



Hochschule für Angewandte Wissenschaften Hamburg  
*Hamburg University of Applied Sciences*

# Seminararbeit

Jan Kuhr

3-Dimensionale Klangpositionierung mit  
Stereokopfhören im Kontext "Erweiterte Realität"

Jan Kuhr

3-Dimensionale Klangpositionierung mit  
Stereokopfhören im Kontext "Erweiterte Realität"

Seminararbeit  
eingereicht im Rahmen der Veranstaltung Seminar/Ringvorlesung  
im Studiengang Master of Science Informatik  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuer: Prof. Dr.-Ing. Bernd Schwarz  
Prüfer: Prof. Dr. rer. nat. Kai von Luck

Abgegeben am 28. Februar 2009

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>4</b>
<b>1 Einleitung und Zielsetzung</b>	<b>5</b>
<b>2 Grundlagen</b>	<b>6</b>
2.1 Der Mensch und Wahrnehmung der Realität . . . . .	6
2.2 Erweiterte Realität / Augmented Reality . . . . .	7
2.3 Realität in Computerspielen . . . . .	7
2.4 Akustische Wahrnehmung . . . . .	8
2.5 Audio-Signalverarbeitung . . . . .	8
2.6 Head Related Transfer Function . . . . .	9
2.7 Digitale Signalfilterung . . . . .	10
<b>3 Anforderungen an ein Beispielsystem</b>	<b>11</b>
3.1 Funktionale Anforderungen . . . . .	11
3.2 Anforderungen an Hard- und Software . . . . .	11
3.2.1 Audioqualität . . . . .	11
3.2.2 Audio Verarbeitungsgeschwindigkeit . . . . .	12
3.2.3 Minimierung von Signalverzögerung . . . . .	12
3.2.4 Positionserkennung und Umsetzung . . . . .	12
3.2.5 Steuerungsinterface . . . . .	12
3.2.6 Mobilität der Komponenten . . . . .	12
<b>4 Lösungsansatz</b>	<b>13</b>
4.1 Hardwareplattform . . . . .	13
4.2 Softwarekomponenten . . . . .	14
<b>5 Fazit und Ausblick</b>	<b>16</b>
<b>Literaturverzeichnis</b>	<b>17</b>

## Abbildungsverzeichnis

2.1	Kunstkopf: horizontale Ebene(1) und Medianebene(2) . . . . .	9
2.2	Differenzgleichung für die FIR Filterantwort . . . . .	10
2.3	FIR Filterstrukturen, links die kanonische Normalform 1 (Direct Form 1, DF1), rechts die kanonische Normalform 2 (Direct Form 2, DF2) . . . . .	10
4.1	Das kopfbezogene Koordinatensystem mit den beiden Winkeln Phi zur horizontalen Verdrehung des Kopfes (auch bezeichnet als Azimuth) und Delta zur Neigung des Kopfes auf der Medianachse (auch bezeichnet als Elevation). Der Abstand des Kopfes zur Klangquelle wird mit r definiert . . . . .	15

# 1 Einleitung und Zielsetzung

Diese Seminararbeit befasst sich mit dem Konzept und der Umsetzung einer in Echtzeit berechneten virtuellen Klangwelt. In einem definierten dreidimensionalen Raum rund um einen Zuhörer soll es zunächst möglich sein, gezielt Geräusche und Klänge zu platzieren. Der Zuhörer soll beim Wahrnehmen dieser Klänge über herkömmliche Stereo-Kopfhörer in der Lage sein die Herkunftsrichtung des Signals zu bestimmen, also die Klangquelle zu orten. Das Prinzip dieses Verfahrens der Audiosignalverarbeitung ist bekannt und weitreichend erforscht.

Die weiterführende Fragestellung mit der ich mich auseinandersetzen werde ist eine Erweiterung dieser Grundfunktionen um die Erhaltung der Position des virtuellen Klanges auch bei Bewegung des Zuhörers in einem dreidimensionalen Raum. Hierbei gilt es nicht nur die absolute Position des Zuhörers zu berücksichtigen, sondern vielmehr auch variablen wie die Verdrehung und Neigung des Kopfes, sowie die Entfernung zur Schallquelle.

Die Adaption der Positionsvariablen und die daraus resultierende Anpassung der Filterung des Klangmaterials soll in Echtzeit geschehen um beim Zuhörer einen möglichst realen Eindruck der virtuellen Klangposition zu erzeugen. Im Hinblick auf die damit einhergehende Mobilität des Hörers ist es darüber hinaus angedacht, die verwendeten Soft- und Hardwarekomponenten möglichst kompakt, transportabel und stromsparend auszuliegen.

Anwendungsgebiete des entwickelten Systems könnten beispielsweise Kopfhörer sein, die unabhängig von der aktuellen Kopfposition und dem Aufenthaltsort in einem Raum, die Schallabstrahlrichtung fest positionierter Lautsprecher simulieren. Dies ist unter anderem im Bereich der Tontechnik und Musikproduktion interessant. Ein weiteres Einsatzgebiet ist die Nutzung in Computerspielen die auf Mobilität des Spielers in seiner realen Umwelt setzen. Die so genannten *Pervasive-Games* erweitern die reale Welt um eine zeit- und ortsgleich existierende virtuelle Spielwelt und werden mit Erreichen einer bestimmten Qualität auf Echtzeit prozessierten 3-D Sound angewiesen sein.

In dieser Ausarbeitung möchte ich verschiedene Ansätze, Techniken und Verfahren zusammentragen und diskutieren, die mit der Entwicklung eines Systems zur virtuellen dreidimensionalen Klangpositionierung und dessen Einsatzgebieten und der kommerziellen Nutzbarkeit in Verbindung stehen.

Die Ergebnisse dieser Arbeit werde ich als Grundlage beim Beginn der Entwicklung dieses Systems im Rahmen des Master-Projektes verwenden. Außerdem dienen die hier gewonnenen Erkenntnisse der Abgrenzung des Themenschwerpunktes meiner Masterarbeit und zur Definition möglicher Unterthemen. Der Fokus liegt hierbei auf Analyse von Komplexität und Sinnhaftigkeit sowie der Identifizierung und Differenzierung einzelner Teil-Aufgaben.

## 2 Grundlagen

In diesem Kapitel möchte ich zunächst einige Sachverhalte, Gegebenheiten und technische Entwicklungen erläutern, welche die Grundlagen für mein Interesse und meine Arbeit in diesem Themengebiet bilden.

Zugleich soll hiermit eine kurze Einführung in das Thema der audiovisuellen Wahrnehmung des Menschen, ihrer Manipulierbarkeit und die damit verbundenen Möglichkeiten, Problemstellungen und Lösungsansätze erfolgen.

### 2.1 Der Mensch und Wahrnehmung der Realität

Audiovisuelle Wahrnehmung hat für die Entwicklung des Menschen und sein Leben im Alltag eine zentrale Bedeutung. Sie dient nicht mehr nur der reinen physikalischen Orientierung eines Organismus in seiner Umwelt, oder der Erkennung von Gefahren, sondern ist vielmehr auch die Grundlage unserer Fähigkeit zur Kommunikation. Durch visuelle und verbale Kommunikation ist es dem Menschen unter Anderem möglich in komplexen sozialen Strukturen zu Leben. Diese Umstände haben maßgeblichen Einfluss auf die Überlebensfähigkeit und die geistige Entwicklung im allgemeinen.

Alle Arten von Wahrnehmung beeinflussen zudem auch die Emotionen und Empfindungen des Menschen. Im Bereich des visuellen sei dies beispielhaft die Freude über das Sehen eines fruchttragenden Baumes bei großem Hunger oder die Angst die aufkommt wenn der Gesichtssinn bei Dunkelheit eingeschränkt ist. Im Bereich des auditiven kann man an dieser Stelle die Angst vor lautem Gebrüll wilder Tiere aufführen, oder die Freude die beim Genuss des lauten Gebrülls eines Rock-Musik Sängers entstehen kann.

Letztere Äußerung soll zugleich verdeutlichen, das sich bestimmte natürliche Reaktionen und Emotionen auf audiovisuelle Reize ändern bzw. weiterentwickeln können. Sowohl die genetische Veranlagung als auch die gesellschaftliche Prägung durch das Erlernen von Sprache und die fortwährende Konfrontation mit bekannten und unbekanntem audiovisuellen Reizen bestimmen einen Großteil der Wahrnehmung im Alltag des menschlichen Individuums.

Die Realität kann man beschreiben als die Wahrnehmung unserer Umgebung mit allen zur Verfügung stehenden Sinnen. Gemeint ist die physische Situation in der sich unser Körper befindet und nicht die Situation in der wir uns gedanklich oder gesellschaftlich gerade befinden können.

## 2.2 Erweiterte Realität / Augmented Reality

Die erweiterte Realität zeichnet sich dadurch aus, dass die Sinnesreize aus einer realen Umgebung mit Hilfe von Computern in Echtzeit ergänzt oder manipuliert werden bevor Sie vom Menschen wahrgenommen und vom Gehirn verarbeitet werden. Im Gegensatz zur virtuellen Realität, in der versucht wird eine möglichst vollständige Simulation einer physisch nicht-realen Umwelt über unsere Sinne an das Gehirn zu übermitteln, geht es in der erweiterten Realität vielmehr darum, die aktuellen Sinneseindrücke mit Zusatzinformationen anzureichern. Ein Anwendungsgebiet hierfür ist beispielsweise das visuelle Einblenden von Informationen über bestimmte Gegenstände die man sich gerade anschaut. Hierfür gibt es entsprechende Technologien wie etwa das *Head-Mounted-Display*.

## 2.3 Realität in Computerspielen

Computerspiele sind nicht mehr nur ein einfacher Zeitvertreib, sondern streben immer mehr nach der Schaffung einer vollständigen virtuellen Realität für den Spieler. Dieses Ziel zu erreichen ist trotz der technologischen Fortschritte und der Rechenleistung moderner Computer immer noch in weiter Ferne. Der Grad der Immersion<sup>1</sup> des Spielers hängt sowohl von der Konzeption bzw. Qualität der virtuellen Welt ab, als auch von der Art ihrer Präsentation.

Das Problem einer real wirkenden Präsentation ist hierbei der hemmende Faktor.

Der Standard von kommerziell erfolgreichen Computerspielen verwendet zur Zeit hauptsächlich audiovisuelle Präsentation, die mit den 2-dimensionalen Einschränkungen von Bildschirmen und Klंगाusgabegeräten auskommen muss. Ein weiteres großes Problem ist die gleichbleibende, bewegungseinschränkende Position des Spielers vor dem Bildschirm bzw. in einem geschlossenen Raum.

Um gerade diese Einschränkungen zu vermeiden gibt es neue Ansätze wie etwa *Location-Based-Games* oder *Pervasive-Games*, deren Ziel es ist, den realen Aufenthaltsort des Spielers zu berücksichtigen, ihm zu sagen wo es als nächstes hingehet und was er dort machen soll. Die virtuelle Welt, in der es bestimmte Ziele zu erreichen gilt, wird gleichzeitig mit der realen Welt durchwandert. Gegebenheiten aus der realen Welt werden dabei in der virtuellen Realität berücksichtigt.

Dies entspricht dem vorher genannten Prinzip der erweiterten Realität genau dann, wenn das Spiel bzw. die Software in der Lage ist dem Spieler in Echtzeit über seine Wahrnehmung einen veränderten Eindruck der Realität zu verschaffen. In der Regel sind hierzu die Methoden der Wahl ebenfalls auditive und visuelle Präsentation.

Die notwendigen Geräte und Techniken um eine derartige audiovisuelle Überblendung in Echtzeit zu gewährleisten befinden sich in verschiedener Form bereits prototypisch im Einsatz. Eingrenzende Faktoren bei der Qualität der Immersion sind die ungenügende Rechenleistung tragbarer Computer und der mangelnde Tragekomfort der Gerätschaften bei mobilen Lösungen.

---

<sup>1</sup>Immersion: Erfahrung eines Spielers, während er sich in einer virtuellen Welt befindet, Gefühl der Realität

## 2.4 Akustische Wahrnehmung

Auch wenn die Qualität einer virtuellen Welt meistens zuerst anhand des visuellen Eindrucks beurteilt wird, so ist die Notwendigkeit einer Ergänzung durch akustische Komponenten naheliegend.

Wenn es darum geht dreidimensional ausgedehnte virtuelle Welten zu erzeugen, was bei den meisten Computerspielen der Fall ist, hat die auditive Komponente einen großen Stellenwert.

Die Orientierung in der VR und die Interaktion mit Objekten basiert oftmals auch auf Geräuschen und Sprache, die wie in der realen Welt aus verschiedenen Richtungen kommen können. Wie gut die Positionsbestimmung der Objekte über Klang funktioniert hat maßgeblichen Einfluss auf die subjektiv wahrgenommene Qualität der virtuellen Realität.

In Spielen die beispielsweise von einem Spannungsbogen leben, der auf der Erzeugung von Emotionen wie Unbehagen und Angst basiert, ist dies besonders eindrucksvoll zu beobachten, weil die Geräusche darin absichtlich bestimmte Gefahrensituationen suggerieren und im Gegensatz zum passiven betrachten eines Filmes auch zur entsprechenden Reaktion des Spielers auffordern.

Der Grund für die starke emotionale Wirkung von Klängen und Geräuschen ist der natürliche Zweck des Gehörs als Schutz vor Gefahren aus der Umwelt. Das Sehen ist beim Menschen hierfür weniger geeignet, da es beispielsweise stark von Lichtverhältnissen beeinflusst wird. Hinzu kommt eine Einschränkung des Gesichtsfeldes des Menschen, also der sehbare Bereich der ohne die Verzögerung durch Bewegung der Augen und des Kopfes abgedeckt werden kann.

Auf der horizontalen Ebene können wir etwa einen Winkel von  $190^\circ$  überblicken, vertikal sind es nur etwa  $150^\circ$ . [Schober:1957]

Der durch das Hören abgedeckte Bereich hingegen ist wesentlich größer und umfasst horizontal und vertikal einen Winkel von annähernd  $360^\circ$  mit kleineren Einschränkungen durch Effekte wie den *Cone-Of-Confusion*. [Sima:2008] Ein weiterer natürlicher Vorteil des Hörens ist die permanente Verfügbarkeit beispielsweise auch während des Schlafens.

## 2.5 Audio-Signalverarbeitung

Die zu verarbeitenden Audiosignale werden zumeist in Form von herkömmlichen digitalen Audiodateien vorliegen. Es kann sich dabei um Klänge, Geräusche oder Sprache handeln die eine ausreichend gute Qualität besitzen um das Hörerlebnis nicht bereits von vornherein einzuschränken. Die Eingangsdaten können prinzipiell sowohl in Mono als auch Stereo vorliegen und eine beliebige zeitliche Länge aufweisen. Denkbar ist auch eine Verwendung von zur Laufzeit erzeugten synthetischen Klängen wie etwa die eines Software-Synthesizers.

Die Ausgangssignale können je nach Art der Implementierung unterschiedlich sein, müssen aber zu mindestens in Stereo vorliegen um den Effekt der räumlichen Wahrnehmung transportieren zu können. In der Regel erfolgt eine Ausgabe an den Kopfhörer bzw. Zuhörer über eine Digital/Analog Wandlung am Ausgang des Systems.

Die gewünschte transformation des Eingangssignals in eine mit räumlichen Informationen ange-reicherte Ausgabe geschieht mittels digitaler Signalfilterung. Eine solche Filterung ist im wesent-lichen eine Mathematische transformation der Schallwellen in ihrer digitalen Form. Dabei wird jedes der Eingangssamples<sup>2</sup> mit bestimmten, für den jeweiligen Zweck dedizierten Berechnun-gen in ein Ausgangssample überführt. Die Gesamtheit der gefilterten Samples ergibt eine neue Wellenform mit den gewünschten Charakteristika.

Für den in diesem Kontext benötigten Anwendungsfall der räumlichen Klangsimation gibt es die im folgenden definierte Funktion und Vorgehensweise.

### 2.6 Head Related Transfer Function

Die *Head-Related-Transfer-Function* oder auch *Kopfbezogene Transferfunktion* abgekürzt *HRTF* Ist eine Beschreibung des natürlichen, physikalischen Filterungsprozesses den das Menschliche Gehör auf eingehende Schallwellen anwendet. Die Filterung ist dabei abhängig von der Rich-tung der eingehenden Schallwellen und deren unterschiedlicher Beugung und Interferenzen mit Ohrmuscheln, Kopf und Schultern. Dabei wird die gesamte lineare Charakteristik des Klages in der *HRTF* abgebildet. Hierzu gehören auch die inteauralen Laufzeit- und Pegeldifferenzen. (*ILD und IPD*)

Die *HRTF* ist der richtungsabhängiger Frequenzgang des Ohres und die Fourier Transformation der Kopfbezogenen Impulsantwort (*engl. HRIR*). [Sima:2008]

Aufgrund dieser Richtungsabhängigkeit benötigt man für jeden Winkel an dem man ein Klan-gobjekt positionieren will eine eigene Variante dieser Transferfunktion. Dies gilt sowohl für die horizontale als auch für die vertikale Ebene. Eine *HRTF* für einen bestimmten Winkel kann man empirisch ermitteln.

Hierzu wird (in einem schalltoten Raum) ein-malig Schall aus der gewünschten Richtung auf einem Kunstkopf gerichtet und mit Mikro-fonen (innerhalb beider Ohren des Kunstkop-fes) aufgezeichnet.

Vergleicht man nun die Signale des Ur-sprungsklages mit den Aufzeichnungen der beiden Mikrofone, lässt sich daraus eine Im-pulsantwort (*HRIR*) für jedes Ohr und daraus wiederum die *HRTF* errechnen. Die gewon-nene Impulsantwort ist eine direkte Abbildung von Filterkoeffizienten die mit einem ent-sprechenden digitalen Filter eine Übertragung der *HRTF* auf beliebiges Klangmaterial ermögli-chen.

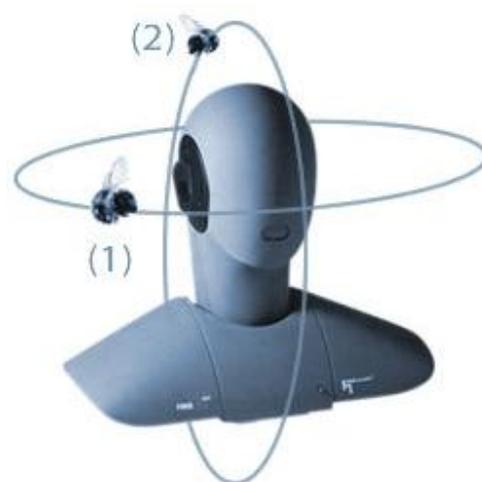


Abbildung 2.1: Kunstkopf: horizontale Ebene(1) und Medianebene(2)

<sup>2</sup>Sample: Im Bereich der Audiodatenverarbeitung ist ein Sample das kleinstmögliche Teilstück einer Wellenform.

## 2.7 Digitale Signalfilterung

Um die Filterwirkung der HRTF digital nachzubilden benötigt man ein geeignetes digitales Filter. Die einfache Struktur der beim *HRTF* Verfahren produzierten Filterkoeffizienten beziehungsweise die Tatsache das die *HRIR* eine endliche Impulsantwort ist legt die Verwendung eines *Finite-Impulse-Response* Filters (FIR) nahe.

Die allgemeine Beschreibung von FIR-Filtern ist im Zeitbereich durch die nachstehende Differentialgleichung gegeben, die jedes Eingangssample  $u_k$  mit einem speziellen Koeffizienten  $h(i)$  gewichtet.

$$y[k] = \sum_{i=0}^m h(i) \cdot u_{k-i}$$

Abbildung 2.2: Differenzgleichung für die FIR Filterantwort

Die Anzahl der zu speichernden vorherigen Eingangssamples  $u[k-i]$  ist die Filterordnung  $m$ , sodass insgesamt für jede Ausgangssampleaktualisierung  $L=m+1$  Produkte zu addieren sind. ( $L$  ist die Filterlänge und zugleich die Anzahl der verwendeten Filterkoeffizienten) [SchwarzUndReichardt:2007]

FIR-Filter lassen sich auf verschiedene Weise realisieren. Man spricht hierbei von Grundformen. Die Normalform 1 lässt sich effizienter mittels sequentiell abgearbeiteter Software implementieren (Native CPU's oder DSP's). In digitalen Schaltungen wie FPGA's in denen parallele Datenverarbeitung möglich ist, kann die Normalform 2 effizienter angewandt werden. Die beiden Normalformen sind äquivalent und können ineinander übergeführt werden.

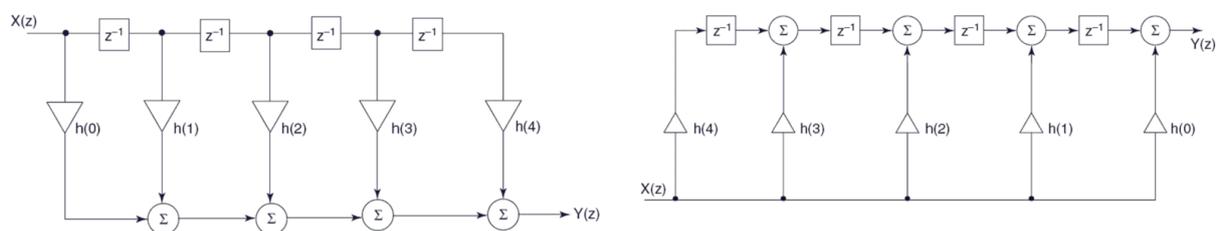


Abbildung 2.3: FIR Filterstrukturen, links die kanonische Normalform 1 (Direct Form 1, DF1), rechts die kanonische Normalform 2 (Direct Form 2, DF2)

## 3 Anforderungen an ein Beispielsystem

### 3.1 Funktionale Anforderungen

Ausgangspunkt für diese Seminararbeit und die daraus folgende Entwicklung eines Soft- und Hardwaresystems ist ein Konzept für die Gestaltbarkeit einer virtuellen Klangwelt durch einen Sound-Designer oder automatisch durch Software.

Diese Klangwelt soll dem Zuhörer über Stereo Kopfhörer dargebracht werden können und ihm ermöglichen die Position von eingespielten Klängen und Geräuschen so zu erleben wie es der Gestalter oder ein gestaltendes Softwaresystem (z.B. ein Computerspiel) vorgesehen hat.

Die Erlebnisqualität der Klangwelt soll außerdem gesteigert werden durch die Erhaltung der räumlichen Bewegungsfreiheit des Zuhörers.

Dies impliziert die kontinuierliche Anpassung des Ausgabesignals bei Kopf- und Körperbewegung in Echtzeit. Um den Echtzeiteindruck möglichst realistisch zu machen darf auch die Verzögerung (Latenz) vom Zeitpunkt des Aktivierens einer Soundquelle bis zu ihrer tatsächlichen Wahrnehmung durch den Hörer nicht zu groß sein. Dies gilt auch für die Latenz der Bewegungskompensation. Es ergeben sich die folgenden funktionalen Anforderungen.

### 3.2 Anforderungen an Hard- und Software

Die Komponenten des zu entwickelnden Hard- und Softwaresystems (im folgenden *System* genannt) sollen in der Lage sein die gewünschte Funktionalität zu bieten. Hierzu müssen sie in diesem Projekt bestimmten Leistungs-, und Mobilitätsanforderungen genügen.

#### 3.2.1 Audioqualität

Die Qualität des Ein- und Ausgangssignals sollte den Anforderungen des Menschlichen Gehörs entsprechen. In der Praxis bedeutet dies eine minimale Abtastfrequenz<sup>3</sup> von 44.1KHz bei Wandlung der analogen in digitalen Signale und zurück. Dieser Qualitätsanspruch resultiert direkt in einen Leistungsanspruch an das System. Es muss in der Lage sein, Audiosamples mit der genannten Abtastfrequenz zu verarbeiten.

---

<sup>3</sup>Die Abtastfrequenz oder auch Samplerate definiert die Anzahl von Audiosamples die pro Sekunde von einem DSP-System verarbeitet werden können.

#### 3.2.2 Audio Verarbeitungsgeschwindigkeit

Die Verarbeitung von Audiosamples im Kilohertz-Bereich ist prinzipiell auf jedem der heutigen Prozessoren im Megahertz-Bereich möglich. Wird allerdings der Algorithmus zur Transformation von Eingangs- in Ausgangssignal zu aufwändig, kann es passieren dass der Prozessor mit der Berechnung eines Samples zu lange braucht. In Folge dessen werden Samples die vom A/D Wandler kommen nicht mehr schnell genug abgerufen und gehen verloren. Dies ist aufgrund der Echtzeitanforderung an das System nicht erlaubt. Die FIR-Signalfilterung anhand der *HRTF*, welche in diesem System zu Einsatz kommen soll, ist nur mit Prozessoren im Bereich einiger hundert Megahertz in Echtzeit möglich.

#### 3.2.3 Minimierung von Signalverzögerung

Eine weitere Echtzeit-Anforderung an das System ist die Vermeidung von Verzögerungen der Ausgangssignale. Solche Verzögerungen entstehen in der Regel an allen Komponenten die in den Signalfluss eingreifen. Einfluss auf die Signalverzögerung haben beispielsweise lange Kabel, A/D und D/A Wandler, Datenpufferung bei der Ein- und Ausgabe in ein Mikroprozessorsystem und vor allem der digitale Filterungsprozess. In dieser Anwendung darf die Verzögerung maximal etwa *10 bis 20 Millisekunden* betragen. Ist sie größer kann sie vom Hörer als störend empfunden werden.

#### 3.2.4 Positionserkennung und Umsetzung

Die Erfassung und Verarbeitung der Kopf- und Körperbewegung des Zuhörers relativ zum virtuellen Raum ist ebenfalls eine essentielle Aufgabe des Systems. Es müssen Positions- und Bewegungsdaten in der realen Umgebung gesammelt, und mit dem Koordinatensystem der virtuellen Klangwelt abgestimmt werden. Die Filterung jedes einzelnen Audiosamples ist dabei abhängig von der Position der Schallquelle und der Bewegung des Hörers. Das zu implementierende Filter sollte also eine schnelle Änderung von Filterkoeffizienten zulassen.

#### 3.2.5 Steuerungsinterface

Zur Steuerung des Systems durch Anwender und Entwickler wird ein Programminterface benötigt. Über dieses Interface sollte es möglich sein Audiodaten an das System zu übergeben sowie Steuersignale zur Einstellung der gewünschten Position einer Schallquelle. Dieses Interface kann bereits in der Entwicklungsphase zum Testen des Systems verwendet werden.

#### 3.2.6 Mobilität der Komponenten

Die Bewegung des Hörers in einer abgegrenzten oder freien Umgebung sollte möglich sein. Aus diesem Grund muss bei der Entwicklung des Systems auf die Mobilität der Komponenten geachtet werden. Hierzu zählen in erster Linie die Größe und das Gewicht, aber soll sich der Hörer auch autonom ohne Stromkabel bewegen können ist die Stromversorgung des Systems ebenfalls einzubeziehen.

## 4 Lösungsansatz

Als Ausgangsbasis für die Entwicklung des digitalen FIR-Filters zur Berechnung der *HRTF* für den linken und rechten Kanal dient eine funktionsfähigen *MATLAB* Referenzimplementierung der HAW Bachelorabsolventin Sylvia Sima. [Sima:08]

Diese Referenzimplementierung beinhaltet Messungen der *HRTF* für verschiedene Winkel auf der horizontalen und der Medianebene und stellt die zur Signalfilterung notwendigen Koeffizientensätze zur Verfügung.

Die Anzahl der Koeffizienten pro Filterung und damit die Filterlänge hat maßgeblichen Einfluss auf die Qualität des Ausgegebenen Audiomaterials. Die Filterlänge in der *MATLAB* Implementierung beträgt **512** Koeffizienten pro Kanal und ist damit ausreichend für diesen Anwendungsfall. Es muss noch erwähnt werden das es sich bei der *MATLAB* Version um eine nicht Echtzeitfähige Offline-Implementierung handelt. Die Filterung findet direkt auf digitalen Audiodaten statt und resultiert in Ausgabedateien im *WAVE*<sup>4</sup> Format.

Die Umsetzung des Projektes gliedert sich in die Zusammenstellung der benötigten Hardwarekomponenten und die Partitionierung der einzelnen Softwaremodule wenn verschiedene Programmiersprachen auf unterschiedlicher Hardware Verwendung finden.

Ist eine solche Partitionierung notwendig, müssen auch die Schnittstellen zwischen den Softwaremodulen identifiziert und definiert werden.

### 4.1 Hardwareplattform

Die Leistungsanforderungen an das zu entwickelnde System werden in erster Linie durch Ansprüche an die digitale Signalverarbeitung in Echtzeit geprägt. [siehe Matthies:2008] Die Filterung von zwei Audiokanälen gleichzeitig mit einer Samplerate von 44.1KHz und die Ansprüche an die Qualität der Filterung mit **512** Koeffizienten pro Kanal legen eine Verwendung eines dedizierten Hardwaremoduls nahe. Dies ist auch im Hinblick auf die gewünschten Mobilität des Systems empfehlenswert.

Da das System zusätzlich auf native Software zur Steuerung der Filteradaption bei Bewegung des Hörers und auf Schnittstellen für externen Zugriff angewiesen ist benötigt man ein Hybrides Soft- und Hardware Design, im allgemeinen bezeichnet als *Hardware/Software Co-Design*.

Als Entwicklungsplattform bietet sich ein an der *HAW* verfügbares Embedded-System Evaluierungsboard der Firma *Xilinx* mit der Bezeichnung *Virtex2Pro* an. Diese Plattform bietet die Möglichkeit eines Co-Designs von parallelisierbaren *FPGA*<sup>5</sup> Modulen und nativ ausgeführter Software und deren direkte Kommunikation miteinander. Native Software wird dabei in einem

---

<sup>4</sup>Waveform Audio Format, Ein standard Dateiformat von Microsoft und IBM um Audiodatenströme zu speichern

<sup>5</sup>FPGA: Field Programmable Gate Array

Softprozessor<sup>6</sup> der *MicroBlaze* Reihe ausgeführt

Es orientiert sich dabei am Konzept und den Möglichkeiten des *System-On-a-Chip* (SoC) Designs. Ein weiterer Vorteil dieser Hardware ist die direkte Integration von Audio Ein- und Ausgabe mittels einer Audio-Codec Erweiterung und Integrierten A/D D/A Wandlern. Kommunikation mit dem laufenden System kann über verschiedene Standardschnittstellen wie, UARTs, LAN, USB und on-board Komponenten wie einem LCD Display und Hardwareschaltern realisiert werden.

### 4.2 Softwarekomponenten

Die Zusammenstellung und Programmierung der *FPGA* Module und die Entwicklung der Software für das *Xilinx* System erfolgt mit einer speziellen Entwicklungsumgebung dem *Embedded-Development-Kit* (EDK)

*FPGA* Module können in der Hardwarebeschreibungssprache *VHDL* programmiert werden und die nativ ausgeführte Software komfortabel in der Hochsprache *C*. Für die Einbindung der On-Board Hardware stehen bereits vorgefertigte *VHDL* Beschreibungen zur Verfügung. Deren Konfiguration ist teilweise über eine grafische Oberfläche möglich. Wie komfortabel die Entwicklung des Systems in dieser Umgebung wirklich ist wird sich im Laufe des Master-Projektes herausstellen wenn mit der Implementierung begonnen wird.

Um die Vorteile der parallelisierten Berechnung des FIR-Filters voll auszuschöpfen muss dieser als *FPGA*-Modul direkt in *VHDL* implementiert werden. Es gibt bereits eine Vielzahl solcher Hardware FIR-Filter und es muss geprüft werden ob die Verwendung einer bereits vorhandenen Lösung bei diesem Anwendungsfall in Frage kommt. Ein vorgefertigtes Filtermodul wäre eine erhebliche Erleichterung bei der Entwicklung dieses Systems.

Die Entwicklung der Software für den Mikroprozessor wird in diesem Projekt den größten Teil der Zeit beanspruchen. Es gilt zunächst Schnittstellen der einzelnen Hardwarekomponenten einzubinden und die Ein- und Ausgabe von Audiodaten sowie die Kommunikation mit dem Benutzer zu realisieren. Der Benutzer ist hierbei in erster Linie der Entwickler selbst, was dazu führt dass die Erstellung von Test- und Debuggingroutinen und deren Nutzung über die On-Board Schnittstellen Priorität haben.

Wenn die Signalfilterung erfolgreich implementiert und verifiziert worden ist und eine erste einfache Positionierung von Klang im Raum möglich ist, kann mit der Entwicklung der Positionsbestimmungskomponente begonnen werden. Hierbei ist es zunächst erforderlich Sensoren an das System bzw. den Benutzer zu binden, die in der Lage sind die Kopf- und Körperposition des Zuhörers im Raum zu bestimmen. Diese Aufgabe muss im Vorfeld gut eingegrenzt werden, da sie eher im komplexen Themengebiet der In und Outdoor Navigation angesiedelt ist.

Ich gehe davon aus die Positionsbestimmung des Hörers im Raum nicht mit einbeziehen zu können. Die Erfassung der Drehbewegung und der Neigung des Kopfes ist mein primäres Ziel, was für eine anschauliche Demonstration der Entwicklungsergebnisse zunächst völlig ausreicht.

---

<sup>6</sup>Softprozessor: Ein Mikroprozessor bzw. Mikrocontroller der selbst ein programmiertes Softwaremodul innerhalb der *FPGA* Logik ist und damit flexibel konfiguriert werden kann.

#### 4 Lösungsansatz

Um die Bewegung des Kopfes zu erfassen benötigt man ein 3-Achsen Kompassmodul, das am obersten zentralen Punkt des Kopfhörerbügels befestigt werden kann. Nach der Kalibrierung eines solchen Kompassmoduls kann man Daten über die verschiedenen Dreh- und Neigungswinkel des Kopfes auslesen und damit die Filterung der Audiosignale steuern.

Die Position der Soundquelle relativ zur Sehachse<sup>7</sup> wird in den gleichen Winkeln angegeben die auch bei der Messung der *HRTF* verwendet werden. Grundlage ist hier das kopfbezogene Koordinatensystem.

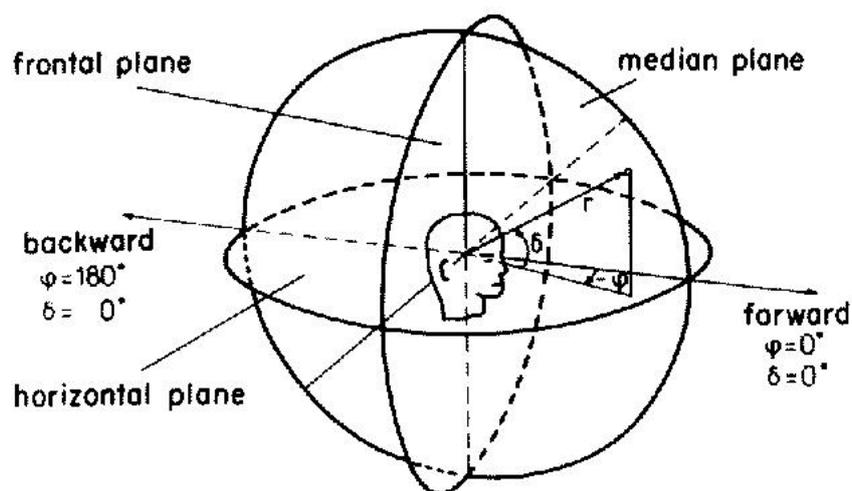


Abbildung 4.1: Das kopfbezogene Koordinatensystem mit den beiden Winkeln Phi zur horizontalen Verdrehung des Kopfes (auch bezeichnet als Azimuth) und Delta zur Neigung des Kopfes auf der Medianachse (auch bezeichnet als Elevation). Der Abstand des Kopfes zur Klangquelle wird mit  $r$  definiert

<sup>7</sup>Sehachse: Gerade die zwischen den Augen entspringt und vom Gesicht weg nach vorne zeigt

## 5 Fazit und Ausblick

Ziel der Seminararbeit war es zunächst, einen Überblick über aktuelle Techniken und Verfahren zur Audiosignalverarbeitung im Bezug auf die Idee einer virtuellen Klangwelt zu bekommen.

Dabei stand das Verständnis der Komplexität des Themengebietes im Vordergrund um die Umsetzbarkeit der erwünschten Funktionalität in ein zusammenhängendes System und damit den erfolgreichen Abschluss des Master-Projektes und der Masterarbeit zu gewährleisten.

Ich sehe mich nun in der Lage das zu Entwickelnde System in konkreten Einzelteilen zu betrachten und dadurch den Umfang von notwendigen Entwicklungsschritten, deren Zeitliche Dauer und Abfolge besser einzuschätzen. Ich habe außerdem gelernt eine Priorisierung der Entwicklungsschritte vorzunehmen und wichtiges von unwichtigem zu separieren.

Während des Ausarbeitungsprozesses wurde nicht nur Wissen angesammelt und gefestigt, sondern auch neue Ideen generiert und weiterführende Fragestellungen aufgeworfen, die in der einen oder anderen Weise ihren Weg in die Masterarbeit finden werden.

Mein ursprünglicher Enthusiasmus in diesem Themengebiet aktiv zu werden hat sich bis heute gehalten und sogar noch verstärkt.

Meine Motivation das beschriebene System bis zu seiner vollständigen Reife zu führen ist hoch. Sogar über den Horizont der Masterarbeit hinaus.

Die in diesem Dokument angedeutete Nutzbarkeit des Systems für mobile Computerspiele fasziniert mich dabei am meisten. Sollten die hier vorgeschlagenen Neuentwicklungen auch zur mobilen Verwendung herangereift sein, ergeben sich dadurch gute Chancen für die Steigerung der Immersion neuartiger Computerspiele.

Da in diesem Geschäftsbereich trotz aktueller wirtschaftlicher Resignation immer noch finanzielle Gewinne generiert werden können ist die Weiterführung des Projektes nach dem Studium für mich von großem Interesse.

Dieses Interesse deckt sich außerdem mit meinem beruflichen Background in der Audiosoftware Branche und meinen persönlichen Interesse an Audiosignalverarbeitung, Klangerzeugung und Musikinstrumenten.

## Literaturverzeichnis

### **Jürgen Reichardt 2007**

JÜRGEN REICHARDT, und Bernd S. (Hrsg.): *VHDL-Synthese Entwurf digitaler Schaltungen und Systeme*. 4. Aufl. München : Oldenbourg Wissenschaftsverlag, 2007. – ISBN 978–3–486–58192–8

### **Matthies 2008**

MATTHIES, Jörn (Hrsg.): *Aktive Geräuschkompensation mit einer FPGA-basierten Signalverarbeitungsplattform*. Hamburg : Masterarbeit an der HAW-Hamburg, 2008

### **Schober 1957**

SCHOBER, H. (Hrsg.): *Das Sehen, Band I*. Leipzig : Fachbuchverlag Leipzig, 1957

### **Sima 2008**

SIMA, Sylvia (Hrsg.): *HRTF Measurements and Filter Design for a Headphone-Based 3D-Audio System*. Hamburg : Bachelorarbeit an der HAW-Hamburg, 2008

### **Zoelzer 1996**

ZOELZER, Udo (Hrsg.): *Digitale Audiosignalverarbeitung*. 1. Aufl. Stuttgart : Teubner, 1996. – ISBN 3–519–06180–5

### **Zoelzer 2002**

ZOELZER, Udo (Hrsg.): *DAFX: Digital Audio Effects*. 1. Aufl. West Sussex, England : Wiley and Sons, 2002. – ISBN 978–0471490784