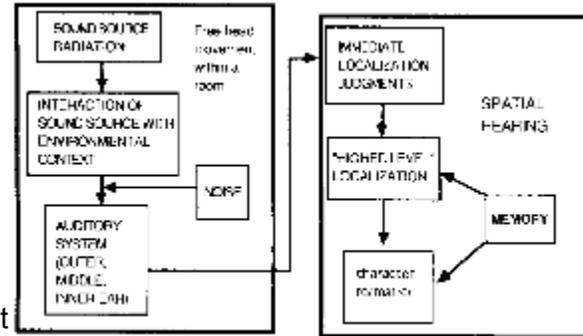
Die Medianebene ist eine Region, in der Schallquellen zu beiden Ohren den gleichen Abstand haben. Die Horizontalebene verläuft waagrecht und geht horizontal durch beide Ohren. Die Frontalebene liegt vertikal und geht durch beide Ohren.

Azimut, Elevation und Distanz

Das Koordinatensystem besteht aus zwei Winkeln und einem Skalar. Der Azimut ist der Winkel in der Horizontalebene und hat einen Bereich von 0° (direkt vor dem Hörer) bis 360° . Er wird von 0° bis zur Projektion des Distanzvektors in die Horizontalebene gemessen. Uneinigkeit herrscht in der Literatur darüber, ob 90° Azimut links oder rechts von dem/r Hörenden sei. Die Elevation ist der Winkel zwischen der Horizontalebene und dem Distanzvektor. Die Werte gehen von 0° bis 90° (nach oben) bzw. -90° (nach unten). Als Skalar wird die Distanz angegeben.

3 Aufbau des Ohrs

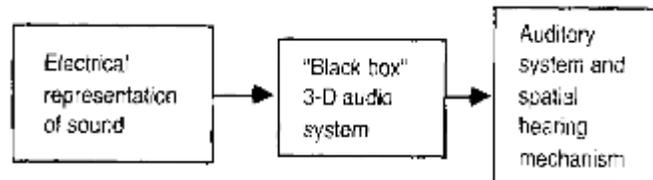
Der Schall (A) wird zuerst durch die Pinna (B), den sichtbaren Teil des Außenohrs, wie Schultern und Kopf, die Effekte des Ohrkanals (oder Mittelohr führt. Dieses besteht aus dem Trommelfell (D) und den Gehörknöchelchen (E). wird durch Bewegung am Flüssigkeitsdruck im Innenohr (oder Cochlea, G) transformiert. Dieser Druck verursacht frequenzabhängige Erregungsmuster auf der Basilarmembran (H). Dadurch werden Haarzellen angeregt, die wiederum elektrische Aktionspotentiale in den Neuronen des Auditiven Systems auslösen. Diese werden auf höheren Niveaus mit den Informationen des anderen Ohrs kombiniert und verarbeitet.



und die nahen Körperteile, beeinflusst. Danach folgen Meatus, C), der zum aus dem Trommelfell (D) Die mechanische Energie ovalen Fenster (F) in Flüssigkeitsdruck im Innenohr (oder Cochlea, G) transformiert. Dieser Druck verursacht frequenzabhängige Erregungsmuster auf der Basilarmembran (H). Dadurch werden Haarzellen angeregt, die wiederum elektrische Aktionspotentiale in den Neuronen des Auditiven Systems auslösen. Diese werden auf höheren Niveaus mit den Informationen des anderen Ohrs kombiniert und verarbeitet.

3 Ein Modell des virtuellen natürlichen und räumlichen Hörens

Unter natürlichen Hörbedingungen gibt es eine Schallquelle, deren abgestrahlter Schall



Unter natürlichen Hörbedingungen gibt es abgestrahlter Schall

zuerst von der Umgebung und dann vom Auditiven System beeinflusst wird. Dieser spezielle Schalleindruck wird vom Gehirn auf mehreren Ebenen verarbeitet. Bei virtuellem räumlichen Hören ist der Ausgangspunkt die elektrische Repräsentation des Schallfeldes. Diese wird in einem 3D-Audiosystem bearbeitet und danach in gleicher Weise vom Gehirn verarbeitet.

4 Die Ziele von virtuellem räumlichen Hören

3D-Sound bedeutet, daß der/die Hörende Schall aus Richtungen hört, die bei konventionellem Stereo nicht wahrgenommen werden. Bei der Nachbildung räumlichen Hörens soll ein möglichst "natürlicher" Raumeindruck entstehen, der die bekannten Effekte simuliert.

Zu den Anwendungsgebieten virtuellen räumlichen Hörens zählen beispielsweise integrierte Audiodisplays (z. B. für die Raumfahrt), Audio-User-Interfaces für Blinde, Head-Trackted-3D-Audio für virtuelle Umgebungen, Computer Workstations mit 3D-Audio für Kommunikation (z. B. Telekonferenzen), Binauralprozessoren für Aufnahme und Wiedergabe, Nutzung binauraler Effekte zur

Verbesserung von Rauschabstand und Sprachverständlichkeit sowie künstlerische Anwendungen.

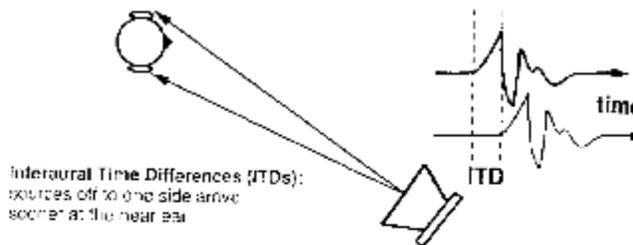
2 Wahrnehmung von Azimut und Elevation

1 Interaurale Zeit- und Intensitätseffekte

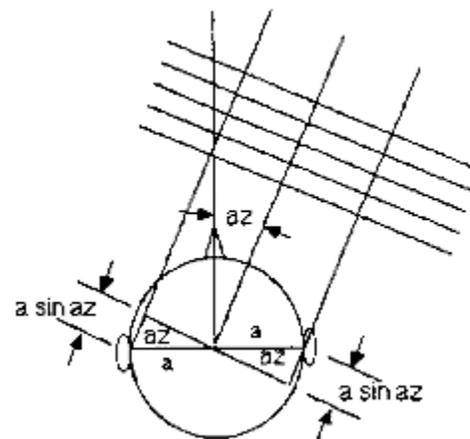
Die frequenzabhängigen Kurven der interauralen Zeitdifferenzen (ITDs) und interauralen Intensitätsdifferenzen (IIDs) sind schon seit über 100 Jahren bekannt und beruhen auf der räumlichen Trennung der beiden Ohren durch den Kopf. Ihre Wirkungsweise beschreibt die Duplex-Theorie.

ITDs entstehen durch Laufzeitdifferenzen des Schalls zwischen ipsi- und kontralateralem Ohr. Bei Wellenlängen, die viel größer als der Kopfdurchmesser sind, wird der Schall am Kopf gebeugt und die Phasenlage detektiert. Daher werden ITDs besonders bei Frequenzen bis zu 1,5 kHz wirksam. Oberhalb dieser Frequenz ist die Phasenlage nicht mehr eindeutig, jedoch kann das Auditive System bei höheren Frequenzen auch ITDs der Einhüllenden wahrnehmen¹.

IIDs bedeuten unterschiedliche Schallintensitäten an den Ohren durch Abschattung des Schalls am kontralateralen Ohr um bis zu 35 dB. Voraussetzung für die Abschattung von Schallwellen ist, daß die Wellenlängen viel kleiner als der Kopfdurchmesser sind. Daher werden IIDs besonders bei Frequenzen ab 1,5 kHz wirksam.



Experimente zur Lateralisation bewirken Wahrnehmung nur durch IIDs bei Dabeı wird aber die entlang der interauralen empfunden. Dennoch über gerade noch *notable differences*, Winkel² (*minimal audible*



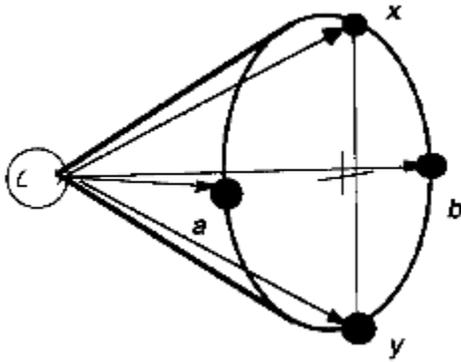
sogenannten "räumliche" (s): Erzeugen von ITDs und Kopfhörerwiedergabe. Schallquelle als im Kopf Achse wandernd können damit Aussagen hörbare Differenzen (*just inds*), minimale hörbare *angles*, MAAs) und das

Zusammenwirken der ITDs und IIDs (*time-intensity-trading*) gemacht werden.

¹Je länger die Anstiegszeit der Einhüllenden ist, desto schlechter wird die Lokalisation, besonders in halliger Umgebung.

²Der MAA beträgt in der Horizontalebene abhängig von der Frequenz 1° bis 3° und steigt rasch an, sobald der Azimut 90° erreicht.

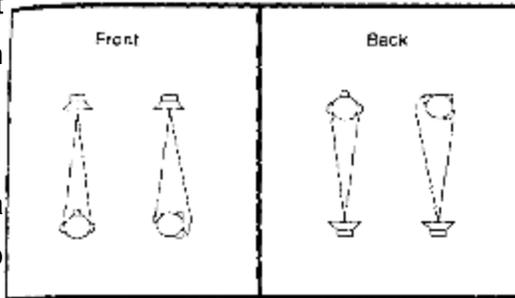
Der Präzedenzeffekt, auch ersten Wellenfront" genannt, des Auditiven Systems, um reflexionsbehafteten Räumen ein Monosignal über verschiedenen ITDs eingespielt, so ergeben sich drei Stadien der Wahrnehmung. Bei ITDs von 0 bis 0,6 ms wandert die Schallquelle scheinbar im Kopf vom Zentrum zu dem Ohr, wo der Schall zuerst eintrifft. Zwischen etwa 0,6 und 35 ms wird die Quelle nur an der Seite des zuerst ankommenden Schalls geortet (Präzedenzeffekt). Bei einer ITD größer als 35 ms werden die beiden Schallereignisse getrennt (als Echo) wahrgenommen.



Haas-Effekt oder "Gesetz der ist ein wichtiger Mechanismus Schall auch in lokalisieren zu können. Wird Kopfhörer nur mit

2 Bewegungen von Kopf und Quelle

Nimmt man einen kugelförmigen Kopf ohne Außenohren an, so kann bei gegebener IID und ITD der Ort der Schallquelle nicht eindeutig bestimmt werden. Abbildung 1.10 zeigt den sogenannten "Cone of Confusion".



eindeutig bestimmt werden. sogenannten "Cone of (vorne) erzeugt dieselbe ITD (hinten). Dies nennt man die Ebenso verhält es sich mit

Ein Schallereignis am Ort a und IID wie eines am Ort b Vorne-Hinten-Ambiguität. Schallquellen an den Orten x (oben) und y (unten), der Elevationsambiguität.

Natürlich ist dieser Ansatz theoretisch, da der Kopf des Menschen nicht sphärisch ist und die Außenohren eine Filterwirkung haben. Dennoch ist, wie Experimente zeigen, eine rein statische Lokalisation weitaus schwieriger als eine dynamische. Denn die unwillkürliche Reaktion des Menschen bei Unsicherheit in der Lokalisation ist ein leichtes Drehen des Kopfes, um durch die Änderungen der IID, ITD und der Filterwirkung eindeutige Rückschlüsse auf die Position der Schallquelle machen zu können (siehe Abb. 1.11). Daher sollten in eine möglichst natürliche Nachbildung räumlichen Hörens Kopfbewegungen miteinbezogen werden.

Einen der wichtigsten Effekte bei bewegten Schallquellen stellt der sogenannte Dopplershift dar. Darunter versteht man die Änderung der Tonhöhe, wenn sich eine Schallquelle an dem/r ruhenden Hörenden vorbeibewegt.

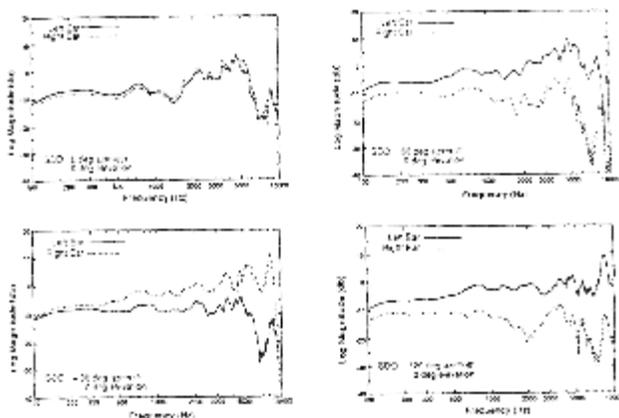
Weiters wurde in Experimenten der minimal hörbare Bewegungswinkel (*minimum audible movement angle*, MAMA) bestimmt. Er beträgt im optimalen Fall (schmalbandige Schallquelle, Geschwindigkeit: 2,8° bis 360°/sec.) etwa 3° und steigt in Abhängigkeit von der Bewegungsgeschwindigkeit, dem Ort der Bewegung und der Art der Schallquelle.

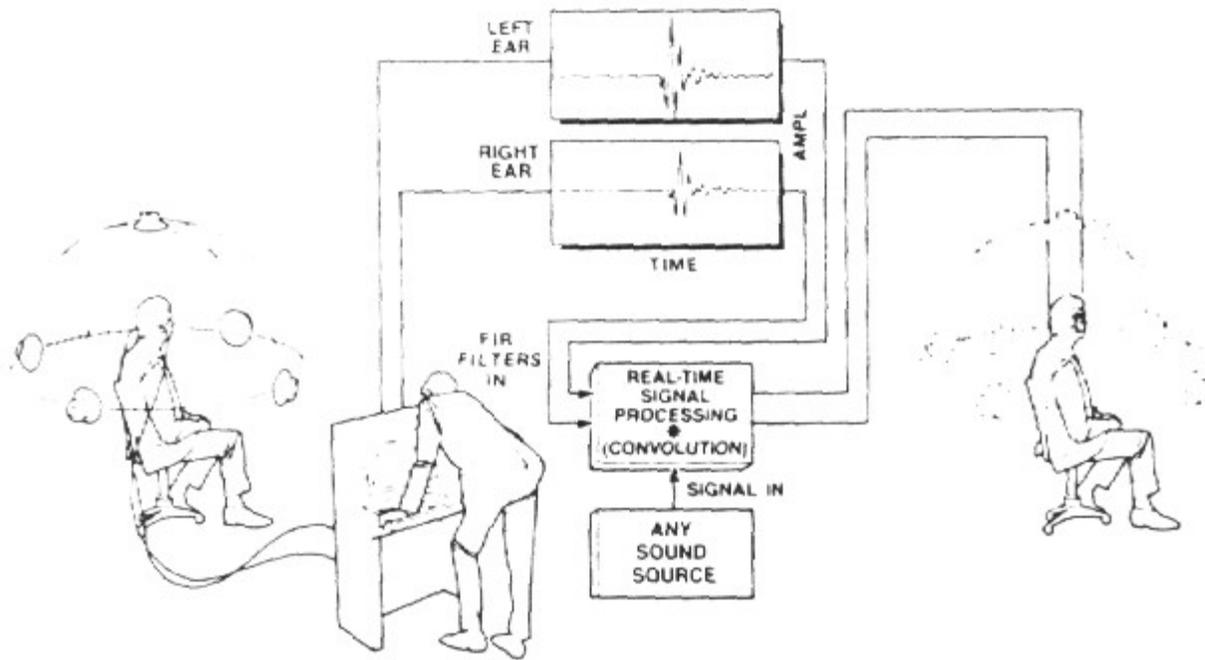
3 Spektrale Einflüsse der Pinnae

Die spektrale Filterung einer Schallquelle, die vor Erreichen des Trommelfells hauptsächlich durch das Außenohr entsteht, nennt man Außenohrübertragungsfunktion (*head-related transfer function*, HRTF), im Zeitbereich entspricht dieser die Außenohrimpulsantwort (*head-related impulse response*, HRIR). Sie sind ein weiteres Mittel zur Auflösung von Mehrdeutigkeiten in der Lokalisation. Die HRTF beschreibt den Schalldruck im Freifeld im Verhältnis zum Druck am Trommelfell und variiert mit Frequenz, Azimut, Elevation und Distanz der Schallquelle (siehe Abb. 1.12). Man kann sich die binaurale HRTF als frequenzabhängige Amplituden- und Zeitverzögerungsdifferenzen vorstellen, die vor allem durch die komplexe Form der Pinnae entstehen. Auch sind beim Menschen linkes und rechtes Ohr nicht exakt gleich, was zu kleinen Unterschieden zwischen linker und rechter HRTF führt. Insgesamt führen also die asymmetrischen, komplex geformten Pinnae zu einem eindeutigen Set von kleinsten Zeitverzögerungen, Resonanz- und Beugungserscheinungen, das in eine eindeutige HRTF für jede Schallquellenposition übertragen wird.

Nachdem HRTFs stark von der Form der Außenohren abhängen, weisen sie große intersubjektive Unterschiede auf. Dies stellt dann ein Problem dar, wenn HRTFs in einem 3D-Audio-System (siehe Abb. 1.13) nachgebildet werden sollen und eine Entscheidung für einen bestimmten Kurvensatz gefällt werden muß.

Weitere Erläuterungen zu den HRTF-Kurven siehe Kapitel 2.





Pinnae (outer ear) responses measured with probe microphones

Pinnae transforms digitized as finite impulse response (FIR) filters

Synthesized cues

3 Auralisation, Schallquellendistanz und Umgebungskontext

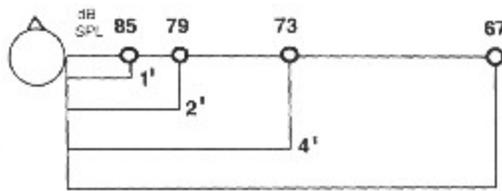
Im Gegensatz zur Lokalisation als "Ortsbestimmung" einer Schallquelle schließt Auralisation die räumlichen Aspekte der Halligkeit mit ein. Distanz und Umgebungskontext sind unter anderem abhängig von der Lautheit, der spektralen Zusammensetzung, der Zusammensetzung der Halligkeit und der kognitiven Bekanntheit der Quelle. Psychoakustische Untersuchungen variieren jedoch meist nur einen oder wenige Parameter. Daher kann es bei Synthese mehrerer Parameter zu unerwarteten Ergebnissen kommen.

1 Distanz

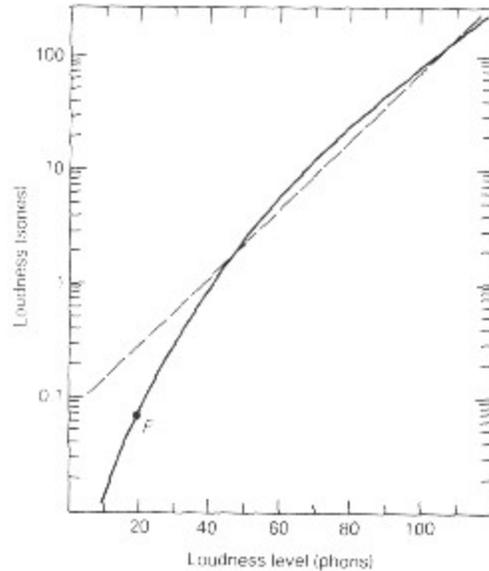
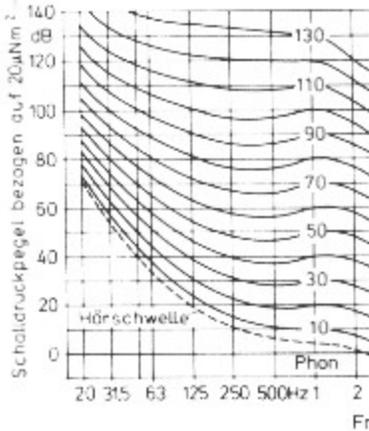
Der wichtigste Parameter zur Bestimmung der Distanz ist die Schallintensität, die als Lautheit interpretiert wird. Der Zusammenhang zwischen Intensität und Distanz entsteht durch die alltägliche Erfahrung in Verbindung mit dem Sehen. Daher kann die Distanz einer bekannten Schallquelle in bekannter Umgebung leicht geschätzt werden, bei unbekanntem Quellen und/oder Umgebungen fällt dies deutlich schwerer.

Bei Freifeldausbreitung einer punktförmigen, omnidirektionalen Schallquelle nimmt die Schallintensität mit $1/r^2$ ab, wobei r die Distanz zur Schallquelle darstellt. Dies entspricht einer Abnahme von 6 dB bei Verdoppelung des Abstands (siehe Abb. 1.14). Bei einer Linienquelle verringert sich die Intensität nur mit $1/r$, entsprechend einer Abnahme von 3 dB pro Verdoppelung der Distanz.

Einen Parameter, der das menschliche Gehör besser berücksichtigt, stellt die Lautheit dar. Diese ist die *empfundene* Amplitude der Intensität. Sie ist frequenzabhängig entsprechend der Empfindlichkeit des Gehörs. Diese Frequenzabhängigkeit wird durch die Isophone (auch Kurven gleicher Lautstärke genannt) dargestellt (siehe Abb. 1.15).



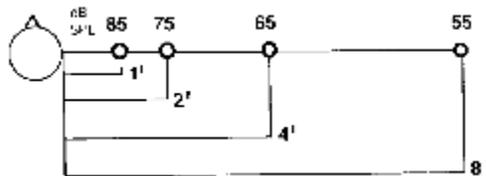
Die Einheit der Lautheit ist Sone, wobei 1 Sone definiert ist als die Empfindungsstärke, die ein 1 kHz-Ton mit 40 Phon Lautstärke, d. h. 40 dB Schalldruckpegel auslöst. Bei Verdoppelung der Distanz muß sich die Lautheit halbieren, da der Ton dann als halb so laut empfunden wird. Eine Halbierung der Lautheit entspricht im Bereich von 400



bis 5000 Hz und 40 bis 100 dB etwa einer Abnahme um 10 dB Schalldruckpegel (siehe Abb. 1.16). Da bei Verdopplung der Distanz zwischen Quelle und Hörendem/r ein Ton als halb so laut empfunden wird, ergibt sich der in Abbildung 1.17 dargestellte Zusammenhang zwischen Schalldruckpegel und Distanzwahrnehmung.

Diese Parameter wurden im schalltoten Raum mit Sinustönen ohne Einfluß anderer Größen bestimmt. Wären Lautheit oder Intensität allein ausschlaggebend, würde die Realisierung obiger Abschätzungen ausreichen. Bei komplexen Signalen kommen aber auch noch andere Effekte zum Tragen. Es wird der Energieanteil innerhalb der kritischen Bandbreite bewertet, bei mehreren Quellen werden die Intensitäten zueinander in Verhältnis gesetzt, ebenso werden Entfernungen bei gleicher Schallintensität in halligen Räumen viel kleiner geschätzt als im schalltoten Raum.

Es zeigt sich weiters, daß in halliger Umgebung das Verhältnis von reflektiertem zu direktem Schall (R/D -ratio) einen viel stärkeren Distanzbestimmung darstellt

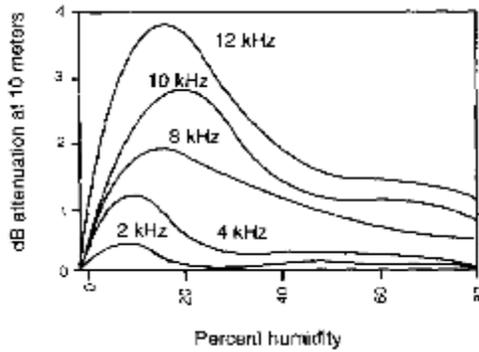


Parameter zur als die Intensität (siehe "Raumeindruck").

Nach der Hypothese von Peter Craven [Gerzon, 1992] wird die scheinbare Distanz d einer Schallquelle durch die relative Zeitverzögerung T und die relative Amplitude g des frühen reflektierten Schalls im Verhältnis zum Direktschall bestimmt. r ist die Absorption pro Zeitverzögerungseinheit. Wendet man diese Parameter auf eine größere Anzahl an frühen Reflexionen an, so kann eine Schätzung der Distanz abgeleitet werden:

Ein schwächerer Effekt bei Distanzabnahme ist die Abschwächung höherer Frequenzen des Quellspektrums durch die Luft, abhängig von Lufttemperatur und -feuchtigkeit. Man kann für große Entfernungen und hohe Frequenzen einen Luftabsorptionskoeffizienten angeben (siehe Abb. 1.18).

Ebenso führt die Art der einer leichten spektralen Person sehr nahe, so werden angehoben (*tone darkening*).



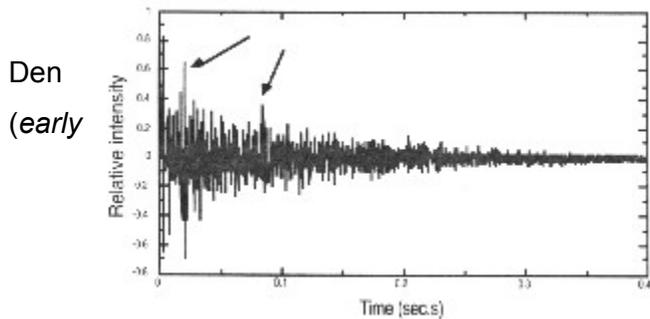
Umgebung (Wind, Boden) zu Färbung. Ist die Quelle der tiefe Frequenzen leicht

Die sogenannte Im-Kopf-Kopfhörer-Entfernungsfehler, Kopfhörer unter wenig natürlichen Bedingungen auftritt (z. B. bei Experimenten zur Lateralisation). Dabei wird die Schallquelle innerhalb des Kopfes geortet. IKL verringert sich unter zunehmend "natürlichen" Bedingungen (binaurale HRTFs, Kopfbewegungen, Raumeindruck).

Lokalisation (IKL) ist der der bei Wiedergabe über

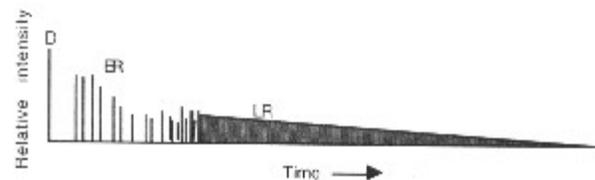
2 Raumeindruck

Wie schon erwähnt, erreicht der Direktschall die Ohren auf direktem Weg, der reflektierte Schall oder Nachhall jedoch erst nach Reflexion an Objekten bzw. den Raumbegrenzungen. Die Eigenschaften eines Raumes können im Zeitbereich durch die Impulsantwort (Abb. 1.19) oder das Reflektogramm (Abb. 1.20) dargestellt werden.



Den (early

Bereich bis 80 ms nennt man frühe Reflexionen



reflections, ER). Es sind dies die ersten Reflexionen z. B. von Boden und Wänden. Sie geben einen Eindruck von der Position im Raum und seinen Abmessungen. Späte Reflexionen (*late reflections*, LR) oder später Nachhall sind diffuser, enthalten weniger Energie und stammen von mehrfachen Reflexionen im Raum.

Das Verhältnis von reflektiertem zu direktem Schall (*reverberant-to-direct sound ratio*, *R/D-ratio*) ist entfernungsabhängig und ein wichtiger Parameter für die Distanzwahrnehmung. Der Hallradius (*critical distance*) ist jener Abstand von der Schallquelle in einem bestimmten Raum, wo dieses Verhältnis eins ist. Die Nachhallzeit (*reverberation time*, t_{60}) ist jene Zeit, in der die Schallenergie des Raumes um 60 dB des ursprünglichen Werts abgenommen hat. Aussagekräftiger sind der zeitliche

Verlauf des Hallpegels und der zeitliche Verlauf des Hallpegels innerhalb der kritischen Bandbreiten
(nach Zwicker).

3 Modellierung von HRTF-Kurven

Ziel der Modellierung ist eine möglichst recheneffiziente Nachbildung der HRTFs, um bei bestehender Rechenleistung möglichst viele Quellen spatialisieren zu können. Ein zusätzliches Problem stellt die Interpolation der HRTFs bei bewegten Quellen dar, da die perzeptiven Merkmale und nicht die einzelnen Abtastwerte interpoliert werden sollen.

Anwendung findet diese Technik z. B. in der virtuellen Akustik, wo Monoaufnahmen "verräumlicht" werden sollen. Eine andere Anwendung stellt der Versuch dar, aus mehrkanaligen Aufnahmen realer Schallsituationen ein Stereosignal für binaurale Wiedergabe zu berechnen.

1 Hauptkomponentenmethoden

1 Allgemeines

Im folgenden werden die verschiedenen Hauptkomponentenmethoden aus der Literatur behandelt. Es handelt sich dabei um dieselbe zugrundeliegende Idee, die verschiedenen Autoren verwenden dabei aber verschiedene Ausgangsmaterialien und Bezeichnungen.

2 Prinzip

Das Ausgangsmaterial stellt ein Satz von Außenohrkurven im Zeit- (HRIRs) oder Frequenzbereich (HRTFs) dar. Dieser Satz enthält Messungen aus n Richtungen zu je p Meßwerten. Daraus wird ein Set von Basisvektoren extrahiert, deren gewichtete Linearkombination die modellierten HRIRs bzw. HRTFs ergeben. Die richtungsabhängigen Gewichte repräsentieren den relativen Anteil jedes Basisvektors an einer Impulsantwort bzw. in einem Spektrum. Diese Art der Zerlegung bedeutet, daß Richtungskomponenten (in Form der Gewichte) und Frequenzkomponenten (in Form der Basisvektoren) getrennt werden.

Zuerst wird aus dem Satz von Außenohrkurven der Mittelwert gebildet:

\mathbf{h}_k stellt die k -te HRIR bzw. HRTF dar. Die Bedeutung des Mittelwerts hängt von den verwendeten Daten ab. Im Fall von HRIRs oder "echten" HRTFs (reell oder komplex in linearem Maßstab) enthält der gemittelte Vektor die subjektabhängigen und richtungsunabhängigen spektralen Eigenschaften,

die alle HRIRs bzw. HRTFs eines individuellen Ohrs gemeinsam haben (z. B. die Ohrkanalresonanz um 2,5 kHz), sowie Meßartefakte (z. B. spektrale Einbrüche durch stehende Wellen). Falls HRTFs in Form von logarithmierten Amplitudengängen verwendet werden, stellt das geometrische Mittel aus den linearen Amplitudengängen dar.

Die gemittelte Funktion wird von den einzelnen HRIRs bzw. HRTFs subtrahiert. Durch Entfernung des Mittelwerts enthalten die Ergebnisfunktionen \mathbf{d}_k primär richtungsabhängige spektrale Effekte. Im Frequenzbereich nennt man die \mathbf{d}_k Richtungsübertragungsfunktionen (*directional transfer functions*, DTFs), um sie von den HRTFs zu unterscheiden, im Zeitbereich seien sie Richtungsimpulsantworten (*directional impulse responses*, DIRs) genannt.

Die Wahl des Datenformats (HRIR, lineare HRTF oder logarithmierter Amplitudengang) wirkt sich auch auf die Filterstruktur für die Spatialisation aus. Um das Ausgangssignal \mathbf{o} für die Richtung (θ_k, φ_k) zu erhalten, wird das Audio-Eingangssignal \mathbf{s} mit \mathbf{h}_k gefiltert. Es gilt:

Im Fall von HRIRs oder linearen HRTFs ergibt das Hinzufügen des Mittelwerts die Überlagerung der Ausgangssignale von zwei Filtern, bei logarithmierten HRTF-Amplitudengängen entspricht das aber einer Filterkaskade.

Für die DIRs bzw. DTFs wird die Kovarianzmatrix berechnet. Die Kovarianzen liefern ein Maß für die Ähnlichkeit zwischen zwei DIRs bzw. DTFs. Sei \mathbf{D} eine $p \times n$ -Matrix, deren Spalten jeweils die DIR bzw. DTF einer bestimmten Richtung enthalten, dann berechnet sich die $p \times p$ -Kovarianzmatrix \mathbf{S} folgendermaßen:

wobei n die Anzahl der gemessenen Richtungen ist.

Seien \mathbf{c}_i die Eigenvektoren von \mathbf{S} und λ_i die korrespondierenden Eigenwerte. Nachdem jede Autokorrelationsmatrix symmetrisch und nichtnegativ definit ist, gibt es p orthogonale Eigenvektoren, und die korrespondierenden Eigenwerte sind reell und nichtnegativ. Ohne Beschränkung der Allgemeinheit wählt man die Indizierung so, daß

Als Basisvektoren nimmt man die q Eigenvektoren von \mathbf{S} , die mit den q größten Eigenwerten korrespondieren. Sie stellen die Spalten der orthonormalen Transformationsmatrix \mathbf{C} dar:

Ist \mathbf{d}_k der k -te DIR- bzw. DTF-Vektor mit Dimension $p \times 1$, so ergeben sich die zugehörigen Gewichte \mathbf{w}_k folgendermaßen:

wobei \mathbf{C} eine $p \times q$ -Matrix ist und \mathbf{w}_k der Gewichtsvektor mit Dimension $q \times 1$. Letzterer repräsentiert den Anteil der Basisvektoren an einer bestimmten \mathbf{d}_k .

\mathbf{C}^T transformiert also \mathbf{d}_k in einen Vektor \mathbf{w}_k , dessen Komponenten paarweise unkorreliert sind. Diese spezielle Art der Orthogonaltransformation nennt man zeitdiskrete Karhunen-Loève Transformation oder Hotelling Transformation (KLT).

Wenn $q=p$ ist, dann erfüllt die reell- bzw. komplexwertige Matrix $\mathbf{T}=\mathbf{C}^T$ die Orthogonalitätsbedingung

Eigenschaften einer Orthogonaltransformation sind:

- Die Distanz zwischen zwei beliebigen Punkten bleibt erhalten.
- Die Länge eines beliebigen Vektors bleibt erhalten.
- Die Determinante der Autokorrelationsmatrix des Eingangsvektors wird nicht verändert.
- Die Eliminierung der Korrelation zwischen den Transformationskoeffizienten reduziert das Produkt der Varianzen der Vektorkomponenten.

Die Autokorrelationsmatrix von \mathbf{w}_k ist gegeben durch

Man sieht, daß der Eingangsvektor tatsächlich dekorreliert wird.

Nach Umformung ergibt sich \mathbf{d}_k als gewichtete Summe der Basisvektoren:

Das Gleichheitszeichen gilt nur für $q=p$. In der Praxis ist $q \ll p$, sodaß die rechte Seite obiger Gleichung nur eine Näherung für \mathbf{d}_k darstellt.

Je größer die Anzahl der Basisvektoren, desto besser die Näherung.

Um das Ausgangssignal \mathbf{o} für die Richtung (θ_k, φ_k) zu erhalten, wird das Audiodaten-Eingangssignal \mathbf{s} mit \mathbf{d}_k gefiltert. Für HRIRs und lineare HRTFs gilt:

Für logarithmierte HRTFs gilt mit

die modifizierte Gleichung:

Nachdem c_i nur frequenz- und nicht richtungsabhängig ist, kann bei bekannten Audiodaten die Filterung mit c_i im voraus berechnet werden.

Die KLT minimiert den mittleren quadratischen Fehler für eine gegebene Zahl an Basisvektoren. Fünf Basisvektoren repräsentieren rund 90% der Varianz der menschlichen HRTF-Amplitude.

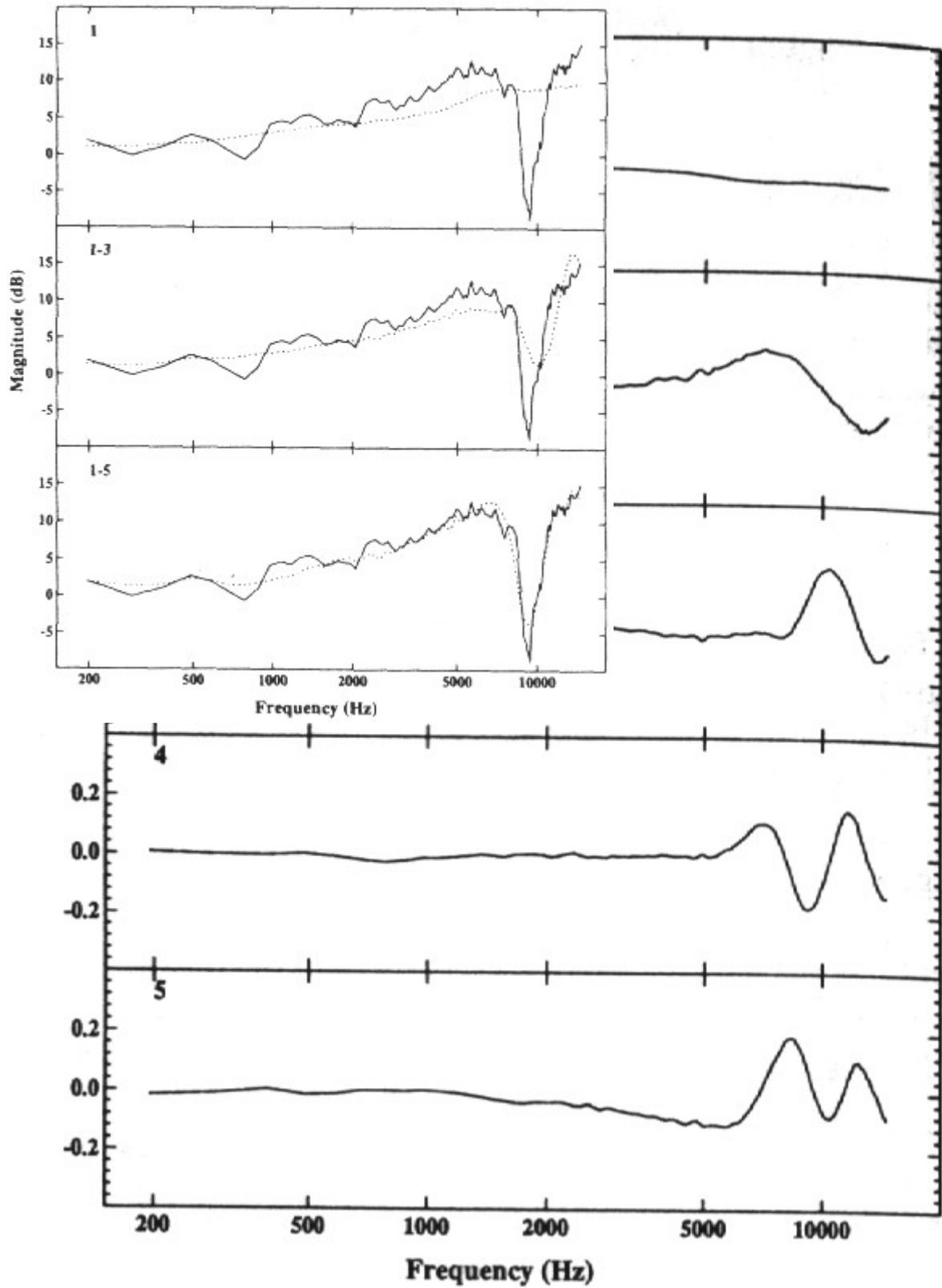
3 Hauptkomponentenanalyse und minimalphasige Rekonstruktion

Bei der Hauptkomponentenanalyse (*principal components analysis*, PCA) nach [Kistler, Wightman, 1995] und [Middlebrooks, Green, 1992] wird als Ausgangsmaterial der logarithmierte HRTF-Amplitudengang genommen. Die HRTF-Phase wird unter Annahme von Minimalphasigkeit und konstanter interauraler Phasen-(Zeit-)Differenz rekonstruiert. Letzteres bedeutet, daß nur eine einfache (konstante) Zeitverzögerung zur Berücksichtigung der ITD der jeweiligen Quellposition nötig ist.

Für die Berechnung der Richtungsübertragungsfunktionen (DTFs) werden alle (hier: 256) logarithmierten Amplitudengänge der HRTFs eines Ohres einer Person gemittelt.

Abbildung 3.1 zeigt die ersten fünf Basisvektoren, Abbildung 3.2 zeigt für eine Richtung die gemessene HRTF (durchgezogene Linien) sowie die Ergebnisse der Rekonstruktion mit 1, 3 und 5 Basisvektoren (punktierte Linien).

Bezüglich der Körpergröße der Testpersonen konnte beobachtet werden, daß bei kleinen Personen die ersten 5 Basisfunktionen im Vergleich zu denen von großen Personen auf der Frequenzachse nach oben verschoben werden, wobei die Einhüllenden ähnlich sind.



4 Extraktion der spatialen Eigenschaften und Regularisierung

Dieser Abschnitt folgt [Chen et al., 1995], [Marolt, 1996] sowie [Gersho, Gray, 1992]. Die Basisfunktionen \mathbf{c}_i werden hier Eigenübertragungsfunktionen (*eigentransfer functions*, EFs), die Gewichte $\mathbf{w}_i(\theta_k, \varphi_k)$ "spatial-charakteristische" Funktionen (*spatial characteristic functions*, SCFs) genannt. Das Ausgangsmaterial bilden die komplexwertigen HRTFs (Amplitude und Phase).

Das Modell für die Nachbildung der HRTFs im Frequenzbereich, $\mathbf{h}_m(\theta_k, \varphi_k)$, ist ein $p \times 1$ Vektor und lautet:

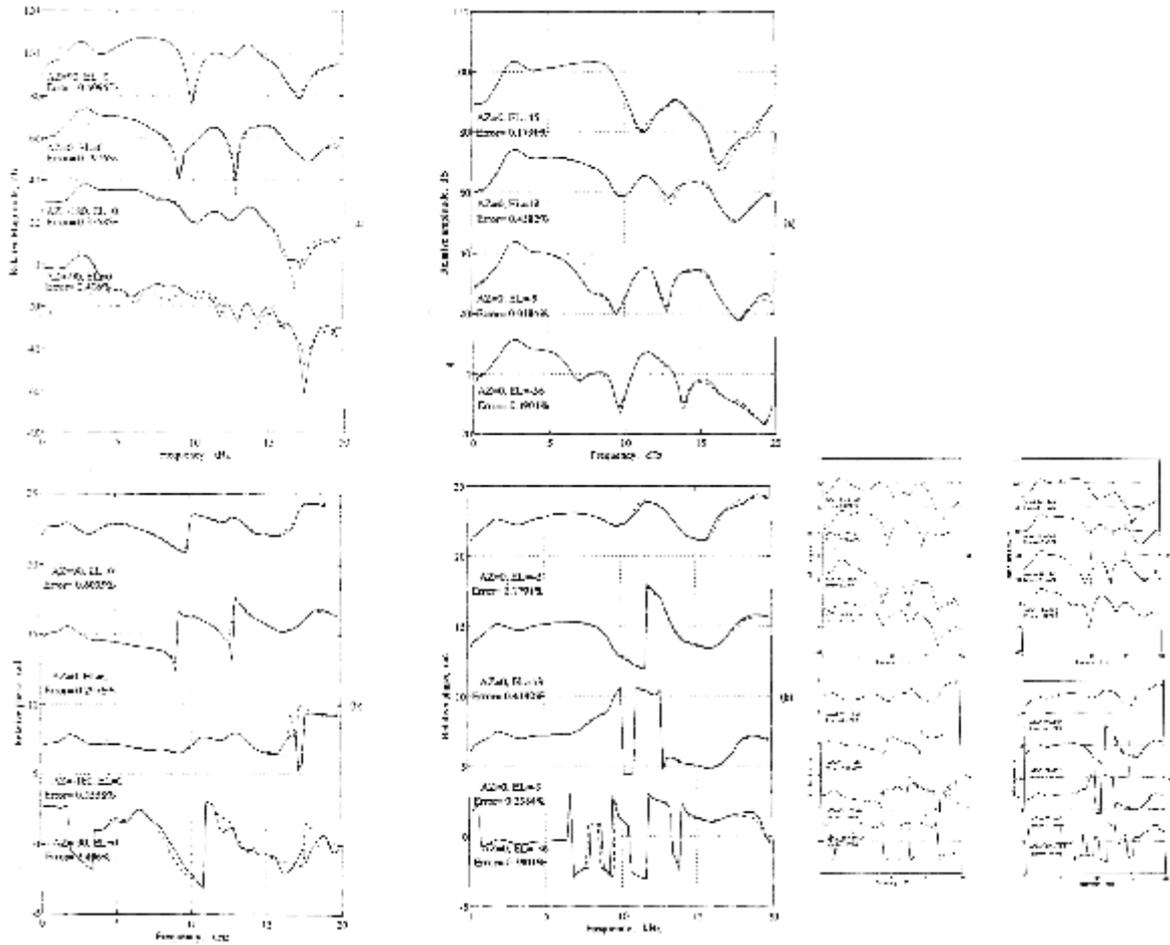
wobei (θ_k, φ_k) die Richtung der virtuellen Quelle, $\mathbf{h}_m(\theta_k, \varphi_k)$ die modellierte HRTF, \mathbf{c}_i ($i = 0, \dots, q$) die Eigenübertragungsfunktion und $\mathbf{w}_i(\theta_k, \varphi_k)$ die spatial-charakteristische Funktion ist.

Der mittlere quadratische Fehler e_k einer bestimmten Richtung (in Abb. 3.3 und 3.4 als "Error" angegeben) ist definiert als

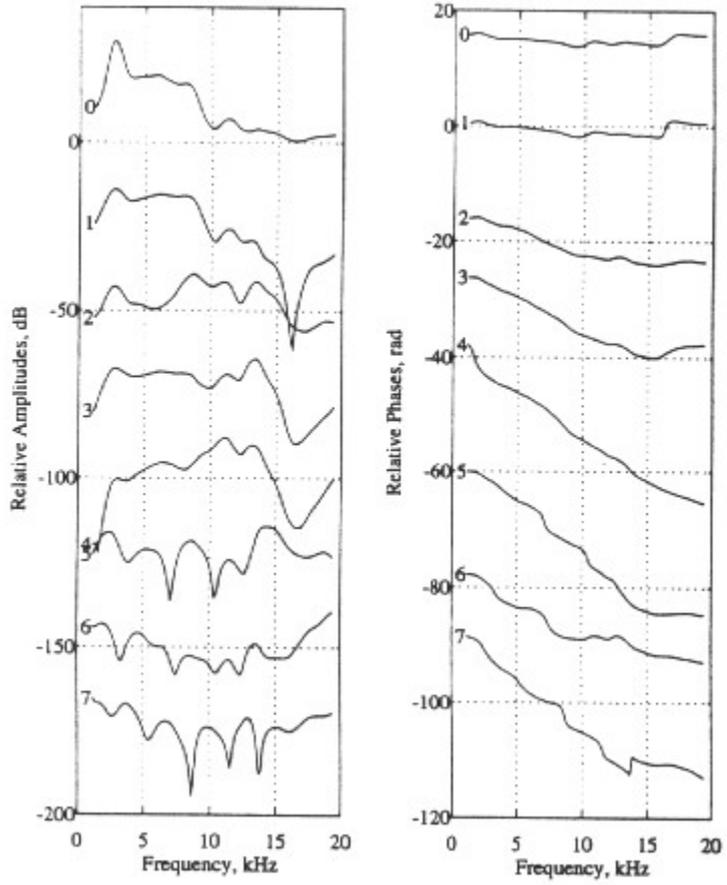
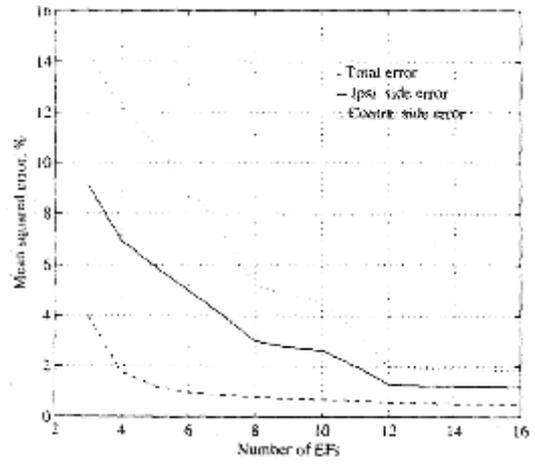
wobei \mathbf{h}_k die k -te gemessene HRTF und $\mathbf{h}_m(\theta_k, \varphi_k)$ durch Gleichung (3.13) gegeben ist. Der gesamte mittlere quadratische Fehler e (Abb. 3.5) sinkt mit der Anzahl q der EFs und ist definiert als

Abbildung 3.6 zeigt die gemittelte HRTF sowie die ersten sieben EFs für KEMAR.

In Abbildung 3.3 sind nur Richtungen abgebildet, die zur Bestimmung der Modellparameter verwendet worden sind. Man sieht, daß die Fehler der interpoliert modellierten HRTFs aus Abbildung 3.4 etwas größer sind als die aus Abbildung 3.3.



Modellierung von HRTF-Kurven

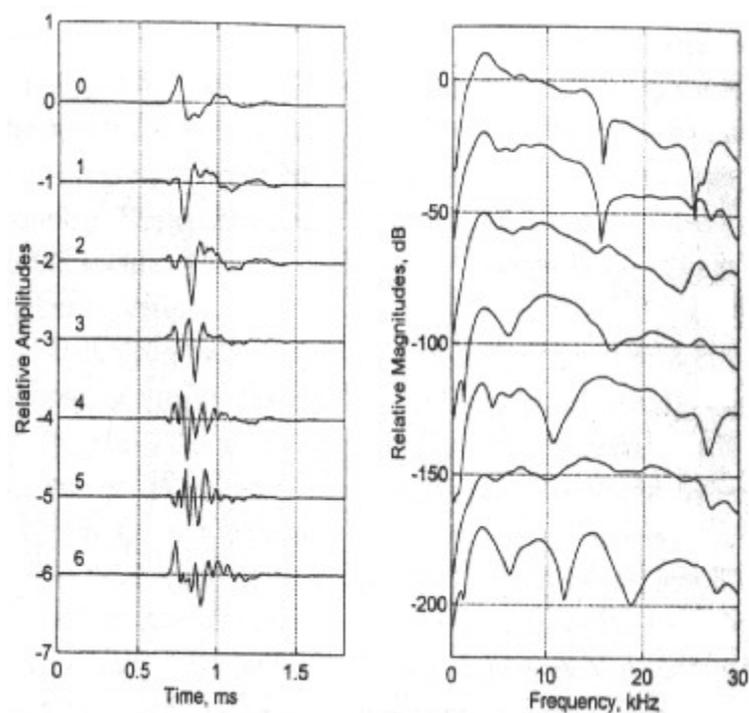


5 Extraktion der spatialen Eigenschaften im Zeitbereich

Bei [Wu et al., 1997] stellen normalisierte Außenohrimpulsantworten (*head-related impulse responses*, HRIRs) das Ausgangsmaterial dar.

Die Gewichte $w_i(\theta, \varphi)$ werden als reelle spatial-charakteristische Funktionen (*real spatial characteristic functions*, RSCFs) bezeichnet und sind reellwertige Funktionen des räumlichen Ortes. Die normalisierten HRIRs werden im kontinuierlichen Raum (θ, φ) als repräsentiert.

Abbildung 3.7 zeigt die gemittelte HRIR und die ersten sechs Basisfunktionen für die HRIRs einer Katze.

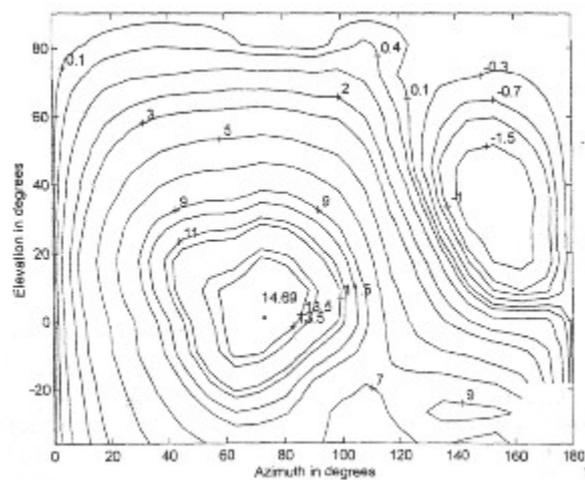


Modellierung von HRTF-Kurven

Schätzungen für beliebige Richtungen erhält man, indem man die RSCFs, ITDs und IIDs der gemessenen Richtungen interpoliert. Verwendet man den Interpolationsalgorithmus für den linearen Raum für eine Richtung (θ, φ) mit Elevation $-36^\circ \dots +90^\circ$, so erhält man als Schätzwerte für die RSCFs

wobei hier $A=9 \times 9$ eine Fläche auf einem 9° -Netz ist und (θ_1, φ_1) , (θ_1, φ_2) , (θ_2, φ_1) und (θ_2, φ_2) die vier benachbarten Meßpunkte von (θ, φ) sind.

Die Abbildungen 3.8 und 3.9 zeigen Höhenliniendarstellungen der ITDs und IIDs.



2 FIR- und IIR-Filterdesign

Dieser Abschnitt folgt [Jot et al., 1995] und [Brandenstein, Unbehauen, 1998].

Folgende Eigenschaften werden beim Design von Synthesefiltern genutzt:

- Lineare Filter können in ein minimalphasiges Filter und ein Allpaßfilter zerlegt werden (*{magnitude, excess phase}* Darstellung). Das Allpaßfilter realisiert eine “excess phase”, die man erhält, wenn man von der Phasenantwort den minimalphasigen Anteil subtrahiert. Den minimalphasigen Phasengang erhält man durch Anwendung der Hilbert-Transformation auf den logarithmierten Amplitudengang.
- Bei HRTFs kann das Allpaßfilter als eine einfache Zeitverzögerung angenähert werden, dies gilt zumindest für den Bereich bis zu 10 kHz. In diesem Bereich ist die “excess phase” eine lineare Funktion der Frequenz. Diese Zeitverzögerung schätzt man durch eine Gerade ab, die durch die “excess phase”-Antwort im Bereich von 1 kHz bis 5 kHz gelegt wird.

1 FIR-Filterdesign

Experimente von D. J. Kistler und F. L. Wightman sowie D. Hammershøi und J. Sandvad konnten zeigen, daß eine minimalphasige Approximation von HRTFs die Lokalisationsgenauigkeit nicht beeinflußt.

Vorgangsweise:

- Repräsentation der HRTFs als *{magnitude, excess phase}* Darstellung.
- Näherung der excess phase als einfache Zeitverzögerung.
- Diffusfeld-Entzerrung des Amplitudengangs.
- Minimalphasige Rekonstruktion der FIR-Filterantwort unter Verwendung von Hilbert-Transformation und inverser Fourier-Transformation.
- HRTF-Approximation durch Kaskadierung der Zeitverzögerung und des minimalphasigen Filters.

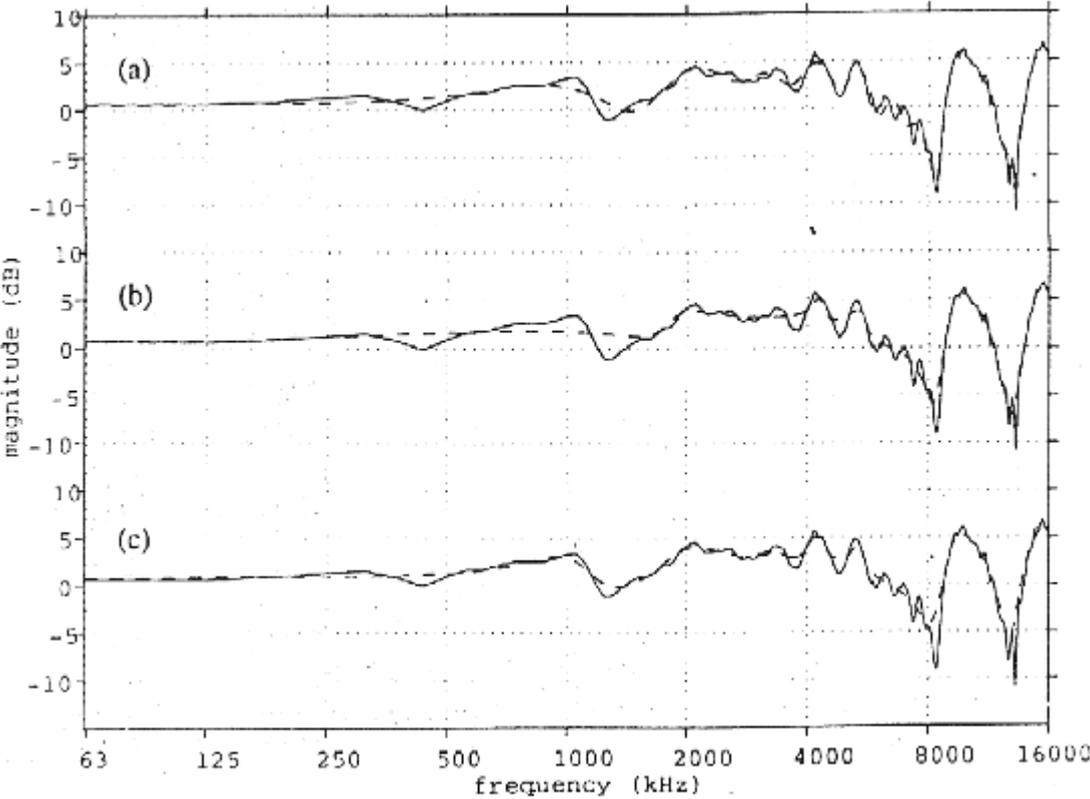
2 IIR-Filterdesign

Vorgangsweise:

- Repräsentation der HRTFs als {magnitude, excess phase} Darstellung.
- Näherung der excess phase als einfache Zeitverzögerung.
- Diffusfeld-Entzerrung des Amplitudengangs.
- "Constant Q smoothing": Glätten des Amplitudenspektrums unter Verwendung eines Hann-Fensters mit frequenzproportionaler Breite.
- "Warping": Resampling des Amplitudenspektrums auf einer verzerrten ("*warped*") Frequenzskala entsprechend der Bilineartransformation $z \rightarrow (z+r)/(1+rz)$ mit $r = -1 \dots +1$.
- Näherung des *verzerrten* Amplitudenspektrums mittels Yule-Walker-Algorithmus ergibt die Übertragungsfunktion .
- "Dewarping": Anwendung der Bilineartransformation $z \rightarrow (z-r)/(1-rz)$ auf die Übertragungsfunktion $H_{m, \text{warp}}(z)$ ergibt $H_m(z)$.

Die Frequenzverzerrung hat den Effekt des Oversampling bei niederen und Undersampling bei hohen Frequenzen, daher muß das Spektrum vorher geglättet werden. r kann so gewählt werden, daß die neue Frequenzachse in etwa der Barkskala entspricht. Der Verlust an Genauigkeit bei hohen Frequenzen wird für eine besserer Näherung bei tieferen Frequenzen in Kauf genommen.

Abbildung 3.10 zeigt die Ergebnisse der HRTF-Modellierung nach [Jot et al., 1995].



Modellierung von HRTF-Kurven

Eine alternative Designmethode wird in [Brandenstein, Unbehauen, 1998] angegeben. Dabei wird ein FIR-Filter durch ein IIR-Filter angenähert:

Gegeben sei die Übertragungsfunktion $F(z)$ eines FIR-Filters mit

Diese soll durch ein IIR-Filter mit der Übertragungsfunktion $H(z)$ angenähert werden, wobei $N < L$ sowie

Das bedeutet, daß man $N+1$ reelle Zählerkoeffizienten p_v und N reelle Nennerkoeffizienten q_v bestimmen muß, sodaß die L_2 -Norm der Differenzfunktion $\Delta(z)$

minimiert wird. Die L_2 -Norm von $\Delta(z)$ ist definiert als

wobei die Integration gegen den Uhrzeigersinn durchgeführt werden muß und "*" für konjugiert-komplex steht. Nachdem das IIR-Filter stabil sein soll, müssen alle Pole von $H(z)$ innerhalb des Einheitskreises ($|z| < 1$) liegen und die Funktion $H(z)$ außerhalb des Einheitskreises ($|z| > 1$) analytisch sein.

Theorem von Walsh

Dieses Theorem wurde ursprünglich für die Näherung von Funktionen formuliert, die analytisch in $|z| < 1$ sind, es kann aber leicht für $|z| > 1$ übertragen werden.

In der Menge der rationalen Funktionen $H(z)$ [nach Gleichungen (3.19) und (3.20)] mit den Polen $\alpha_1, \dots, \alpha_N$ innerhalb des Einheitskreises ist die beste Approximation für $F(z)$ (analytisch in $|z| > 1$ und kontinuierlich in $|z| \geq 1$) im Sinne der kleinsten quadratischen Fehler jene eindeutige Funktion, die $F(z)$ in den Punkten $z = \infty, 1/\alpha_1^*, \dots, 1/\alpha_N^*$ interpoliert.

Dieses Theorem kann man verwenden, um die Bestimmung der Zählerkoeffizienten p_v und der Pole α_v zu entkoppeln. Hat man die Pole bestimmt, ergeben sich die Zählerkoeffizienten als Lösung eines Interpolationsproblems. Es können immer $N+1$ lineare Gleichungen in $N+1$ Unbekannten p_v für die Interpolation angegeben werden, was zu einer eindeutigen Lösung führt. Sind alle α_v verschieden, so ist dies offensichtlich. Definiert man so lautet die Bedingung einfach

Nachdem alle Pole paarweise verschieden sind, erhält man $N+1$ linear unabhängige Gleichungen in $N+1$ Unbekannten p_v .

Bei mehrfachen Polen oder mindestens einem Pol an der Stelle $z=0$ wird für die Aufstellung der Bestimmungsgleichungen die der Multiplizität des jeweiligen Pols entsprechende Anzahl von Ableitungen der Gleichung (3.24) nach z^{-1} an der Stelle des Pols verwendet.

3 Zeitvariante Implementation

Hier kann man zwei verschiedene Prozesse der zeitvarianten Implementation unterscheiden:

Interpolation:

Unter Interpolation versteht man die Synthetisierung einer Übertragungsfunktion unter Verwendung einer Datenbank aus vordefinierten Filtern. In unserem Fall besteht die Datenbank aus Richtungsfilterpaaren, die durch Näherung der gemessenen HRTF-Kurven gewonnen worden sind.

Kommutation:

Kommutation bedeutet Update der Filterkoeffizienten, während das Filter *"läuft"*. Der Prozessor realisiert das Update sofort zwischen der Berechnung zweier aufeinanderfolgender Output-Samples. Das Update von Koeffizienten in einem digitalen Filter kann hörbare Klicks verursachen.

In einem synchronen Operationsmodus wird typischerweise in regelmäßigen Zeitintervallen die neue Zielposition angegeben. Das Update der Filterkoeffizienten kann durch folgende zyklische Prozedur vorgenommen werden:

1. Lesen eines neuen Filterpaares entsprechend der neuen Richtung aus der Datenbank bzw. Berechnen eines neuen Filterpaares durch Interpolation.
2. Initialisierung des Kommutationsprozesses: Berechnen der elementaren Inkremente für alle Koeffizienten, die die Kommutation kontrollieren.
3. Inkrementieren und Updaten der Filterkoeffizientensets in regelmäßigen Zeitabständen, bis die nächste gewünschte Richtung erreicht ist. Danach zurück zu Punkt 1.

3 Modellierung durch Richtstrahlbildung

1 Allgemeines

Die Methode der Richtstrahlbildung (*Beamforming*) wird in [Chen et al., 1992] behandelt.

Am Ausgang eines Richtstrahlers (*Beamformer*) liegt eine gewichtete Kombination der Daten an, die an einem Feld von räumlich verteilten Sensoren empfangen worden sind. Die Gewichte des Richtstrahlers und die Sensor-Feldgeometrie determinieren die räumliche und zeitliche Filtercharakteristik. Im vorliegenden Fall ergibt sich die Sensorgeometrie aus der Außenohrakustik. Die Gewichte werden so gewählt, daß der mittlere quadratische Fehler zwischen der Richtstrahler-Antwort und den gemessenen Außenohrantworten minimiert wird. Die Modellantwort ist das innere Produkt aus dem Gewichtsvektor und einem Feldantwortvektor (*array response vector*). Letzterer repräsentiert explizit die Feldgeometrie und den Ort der Quelle.

Es handelt sich um eine Modellierung der Transformationseigenschaften des Außenohres als Eingangs-Ausgangs-Beziehung.

Die Vorteile dieses Verfahrens sind:

- Es liefert einen expliziten mathematischen Ausdruck, der die Richtungsabhängigkeit der Schallwelle am Trommelfell enthält.
- Es repräsentiert die Außenohrübertragungscharakteristik als eine kontinuierliche Funktion der räumlichen Richtung, auch wenn die Modellparameter durch diskrete Messungen ermittelt wurden.
- Substantielle Datenkompression bei Repräsentation von HRTFs mit vielen Richtungen und Frequenzen.

2 Prinzip

Wie aus
ersichtlich,
Richtstrahler M
sich in jedem
FIR-Filter der
befindet. Die
des
modellierte
durch den

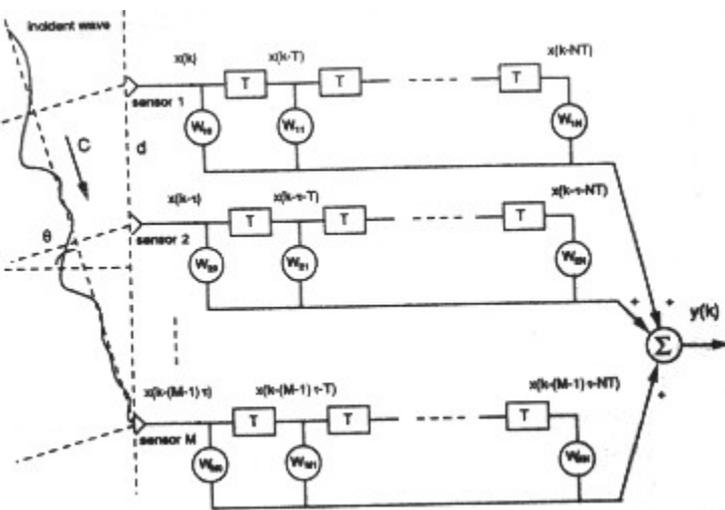


Abbildung 3.11
enthält der
Sensoren, wobei
Sensorkanal ein
Länge N+1
Antwort $h(\omega, \theta)$
Richtstrahlers (= HRTF) wird

Feldantwortvektor $\mathbf{d}(\omega, \theta)$ und den Gewichtsvektor \mathbf{w}^H dargestellt:
Dabei ist ω die Kreisfrequenz und θ der Azimut sowie
und

Dabei ist N+1 die FIR-Filterordnung, T das Abtastintervall (im folgenden sei der Einfachheit wegen $T=1$) und τ_i die Zeitverzögerung zufolge Ausbreitung zwischen erstem und i-tem Sensor. τ_i enthält die Sensorfeldgeometrie sowie die Richtung θ der Quelle. In Gleichung (3.27) bezieht sich das erste Subskript auf den Kanal-Index und das zweite auf den Tap-Index.

3 LMS-Richtstrahler-Design

Ziel ist die Bestimmung von \mathbf{w} für eine gegebene Feldgeometrie, sodaß der Richtstrahler eine gewünschte Antwort annähert. Eine Optimierung der Feldgeometrie wird nicht in Betracht gezogen, da dies ein schwieriges Problem ist aufgrund der komplizierten Art, mit der die Geometrie in $\mathbf{d}(\omega, \theta)$ eingeht.

Wenn $h_d(\omega, \theta)$ die gewünschte Antwort (hier: Außenohrantwort) darstellt, so beträgt der Fehler

Die Minimierung des mittleren quadratischen Fehlers bedeutet

Dabei ist $\gamma(\omega, \theta)$ eine nicht-negative Gewichtsfunktion, um bestimmte Frequenzen und Richtungen hervorzuheben. Nachdem von den n HRTFs aus Richtung θ_i ($i=1, \dots, n$) nur je p diskrete Meßwerte bei den Frequenzen ω_j ($j=1, \dots, p$) vorliegen, reduziert sich die Lösung auf

Der Fehlervektor \mathbf{e}^H ist definiert als:

Sei

wobei

und

Nun schreibt man $\gamma(\omega, \theta_i)$ als diagonale Gewichtsmatrix der Dimension $p \times n$ um:

Gleichung (3.30) wird nun geschrieben als

was äquivalent ist zu einer Lösung von $S=p \times n$ linearen Gleichungen in $L=M \times (N+1)$ Unbekannten

Im allgemeinen ist $S > L$.

Bisher wurde \mathbf{w} komplex angenommen. Die Antwort der zu modellierenden physikalischen Systeme ist aber konjugiert-komplex symmetrisch auf der Frequenzachse, das heißt

Nun sind

Nachdem $\mathbf{d}^*(\omega, \theta) = \mathbf{d}(-\omega, \theta)$, ist

Dies impliziert, daß $\mathbf{w}^T = \mathbf{w}^H$. Daher ist \mathbf{w} reellwertig und es gilt:

wobei die Indizes I und R Imaginär- bzw. Realteil der jeweiligen Größe bezeichnen.

Die Euklidische Norm eines Vektors ist folgendermaßen definiert:

Nachdem

ist leicht zu zeigen, daß Gl. (3.36) äquivalent ist zum reellwertigen Problem

Zur Lösung dieser Gleichung kann man \mathbf{C} mittels Singulärwertzerlegung (*singular value decomposition*, SVD) als Produkt seiner Singulärwerte und -vektoren darstellen:

r ist der Rang von \mathbf{C} , \mathbf{u}_i und \mathbf{v}_i sind die Links- bzw. Rechtssingulärvektoren und σ_i die Singulärwerte. Unter der Annahme von $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq \sigma_{r+1} = \dots = \sigma_L = 0$ ergibt sich die Lösung von \mathbf{w} als

Ist die Matrix \mathbf{C} "rank deficient", so ergibt eine numerische Auswertung der SVD Singulärwerte, die nicht exakt gleich Null sind. Diese sehr kleinen Singulärwerte bewirken eine sehr große Norm von \mathbf{w} . Daher verwendet man meist nur ρ signifikante Singulärwerte, um \mathbf{w} zu lösen:

ρ wird entsprechend der Anzahl der Singulärwerte gewählt, die über einer bestimmte Schwelle δ liegen. Der mittlere quadratische Fehler von Gl. (3.47) ist

Die Norm von \mathbf{w} ist

Mit sinkendem ρ steigt der Fehler und sinkt die Norm von \mathbf{w} .

4 Beschreibung der experimentellen Daten

Das Modell wurde mit der Außenohrübertragungsfunktion einer Katze verifiziert. Diese hat einen größeren Frequenzbereich und eine ähnliche spektrale Einhüllende wie die eines Menschen.

Es gab drei Ansätze zur Phasenrekonstruktion:

- Linearphasen-Methode: Statt der originalen Phase wird eine lineare Phase mit einer Gruppenverzögerung proportional zum zeitlichen Öffnungszentrum (*temporal aperture center*) des Richtstrahlers verwendet.
- Minimalphasen-Methode: Die originale Phase wird durch ihre minimalphasige Komponente ersetzt. Dies garantiert eine eindeutige Darstellung der originalen Außenohrübertragungsfunktion mit minimaler Zeitverzögerung.
- Allphasen-Methode: Zerlegung der Originalphase in eine minimalphasige und eine Allpaß-Komponente. Die Allpaß-Komponente wird geglättet und daraus die lineare Komponente, die eine reine Zeitverzögerung darstellt, entfernt. Die verbleibende nichtlineare Komponente wird zur minimalphasigen Komponente hinzuaddiert.

Die Linearphasen-Methode liefert für eine gegebene Feldstruktur die beste Näherung der gewünschte Amplitudenantwort, sie ignoriert jedoch die Phasenbeziehungen. Daher gibt es erhebliche Unterschiede zwischen der modellierten und der gemessenen Impulsantwort. Die Allphasen-Methode liefert im Vergleich zur gemessenen Impulsantwort die geringste morphologische Zerstörung der modellierten Impulsantwort. Sie ergab aber den größten Fehler zwischen modellierter und gemessener Frequenzantwort. Die Minimalphasen-Methode stellt einen Kompromiß zwischen den beiden Methoden dar.

5 Experimentelle Ergebnisse

Es werden sowohl verschiedene Feldstrukturen bezüglich des mittleren Näherungsfehlers untersucht als auch der Interpolationsfehler und der Fehler im Zeitbereich.

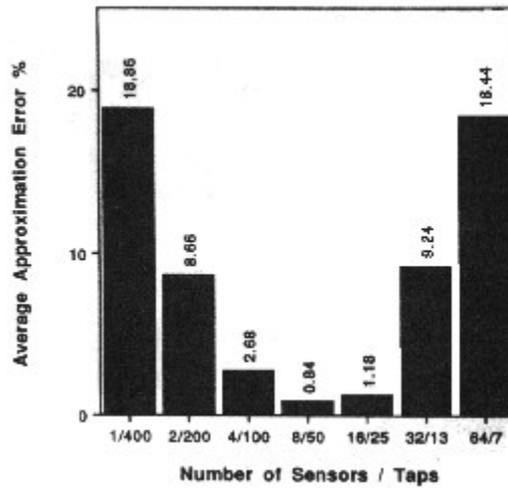
Gute Ergebnisse lieferte eine 5-1-5 L-förmige Sensorgeometrie mit 8 mm Sensorabstand und je 40...50 Filter-Taps.

Allerdings wurden nur verschiedene Außenohrkurven gleicher Elevation untersucht. Will man die binaurale Außenohrantwort nachbilden, so muß man auf eine dreidimensionale Feldgeometrie erweitern sowie die interauralen Zeitdifferenzen berücksichtigen.

Abbildung 3.12 zeigt den mittleren Näherungsfehler als Funktion der Anzahl der Sensoren und der FIR-Filterlänge. Der mittlere Näherungsfehler ist definiert als wobei

Modellierung von HRTF-Kurven

Abbildung 3.13 zeigt den gemessenen und Amplitudenantworten in 4 Richtungen für Sensoren und FIR-



Vergleich von modellierten verschiedenen verschiedene Anzahl von Filterlängen.

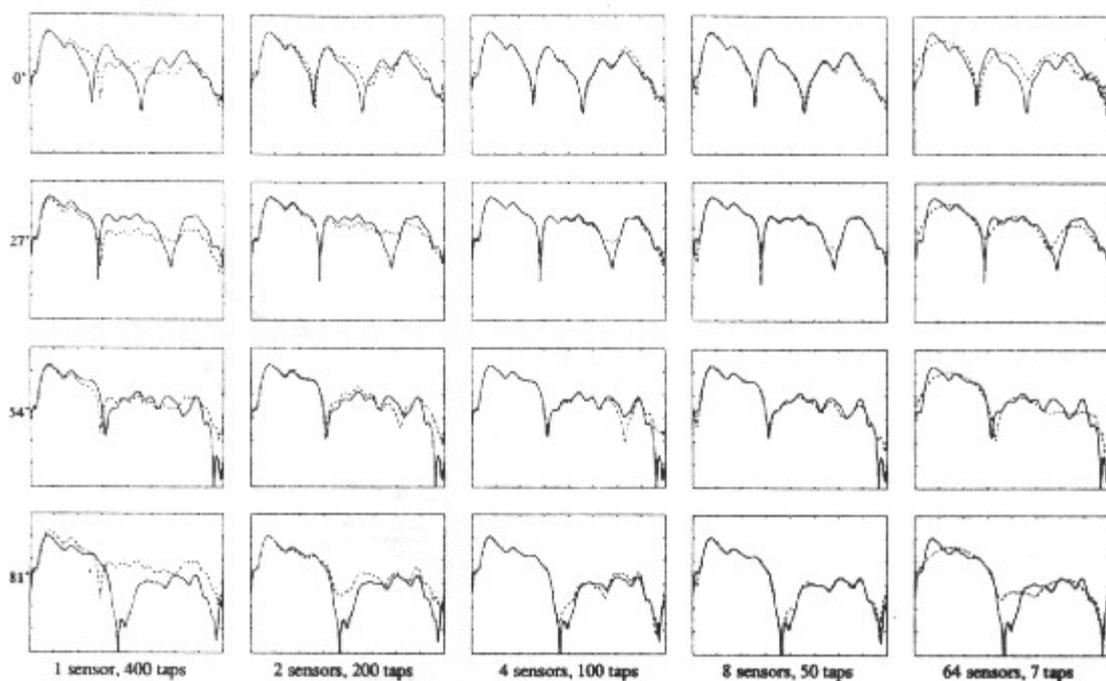
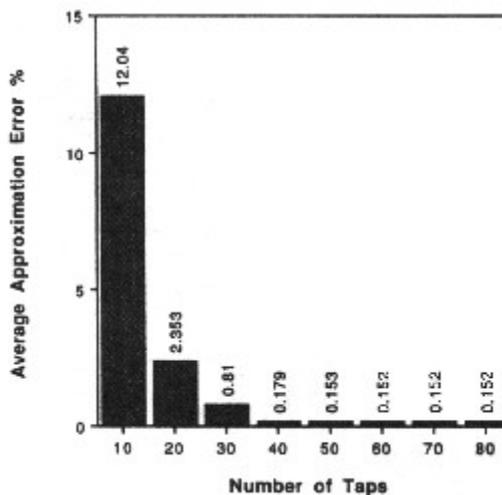
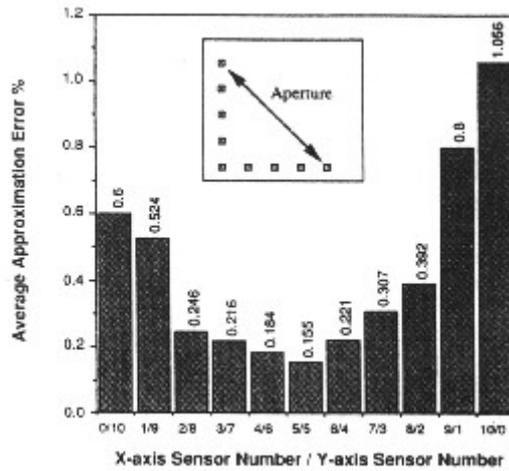
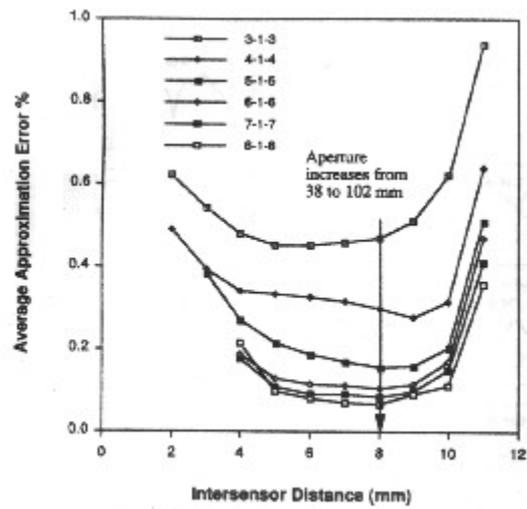


Abbildung 3.14 zeigt den Näherungsfehler als Funktion der Anzahl von Sensoren an jeder Achse einer L-förmigen Feldgeometrie. Jeder Sensorkanal enthält ein FIR-Filter der Länge 64, der Abstand zwischen den Sensoren beträgt 8mm. Abbildung 3.15 zeigt den Näherungsfehler in Abhängigkeit von der Filterlänge in jedem Kanal eines 5-1-5 L-förmigen Feldgeometrie. Der Abstand der Sensoren beträgt jeweils 8 mm.



Modellierung von HRTF-Kurven

Abbildung 3.16 zeigt den mittleren Näherungsfehler in Abhängigkeit vom Abstand der Sensoren für sechs verschiedene Sensorkonfigurationen. Alle Konfigurationen haben FIR-Filter der Länge 64 in jedem Sensorkanal.



4 Balanced Model Truncation (BMT)

1 Allgemeines

Eine balancierte Realisierung ist eine spezielle Zustandsraumdarstellung eines Systems mit bestimmten Eigenschaften, die es ermöglichen, durch Partitionierung die Ordnung des Systems zu reduzieren. Das ordnungsreduzierte System stellt dann eine Näherung für das Originalsystem dar. Definiert man in geeigneter Weise einen Näherungsfehler, so kann man für diesen eine obere Schranke angeben. Man nützt diese Erkenntnisse, um für ein gegebenes FIR-Filter eine spezielle Zustandstransformation so anzugeben, daß das System in eine balancierte Realisierung übergeführt wird und anschließend reduziert werden kann. (Die Methode der balancierten Realisierung basiert auf der Wahl der Zustandsvariablen.)

Diese Methode wird in [Mackenzie et al., 1997], [Beliczinski et al., 1992] und [Kale, 1993] behandelt.

In Abschnitt 3.4.5 werden weitere mathematische Zusammenhänge angeführt, auf die im Text mit "siehe Zusatz X in Abschnitt 3.4.5" verwiesen wird.

2 Prinzip der balancierten Realisierung und Modellreduktion

Zu einer Übertragungsfunktion $F(z)$ eines Übertragungssystems kann ein Zustandsraummodell der Form angegeben werden. Hierbei ist \mathbf{x} der Zustandsvektor, u die skalare Eingangsgröße und y die skalare Ausgangsgröße.

Zu einer Übertragungsfunktion gibt es bekanntlich unendlich viele Eingangs-/Ausgangs-äquivalente Realisierungen im Zustandsraum. Von einer Zustandsraumdarstellung gelangt

man mittels einer regulären Zustandstransformation \mathbf{T} in eine andere:

$$\begin{aligned} \mathbf{x}(k) &= \mathbf{T} \\ \mathbf{z}(k) & \end{aligned}$$

Somit sind

die entsprechenden Größen der Zustandsraumdarstellung des Modells in \mathbf{z} .

Die zugehörige Übertragungsfunktion ändert sich nicht, d. h. es gilt:

Definition: Balancierte Realisierung

Einer Zustandsraumdarstellung des Systems $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ können zwei spezielle Matrizen \mathbf{P} und \mathbf{Q} zugeordnet werden. Sie sind die Lösungen der Ljapunow-Gleichungen:

und

Diese Matrizen \mathbf{P} und \mathbf{Q} sind als Steuerbarkeits- bzw. Beobachtbarkeits-Gramsche Matrizen bekannt. Wenn das Übertragungssystem $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ asymptotisch stabil ist und das Paar (\mathbf{A}, \mathbf{b}) vollständig steuerbar und (\mathbf{A}, \mathbf{c}) vollständig beobachtbar ist, dann sind \mathbf{P} und \mathbf{Q} positiv-definite Matrizen. Diese sind dann eindeutig bestimmbar.

Eine Zustandsraumdarstellung mit der Eigenschaft, daß für die Matrizen \mathbf{P} und \mathbf{Q} gilt: nennt man balancierte Realisierung (*balanced realization*).

Die Wahl der Transformation \mathbf{T} , die zu einer balancierten Realisierung führt, wird in Zusatz 1 in Abschnitt 3.4.5 beschrieben.

Eigenschaften einer balancierten Realisierung

Wendet man eine Transformation \mathbf{T} auf das System $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ an, so besitzt das neue System neue Gramsche Matrizen \mathbf{P} und \mathbf{Q} , die sich aus folgenden Gleichungen ergeben:

und

Setzt man nun Gleichung (3.55) in die Gleichungen (3.60) und (3.61) ein, dann erhält man $\mathbf{P} = \mathbf{T}^{-1} \mathbf{P}_0 \mathbf{T}$ und $\mathbf{Q} = \mathbf{T} \mathbf{Q}_0 \mathbf{T}^{-1}$. Die Gramschen Matrizen \mathbf{P} und \mathbf{Q} hängen stark von der Wahl der Zustandsgrößen ab. Die Eigenwerte ihres Produkts $\lambda_i(\mathbf{PQ})$ aber sind invariant unter einer Zustandstransformation und daher eine Eingangs-/Ausgangs-Invariante.

Sind alle Eigenwerte der Matrix \mathbf{A} negativ, d. h. $\text{Re}(\lambda_i(\mathbf{A})) < 0$, dann sind die Hankel-Singulärwerte (*Hankel Singular Values*, HSVs) von $F(z)$ definiert als

Diese Werte sind üblicherweise in absteigender Reihenfolge geordnet:

Der maximale Singulärwert definiert die sogenannte Hankel-Norm einer Übertragungsfunktion:

Theorem: Balanced Model Truncation (BMT)

Ein balanciertes System $(\mathbf{A}_b, \mathbf{b}_b, \mathbf{c}_b)$ der Ordnung n kann im Zustandsraum in zwei Untersysteme partitioniert werden, nämlich in ein gestutztes (*truncated*) System $(\mathbf{A}_{11}, \mathbf{b}_1, \mathbf{c}_1)$ der Ordnung k und ein verworfenes (*rejected*) System $(\mathbf{A}_{22}, \mathbf{b}_2, \mathbf{c}_2)$ der Ordnung $n-k$:

Die zum Übertragungssystem gehörigen Matrizen \mathbf{P}_b und \mathbf{Q}_b (nach Gl. (3.57) und (3.58)) sind dann durch $\mathbf{\Sigma}$ folgendermaßen darstellbar:

Ist das System $(\mathbf{A}_b, \mathbf{b}_b, \mathbf{c}_b)$ asymptotisch stabil, balanciert, vollständig steuerbar und vollständig beobachtbar, so ist das gestutzte Subsystem $(\mathbf{A}_{11}, \mathbf{b}_1, \mathbf{c}_1)$ der Ordnung k ebenfalls asymptotisch stabil, balanciert, vollständig steuerbar und vollständig beobachtbar.

$F_k(z)$ ist nun die Übertragungsfunktion, die man durch Stutzung (*truncation*) der balancierten Realisierung $F(z)$ auf die ersten k Zustände erhält:

k kann beliebig gewählt werden. Eine wichtige Entscheidungsgrundlage bildet aber die folgende Eigenschaft:

Durch die in Gleichung (3.65) definierte Hankel-Norm kann eine Abschätzung für die reduzierte Übertragungsfunktion $F_k(z)$ folgendermaßen angegeben werden¹:

Das heißt, die Hankel-Norm der Differenz zwischen der originalen Systemübertragungsfunktion $F(z)$ und der angenäherten $F_k(z)$ ist kleiner oder gleich der Summe der HSVs des verworfenen Systems. Wenn die HSVs des verworfenen Systems "viel kleiner" als die des gestutzten Systems sind, wird der Näherungsfehler klein. In der Praxis ist die Reduzierbarkeit der Ordnung des Systems von der Verteilung der HSVs der balancierten Form des Originalsystems abhängig.

Berechnung der Hankel-Singulärwerte und Bestimmung einer balancierten Realisierung

Die Hankel-Norm kann zur Berechnung der HSVs verwendet werden. Für das Übertragungssystem $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ sind die Markow-Parameter H_k definiert als

Die Markow-Parameter repräsentieren die Impulsantwortfolge des Übertragungssystems $(\mathbf{A}, \mathbf{b}, \mathbf{c})$. Die Hankel-Matrix \mathbf{H} ist folgendermaßen definiert:

Die Hankel-Singulärwerte von $F(z)$ sind die Eigenwerte von \mathbf{H} und es gilt:

Weitere Zusammenhänge zwischen \mathbf{P} , \mathbf{Q} und \mathbf{H} werden in Zusatz 2 in Abschnitt 3.4.5

¹tr \mathbf{A} ist die Spur (*trace*) der Matrix \mathbf{A} . Sie ist definiert als die Summe der Elemente in der Hauptdiagonale von \mathbf{A} und ist weiters gleich der Summe der Eigenwerte von \mathbf{A} .

Modellierung von HRTF-Kurven
angegeben.

Nachdem \mathbf{H} eine symmetrische Matrix ist, kann sie wie folgt faktorisiert werden:

wobei

Dabei ist $\mathbf{\Lambda}$ eine Diagonalmatrix und \mathbf{I} die Einheitsmatrix.

Ein Folgesatz betreffend \mathbf{P} und \mathbf{W}_c (siehe Zusatz 2) wird in Zusatz 3 in Abschnitt 3.4.5 angegeben.

Die Zustandstransformation

führt zu einer balancierten Realisierung des Systems, wobei $|\cdot|$ die absoluten Werte der Matrixelemente symbolisiert.² Weiters kann man zeigen, daß

Die Zusammenhänge zwischen \mathbf{H} , \mathbf{P} , \mathbf{Q} , \mathbf{V} und $\mathbf{\Lambda}$ werden in Zusatz 4 in Abschnitt 3.4.5 angeführt.

Verwendet man die Zustandstransformation aus Gleichung (3.75) unter Anwendung von Gleichung (3.62), so erhält man

Hat man diese balancierte Realisierung, so kann man die Ordnung des Systems reduzieren, indem man ein gestutztes System bildet.

Falls diese Realisierung aufgrund der stark variierenden Diagonalelemente von $\mathbf{\Lambda}$ nicht gut ist, kann man auch die folgende Transformation verwenden:

Man erhält dadurch das System in einer Form, die als "input normal" bezeichnet wird. Die Übertragungsfunktion, die man durch Transformation nach Gleichung (3.78) erhält, ist identisch zu jener, die sich durch Transformation nach Gleichung (3.75) ergibt.

Die wichtigen (und praktischen) Zusammenhänge zwischen dem System $(\mathbf{A}_k, \mathbf{b}_k, \mathbf{c}_k)$ bzw. $(\mathbf{A}_b, \mathbf{b}_b, \mathbf{c}_b)$ und \mathbf{V} und $\mathbf{\Lambda}$ enthält Zusatz 5 in Abschnitt 3.4.5.

²Beweis: Unter Verwendung der Gleichungen (3.81) und (3.94) erhält man (3.98).

3 Konversionsalgorithmus FIR -> IIR

Gegeben sei ein FIR-Filter mit N Koeffizienten (ein System $(N-1)$ -ter Ordnung, d. h. $n=N-1$), geschrieben in der folgenden Form als Übertragungsfunktion:

Das Filter $F(z)$ kann (wie in Abschnitt 3.4.2 beschrieben) als Zustandsraummodell angegeben werden:

wobei

Zu bemerken ist, daß das Filter $F_1(z)$ allein durch \mathbf{A} , \mathbf{b} und \mathbf{c} dargestellt werden kann.

Da das System $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ eine endliche Impulsantwort besitzt, wird nun die finite Hankel-Matrix nach Gleichung (3.70) gebildet:

Danach wird mit \mathbf{H} eine Singulärwertzerlegung (*singular value decomposition*, SVD) durchgeführt, als Ergebnis erhält man die HSVs.

Bei Darstellung der HSVs als Plot kann man sofort die voraussichtliche Ordnung k der Näherung ablesen: Sie entspricht der Anzahl der k größten HSVs, die im Vergleich zu den nachfolgenden nicht vernachlässigt werden können. Der zu erwartende Fehler kann mit Gleichung (3.69) abgeschätzt werden.

Danach erfolgt die Berechnung des gestutzten Systems:

wobei $\mathbf{V}(i:j, k:m)$ die Extraktion der Reihen i bis j und der Spalten k bis m der Matrix \mathbf{V} bedeutet.

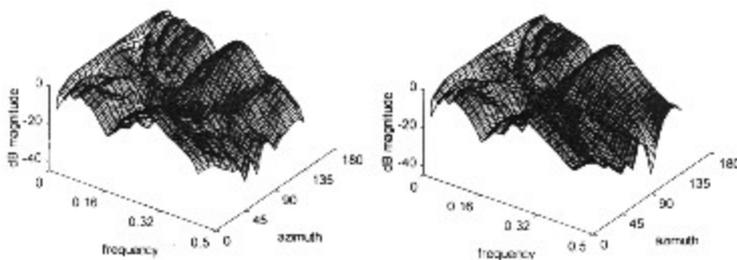
Die Übertragungsfunktion $F_r(z)$ des reduzierten Systems $(\mathbf{A}_k, \mathbf{b}_k, \mathbf{c}_k, d)$ mit $d=c_0$ aus Gleichung (3.79) ergibt sich aus

4 Experimentelle Ergebnisse

Vorgangsweise nach [Mackenzie et al., 1997]:

- Ausgangspunkt sind die gemessenen HRTF-Impulsantworten (Kunstkopf) von 512 Abtastwerten bei 44,1 kHz.
- Entfernen der Zeitverzögerung am Anfang
- Diffusfeld-Entzerrung der HRTFs
- Glättung der Amplitudenantwort
- Rekonstruktion einer minimalphasigen Version => 128-Punkt-Antwort
- Interpretation der 128-Punkt-Antwort als FIR-Filter mit den Koeffizienten c_0, c_1, \dots, c_n .
- BMT wie in Abschnitt 3.4.3 beschrieben

In [Mackenzie et al., 1997] wurden IIR-Filter 10. Ordnung mittels BMT von FIR-Filtern 128. Ordnung eines HRTF-Satzes abgeleitet. Die Abbildung 3.17 zeigt die Amplitudenantworten für 19 HRTFs von FIR-Filtern 128. Ordnung (linke Seite) bzw. IIR-Filtern 10. Ordnung (rechte Seite). Die Autoren erhalten bessere Ergebnisse als mit nach der Prony- und Yule-Walker-Methode (wie sie in der Matlab-Toolbox implementiert sind) entworfenen IIR-Filtern (Abb. 3.18).

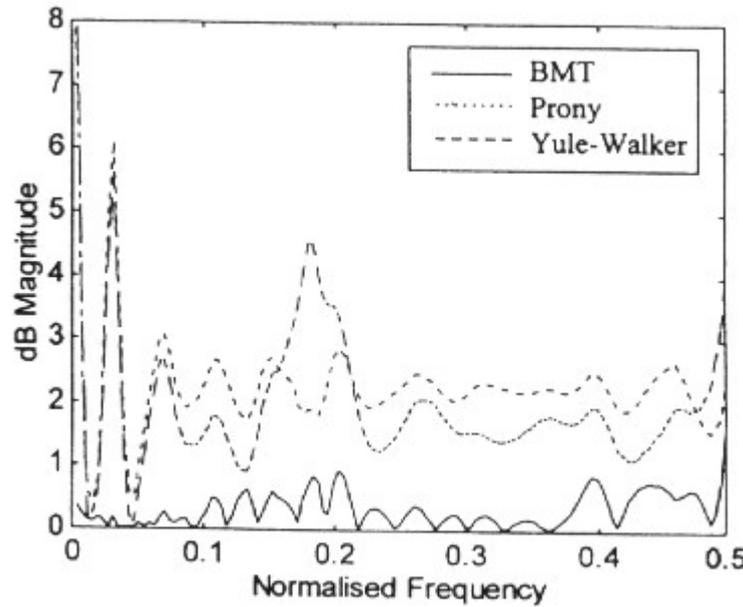


[Beliczinski et al.,
folgende
Beobachtungen

· Enge
konnten
breitere
desselben
reduziert
Bei
Filtern mit

Amplitudenantwort wurde die beste Reduktion für minimalphasige Filter erzielt. Bei Maximalphasigkeit ergab sich eine signifikant schlechtere Reduktion.

· Die präsentierte Technik ist am geeignetsten für "nahezu linearphasiges" IIR-Filterdesign. Die effiziente Reduktion ist möglich, weil die Prototyp-FIR-Filter in der gesamten Bandbreite linear sind, die reduzierten IIR-Filter jedoch sind nichtlinear in den irrelevanten Stopbändern, behalten die Linearphasigkeit in den Durchlaßbändern aber bei.



1992] konnten
praktische
machen:

Bandpaßfilter
effizienter als
Bandpaßfilter
Phasentyps
werden.

verschiedenen
gleicher

5 Weitere mathematische Zusammenhänge

Zusatz 1

Dies ist (neben der Bestimmung der Hankel-Singulärwerte in Abschnitt 3.4.2 und 3.4.3) eine weitere Möglichkeit zur Bestimmung einer balancierten Realisierung.

Man bestimmt nun eine spezielle Transformation \mathbf{T} durch die Matrizen \mathbf{S} , \mathbf{U} und $\mathbf{\Sigma}$, die zu einer balancierten Realisierung führt. Es gilt:

Dabei sind die Matrizen \mathbf{S} , \mathbf{U} und $\mathbf{\Sigma}$ durch folgende Zusammenhänge gegeben:

Wenn das System $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ stabil ist und außerdem steuerbar und beobachtbar, dann sind die Matrizen \mathbf{P} und \mathbf{Q} positiv-definit und können faktorisiert werden:

Dann ist die Matrix

eine symmetrische Matrix, die folgendermaßen faktorisiert werden kann:

wobei

Daß diese Wahl von \mathbf{T} tatsächlich auf eine balancierte Realisierung führt, zeigt folgende Umformung:

und

Die Transformation \mathbf{T} , auf das System $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ angewandt, ergibt also ein balanciertes System.

Eine balancierte Realisierung ist eindeutig bezüglich einer beliebigen Zustandsraumtransformation \mathbf{T}_b , wenn gilt

Zusatz 2

Die Hankel-Matrix kann geschrieben werden als

wobei

Das Kronecker-Theorem besagt, daß der Rang (\mathbf{H}) gleich der Ordnung von $F(z)$ ist. Nachdem Gl. (3.57) so geschrieben werden kann

ist

Analog kann man zeigen, daß

Die Matrizen \mathbf{W}_o und \mathbf{W}_c definiert in Gleichung (3.94) haben ebenfalls endliche Dimensionen:

. $\mathbb{R}^{n \times n}$ ist die Menge reeller Matrizen, \mathbf{W}_o bzw. \mathbf{W}_c wird Beobachtbarkeits- bzw.

Steuerbarkeitsmatrix genannt.

Zusatz 3

Folgesatz:

Für das oben angeführte System $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ sind die Steuerbarkeitsmatrix und die Steuerbarkeits-Gramsche Matrix Einheitsmatrizen, das heißt

Zusatz 4

In Übereinstimmung mit den Gleichungen (3.93), (3.98), (3.97) und (3.73) sind
und
sowie

Zusatz 5

Theorem:

Ist die Hankel-Matrix \mathbf{H} eines FIR-Filters $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ n-ter Ordnung (nach Gl. (3.101)) wie in Gleichung (3.73) faktorisiert, dann ist ein reduziertes balanciertes System k-ter Ordnung Eingangs-/Ausgangs-äquivalent zum System $(\mathbf{A}_k, \mathbf{b}_k, \mathbf{c}_k)$, wobei
und \mathbf{V}_k eine rechteckige $n \times k$ -Matrix ist, erhalten durch die folgenden Partitionierung:

Beweis:

Partitioniert man \mathbf{A} wie folgt

dann erhält man unter Verwendung der Zustandstransformation die balancierte Realisierung

Das gestutzte System hat dann die Übertragungsfunktion

Aufgrund der speziellen Form von \mathbf{A} und \mathbf{b} kann man zeigen, daß

wobei $\mathbf{V}(i:j, k:m)$ die Extraktion der Reihen i bis j und der Spalten k bis m der Matrix \mathbf{V} bedeutet.

4 Implementierung von HRTFs nach der Hauptkomponentenmethode

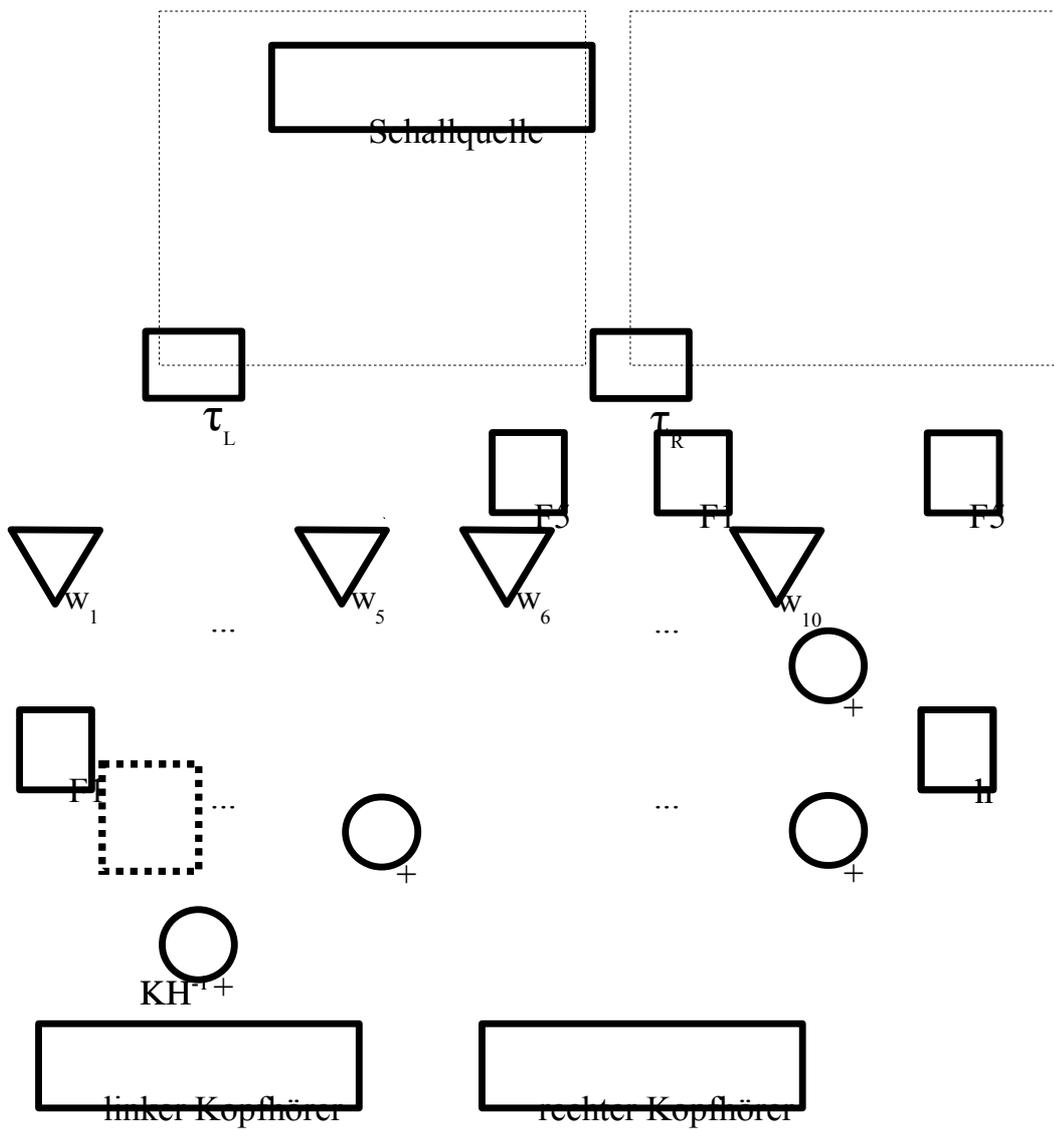
Im folgenden wird die Implementierung von HRTFs nach der Hauptkomponentenmethode beschrieben. Diese Implementierung wurde als Prototyping in MATLAB mit den Daten von Gardner/Martin (KEMAR Kunstkopf) durchgeführt.

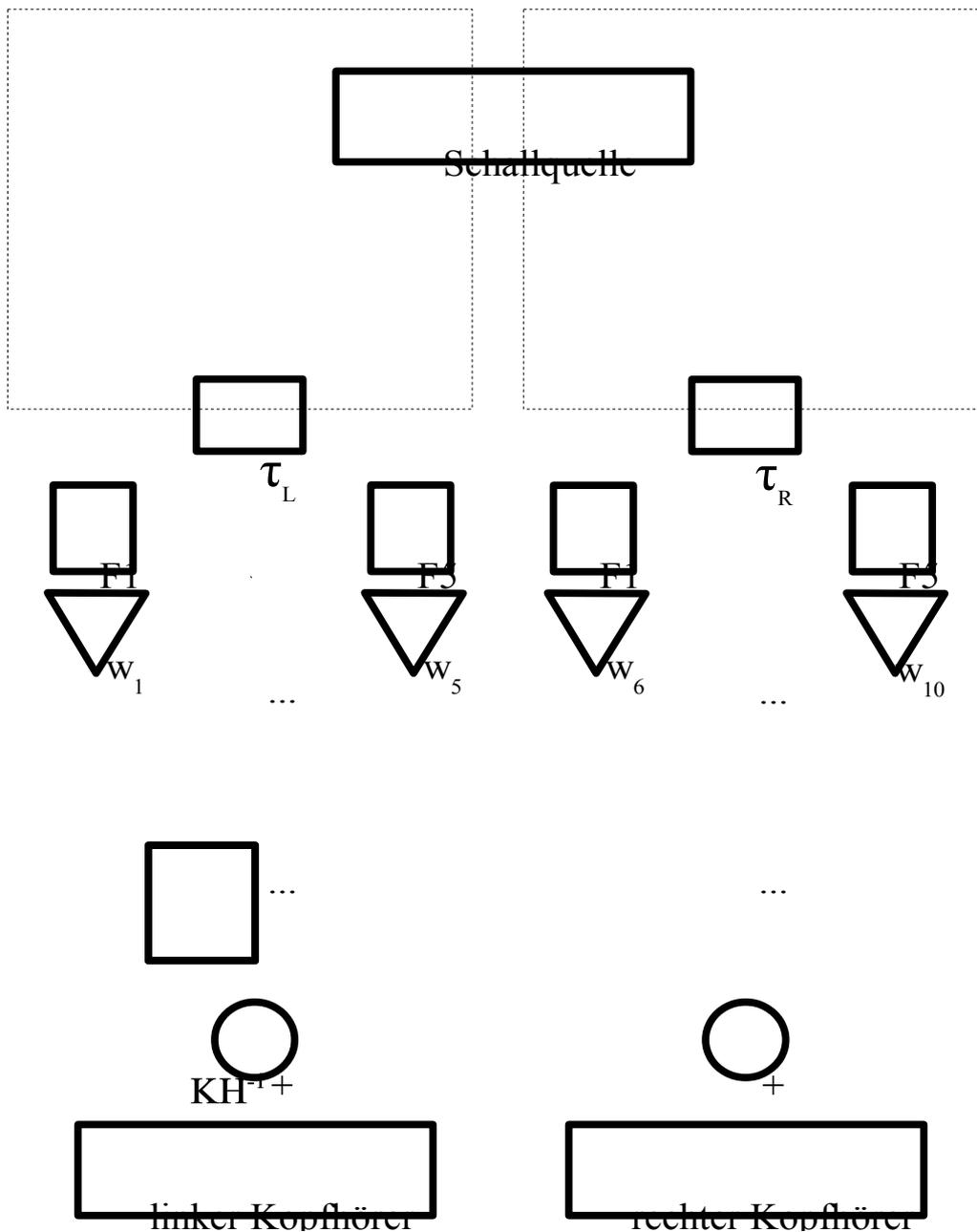
Nach Bearbeitung der Rohdaten wird die Hauptkomponentenmethode (wie in Abschnitt 3.1 beschrieben) sowohl im Frequenz- als auch im Zeitbereich gerechnet.

Das Prinzip der Implementierung der Hauptkomponentenmethode zeigen die Abbildungen 4.1 und 4.2. Geht man von einer hinreichend genauen Näherung mit 5 Basisvektoren aus, so benötigt man bei einer Schallquelle für jedes Ohr eine Zeitverzögerung τ , 5 Gewichte und 5 Filter für die Basisvektoren. Die gemittelte Impuls- bzw. Frequenzantwort wird im Fall von logarithmischen Daten im Frequenzbereich in Kaskade geschaltet (Abb. 4.2), im Fall von linearen Daten im Zeit- und Frequenzbereich parallel (Abb. 4.1).

Will man mehrere Schallquellen gleichzeitig bearbeiten, so bleibt die Anzahl der benötigten Filter gleich, es kommen pro Quelle und Ohr eine Zeitverzögerung und 5 Gewichte hinzu (in den Abbildungen strichliert eingefaßt).

Implementierung von HRTFs nach der Hauptkomponentenmethode





1 **Bearbeitung und Korrektur der Rohdaten**

Als Ausgangsmaterial wurden die Impulsantworten mit 512 Punkten Länge der linken KEMAR-Pinna gewählt. Die Abtastrate beträgt 44,1 kHz, es liegen 710 verschiedene Richtungen vor.

Beim Einlesen der Daten wird die ITD extrahiert. Nachdem die Werte der Impulsantworten im Bereich von ± 1 liegen, wird nach dem ersten Wert gesucht, dessen Betrag größer als 0,01 ist. Als Impulsantwort werden, beginnend beim 4. Abtastwert davor, die folgenden 256 Werte genommen. Die ermittelte ITD wird gespeichert, um sie später wieder hinzuzufügen. Der Verlauf der HRTF-Meßdaten für die Richtung $(0^\circ, 0^\circ)$ ist aus den Abbildungen 4.3 und 4.4 ersichtlich, der Verlauf der ITDs der linken Pinna aus Abbildung 4.5.

Am Verlauf der ITDs sieht man, daß die Abweichungen von einem glatten Kurvenverlauf maximal einen Abtastwert betragen, was 0,023ms entspricht. Da dieser Unterschied nicht wahrnehmbar ist, ist keine Korrektur des Kurvenverlaufs nötig.

Der Funktionsverlauf der ITDs folgt für die dem Ohr zugewandte Seite einer Sinusfunktion, für die abgewandte Seite ist der Zusammenhang linear. Letzteres erklärt sich durch die Verlängerung des Schallwegs durch Beugung.

Die Impulsantworten für den Kopfhörer (AKG 240) und den verwendeten Lautsprecher werden auf die gleiche Art mit einer Länge von 128 Punkten eingelesen. Allerdings werden nur so viele Werte genommen, daß in der resultierenden Impulsantwort 99,99% des Energieinhalts enthalten sind. Die restlichen Werte werden gleich Null gesetzt. Die ersten 4 und letzten 10 Werte ungleich Null werden linear ein- bzw. ausgeblendet. Die Abbildungen 4.6 bis 4.9 zeigen diese Lautsprecher- und Kopfhörer-Impulsantworten.

Nimmt man bei den Kopfhörer- und Lautsprecher-Impulsantworten Werte ungleich Null nur bis 99% oder 99,8% Energieanteil, so ergibt sich dadurch ein Fehler im Bereich bis 500 Hz, der sich bei der Invertierung des Frequenzgangs und der weiteren Datenkorrektur als ungünstig erweist. Die Abbildungen 4.10 und 4.11 zeigen die Lautsprecher-Frequenzgänge für 99% und 99,8% Energieinhalt. Der Bereich bis 500 Hz würde bei Invertierung stärker angehoben, was für die Korrektur der Rohdaten unerwünscht ist.

Für die weitere Bearbeitung der HRTF-Rohdaten werden die Lautsprecher- und Kopfhörerdaten mittels einer 512-Punkte FFT in den Frequenzbereich transformiert. Nachdem eine Invertierung der Frequenzgänge durch einfache Kehrwertbildung nur für minimalphasige Systeme zulässig ist, muß eine andere Vorgangsweise gewählt werden.

Der Amplitudengang wird durch Kehrwertbildung invertiert. Um den inversen Amplitudengang für besonders hohe und niedrige Werte zu begrenzen, wird folgende Funktion verwendet:

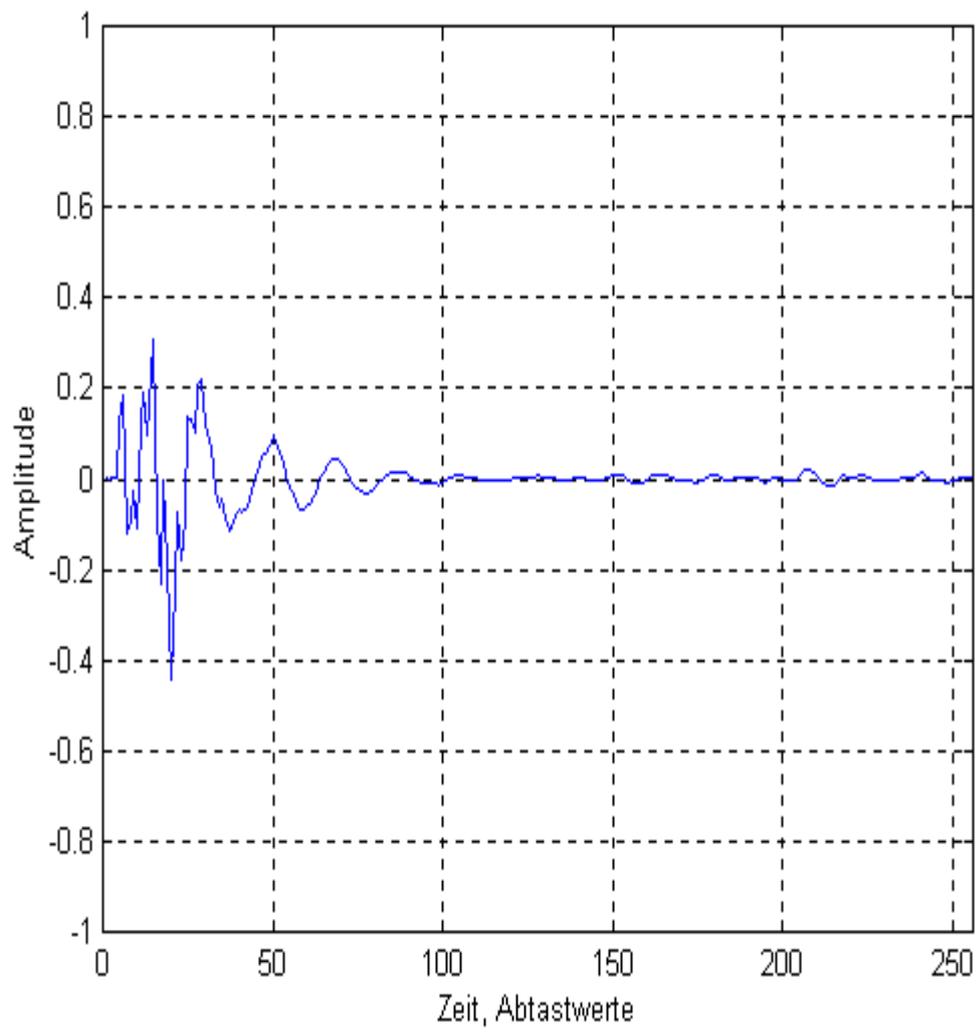
Dabei ist x der zu invertierende Amplitudengang, x_{lim} der beschränkte, invertierte Amplitudengang und k eine Konstante. Die Wahl des Werts für k bestimmt die Stärke der Stauchung und Beschränkung: je größer k , desto geringer der Dynamikbereich von x_{lim} . Abbildung 4.12 zeigt den invertierten Lautsprecher-Amplitudengang für $k = 0,05, 0,25$ und $0,5$. Für die Invertierung wurde $k = 0,25$ gewählt.

Der Phasengang wird minimalphasig rekonstruiert.

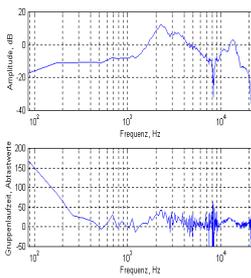
Abbildungen 4.13 bis 4.16 zeigen die resultierenden, invertierten Kopfhörer- und Lautsprecher-Impuls- bzw. Frequenzantworten.

Anschließend werden die Rohdaten mit der invertierten Lautsprecherimpulsantwort gefaltet. Die resultierenden Impulsantworten mit einer Länge von 383 Punkten werden nun wieder gekürzt und mit einer Länge von 128 Punkten verwendet. Diese Daten werden für die Freifeldsimulation verwendet und sind die Ausgangsbasis für alle weitere Berechnungen.

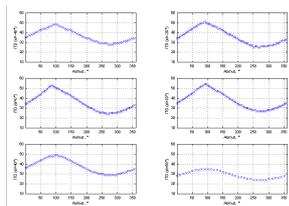
Die Abbildungen 4.17 und 4.18 zeigen Impulsantwort bzw. Frequenzgang der Freifeldantworten für die Richtung $(0^\circ, 0^\circ)$.



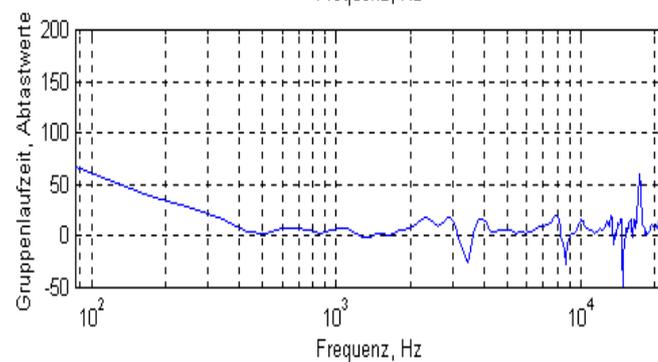
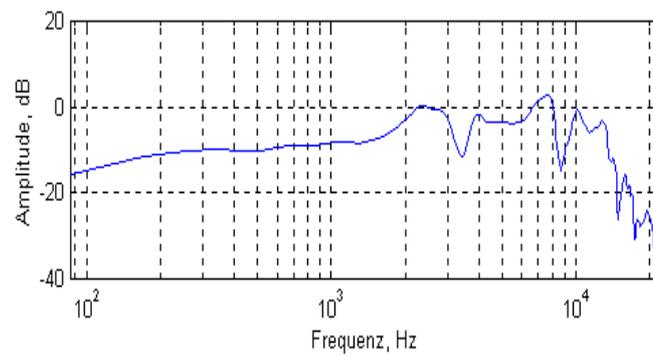
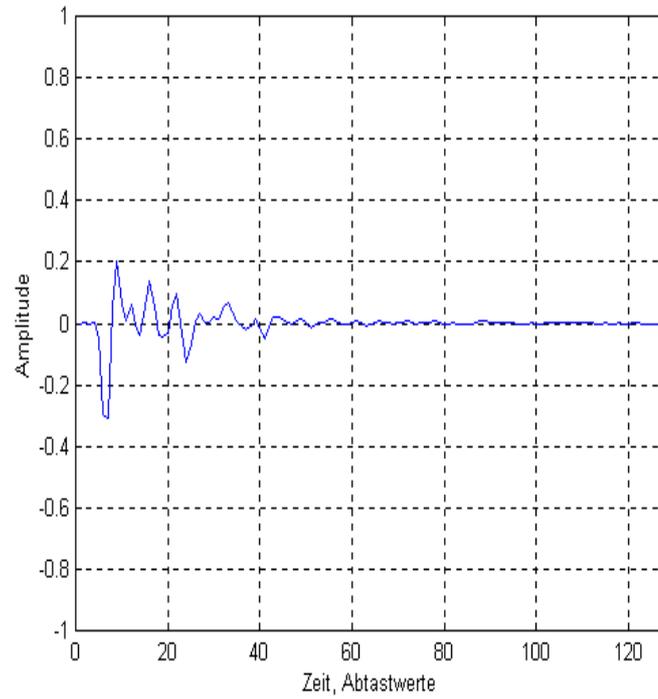
Implementierung von HRTFs nach der Hauptkomponentenmethode



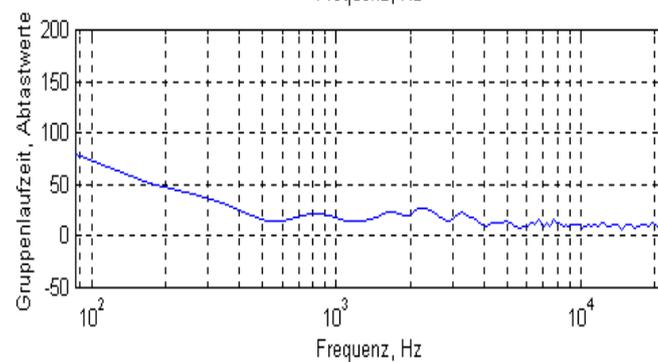
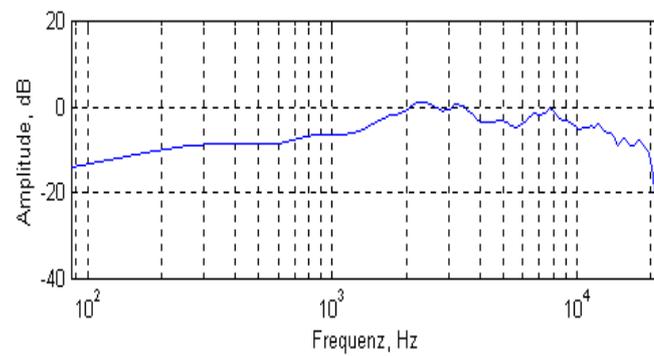
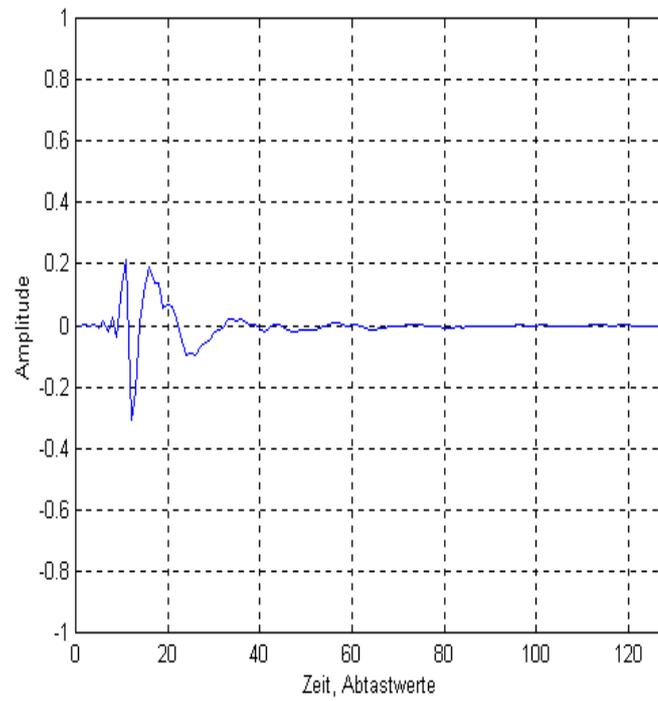
Implementierung von HRTFs nach der Hauptkomponentenmethode



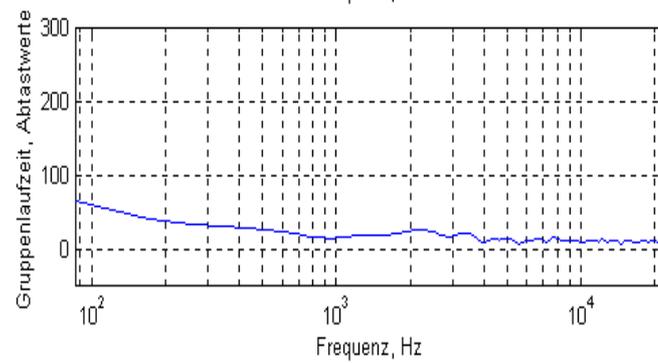
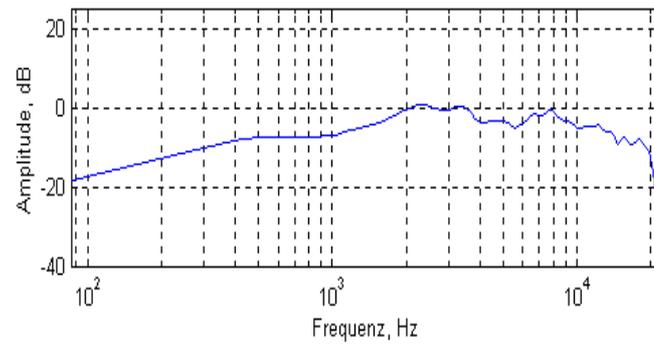
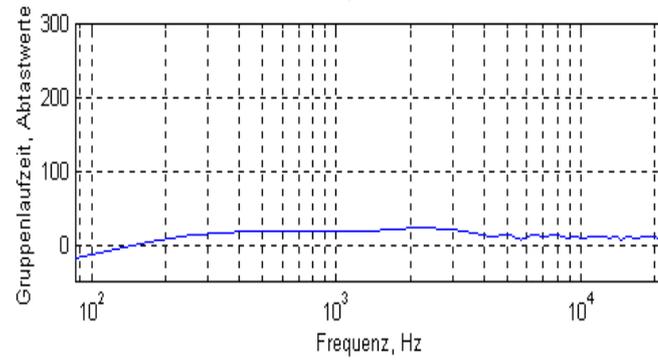
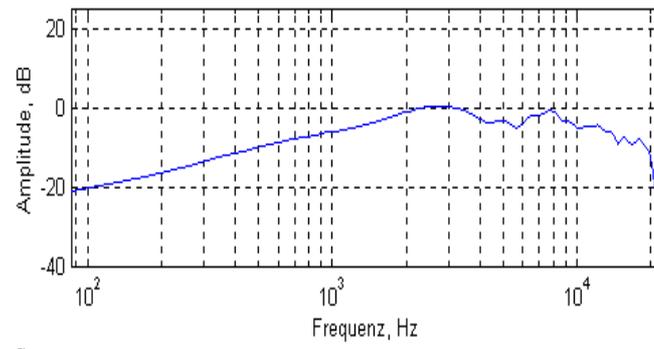
Implementierung von HRTFs nach der Hauptkomponentenmethode



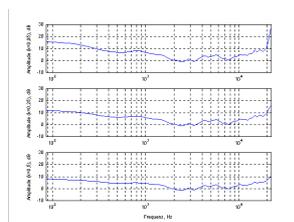
Implementierung von HRTFs nach der Hauptkomponentenmethode



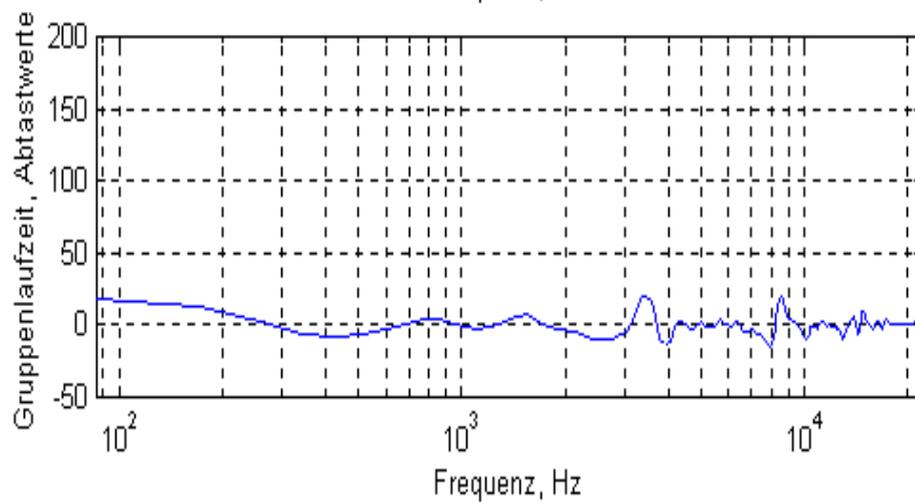
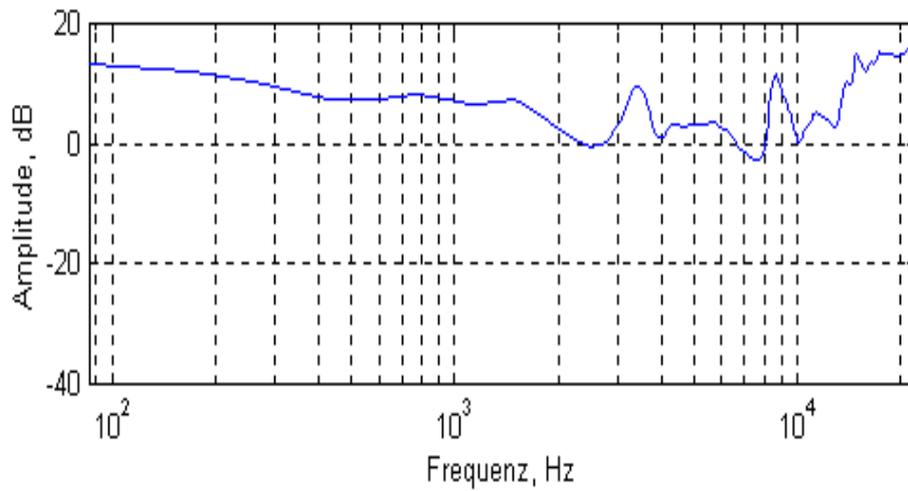
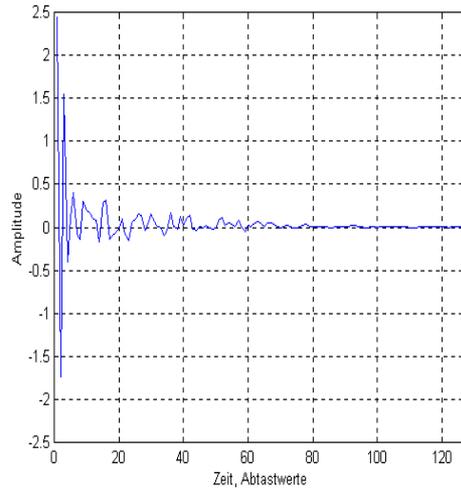
Implementierung von HRTFs nach der Hauptkomponentenmethode



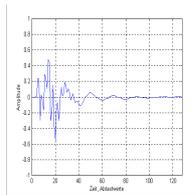
Implementierung von HRTFs nach der Hauptkomponentenmethode



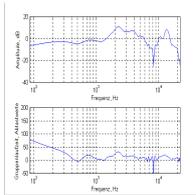
Implementierung von HRTFs nach der Hauptkomponentenmethode



Implementierung von HRTFs nach der Hauptkomponentenmethode



Implementierung von HRTFs nach der Hauptkomponentenmethode



2 Berechnung der Hauptkomponenten

Für die Rechnung im Zeitbereich stellen die Impulsantworten (HRIRs) das Ausgangsmaterial dar.

Für die Rechnung im Frequenzbereich werden die Daten zuerst mit einer 512-Punkte FFT in den Frequenzbereich transformiert (HRTFs). Vor Berechnung des Mittelwerts nach Gleichung (3.1) werden die Daten logarithmiert: für die reelle Rechnung nimmt man den Absolutbetrag der Daten, die komplexe Rechnung erfordert den komplexen Logarithmus. Nach Berechnung der DTFs laut Gleichung (3.2) werden die Daten wieder entlogarithmiert.

Entsprechend den Gleichungen (3.1) bis (3.6) werden der Mittelwert \bar{h} , die Kovarianzmatrix \mathbf{S} , die Basisvektoren \mathbf{c}_i und die Gewichte \mathbf{w}_i berechnet. Der Mittelwert wird minimalphasig realisiert (\bar{h}_m). Mit 5 Basisvektoren erhält man gute Näherungen der HRTFs, für eine hochgenaue Nachbildung sind ca. 30 Basisvektoren notwendig. Bei Verwendung aller Basisvektoren würden die ursprünglichen Kurven identisch nachgebildet.

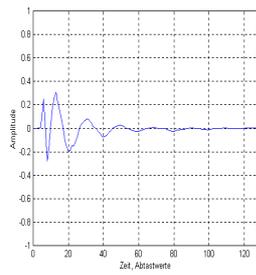
1 Rechnung im Zeitbereich

Die Abbildungen 4.19 bis 4.22 zeigen die gemittelte (\bar{h}) bzw. die minimalphasig realisierte, gemittelte Impulsantwort (\bar{h}_m) sowie ihre Darstellungen im Frequenzbereich. Um 2,5 kHz ist die Ohrkanalresonanz deutlich zu sehen.

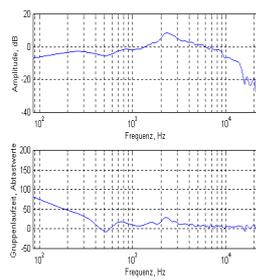
Die Abbildungen 4.23 und 4.24 zeigen die DIR für die Richtung $(0^\circ, 0^\circ)$ im Zeit- und Frequenzbereich.

Die Abbildungen (4.25 bis 4.30 zeigen die Näherungen der HRIR für die Richtung $(0^\circ, 0^\circ)$ mit 5, 10 und 20 Basisvektoren. Die Abbildung 4.31 zeigt die ersten 5 Basisvektoren.

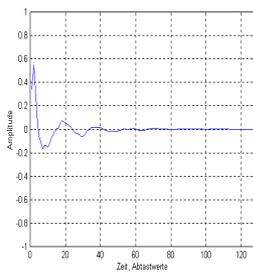
Implementierung von HRTFs nach der Hauptkomponentenmethode

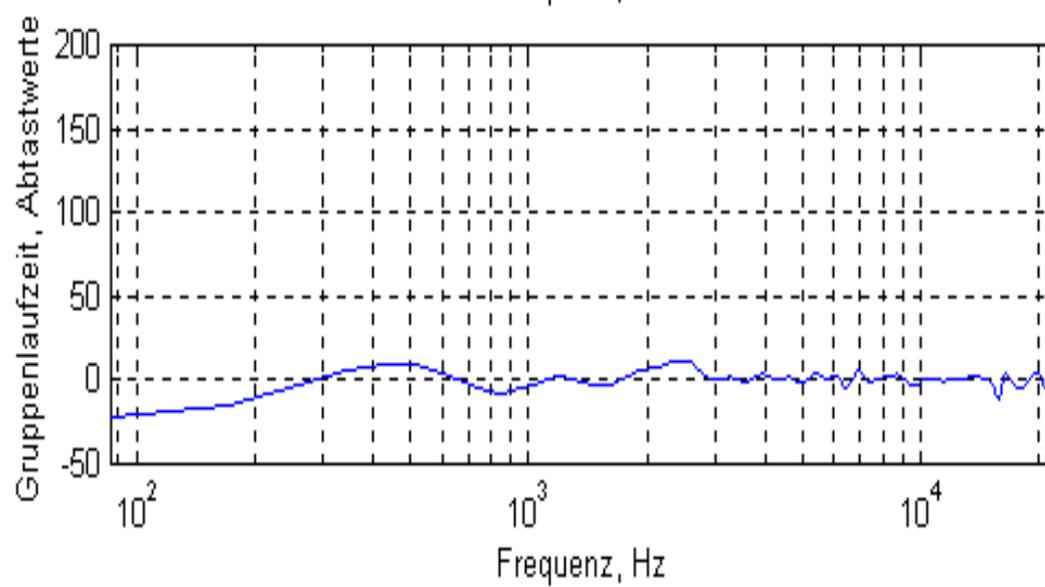
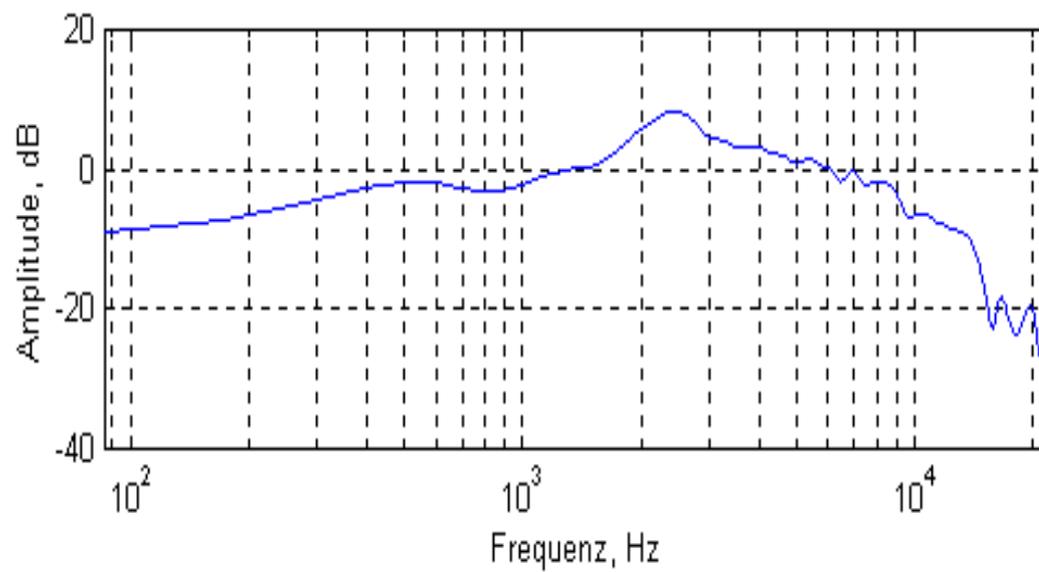


Implementierung von HRTFs nach der Hauptkomponentenmethode

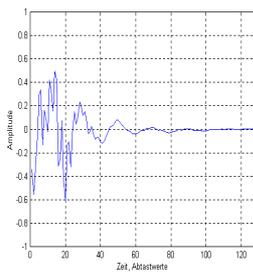


Implementierung von HRTFs nach der Hauptkomponentenmethode

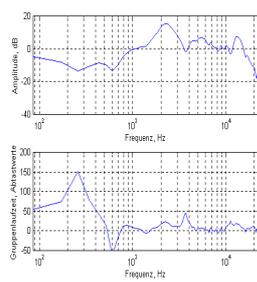




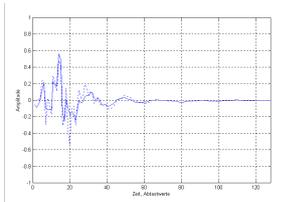
Implementierung von HRTFs nach der Hauptkomponentenmethode



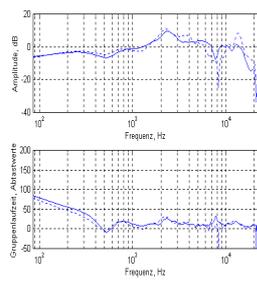
Implementierung von HRTFs nach der Hauptkomponentenmethode



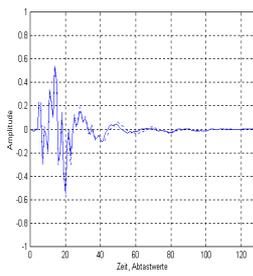
Implementierung von HRTFs nach der Hauptkomponentenmethode



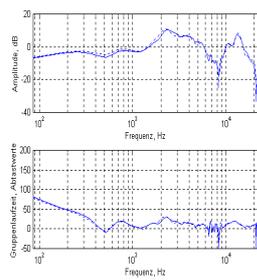
Implementierung von HRTFs nach der Hauptkomponentenmethode

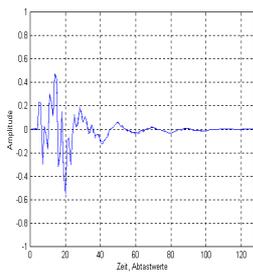


Implementierung von HRTFs nach der Hauptkomponentenmethode

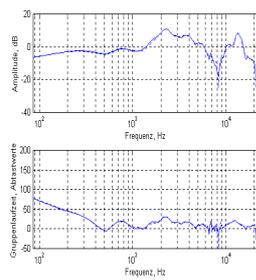


Implementierung von HRTFs nach der Hauptkomponentenmethode

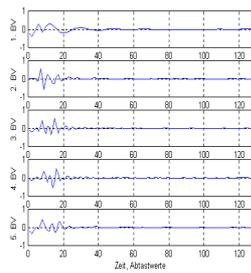




Implementierung von HRTFs nach der Hauptkomponentenmethode



Implementierung von HRTFs nach der Hauptkomponentenmethode

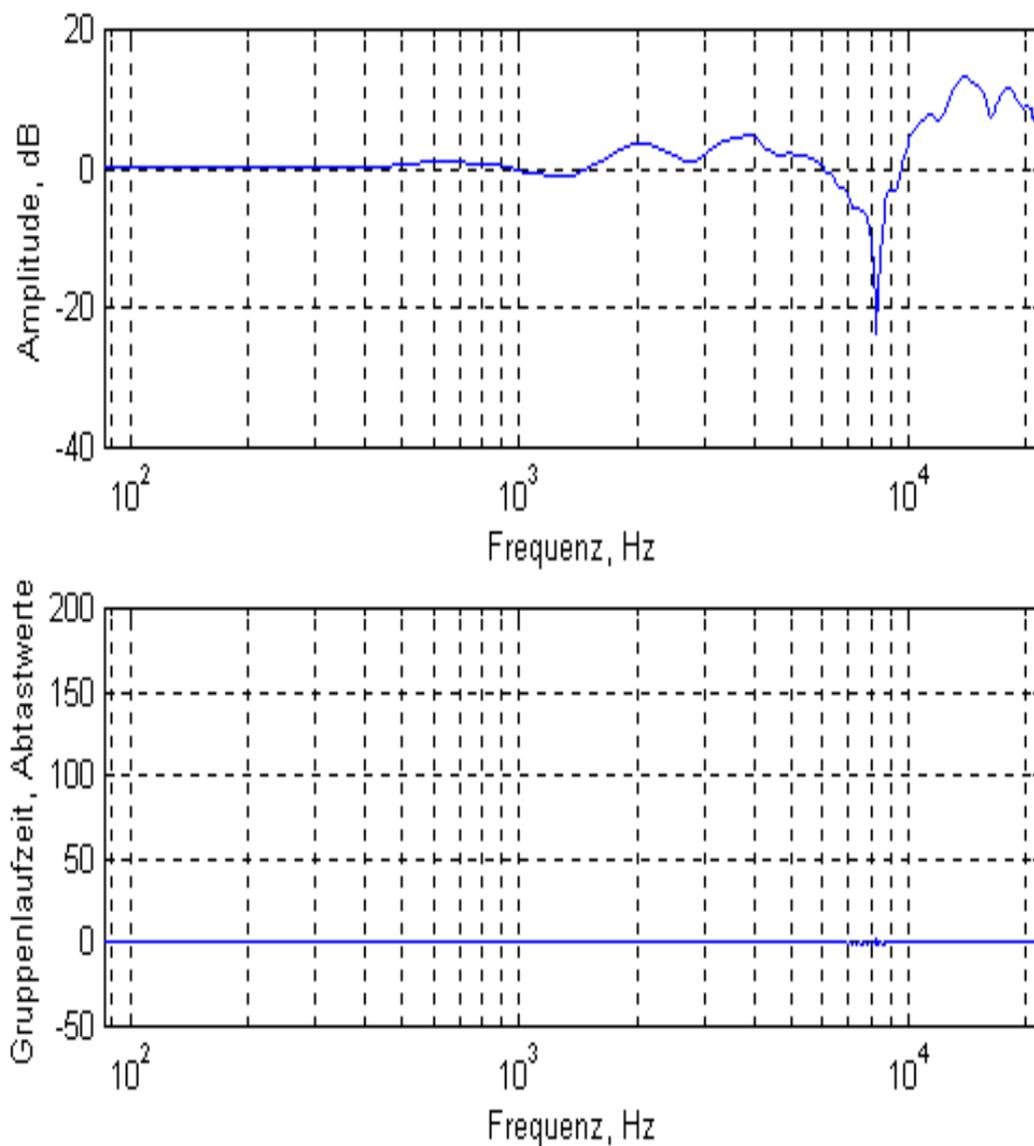


2 Reelle Rechnung im Frequenzbereich

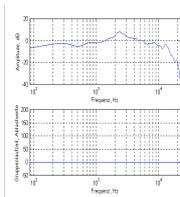
Die Abbildungen 4.33 und 4.34 zeigen die gemittelte () bzw. die minimalphasig realisierte, gemittelte Frequenzantwort ().

Abbildung 4.32 zeigt die DTF für die Richtung $(0^\circ, 0^\circ)$.

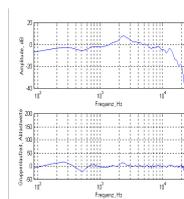
Die Abbildungen 4.35 bis 4.37 zeigen die Näherungen der HRTF für die Richtung $(0^\circ, 0^\circ)$ mit 5, 10 und 20 Basisvektoren. Die Abbildung 4.38 zeigt die ersten 5 Basisvektoren.



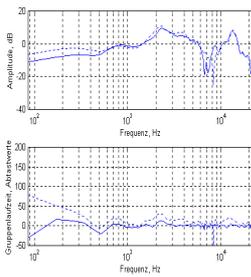
Implementierung von HRTFs nach der Hauptkomponentenmethode



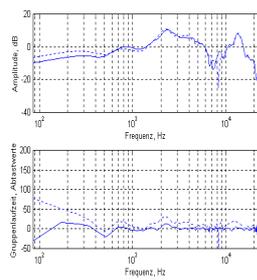
Implementierung von HRTFs nach der Hauptkomponentenmethode



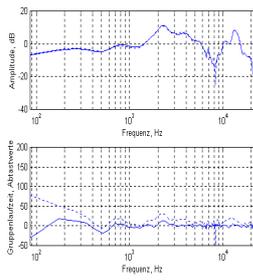
Implementierung von HRTFs nach der Hauptkomponentenmethode



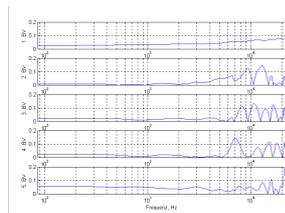
Implementierung von HRTFs nach der Hauptkomponentenmethode



Implementierung von HRTFs nach der Hauptkomponentenmethode



Implementierung von HRTFs nach der Hauptkomponentenmethode

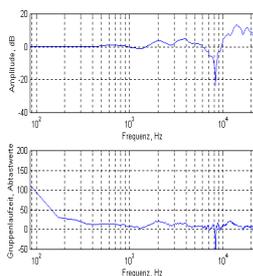


3 Komplexe Rechnung im Frequenzbereich

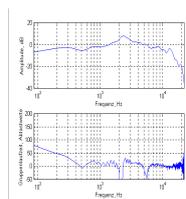
Die Abbildungen 4.40 und 4.41 zeigen die gemittelte () bzw. die minimalphasig realisierte, gemittelte Frequenzantwort.

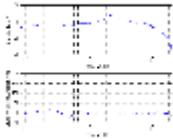
Abbildung 4.39 zeigt die DTF für die Richtung $(0^\circ, 0^\circ)$.

Die Abbildungen 4.42 bis 4.44 zeigen die Näherungen der HRTF für die Richtung $(0^\circ, 0^\circ)$ mit 5, 10 und 20 Basisvektoren. Die Abbildung 4.45 zeigt die ersten 5 Basisvektoren.

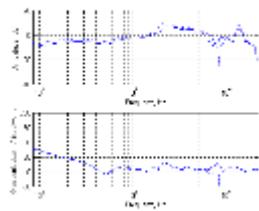


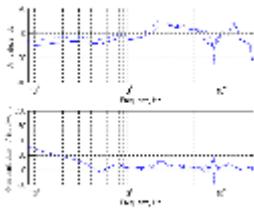
Implementierung von HRTFs nach der Hauptkomponentenmethode



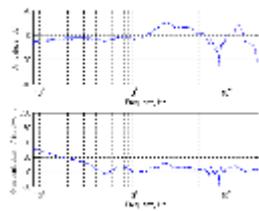


Implementierung von HRTFs nach der Hauptkomponentenmethode





Implementierung von HRTFs nach der Hauptkomponentenmethode



```

1  #include <math.h>
2  #include <stdio.h>
3  #include <stdlib.h>
4  #include <string.h>
5  #include <unistd.h>

```

5 Zusammenfassung und Ausblick

1.1 Zusammenfassung

Die vorliegende Arbeit behandelt Modelle für die recheneffiziente Nachbildung der Außenohrübertragungsfunktionen. Diese sind für einen möglichst natürlichen Höreindruck bei der Simulation räumlichen Hörens von großer Bedeutung. Nach einer Einführung über räumliches Hören und die speziellen Eigenschaften der menschlichen Außenohrübertragungsfunktionen werden vier mathematische Modelle vorgestellt: FIR- und IIR-Filterdesign, die Hauptkomponentenmethoden, Modellierung durch Richtstrahlbildung und Balanced Model Truncation. Die letzten drei Methoden stellen gegenüber direktem FIR- und IIR-Filterdesign eine große Effizienzsteigerung dar. Anschließend wird die Hauptkomponentenmethode in MATLAB implementiert. Dabei wird die Rechnung sowohl im Zeitbereich als auch reell und komplex im Frequenzbereich durchgeführt. Jede dieser Vorgangsweisen führt bereits bei Verwendung von nur 5 Basisvektoren zu guten Näherungen. Dies bedeutet, daß bei einer Realisierung für beide Ohren 12 Filter, 2 Zeitverzögerungen und 10 Gewichte nötig sind. Pro zusätzlich zu spatialisierender Schallquelle kommen aber nur 2 Zeitverzögerungen und 10 Gewichte hinzu.

1.2 Ausblick

Eine Erweiterung der vorliegenden Arbeit kann in der spezielleren Gestaltung der der Kovarianzmatrix zugrundeliegenden Daten für die Hauptkomponentenmethode bestehen. Um etwa bestimmte Richtungen zu bevorzugen, kann man das Gitter der vorliegenden Richtungen unterschiedlich dicht gestalten, entweder durch Verwendung von weniger Meßrichtungen oder durch Interpolation nicht gemessener Richtungen mit einer großen Anzahl von Basisvektoren. Für eine stärkere Gewichtung von bestimmten Frequenzen dem menschlichen Ohr entsprechend kann die Abtastung der Daten im Frequenzbereich ungleichmäßig erfolgen. Eine sinnvolle Möglichkeit wäre etwa, pro Oktave eine konstante Anzahl von Abtastpunkten zu wählen.

Literaturverzeichnis

Kapitel 1

Batteau D. W. (1967): "The role of the pinna in human localization", Proc. R. Soc. London Ser. B 168, pp. 158-180.

Begault D. R. (1994): "3-D Sound for Virtual Reality and Multimedia", AP Professional, USA 1994.

Blauert J. (1974): "Räumliches Hören", S. Hirzel Verlag Stuttgart, 1974.

Duda R. O. (1997): "Sound Localization Research", <http://www-engr.sjsu.edu/~duda/Duda.html>

Gerzon M. A. (1992): "The Design of Distance Panpots", Preprint 3308 of the 92nd Audio Engineering Society Convention, Wien, März 1992.

Hall D. E. (1980): "Musical Acoustics", Brooks/Cole Publishing Company, Pacific Grove, California, 1980.

Hebrank J. H., Wright D. (1974a): "Are two ears necessary for localization of sound sources on the median plane?", Journal of the Acoustical Society of America, 56, pp. 935-938.

Hebrank J. H., Wright D. (1974b): "Spectral cues in the localization of sound sources on the median plane", Journal of the Acoustical Society of America, 56, pp. 1829-1834.

Kendall G. S. (1995): "A 3-D Sound Primer: Directional Hearing and Stereo Reproduction", Computer Music Journal 19(4), pp. 23-46.

Thompson S. P. (1882): "On the function of the two ears in the perception of space", Philos. Mag. 13, pp. 406-416.

Wenzel E. M. (1994): "Spatial Sound and Sonification", in Kramer G. (Hrsg.): "Auditory Display", pp. 127-150, Addison-Wesley, 1994.

Wightman F. L., Kistler D. J. (1993): "Sound Localization", in Yost W. A., Popper A. N., Fay R. R. (Hrsg.): "Human Psychophysics", pp. 155-192, Springer Verlag New York, 1993.

Kapitel 2

Begault D. R. (1994): "3-D Sound for Virtual Reality and Multimedia", AP Professional, USA 1994.

Gardner B., Martin K. (1994): "HRTF Measurements of a KEMAR Dummy-Head Microphone", <http://sound.media.mit.edu/KEMAR.html>.

Wightman F. L., Kistler D. J. (1997): "Description of HRTF data set", Textdatei zu den HRTF-Daten.

Kapitel 3

- Beliczynski B., Kale I., Cain G. D. (1992): "Approximation of FIR by IIR Digital Filters: An Algorithm Based on Balanced Model Reduction", IEEE Transactions on Signal Processing, vol. 40, no. 3, pp. 532-542.
- Brandenstein H., Unbehauen R. (1998): "Least-Squares Approximation of FIR by IIR Digital Filters", IEEE Transactions on Signal Processing, vol. 46, no. 1, January 1998, pp 21-30.
- Chen J., Van Veen B. D., Hecox K. E. (1992): "External ear transfer function modeling: A beamforming approach", Journal of the Acoustical Society of America, 92(4), Pt. 1, pp. 1933-1944.
- Chen J., Van Veen B. D., Hecox K. E. (1995): "A spatial feature extraction and regularization model for the head-related transfer function", Journal of the Acoustical Society of America, 97(1), pp. 439-452.
- Gersho A., Gray R. M. (1992): "Vector Quantization and Signal Compression", Kluwer Academic Publishers, Boston/Dordrecht/London, 1992.
- Jot J.-M., Larcher V., Warusfel O. (1995): "Digital Signal Processing Issues in the Context of Minaural and Transaural Stereophony", Preprint 3980 of the 98th Audio Engineering Society Convention, Paris, Februar 1995.
- Kale I. (1993): "High Fidelity Digital Filter Design and Implementation Techniques", ERK '93, Portoro_
- Kistler D. J., F. L. Wightman (1992): "A model of head-related transfer functions based on principal components analysis and minimum phase reconstruction", Journal of the Acoustical Society of America, 91(3), pp. 1637-1647.
- Ludyk G. (1995): "Theoretische Regelungstechnik 1+2", Springer Verlag Berlin Heidelberg, 1995
- Mackenzie J., Huopaniemi J., Välimäki V., Kale I. (1997): "Low-Order Modeling of Head-Related Transfer Functions Using Balanced Model Truncation", IEEE Signal Processing Letters, vol. 4, no. 2, pp. 39-41.
- Marolt M. (1996): "A New Approach to HRTF Audio Spazialization", Proceedings of the International Computer Music Conference 1996, pp. 365-367, San Francisco, International Computer Music Association.
- Middlebrooks J. C., Green D. M. (1992): "Observations on a principal components analysis of head-related transfer functions", Journal of the Acoustical Society of America, 92(1), pp. 597-599.
- Oppenheim A. V., Shafer R. W.: "Discrete-Time Signal Processing", Prentice Hall, 1989.
- Wu Z., Chan F. H. Y., Lam F. K., Chan J. C. K. (1997): "A time domain binaural model based on spatial feature extraction for the head-related transfer function", Journal of the Acoustical Society of America, 102(4), pp. 2211-2218.