

Abschlussarbeit zum TI-Projekt

Binauralsignale zum Tracking von Position und Orientierung in mobilen Augmented Reality Audio Systemen

Nadine Kroher

nadine.kroher@gmx.de

Graz, 15. Oktober 2010

Betreuer: Univ.Prof. Dr.phil. Gerhard Eckel



UNIVERSITÄT
FÜR MUSIK UND
DARSTELLENDEN KUNST
GRAZ - AUSTRIA



Institut für elektronische Musik und Akustik

Inhalt

1. Einleitung	3
2. Mobile Augmented Reality Audio Systeme	3
2.1. Funktionsweise	3
2.2. Anwendungsgebiete	4
3. Binauralsignale zum Tracking von Position und Orientierung	5
3.1. Hintergrund	5
3.2. Vor- und Nachteile gegenüber anderen Tracking-Verfahren	5
3.3. Grundlegende Funktionsweise	6
3.4. Bestimmung der Orientierung	6
3.5. Bestimmung der Position	7
3.6. Verwendung mehrerer Referenzlautsprecher	8
3.7. Referenzsignale und Korrelation	8
3.8. Maskierung des Referenzsignals	10
3.9. Begrenzung der Bandbreite	10
4. Messungen	10
4.1. Messaufbau	10
4.2. Software	11
4.3. Kalibrierung	12
5. Messergebnisse	13
6. Diskussion und Ausblick	21
7. Referenzsignale in praktischen Anwendungen	21

1. Einleitung

Audio Augmented Reality Systeme schaffen eine für den User nicht differenzierbare Synthese aus tatsächlich in der Umgebung vorhandenen und künstlich erzeugten Schallquellen. Letztere werden durch die Faltung mit individuellen HRTFs virtuell im Umfeld des Users positioniert. Das resultierende System besteht aus Kopfhörern mit eingebauten, nach außen gerichteten, Mikrofonen und einem Mixer, welcher das Verhältnis der direkt von den Mikrofonen zum Kopfhörer durchgeschalteten Außengeräusche zu den virtuellen Schallquellen regelt.

Für die Binauralsynthese des künstlich erzeugten Schallfeldes ist ein genaues Tracking von Position und Orientierung des Users notwendig. Eine akustische Lokalisation unter der Verwendung von Binauralsignalen in Kombination mit Referenzschallquellen bietet hierbei den großen Vorteil, dass die benötigten In-Ear-Mikrofone bereits Teil des Systems sind und zum anderen die Tracking-Daten direkt beim User für die weitere Verarbeitung vorliegen.

Die Projektarbeit umfasst den Aufbau eines solchen Tracking-Systems, die Messung von Position und Orientierung mittels verschiedener Ankersignale und den Einfluss von Umgebungsgeräuschen auf die Genauigkeit der Messergebnisse. Zudem wird untersucht, wie sich das Vorhandensein der Referenzschallquelle auf die künstlerische Produktion von Inhalten für Audio Augmented Reality Systeme auswirkt.

2. Mobile Augmented Reality Audio Systeme

Funktionsweise

In „Mobile Augmented Reality Audio“ (MARA) Systemen entsteht eine Erweiterung des real existenten Schallfeldes mit virtuellen Schallquellen. In einem aus In-Ear-Kopfhörern mit eingebauten Mikrofonen bestehenden Headset werden die Umgebungsgeräusche mit den Mikrofonen abgenommen und direkt über die Kopfhörer wiedergegeben. Da der so entstehende Höreindruck von der Wahrnehmung ohne Headset aufgrund der Signalkette Mikrophon-Kopfhörer abweichen kann, wird dieser Teil des Systems als pseudo-akustische Umgebung bezeichnet. Die virtuellen Schallquellen werden mit Hilfe von Head Related Transfer Functions (HRTFs) mittels Binauralsynthese virtuell in der Umgebung des Users platziert. Ein Mixer

regelt hierbei das Verhältnis zwischen der pseudo-akustischen und der virtuellen Umgebung. Ziel ist es, dass der User nicht zwischen realen und virtuellen Quellen unterscheiden kann.

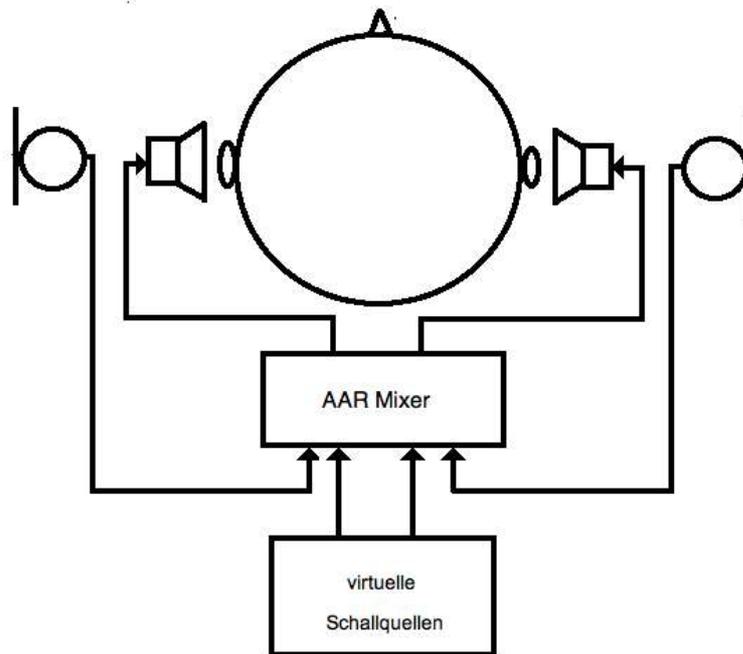


Abb. 2.1.: Schematischer Aufbau eines MARA-Systems

Anwendungsgebiete

MARA-Systeme haben in den vergangenen Jahren sowohl in der Telekommunikation und in Informationssystemen, als auch in der Klangkunst Anwendung gefunden.

Ein grundlegendes Einsatzgebiet ist die binaurale Telefonie, bei der sowohl die Stimme des Gesprächspartners, als auch dessen Umgebungsgeräusche übertragen werden. Zudem bieten einige Systeme die Möglichkeit, die Stimme des Gesprächspartners als virtuelle Schallquelle in der Umgebung zu platzieren und somit die Im-Kopf-Lokalisation zu vermeiden und bei Telefonkonferenzen eine bessere Differenzierung der Gesprächsteilnehmer zu erreichen. Projekte wie „Hear&There“ [7] oder „Kamara“ [8] verwenden das Prinzip der Auditory Post-Its. Der User kann lokal gebundene auditive Nachrichten an einem bestimmten Ort hinterlassen, welche dann später von ihm selbst oder von anderen Usern am selben Ort abgehört werden können. Auf dem gleichen Prinzip basierend wurde im Projekt „Guided by Voices“ [6]

ein rein auditives und interaktives Fantasy-Spiel entwickelt. Automated Tour Guides [5] bieten u.a. in Museen den Besuchern die Möglichkeit, begleitende Beschreibungen in Form von Audio-Files interaktiv aufzurufen, indem sie sich dem entsprechenden Ausstellungsobjekt nähern. Die Implementierung der pseudo-akustischen Umgebung ermöglicht zudem gleichzeitig eine ungestörte Kommunikation mit anderen Besuchern. Im Bereich der Klangkunst wurde im Projekt „LISTEN“ [9] eine Klanginstallation realisiert, in der virtuelle Schallquellen mittels Binauralsynthese implementiert wurden. Zudem hat der Besucher die Möglichkeit abhängig von seiner Position im Raum einzelne Parameter einer algorithmischen Komposition zu beeinflussen.

Weitere Ansätze aus dem Bereich der Telekommunikation versuchen durch die Einbindung der Umgebungsgeräusche des Telefonierenden eine sichere Navigation im Strassenverkehr zu ermöglichen.

3. Binauralsignale zum Tracking von Position und Orientierung

Hintergrund

Die Binauralsynthese, die virtuelle Positionierung von Schallquellen in der Umgebung des Users, basiert auf der Filterung des Signals mit der vom Einfallswinkel abhängigen Übertragungsfunktion des Außenohrs (Head Related Transfer Funktion, HRTF). Durch dieses Verfahren werden die Schallquellen unabhängig von der Kopfbewegung des Users immer aus der gleichen Richtung geortet. Eine Simulation der Entfernung der Schallquelle geschieht meist durch Anpassung der Lautstärke und Hinzufügen von Reflexionen. Zur Realisierung der Binauralsynthese werden also die Ortskoordinaten und die Orientierung des Users benötigt. Im betrachteten Verfahren werden diese Daten direkt aus den an den InEar-Mikrofonen vorliegenden Signalen gewonnen.

Vor- und Nachteile gegenüber anderen Tracking-Verfahren

Der wohl größte Vorteil bei der Verwendung von Binauralsignalen für die Bestimmung von Position und Orientierung liegt in der Tatsache dass die InEar-Mikrofone bereits Teil des Systems sind und somit die Mobilität nicht unnötig einschränken. Zudem liegen im Vergleich zu anderen Verfahren, wie z.B. bei elektromagnetischen Trackern, die Daten direkt am User vor, wo sie für die Synthese benötigt werden. Die Messungen aus [3] zeigen außerdem sehr genaue und robuste Ergebnisse.

Ein Nachteil des Systems ist jedoch die geringe Reichweite, welche durch die Amplitudendämpfung und die Absorption durch Objekte im Umfeld bei der Ausbreitung des Schalls begrenzt ist. Tracking-Verfahren mit größerer Reichweite, wie z.B. GPS, liefern jedoch meist weniger genaue Ergebnisse und können in der Regel die Position, nicht aber die Orientierung, bestimmen. Ein kritischer Aspekt ist zudem das unter Umständen als störend empfundene Vorhandensein von Referenzschallquellen.

Grundlegende Funktionsweise

Prinzipiell beruht das Tracking der Position auf der Messung der Laufzeit des Schalls von einer Referenzschallquelle zu den Mikrofonen. Hieraus kann der Abstand des Users zum Lautsprecher bestimmt werden. Die Orientierung wird aus der Laufzeitdifferenz (Interaural Time Difference, ITD) zwischen den an den beiden Mikrofonen ankommenden Referenzsignalen ermittelt. Grundsätzlich werden hier 3 verschiedene Setups unterschieden:

1. Synchrone Messung: Die Referenzsignale und ihre Beginnzeiten sowie die Positionen der Referenzschallquellen sind bekannt. In diesem Fall ist eine exakte Bestimmung von Position und Orientierung möglich.
2. Asynchrone Messung: Im Gegensatz zum ersten Fall sind die Beginnzeiten der Referenzsignale nicht bekannt. Es ist eine Bestimmung der Orientierung und der relativen Änderung der Position möglich.
3. „Blinde“ Messung: Als Referenzsignale dienen Umgebungsgeräusche, von denen weder die Position, noch die Signalform bekannt ist. Es kann die Orientierung geschätzt werden.

Die folgenden Ausführungen beziehen sich auf den ersten Fall. Zu einem bekannten Zeitpunkt wird das Referenzsignal über einen Lautsprecher reproduziert und gleichzeitig die Signale an den Mikrofonen aufgenommen.

Bestimmung der Orientierung

Zur Bestimmung der Orientierung wird zunächst jeweils die Kreuzkorrelation eines Mikrofonsignals ($x_l(t)$ und $x_r(t)$) mit dem Referenzsignal ($Ref(t)$) bestimmt.

$$(3.1) \quad R_{l, Ref}(T) = \int x_l(t) * Ref(t+T) dt$$

Kreuzkorrelation zwischen linkem Eingangssignal und Referenzsignal

$$(3.2) \quad R_{r,Ref}(T) = \int x_r(t) * Ref(t+T) dt$$

Kreuzkorrelation zwischen rechtem Eingangssignal und Referenzsignal

Die ITD ergibt sich aus der Differenz der Indexwerte der Peaks beider Korrelationsfunktionen.

$$(3.3) \quad ITD = \maxarg(R_{l,Ref}) - \maxarg(R_{r,Ref})$$

ITD aus den Kreuzkorrelationsfunktionen

Mit Hilfe der Schallgeschwindigkeit c kann aus der ITD die Wegdifferenz bestimmt werden.

$$(3.4) \quad dx = c * ITD$$

Wegdifferenz aus ITD

Mit geometrischen Mitteln ergibt sich an Hand des ITD-Modells und der Vereinfachung des parallelen Schalleinfalls aus dx und dem Kopfradius r der Winkel Φ der Orientierung.

$$(3.5) \quad \Phi = \arcsin\left(\frac{dx}{2 * r}\right)$$

Orientierungswinkel aus der Wegdifferenz

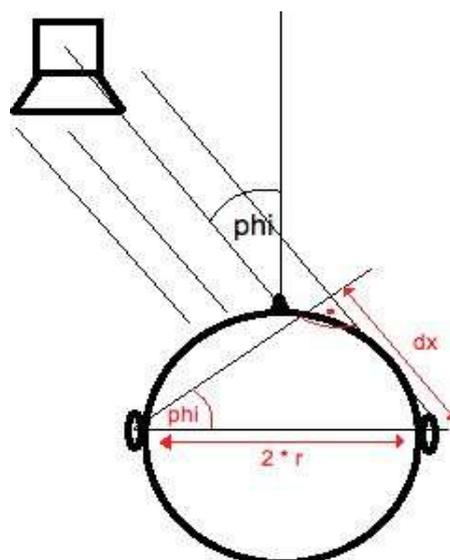


Abb.3.1.: Winkelberechnung im ITD-Modell

Bestimmung der Position

Die Position ergibt sich aus der Laufzeit des Schalls vom Lautsprecher zu den Mikrofonen. Aus den zuvor ermittelten Zeitpunkten der Peaks der Kreuzkorrelationen wird der Mittelwert gebildet, welcher der mittleren Verzögerung der Signale an den Mikrofonen gegenüber dem Zeitpunkt des Abspielens ab Lautsprecher entspricht, also der Schalllaufzeit.

$$(3.6) \quad T = 0.5 * (\maxarg(R_{l,Ref}) + \maxarg(R_{r,Ref}))$$

Schalllaufzeit vom Lautsprecher zu den Mikrofonen

Mit Hilfe der Schallgeschwindigkeit c kann die mittlere Distanz zwischen den Mikrofonen und dem Lautsprecher berechnet werden.

$$(3.6) \quad d = c * T$$

Distanz zwischen Mikrofonen und Lautsprecher

Verwendung mehrerer Referenzlautsprecher

Mit der oben beschriebenen Methode kann unter der Verwendung eines Referenzlautsprechers (Anker) lediglich der Abstand zu diesem berechnet werden. Aus der Distanz zu zwei Ankern kann die Position auf einer durch die Verbindungslinie der beiden Lautsprecher getrennten Halbebene bestimmt werden. Drei Anker erlauben die Positionsbestimmung innerhalb der von den Lautsprechern aufgespannten Fläche. Zudem kann durch die Verwendung mehrerer Anker und die Mittelung der gewonnenen Orientierungswinkel eine robustere Positionsbestimmung gewährleistet werden.

Referenzsignale und Korrelation

Das zuvor beschriebene Prinzip beruht auf der Tatsache, dass die Kreuzkorrelation einen deutlichen Peak an der Stelle der Zeitverzögerung durch den Schallweg aufweist. Dies setzt voraus, dass das verwendete Signal stark unkorreliert sein muss bzw. möglichst wenig periodische Anteile aufweist. Ideal ist hierfür weißes Rauschen geeignet, da dessen Autokorrelation nur einen Spitzenwert, im Idealfall einen Dirac-Puls, an der Stelle $t=0$ aufweist. Im Idealfall ohne Störgeräusche verschiebt sich dieser Peak in der Kreuzkorrelation zwischen Mikrofon- und Referenzsignal um die Dauer des Schallwegs, wobei alle anderen Werte der Funktion null sind.

Bei Signalen mit ausgeprägten periodischen Anteilen, wie zum Beispiel bei Musiksignalen, zeigen sich in der Autokorrelation mehrere ggf. sogar gleich

starke Peaks. Ist der durch die Zeitverzögerung entstandene Peak weniger stark ausgeprägt, kann er nicht mehr als Spitzenwert detektiert werden.

Um dieses Prinzip zu verdeutlichen, zeigt die folgende Grafik jeweils für Musik und Rauschen die Kreuzkorrelationen zwischen einem Referenzsignal und einem Ausschnitt aus diesem. Bei dem unkorrelierten Rauschsignal ist anhand des Peaks der Kreuzkorrelation deutlich der Ort des Signalausschnitts zu erkennen. Das Musiksignal hingegen zeigt mehrere stark ausgeprägte Peaks.

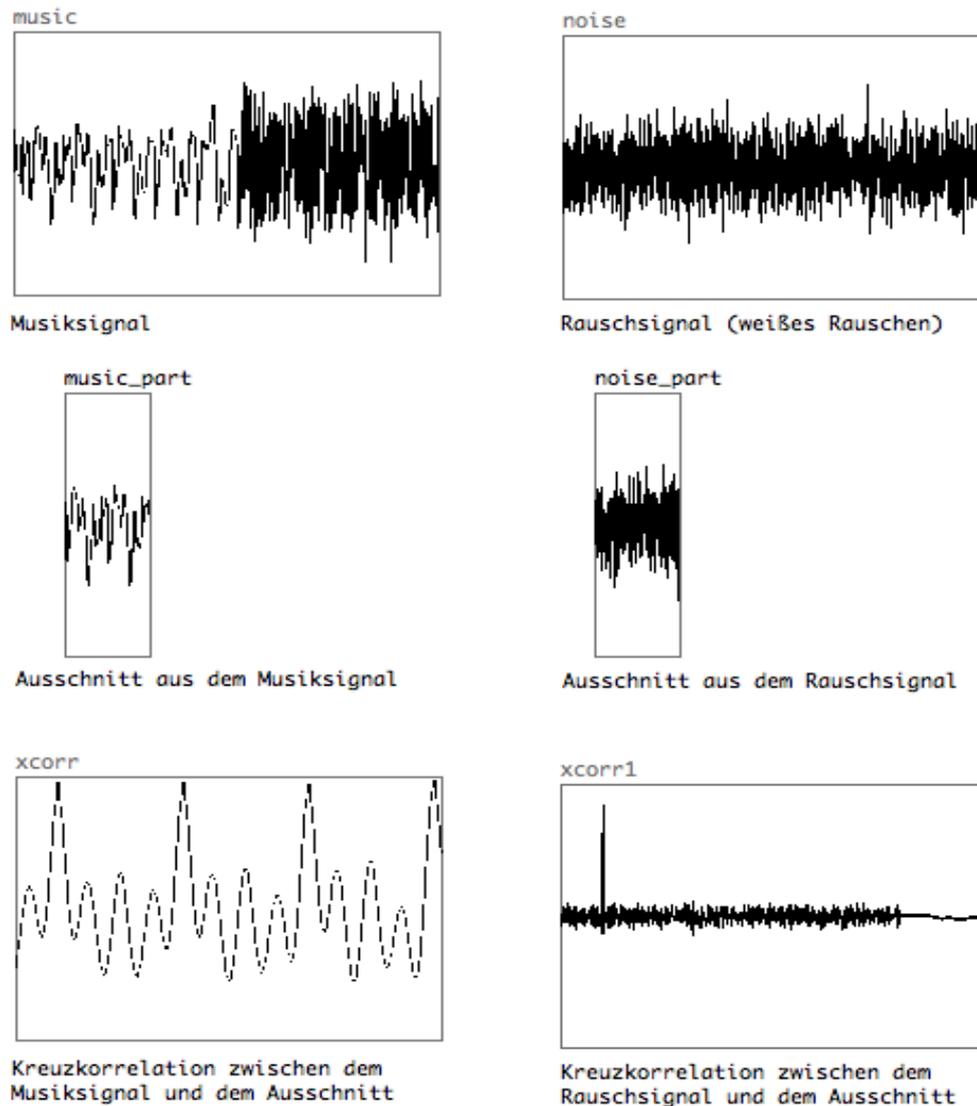


Abb.3.1.: Vergleich Kreuzkorrelationen Musik und Rauschen

Bei der Verwendung mehrerer Anker tritt die Problematik auf, die Referenzsignale der verschiedenen Anker an den Mikrofonen zu unterscheiden. Eine Möglichkeit besteht in der Verwendung von Bandrauschen in unterschiedlichen Frequenzbändern. Die Mikrofonsignale

können anschließend entsprechend gefiltert und im jeweiligen Frequenzbereich des Signals für den dazugehörigen Anker ausgewertet werden. Ein anderer Ansatz basiert auf der Verwendung von frequenzmodulierten Signalen. Das Referenzsignal moduliert für jeden Anker eine eigene Frequenz. Die Trennung an den Mikrofonen erfolgt über entsprechende Filterung des Gesamtsignals und anschließende Demodulation der Teilbänder.

Maskierung des Referenzsignals

Prinzipiell kann das Rauschsignal von anderen Geräuschquellen, wie z.B. Hintergrundmusik oder Umgebungsgeräuschen überlagert werden. Der Peak in der Kreuzkorrelation besteht auch, wenn das Signal nicht hörbar ist. Die oben beschriebenen Peaks in periodischen Signalen dürfen jedoch nicht größer sein bzw. über mehr Energie verfügen als der durch die Zeitverzögerung entstandene Spitzenwert. Durch eine Analyse der Umgebungsgeräusche in Echtzeit könnte theoretisch die Maskierungsschwelle des Rauschen berechnet und dieses entsprechend gefiltert werden. Um die Hörbarkeit des Referenzsignals einzuschränken kann dieses ausserdem mit der inversen Hörkurve gefiltert werden.

Begrenzung der Bandbreite

Wird weißes Rauschen gefiltert, ändert sich die Form der Autokorrelation und somit auch der Bereich um den Peak Kreuzkorrelation Mikrofon- und Referenzsignal. Wird beispielsweise ein Tiefpassfilter mit der Verstärkung W_0 und der Grenzfrequenz f_0 verwendet, wird die Autokorrelation im Bereich um $t=0$ zu einer sinc-Funktion, wobei die Hauptkeule mit sinkender Grenzfrequenz und somit geringerer Bandbreite, breiter wird. Es wird daher schwerer, den Peak der Kreuzkorrelation zu detektieren.

4. Messungen

Messaufbau

Zur Messung wurde zunächst ein Kunstkopf mit Binauralmikrofonen ausgestattet. Die internen Mikrofone wurden nicht verwendet, um den Fehler, der durch die Länge des Ohrkanals entsteht, zu vermeiden. Um den Winkel zwischen Kopf und Lautsprecher einstellen zu können, wurde der Kunstkopf auf einem steuerbaren Drehteller platziert. Als Referenz für den Abstand zwischen Kunstkopf und Lautsprecher diente ein Laser-Abstandsmessgerät. Die Steuerung des Systems und die Messung erfolgte über PureData-Patch.

Die Mikrofonsignale wurden über ein Audio-Interface aufgezeichnet, über welches auch die Referenzsignale abgespielt wurden. Als Referenzsignal wurde zunächst weißes Rauschen verwendet.

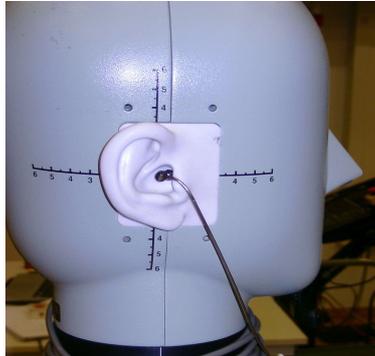


Abb.4.1.: InEar-Mikrofone



Abb.4.2.: Messaufbau

Software

Zur Steuerung und Messungen wurde ein PureData-Patch entwickelt. Die Steuerung des Drehtellers erfolgt hierbei via Netzwerkkabel über UDP und wurde ebenfalls in das Patch integriert. Zur Berechnung des Winkels und der Distanz werden zunächst die Kreuzkorrelationen zwischen dem abgespielten Referenzsignal und den aufgenommenen Mikrofonsignalen berechnet. Eine Peak Detection extrahiert daraus den jeweiligen Sample-Index des

Maximalwertes. Für die Distanzbestimmung erfolgt eine Mittelung beider Werte. Über die Sample-Rate kann hieraus zunächst die Schalllaufzeit und dann wie oben beschrieben der Schallweg und somit die Distanz bestimmt werden. Aus der Differenz der Indizes der Peaks beider Mikrofone wird die ITD und dann auf oben beschriebenen Weg die Orientierung bestimmt.

Kalibrierung

Bei der Berechnung der Kreuzkorrelationen verschiebt die Latenz des Gesamtsystems die Spitzenwerte. Die durch die Verarbeitung verstrichene Zeit fließt also in die Messergebnisse ein. Die Latenz ist zeitlich weitgehend konstant, hängt aber von der verwendeten Hardware ab. Um die Software von der Peripherie möglichst unabhängig zu halten, enthält das Programm die Möglichkeit der Kalibrierung. Der Lautsprecher wird hierzu im Abstand von 1m zum linken Mikrofon platziert. Das Programm errechnet die Differenz des Sample-Index des ersten Spitzenwertes der Kreuzkorrelation und der für den Schallweg von einem Meter benötigten Zeit in Samples. Dieser Wert wird für weitere Messungen als Latenzkorrektur verwendet und vom jeweiligen Sample-Index des Peaks subtrahiert. Die Bestimmung der Winkeländerung kann mit Hilfe des Drehtellers auf $0,5^\circ$ genau erfolgen. Der Nullwinkel muss jedoch händisch gefunden werden.

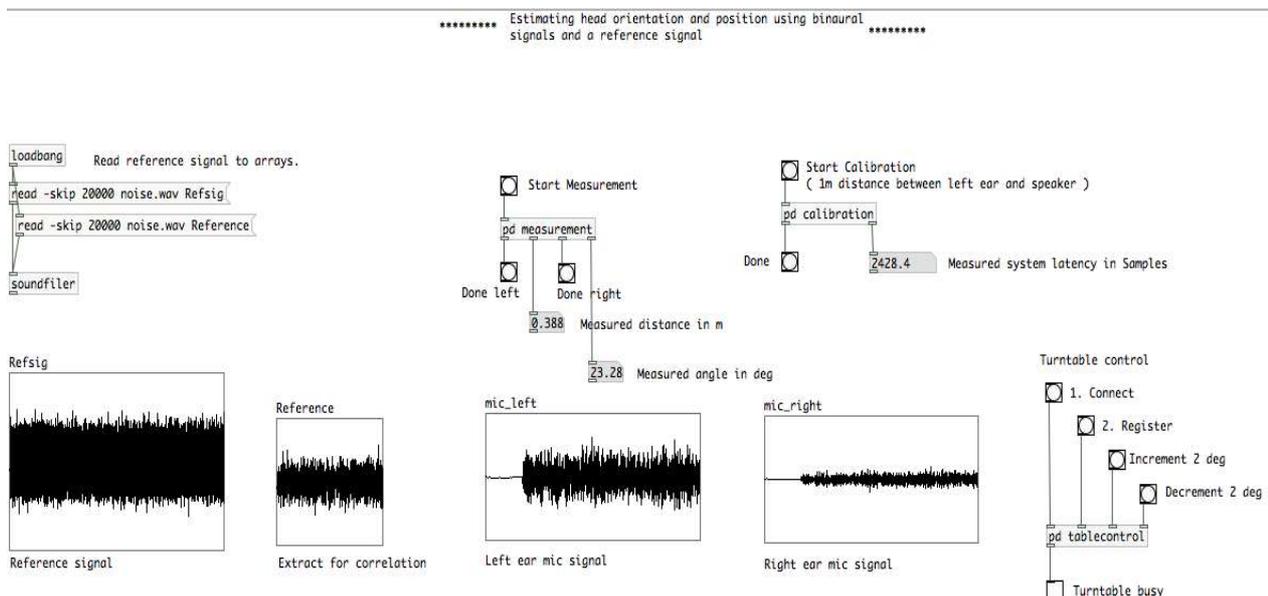


Abb. 4.3.: Screenshot PD-Patch

Genauigkeit, Messdauer und Interpolation

Um eine möglichst hohe Genauigkeit des Systems zu erreichen wurde eine

Abtastrate von 96kHz gewählt. Bei einer Schallgeschwindigkeit von 343 m/s ergibt sich hieraus eine vom Schall während eines Samples zurückgelegte Wegstrecke von 3,57mm, was der theoretisch Empfindlichkeit der Distanzbestimmung entspricht. Aus Formel (3.4) ergibt sich somit eine Winkelauflösung von $1,28^\circ$. Um gerade bei der Bestimmung der Kopfdrehung genauere Ergebnisse erzielen zu können, wurden die berechneten Kreuzkorrelationen in der Software interpoliert, wodurch die Genauigkeit um den Faktor 10 auf $0,128^\circ$ erhöht wurde.

Um eine Distanz von 5 Metern noch bestimmen zu können, müssen mindestens 1400 Samples aufgenommen werden, damit das am Mikrofon eintreffende Signal noch aufgenommen wird. Hinzu kommt die Systemlatenz, die bei dem verwendeten Equipment bei ca. 2000 Samples lag. Um noch Headroom für größere Abstände und höhere Latenzen zu gewährleisten, werden in der Software die ersten 6000 Samples nach Abspielen des Referenzsignals berücksichtigt. Wird ein zu großer Zeitraum betrachtet, erhöht dies zum einen die Messdauer und zum anderen können starke Reflexionen störende Peaks in der Kreuzkorrelation verursachen, was ggf. zu Messfehlern führt.

Das für die Kreuzkorrelation verwendete PD-Objekt *tab_cross_corr* berechnet die Kreuzkorrelation im Zeitbereich, was zu deutlich längeren Messzeiten führt als bei der Berechnung im Frequenzbereich. Die Berechnung von Position und Winkel für beide Kanäle dauert im vorliegenden Programm ca. 1,7s.

5. Messergebnisse

In einer ersten Messung wurde zunächst der Abstand bei 1,0m konstant gehalten und der Winkel im Bereich von -45° bis 45° in 2° - bzw. 5° -Schritten verändert.

Die gemessenen Ergebnisse stimmen im Bereich geringer Auslenkungen sehr gut mit den eingestellten Winkeln überein. Bei größeren Winkeln kommt es zu größeren Abweichungen. Diese Tendenz wurde auch in (3) beobachtet und ist durch die Beugung des Schalls um den Kopf zu erklären. Der daraus resultierende vergrößerte Wegunterschied wird in der Berechnung nicht berücksichtigt, könnte aber theoretisch durch die Verwendung eines anderen ITD-Modells korrigiert werden.

Winkel eingestellt	Winkel gemessen	Winkel eingestellt	Winkel gemessen
-45°	-48,4°	45°	49,2°
-40°	-42,2°	40°	44,2°
-35°	-36,1°	35°	37,6°
-30°	-29,4°	30°	31,4°
-28°	-27,5°	28°	28,9°
-26°	-25,6°	26°	26,8°
-24°	-23,9°	24°	24,7°
-22°	-21,8°	22°	22,4°
-20°	-20,1°	20°	20,3°
-18°	-18,0°	18°	18,2°
-16°	-16,0°	16°	16,1°
-14°	-14,2°	14°	14,2°
-12°	-12,2°	12°	12,2°
-10°	-10,1°	10°	10,0°
-8°	-7,9°	8°	8,2°
-6°	-6,0°	6°	6,2°
-4°	-4,1°	4°	4,1°
2°	-2,0°	2°	1,9°
0°	0,0°		

Tabelle 5.1.: Messergebnisse bei d=1,0m und variablem Winkel

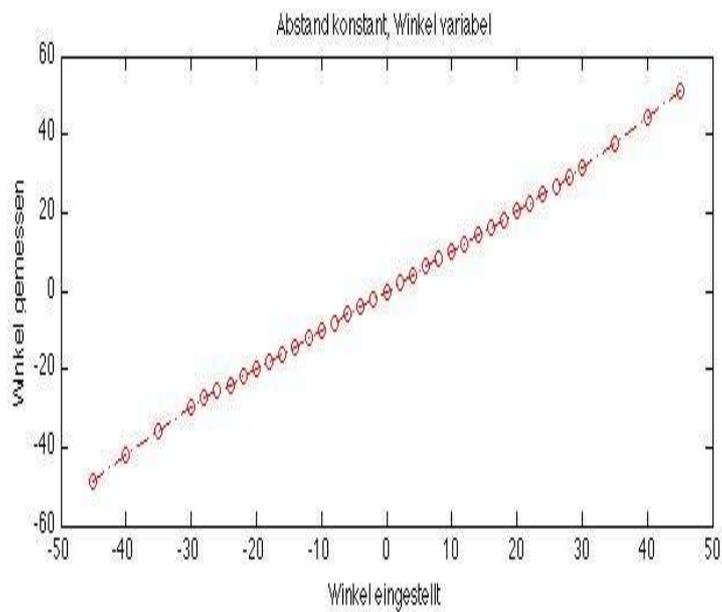


Abb. 5.1.: Messergebnisse bei d=1,0m und variablem Winkel

Um die Genauigkeit der Distanzbestimmung zu ermitteln, wurde in einer weiteren Messung bei einem konstanten Winkel von 0° der Abstand zwischen Lautsprecher und Kunstkopf im Bereich von 1m bis 3m in Schritten von 0,1m verändert.

Abstand eingestellt	Abstand gemessen	Abstand eingestellt	Abstand gemessen
1,000 m	1,000 m	1,810 m	1,797 m
1,134 m	1,136 m	1,892 m	1,887 m
1,212 m	1,207 m	2,014 m	1,995 m
1,296 m	1,291 m	2,194 m	2,172 m
1,401 m	1,392 m	2,398 m	2,378 m
1,511 m	1,502 m	2,635 m	2,618 m
1,607 m	1,599 m	2,843 m	2,825 m
1,709 m	1,696 m	3,112 m	3,089 m

Tabelle 5.2.: Messergebnisse bei $\phi=1,5^\circ$ und variabler Distanz

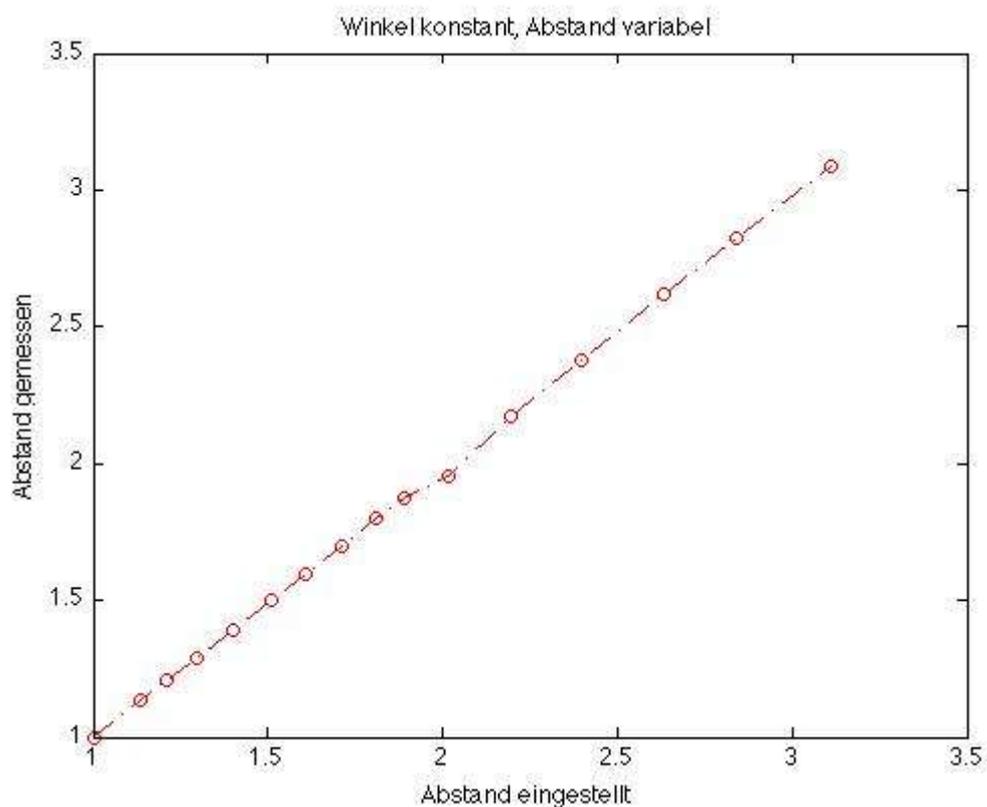


Abb. 5.2.: Messergebnisse bei $\phi=1,5^\circ$ und variablem Abstand

Aus den Ergebnissen ist ersichtlich, dass das System auch bei der

Abstandsbestimmung brauchbare Ergebnisse liefert. Es ist zu bemerken, dass die Referenzmessung mittels Laser-Distanzmeter nur mit Abweichungen möglich war.

In der folgenden Messung wurde der Einfluss von Hintergrundgeräuschen, in diesem Fall Musik, auf die Robustheit des Systems untersucht. Hierzu wurden bei einem konstanten Abstand von 1m und einem ebenfalls konstanten Winkel Messungen durchgeführt, bei denen der Pegel des Referenzsignals im Verhältnis zur Musik verändert wurde.

Es zeigt sich, dass wenn das Rauschen um 9dB leiser ist als die Musik, noch immer der richtige Wert ermittelt werden kann. Liegt der Pegel des Rauschsignals bei -17dB, wird ein falscher Peak detektiert und das Ergebnis verfälscht.

Winkel eingestellt	Winkel gemessen	Pegelverhältnis
4,0°	4,09°	nur Rauschen
4,0°	4,09°	gleich laut
4,0°	4,09°	Rauschen -6dB
4,0°	3,96°	Rauschen -9dB
4,0°	-2,81°	Rauschen -17dB
4,0°	-2,40°	nur Musik

Tabelle 5.3.: Messergebnisse bei $\phi=4^\circ$, $d=1\text{m}$ und variablem Pegelverhältnis von Musik und Rauschen

In diesem Versuch wurde ebenfalls gezeigt, dass ein Musiksignal nicht als Referenzsignal geeignet ist. Die folgenden Abbildungen zeigen die Kreuzkorrelation in zu den jeweiligen Pegelverhältnissen.

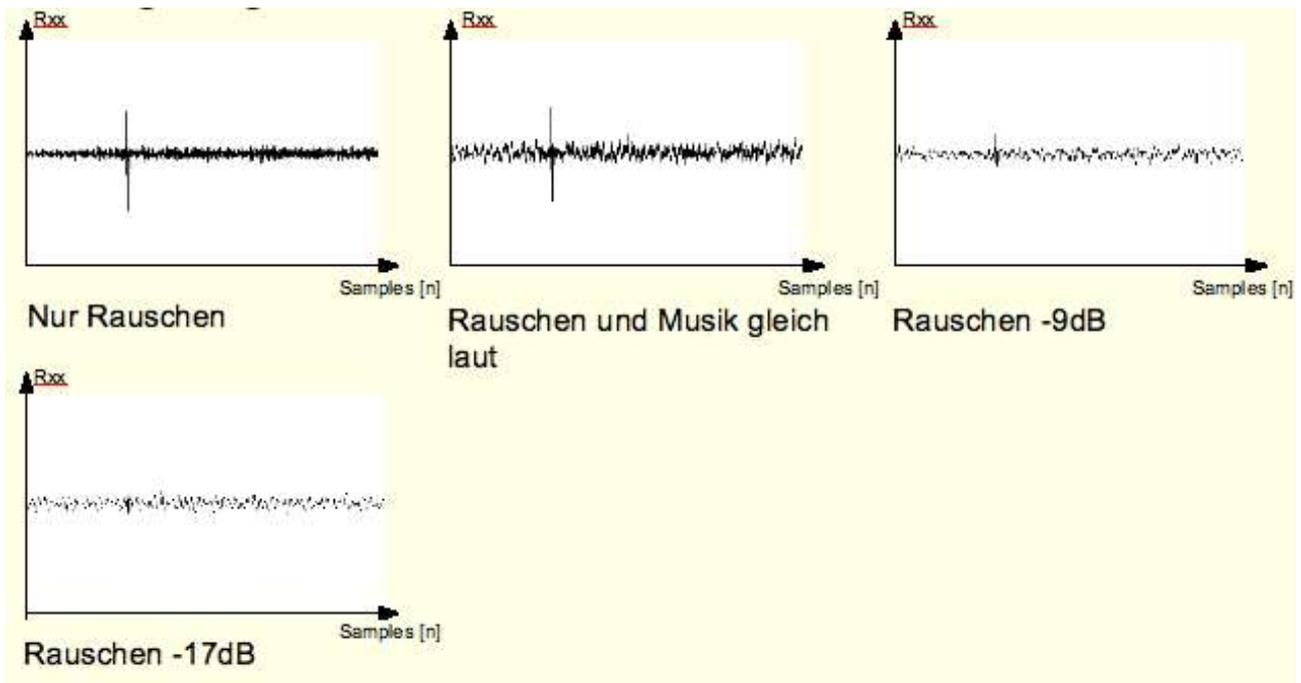


Abb. 5.3.: Kreuzkorrelationen bei Überlagerung mit Musik

Nach dem gleichen Prinzip wurde eine weitere Messung durchgeführt, bei der das Referenzsignal mit Aufnahmen von Umgebungsgeräuschen aus einem Museum überlagert wurde. Da dieses Signal weniger korreliert ist als das vorher verwendete Musiksignal, kann auch bei einer Pegeldifferenz von -17dB noch Position und Orientierung bestimmt werden. Erst wenn das Referenzsignal 20dB leiser ist als das überlagerte Umgebungsgeräusch, ist der Peak der Kreuzkorrelation nicht mehr erkennbar.

Winkel eingestellt	Winkel gemessen	Pegolverhältnis
4,0°	3,96°	nur Rauschen
4,0°	3,96°	gleich laut
4,0°	3,96°	Rauschen -4dB
4,0°	3,96°	Rauschen -10dB
4,0°	3,96°	Rauschen -17dB
4,0°	-5,81°	Rauschen -20dB
4,0°	-5,81°	nur Umgebungsgeräusche

Tabelle 5.4.: Messergebnisse bei $\phi=4^\circ$, $d=1\text{m}$ und variablem Pegolverhältnis von Umgebungsgeräuschen und Rauschen

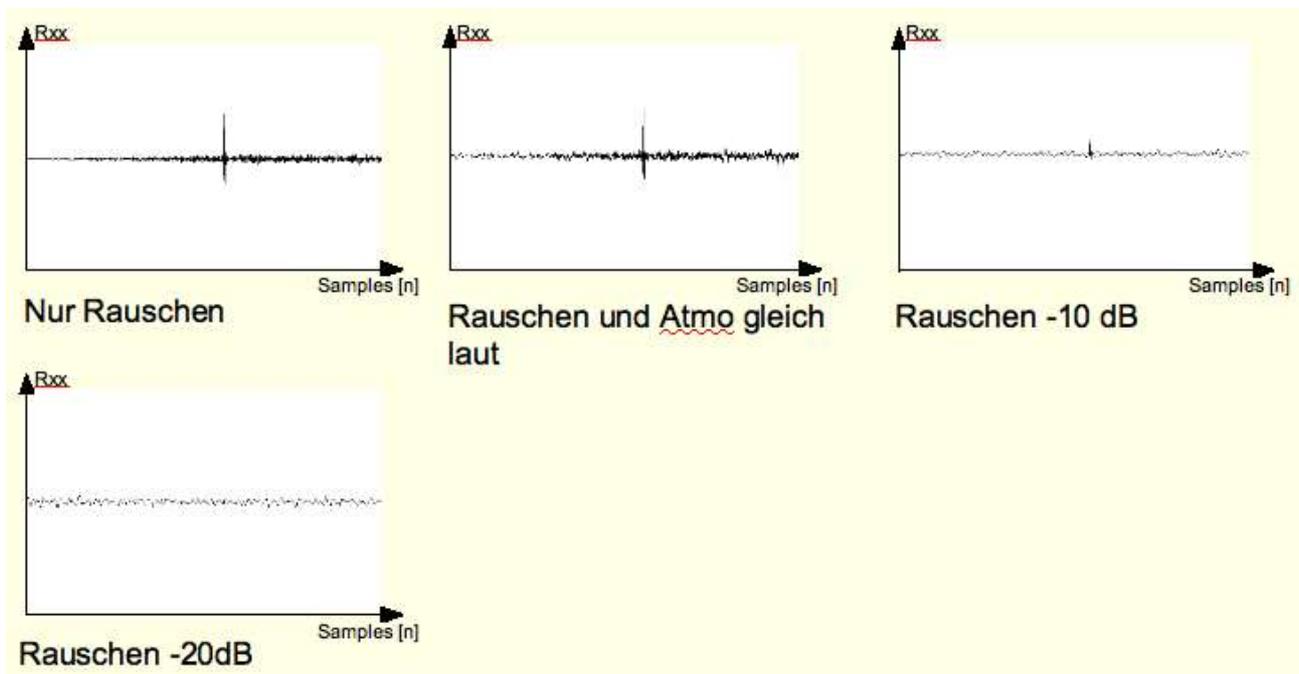


Abb. 5.4.: Kreuzkorrelationen bei Überlagerung mit Umgebungsgeräuschen

In einem weiteren Versuch wurde das Rauschsignal zunächst mit einem Hochpass und dann mit einem Tiefpass gefiltert. Der Winkel wurde konstant bei 0° und der Abstand bei 1m gehalten.

Winkel eingestellt	Winkel gemessen	Filter
0°	0°	ohne
0°	0°	Hochpass, $f_g=1\text{kHz}$
0°	$-23,7^\circ$	Tiefpass, $f_g=8\text{kHz}$

Tabelle 5.4.: Messergebnisse bei $\phi=0^\circ$, $d=1\text{m}$ und gefiltertem Rauschen

Die Ergebnisse zeigen einen deutlichen Messfehler bei der Verwendung des Tiefpass-gefilterten Signals. Eine Analyse der Absolutbeträge der interpolierten Kreuzkorrelationen zeigt, dass in diesem Fall ein späterer, wahrscheinlich durch eine frühe Reflexion entstandener, Peak als Maximum detektiert wurde.

Eine mögliche Erklärung hierfür ist, dass hohe Frequenzen direkt auf die Membran des Mikrofons treffen, während tiefe Frequenzen zunächst um die Membran gebeugt und dann an der Ohrmuschel reflektiert werden, bevor sie auf die Membran treffen.

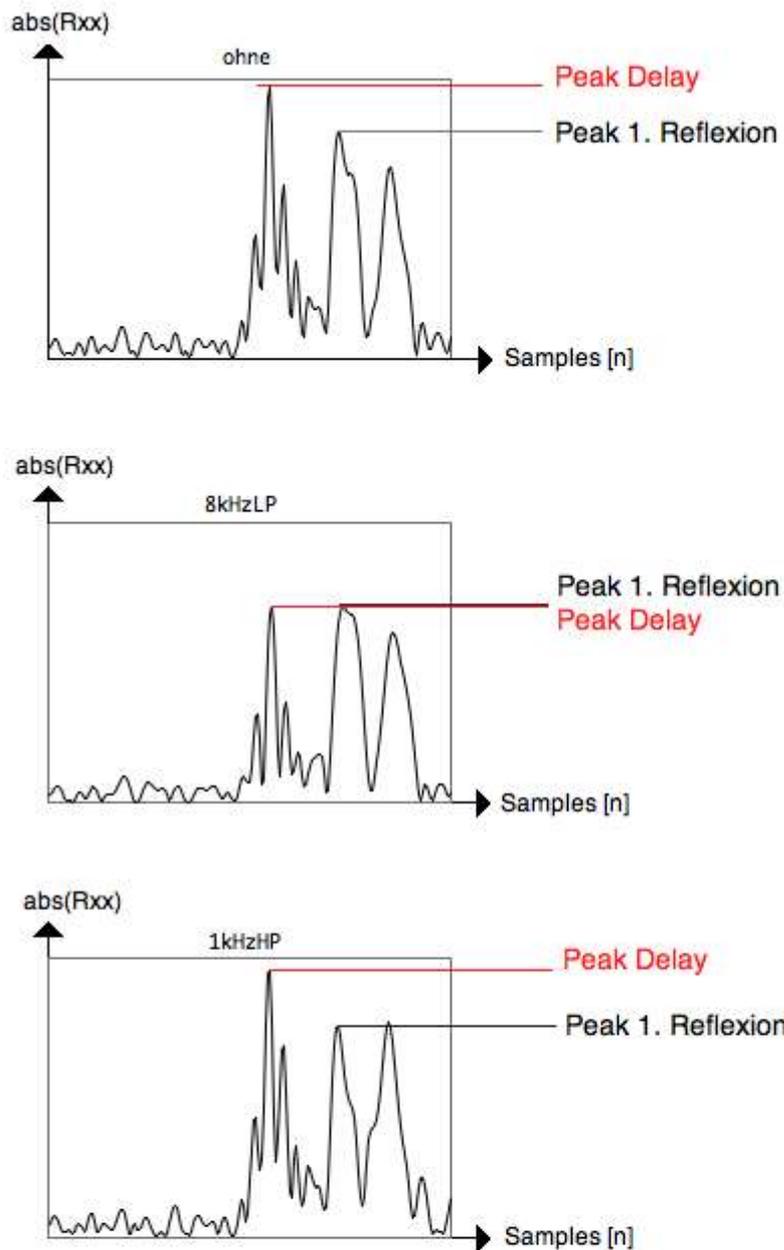


Abb 5.5.: Kreuzkorrelationen bei unterschiedlich gefilterten Signalen

6. Diskussion und Ausblick

Das aufgebaute System liefert für eine Binauralsynthese ausreichend genaue Trackingdaten. Bei größeren Winkeln ist zwar ein deutlich geringerer SNR beim von der Quelle abgewandten Ohr sichtbar, der Peak in der Kreuzkorrelation ist jedoch deutlich erkennbar und kann problemlos detektiert

werden. Durch eine geeignete Wahl des betrachteten Bereichs der Kreuzkorrelation sind Messfehler durch den Einfluss von Raumreflexionen auszuschließen. Reflexionen an der Ohrmuschel führen jedoch zu Peaks in der Kreuzkorrelation, welche das Ergebnis verfälschen können. Hierzu sollte ein geeigneter Frequenzbereich des Referenzsignals gewählt werden.

Für ein echtzeitfähiges System mit größeren Entfernungsänderungen zwischen Lautsprecher und Mikrofonen sollte die Auswahl des betrachteten Bereiches der Kreuzkorrelation dynamisch angepasst und diese im Frequenzbereich berechnet werden.

Als Referenzsignal eignet sich Rauschen, welches bis zu einer messbaren Pegeldifferenz von anderen Signalen mit periodischen Anteilen überlagert werden kann. Um die mögliche Pegeldifferenz noch zu vergrößern, könnte eine Peak-Detection verwendet werden, welche auch die Steilheit der Peaks mit einbezieht und so den von der Verzögerung des Rauschsignals verursachten, schmalen, Peak von den deutlich breiteren Peaks der periodischen Anteile des überlagerten Signals trennt.

7. Referenzsignale in praktischen Anwendungen

Das auffälligste Problem bei dem Einsatz akustischer Trackingverfahren in der Praxis stellt die Abhängigkeit von Referenzschallquellen dar. Es ist zwar meist mit wenig Aufwand möglich, Lautsprecher vor Ort zu installieren, die Referenzsignale sind jedoch für den Anwender außer im Falle einer totalen Maskierung grundsätzlich hörbar. Wie bereits eingehend erwähnt, ist zudem die Regelung des Pegels des Referenzsignals unter die durch die Umgebungsgeräusche verursachte Mithörschwelle sehr aufwändig.

Gerade bei künstlerischen Arbeiten – wie z.B. Klanginstallationen – stellt diese Tatsache prinzipiell einen massiven Eingriff in die vom Komponisten kreierte Hörwahrnehmung dar. Ist das Vorhandensein der Referenzschallquellen jedoch zuvor bekannt, kann dieser Aspekt jedoch als Herausforderung, ähnlich wie z.B. eine außergewöhnliche Raumakustik oder andere am Aufführungsort vorhandene Hintergrundgeräusche, aufgefasst und das Werk daran angepasst werden. Da die Referenzsignale bis zu einem bestimmten Grad was ihren Frequenzgehalt und die Abspielintervalle betrifft variabel sind, könnten diese sogar künstlerisch eingebunden werden.

In anderen Einsatzgebieten, etwa als Automated Tour Guide, besteht die

Möglichkeit, eine künstliche Geräuschkulisse mit geringer Dynamik im Hintergrund, wie z.B. Musik, zu schaffen und somit die Referenzsignale weitestgehend zu maskieren. Zudem trägt eine Filterung des Referenzsignals mit der inversen Hörkurve zur geringeren Wahrnehmung bei.

Literaturverzeichnis

- (1) A. Härma et al. „Techniques and applications of wearable augmented reality audio“; *Proc. of 114th AES Convention*, preprint 5768, Amsterdam, 22-25. März 2003
- (2) A. Härma, M. Karjalainen, M. Tikander „Head-Tracking and subject positioning using binaural headset microphones and common modulation anchor sources“; *Proc. Of the IEEE Workshop on Appl. Of Sig. Proc. To Audio and Acoustics*, New Paltz, New York, USA, Oktober 2003
- (3) M. Tikander, A. Härmä, M. Karjalainen „Acoustic Positioning and head tracking based on binaural signals“; *Proc. 116th AES Convention*, Berlin, 8.-11. Mai 2004
- (4) J. H. DiBiase, H. F. Silverman, M.S. Brandstein „Robust Localization in Reverberant Rooms“; *Microphone Arrays: Signal Processing Techniques and Applications*, M.s. Brandstein, D.Ward, Eds. Springer-Verlag, 2001, Kapitel 7, pp. 131-154
- (5) B. Benderson „Audio Augmented Reality: A Prototype Automated Tour Guide“; *ACM Human Computer in Computing Systems conference companion*, pp 210-211, 1995
- (6) K. Lyons, M. Gandy, T. Starner „Guided by voices: An Audio Augmented Reality System“; *Proc. of International Conference on Auditory Display (ICAD) 2000*, Atlanta, GA, April 2000
- (7) J. Rozier, K. Karahalios, J. Donath „Hear & There: An Audio Augmented Reality System of Linked Audio“; <http://smg.media.mit.edu/projects/HearAndThere/>
- (8) M. Peltola „Augmented Reality Audio Applications in Outdoor Use“; *Helsinki University of Technology, Master´ Thesis*, Februar 2009
- (9) G. Eckel „Audio-Augmented Reality – Ein neues Medium für die Klangkunst“; *Sonambiente Berlin 2006*, ed. Helga de la Motte-Haber et al., Kehrer, Heidelberg, Germany, 2006, pp. 348-349