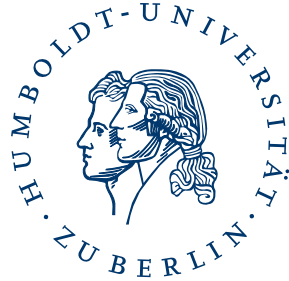


Humboldt-Universität zu Berlin
Mathematisch-Naturwissenschaftliche Fakultät II
Institut für Informatik



Schallquellen-Lokalisation für humanoide Roboter unter Verwendung eines binauralen Audiosystems

Diplomarbeit

im Fachgebiet Informatik

vorgelegt von: Robert Wolff
Matrikelnummer: 196697
Betreuer: Dr. Manfred Hild
Gutachter: Prof. Dr. Hans-Dieter Burkhard
Prof. Dr.-Ing. Beate Meffert

Berlin, den 16. Juni 2010

Zusammenfassung

In der vorliegenden Arbeit wurde ein einfaches, dreistufiges Verfahren zur Schallquellen-Lokalisation auf Basis eines binauralen Systems entwickelt und untersucht.

In einem ersten Schritt wurde das binaurale System, der Prototyp eines humanoïden Roboterkopfes, hinsichtlich seiner Eignung zur Schallquellen-Lokalisation evaluiert und um die Elektronik zur Stereosignalverarbeitung erweitert.

Im Vergleich zu bekannten, leistungsfähigen, aber rechen- und speicherintensiven Methoden der Schallquellen-Lokalisation, wurde das entwickelte Verfahren an eine beschränkte Rechenleistung und einen beschränkten Arbeitsspeicher angepasst. So kann es auf einem eingebetteten System oder einem digitalen Signalprozessor (DSP) ausgeführt werden. Dazu wurde ein einfaches, im Zeitbereich anwendbares Verfahren zur zweidimensionalen Schallquellen-Lokalisation entwickelt und implementiert. Die zentralen Schwerpunkte des Modells waren die Onset-Detektion sowie die horizontale und vertikale Richtungsanalyse.

Um das entwickelte Verfahren anschließend untersuchen zu können, wurden die Anforderungen eines möglichen Anwendungsszenarios analysiert und daraus ein Experimentalaufbau abgeleitet, welcher die Grundlage zur Erzeugung eines repräsentativen Audiodatensatzes darstellte. Der humanoïde Roboterkopf diente als Messobjekt. Für die Messungen wurde ein spezielles Audiosignal entwickelt, das eine Analyse der Schallquellen-Richtung im Zeitbereich ermöglichte. Mit Hilfe des Audiodatensatzes konnte die korrekte Funktionsweise der algorithmischen Umsetzung des Verfahrens nachgewiesen werden. Berechnungen der Genauigkeit zeigten, dass das Verfahren ebenfalls in der Praxis mit hinreichender Genauigkeit funktionieren kann.

Inhaltsverzeichnis

1. Einleitung und Begriffsdefinitionen	1
1.1. Zielsetzung	2
1.2. Existierende Ansätze	3
1.3. Einführung grundlegender Begriffe	4
1.4. Allgemeine Verfahren und Konzepte	8
1.5. Gliederung der Arbeit	11
2. Beschreibung und Evaluation der Hardware	13
2.1. Humanoider Roboter	13
2.2. Experimentalsystem	15
2.2.1. Prototyp des Roboterkopfes	16
2.2.2. Wiedergabesystem und zusätzliche Messtechnik	22
3. Experimentaldesign und Reizgenerierung	23
3.1. Experimentalanforderung	23
3.2. Design des Audiosignals zur Lokalisation	26
3.2.1. Analyse der Anforderungen	26
3.2.2. Synthese des Audiosignals	28
3.3. Experimentaldesign und Messdurchführung	30
3.3.1. Versuchsaufbau	31
3.3.2. Messdurchführung	33
4. Konzeption der Modelle und Verfahren	35
4.1. Verfahren zur Detektion des Onsets	35
4.2. Geometrisches Modell zur Bestimmung des Azimuts	38
4.2.1. Minimales Modell eines Stereosystems	39
4.2.2. 1D-Modell des Roboterkopfes	41
4.2.3. 2D-Modell des Roboterkopfes	44

4.3.	Verfahren zur Bestimmung der Elevation	47
4.3.1.	Grundkonzepte des Verfahrens	47
4.3.2.	Synthese des Verfahrens	51
5.	Untersuchung und Auswertung	57
5.1.	Detektion des Onsets	57
5.1.1.	Anpassung der Modellparameter	57
5.1.2.	Untersuchung	59
5.1.3.	Auswertung	61
5.2.	Bestimmung des Azimuts	61
5.2.1.	Untersuchung und Vervollständigung des Modells	62
5.2.2.	Auswertung	64
5.3.	Bestimmung der Elevation	65
5.3.1.	Erläuterung der Implementierung	66
5.3.2.	Nachweis der korrekten Funktionsweise	68
5.3.3.	Analyse der Richtungsgenauigkeit	73
5.3.4.	Auswertung	77
6.	Zusammenfassung und Ausblick	79
6.1.	Zusammenfassung	79
6.2.	Ausblick	82
A.	Anhang	85
	Literaturverzeichnis	93

1. Einleitung und Begriffsdefinitionen

Eine grundlegende Fähigkeit fast aller Lebewesen ist die auditive Wahrnehmung. Sie erlaubt beispielsweise dem Menschen, mit Hilfe seiner beiden Ohren Schallsignale der Umwelt aufzunehmen und zu verarbeiten. Somit ist eine Kommunikation auf Basis von Lauten oder einer Sprache möglich. Die auditive Wahrnehmung ist aber auch eine Voraussetzung für die räumliche Wahrnehmung der Umwelt.

Beim Menschen basiert die räumliche Wahrnehmung nicht nur auf der auditiven, sondern in erster Linie auf der visuellen Wahrnehmung. Beide sind für unterschiedliche Aspekte zuständig: Die Augen sind nach vorn gerichtet und dienen dazu, innerhalb des Raumes Objekte detailliert zu fokussieren und räumlich voneinander zu unterscheiden. Da das Sichtfeld beim Menschen horizontal auf circa 170 Grad und vertikal auf circa 130 Grad [AP80] beschränkt ist, dient die visuelle Wahrnehmung nur zur frontalen Orientierung nach vorn. Dagegen erlauben die Ohren eine vollständige räumliche Orientierung sowohl in der horizontalen als auch in der vertikalen Ebene. Die Orientierung auf Grundlage akustischer Signale geschieht in der Regel passiv durch die Lokalisation von Schallquellen und die Auswertung von Schallreflexionen und Hall, was auch die Selbst-Lokalisation im Raum erlaubt. Sehbehinderte oder blinde Menschen können trotz ihrer Einschränkung allein mit Hilfe ihres geschulten Gehörs sehr gut im Raum orientieren. In der Natur gibt es noch andere Beispiele für die Nützlichkeit des *räumlichen Hörens*. Die Fähigkeit, Schallquellen zu lokalisieren, bringt vielen Säugetieren einen lebensnotwendigen Vorteil. So ist aus der Biologie bekannt, dass auf diese Weise in der Dunkelheit Beute geortet oder einem getarnten beziehungsweise sich außerhalb des Sichtfeldes nähernden Jäger rechtzeitig ausgewichen werden kann. Zusätzlich ist die Schallquellen-Lokalisation für einige Insekten zur Arterhaltung notwendig. Auf diese Weise sind zum Beispiel Laubheuschrecken in der Lage, einen paarungswilligen Fortpflanzungspartner, welcher ein Zirp-Geräusch aussendet, über große Entfernungen zu orten und sich diesem zu nähern [Jat99].

In Anbetracht der Bedeutung für den Menschen und andere Lebewesen stellt die Schallquellen-Lokalisation in der Robotik einen interessanten Forschungsschwer-

punkt dar. In vergangenen und aktuellen Projekten werden insbesondere humanoide Roboter neben zahlreichen Sensoren wie Tast-, Winkel-, Lage-, Beschleunigungssensoren und einer Kamera zur visuellen Wahrnehmung [Ben10] ebenfalls mit einem Audiosystem ausgestattet. Somit kann die Schallquellen-Ortung erforscht und vom Roboter zum Beispiel zur autonomen räumlichen Orientierung genutzt werden. Einige Beispiele für solche Projekte sind [CNK01], iCub [BLP⁺06] und Cb [CHM⁺06]. Aber auch in nicht-humanoiden Robotern wie in Do:Little [Hil07] oder dem Aibo [LM08] existieren Mikrofonsysteme, die zur Ortung genutzt werden können.

1.1. Zielsetzung

Der im Rahmen des Projektes *Artificial Language Evolution on Autonomous Robots* (ALEAR) entwickelte autonome, humanoide Roboter soll ebenfalls mit einem binauralen System zur Ortung von Schallquellen ausgestattet werden. Gegenstand dieser Arbeit ist die Konzeption und Realisierung eines solchen Systems. Voraussetzung ist ein binaurales Audiosystem, das zur Schallquellen-Lokalisation geeignet ist und in einem autonomen, mobilen Roboter verwendet werden kann. Die Autonomie und Mobilität des Roboters ist durch eine eigenständige Steuerung möglich. Es wird auch gefordert, dass alle Berechnungen zur Laufzeit auf der eingebetteten, elektronischen Hardware erfolgen sollen. Hinzu kommt, dass eine unabhängige, akkubasierte Stromversorgung Bestandteil des Roboters sein sollte. Eingesetzt werden daher bevorzugt stromsparende Systeme, wie Mikrocontroller, DSPs oder FPGAs¹, welche möglichst kostengünstig sein sollten. Für die ersten beiden eingebetteten Prozessoren ergeben sich häufig technische Beschränkungen bezüglich des Speichers und der Rechenleistung, die in dieser Arbeit berücksichtigt werden sollen.

Das Hauptziel dieser Arbeit ist die Konzeption und Untersuchung eines geeigneten Verfahrens oder Modells zur Schallquellen-Lokalisation unter Berücksichtigung der technischen Beschränkungen eines eingebetteten Systems. Dabei soll auf die Transformation der Audiodaten in den Frequenzbereich und somit auf die Verarbeitung eines komplexwertigen Signalspektrums verzichtet werden. Die Analyse der zur vertikalen Ortung benötigten Frequenzinformationen ist unter bestimmten Umständen auch sehr einfach im Zeitbereich möglich, was bei sehr kurzen Signalen zu einer Verringerung der Komplexität führt. Die genauen Anforderungen an die Struktur des

¹FPGA ist die englische Abkürzung für *Field Programmable Gate Array*.

benötigten Signals und die verwendeten Analyseverfahren werden erarbeitet. Um die Komplexität der Ortungsaufgabe zu reduzieren, werden die Anforderungen des möglichen Anwendungszenarios analysiert und Rückschlüsse für die Entwicklung eines reduzierten Experimentalaufbaus gezogen.

1.2. Existierende Ansätze

Die Anfänge der Erforschung der Schallquellen-Lokalisation reichen bereits über 100 Jahre zurück, als Anfang des 20. Jahrhunderts erste psychologische Untersuchungen beispielsweise zur monauralen Lokalisation von Schall am Menschen durchgeführt wurden [AF01]. Zahlreiche Studien zur Lokalisation von Schallquellen wurden seitdem veröffentlicht. Diese werden bis heute experimentell am Menschen [WNR49], [Bla69], [ZH06] und seit den 70er Jahren zusätzlich auf Basis binauraler Aufnahmen mit künstlichen Ersatzsystemen, wie dem *Knowles Electronic Manikin for Acoustic Research* (KEMAR) [GK94], [WB06], durchgeführt. Die Verfügbarkeit digitaler binauraler Aufnahmen erlaubte es, verschiedene Modelle und Verfahren zur akustischen Ortung unter vielfältigen Gesichtspunkten zu entwickeln und zu analysieren (Beispiele [Bla74], [Beg94], [WB06]). Bei der Untersuchung standen meist die Genauigkeit der Richtungsbestimmung und die Robustheit bezüglich äußerer Einflüsse, wie beispielsweise Umgebungsrauschen² [Bla69], Schall-Reflexionen [FM04] oder bewegte [KKOO08] und konkurrierende Schallquellen [CNNU03] im Vordergrund. Die Modelle werden zunehmend komplexer und aufwendiger in der Realisierung.

Im Rahmen der Robotik spielt die effiziente Umsetzbarkeit und die Ausführung zur Laufzeit eine wichtige Rolle. Es sind jedoch keine Quellen zu finden, in denen bekannte Verfahren zur zweidimensionalen Schallquellen-Lokalisation auf die schnelle Ausführbarkeit im realen Einsatz auf stark beschränkten Systemen mit Prozessor-Taktraten unter 100 MHz oder wenigen Kilobyte Arbeitsspeicher getestet wurden. Von der Implementierung der komplexen Modelle und der Untersuchung auf ihre Einsetzbarkeit auf beschränkten Systemen wird in dieser Arbeit abgesehen. Ziel ist es stattdessen, auf Basis von einfachsten mathematischen und geometrischen Modellen und Konzepten, welche die Anforderung der Hardware mit hoher Wahrscheinlichkeit erfüllen, ein eigenes Verfahren zur Schallquellen-Lokalisation zu entwickeln, zu implementieren und zu untersuchen.

²Das Umgebungsrauschen bezeichnet das Grundrauschen in einem nicht-schallisolierten Raum.

Die benötigten grundlegenden Modelle und Verfahren werden in den Kapiteln hergeleitet und erläutert, in denen sie angewendet werden. Dies betrifft die Verfahren zur Onset-Detektion³ und grundlegende Modelle zur horizontalen und vertikalen Schallquellen-Lokalisation. Grundlegende Begriffe, die zum allgemeinen Verständnis beitragen, und allgemeine Konzepte, welche Hilfsmittel zur Modellbildung der vertikalen Schallquellen-Lokalisation darstellen, werden in den folgenden zwei Abschnitten kurz erklärt.

1.3. Einführung grundlegender Begriffe

In dieser Arbeit wird die Kenntnis einiger physikalischer Begriffe und Konzepte vorausgesetzt, die nachfolgend erklärt werden.

Im ersten Abschnitt wird das Koordinatensystem beschrieben, das zur Angabe der Schallquellen-Position verwendet wird. Des Weiteren geht es um die Einführung der Begriffe Schall und Schallreflexion, die insbesondere in der Schallanalyse in geschlossenen Räumen von großer Bedeutung sind. Anschließend wird der klassische Amplitudenverlauf von Schallsignalen dargelegt, welcher für die Analyse des Onsets, also des Beginns eines Audiosignals, von Bedeutung ist.

Geometrisches Modell der Position

Die Position einer Schallquelle in Bezug auf einen Kunstkopf mit einem binauralen System wird mit einem Koordinatensystem aus zwei Winkeln beschrieben, ähnlich dem Konzept der Kugelkoordinaten [Bla69], [Beg94]. Der Kunstkopf befindet sich im Zentrum (siehe Abbildung 1.1). Die Position einer Schallquelle wird durch ihre Richtung beschrieben, welche sich aus dem Seiten- und Höhenwinkel zusammensetzt. Der Seitenwinkel beziehungsweise Azimut wird durch Polarkoordinaten von 0 bis 360 Grad mit einem dem Uhrzeigersinn entgegengesetzten Richtungssinn angegeben. Die Seitenwinkel 0 und 180 Grad befinden sich in der Median- beziehungsweise Sagittalebene, die den Kopf genau in eine rechte und eine linke Hälfte gliedert. Die Ohren befinden sich in der Frontalebene, welche den Raum in den Bereich vor und hinter dem Kopf teilt. Der Höhenwinkel, auch Elevation genannt, wird von -90 Grad (senkrecht unter dem Kopf) bis $+90$ Grad (senkrecht über dem Kopf) angegeben, ähnlich den Breitengraden der Erde. Der Höhenwinkel von 0 Grad befindet sich in der Transversalebene, die den Raum in ober- und unterhalb des Kopfes aufteilt.

³*Onset* ist die englische Bezeichnung für Anfang.

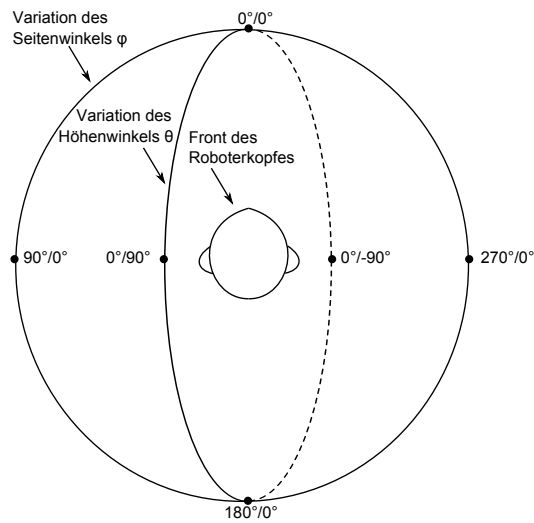


Abbildung 1.1. – Das Koordinatenpaar aus Seitenwinkel φ und Höhenwinkel ϑ gibt die Position der Schallquelle bezogen auf den Kunstkopf an. Der Azimut nimmt Werte zwischen 0 und 360 Grad und die Elevation zwischen -90 und $+90$ Grad an.

Schall und Schallreflexion

Die folgenden allgemeingültigen Grundlagen wurden basierend auf den Standardwerken [Bla74], [Bla85], [Beg94] zusammenfassend dargestellt, ohne im Einzelnen auf die Quellen zu verweisen: Eine Schallwelle ist die zeitliche Änderung der Dichte eines Mediums, welche sich wellenförmig ausbreitet. Das Schallfeld, auch akustisches Feld genannt, ist im Allgemeinen ein luftgefülltes Raumgebiet, in dem sich Schallwellen ausgehend von einer Schallquelle ausbreiten. Es wird zwischen einem freien Schallfeld, in dem die Schallausbreitung kugelförmig geschieht, und einem ebenen Schallfeld unterschieden. Die Ausbreitung der Schallwellen erfolgt im letzten Fall in senkrecht zur Fortpflanzung stehenden Wellenfronten. In der näheren Umgebung einer punktförmigen Schallquelle, beispielsweise einem Lautsprecher, breitet sich der Schall kugelförmig aus. Ab einem Abstand von circa 3m handelt es sich um ein ebenes Schallfeld, da die Krümmung der Wellenfront nur noch zu Pegeldifferenzen unter einem Dezibel führt [Bla85]. Es wird in diesem Fall auch von einer ebenen beziehungsweise linearen Welle gesprochen.

In geschlossenen Räumen wird das Schallfeld einer Quelle durch Reflexionen der Schallwellen überlagert und somit gestört. Die Reflexionen können an beliebigen Oberflächen zum Beispiel an Wänden, Decke, Boden oder anderen Objekten im

Raum entstehen. Vor einer ebenen, reflektierenden Fläche kann das Schallfeld einer Quelle als Überlagerung des ungestörten Schallfeldes mit dem gedämpften einer Spiegel- beziehungsweise Phantomschallquelle betrachtet werden. Sind Informationen über den Abstand der Schallquelle und des Messpunktes zur Reflexionsfläche bekannt, kann die zeitliche Differenz zwischen dem Eintreffen des Direktschalls und der ersten Reflexion berechnet werden. In Abbildung 1.2 wird ein charakteristischer

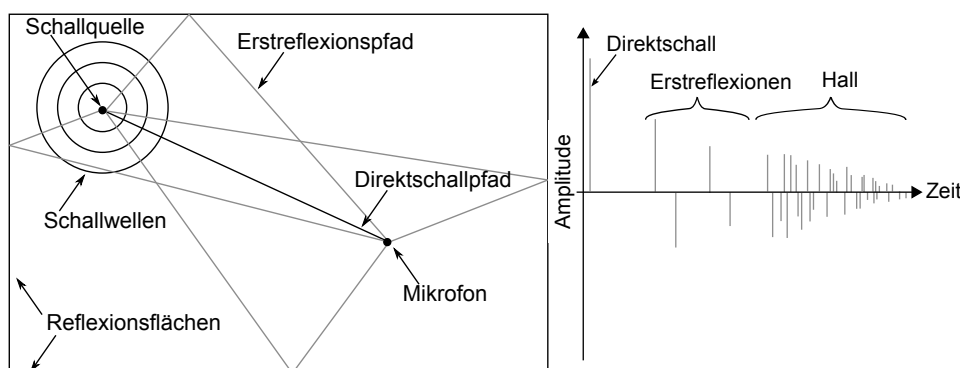


Abbildung 1.2. – Links: Abbildung verschiedener Signalpfade in der Ebene zwischen Quelle und Empfänger. Der kürzeste Pfad entspricht dem Direktschallpfad. Reflexionen an den Wänden beziehungsweise Boden und Decke folgen darauf. Rechts: Abbildung der Impulsantwort des Schalls bestehend aus Direktschall, gefolgt von unterscheidbaren Erstreiflexionen sowie diffusem Hall mit abklingender Intensität.

zeitlicher Verlauf einer Impulsantwort an einem Messpunkt im geschlossenen Raum dargestellt. Die ersten, nach Ankunft des primären Impulses gemessenen Signale, werden Erstreiflexionen genannt. Die anschließend auftretenden teilweise mehrfach reflektierten Impulse, werden als Hall bezeichnet. Da die Geschwindigkeit, mit der die Amplitude des Halls abklingt, von der Raumgröße abhängt, kann auf diese Weise die Raumabklingzeit bestimmt werden. Das zeitliche Auftreten der Impulse in der Impulsantwort ist identisch mit dem Auftreten der ersten Wellenfront, wenn ein zeitlich ausgedehntes Schallsignal anstelle des Impulses verwendet wird. So wird der Direktschall mit Ankunft der ersten Reflexion durch diese überlagert. Eine vollständig verlustfreie Rekonstruktion des kompletten Direktschalls ist dann im Allgemeinen nicht mehr möglich.

Signalstruktur und Onset

Für die folgenden Begriffserklärungen wird auf [BDDS04], [BDA⁺05] und [Pek06] zurückgegriffen, welche zusammenfassend dargestellt werden: In der Akustik wird der Onset als der Beginn eines diskreten zeitlichen Ereignisses definiert, welches durch die Änderung der psychoakustischen Effekte Intensität, Tonhöhe oder Klangfarbe markiert wird. Die mit der menschlichen Wahrnehmung in Zusammenhang stehenden Effekte lassen sich ebenfalls grob mit den physikalischen Parametern Schallpegel, Grundfrequenz und Schallspektrum beschreiben und somit digital verarbeiten [Pek06]. In Abbildung 1.3 ist der allgemeine Verlauf eines Schallsignals anhand eines instrumental erzeugten Tons dargestellt.

Die Struktur eines klassischen Schallsignals wird mit dem Konzept der Transienten beschrieben. In der Wellenlehre wird damit ein sehr schneller, impulsartiger Einschwingvorgang bezeichnet. Die Transienten beginnen mit einem steilen Anstieg der Amplitude, auch als „*attack-Phase*“⁴ bezeichnet. Die maximale Amplitude des Signals wird bei der Signalerzeugung über einen kurzen Zeitraum aufgebaut. Anschließend folgt ein Intervall, in dem das Signal erhalten bleibt und nur leicht im Pegel fällt. In der anschließenden „*decay-Phase*“⁵ klingt die Signalamplitude ab und fällt langsam auf den Ruhepegel zurück.

⁴ *Attack* ist die englische Bezeichnung für Anschlag eines Instruments, beispielsweise der Anriss der Gitarrensaite.

⁵ *Decay* ist die englische Bezeichnung für Abklingen.

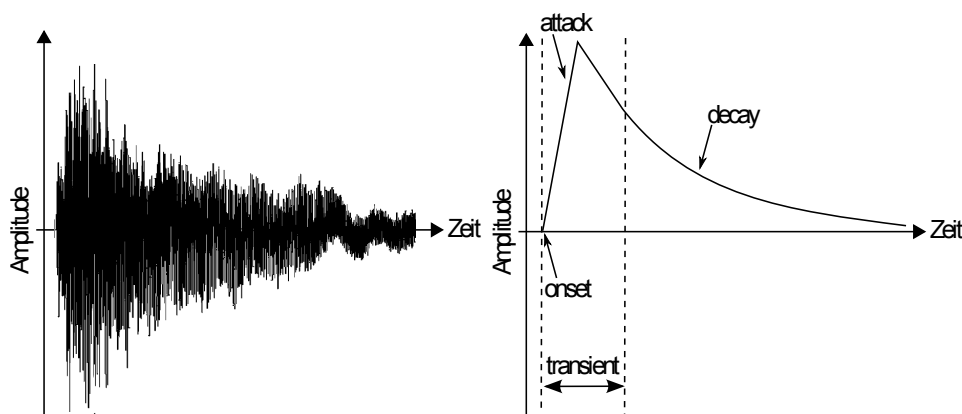


Abbildung 1.3. – Links: Typische Wellenform eines akustischen Signals, in diesem Beispiel erzeugt durch den Anriss der Gitarrensaite. Rechts: Schematischer Aufbau der zugehörigen Hüllkurve nach Bello et al. [BDDS04].

1.4. Allgemeine Verfahren und Konzepte

Für die Entwicklung eines Verfahrens zur vertikalen Schallquellen-Lokalisation wird das Verfahren der *Principal Component Analysis*⁶ verwendet. Die generische Suche nach optimalen Parametern für die Datenreduktion wird mit Hilfe der *künstlichen Evolution* realisiert. Diese kommt im Rahmen des *maschinellen Lernens* zum Einsatz. Alle drei Verfahren werden nachfolgend vorgestellt.

Principal Component Analysis

Die folgenden Grundlagen zur *Principal Component Analysis* werden basierend auf dem Nachschlagewerk [MH04] zusammenfassend dargestellt:

Die *Principal Component Analysis* (PCA), ist ein Verfahren der multivariaten Statistik. Die Eingangsgrößen sind N korrelierte Zufallsvektoren oder Messreihen \mathbf{f}_ν mit jeweils M Elementen $f_{\nu\mu}$. Der Index ν (mit $\nu = 1, \dots, N$) bezeichnet ein Element der Menge der Variablen beziehungsweise eine Dimension der Datenbasis, wohingegen μ (mit $\mu = 1, \dots, M$) ein Element der Menge der Messungen angibt.

Ziel ist es, die korrelierten Eingangsgrößen in eine Menge von N unkorrelierten Messreihen \mathbf{g}_ν zu transformieren. Dazu werden die Achsen des N -dimensionalen Raums in Richtung der „Hauptachsen“ gedreht, die auch als Hauptkomponenten oder *Principal Components* bezeichnet werden. Hier weisen die Messwerte die größte Streuung beziehungsweise Varianz auf, welche bei Zufallssignalen mit einem hohen Informationsgehalt gleichgesetzt werden kann.

Die Berechnung der PCA setzt sich entsprechend aus folgenden Schritten zusammen: Für die Transformation wird die Kovarianzmatrix der Eingangsgrößen $\mathbf{COV}(f) = \mathbf{S}_f$ der Dimension $N \times N$ benötigt, welche in die Kovarianzmatrix der Ausgangsgrößen $\mathbf{COV}(g) = \mathbf{S}_g$ überführt wird. Damit die Ausgangsgrößen de-korreliert sind, werden die Matrixelemente von \mathbf{S}_g außerhalb der Hauptdiagonalen Null. Die Transformation geschieht durch die Beziehung:

$$\mathbf{S}_g = \mathbf{X} \cdot \mathbf{S}_f \cdot \mathbf{X}^T. \quad (1.1)$$

Diese wird unter der Bedingung, dass die Transformationsmatrix \mathbf{X} orthogonal ist, aus der allgemeinen Beziehung

$$\mathbf{S}_g = \mathbf{X} \cdot \mathbf{S}_f \cdot \mathbf{X}^{-1} \quad (1.2)$$

⁶*Principal Component Analysis* ist die englische Bezeichnung für Hauptkomponentenanalyse.

hergeleitet. Die Matrix \mathbf{X} der Dimension $N \times M$ wird aus den Eigenvektoren \mathbf{V}_ν von \mathbf{S}_f gebildet, damit die Matrix \mathbf{S}_g aus Hauptdiagonalelementen, in diesem Fall den Eigenwerten λ_ν , besteht. Für die Eigenvektoren von \mathbf{S}_f gilt:

$$\mathbf{S}_f \mathbf{V}_\nu = \lambda_\nu \mathbf{V}_\nu. \quad (1.3)$$

Diese Gleichung ist lösbar, wenn folgende Bedingung erfüllt ist:

$$\det(\mathbf{S}_f - \lambda_\nu \mathbf{E}) = 0 \quad (1.4)$$

Vektor \mathbf{E} ist ein Einheitsvektor der Dimension M . Die einzelnen Eigenvektoren \mathbf{V}_ν können bestimmt werden. Daraus leitet sich die Berechnung der N dekorrelierten Zufallsvektoren \mathbf{g}_ν ab:

$$\mathbf{g}_{\nu\mu} = \mathbf{V}^T(\lambda_\nu) \cdot [f_{1\mu} \ f_{2\mu} \ \dots \ f_{N\mu}]^T. \quad (1.5)$$

Gewöhnlich werden die Ausgangsgrößen \mathbf{g}_ν absteigend nach ihrem Informationsgehalt sortiert, der durch die Größe des Eigenwerts λ_ν identifiziert wird. So kann durch Verwendung der R relevantesten, unkorrelierten Ausgangsgrößen mit $R < N$ die Datenmenge reduziert werden. Die Eingangsvektoren lassen sich durch Rücktransformation wiederherstellen, wobei der Rekonstruktionsfehler von der Anzahl der Ausgangsvektoren abhängt.

Die Hauptachsentransformation wird als problem- oder prozessabhängige Transformation bezeichnet, da die Transformationsmatrix an die statischen Eigenschaften des signalgenerierenden Prozesses angepasst wird [MH04]. So muss für jeden Datensatz eine eigene Transformationsmatrix berechnet werden. Für nicht normal verteilte Datensätze ist die PCA keine optimale Lösung, da die Daten nach Anwendung der Principal Component Analysis nicht vollständig dekorreliert sind. Anwendungsbeispiele der Hauptkomponentenanalyse sind die Bildanalyse und Signalkomprimierung sowie die automatische Trennung und Klassifizierung von Merkmalen.

Maschinelles Lernen

Nach Nilsson wird unter maschinellem Lernen verstanden, dass ein künstliches System Änderungen an einer Struktur, einem Programm oder Daten basierend auf Eingaben oder externen Informationen in der Weise durchführt, dass es seine zukünftige Leistung verbessert [Nil96]. Das maschinelle Lernen wird immer dann be-

nötigt, wenn für ein gegebenes Problem kein exaktes Modell existiert, aber die Eingaben und in Abhängigkeit der Methode auch die zu erzielenden Ausgaben bekannt sind. Es sind einige Belegungen der Ein- und Ausgaben, in ihrer Gesamtheit als Lerndaten bezeichnet, bekannt, auf deren Basis das System trainiert wird. Für den späteren Einsatz muss das Gelernte soweit verallgemeinerbar sein, dass für unbekannte Eingabe eine korrekte Ausgabe oder Vorhersage getroffen werden kann. Die Stärke des Lernens liegt daher darin, Gesetzmäßigkeiten zu erkennen und daraus resultierend unbekannte Daten korrekt zu bewerten. Allgemein wird in drei Schemata untergliedert, nach denen Lernalgorithmen entwickelt werden. Zum einen gibt es das überwachte Lernen, bei dem ein Algorithmus auf Grundlage vorgegebener Paare von Ein- und Ausgabewerten lernt. Dagegen wird beim unüberwachten Lernen mit Hilfe einer Menge von Eingabewerten ein Modell erzeugt, welches die Eingaben selbst beschreibt und Zusammenhänge zwischen unterschiedlichen Dimensionen der Eingaben erkennt, die nicht beobachtbar sind. Als drittes Schema ist das bestärkende Lernen zu nennen, wobei ein Algorithmus durch Belohnung und Bestrafung Verhaltensmuster eines Agenten in bestimmten Situationen erlernt.

Künstliche Evolution

Das Konzept der künstlichen Evolution wird nachfolgend basierend auf dem Nachschlagewerk [NF00] zusammenfassend dargestellt:

Die künstliche Evolution bezieht sich auf die vereinfachte Anwendung evolutionärer Prozesse auf Probleme der Informatik. Im Bereich der Steuerungssysteme von Robotern werden häufig generische Algorithmen auf eine der folgenden drei Strukturen angewendet: neuronale Controller, Parameter vordefinierter Programme oder die Struktur der Programme selbst.

Der evolutionäre Algorithmus arbeitet auf einer Population künstlicher Chromosomen beziehungsweise Genen von Individuen. Es handelt sich um eine Folge von Parametern (Genotyp), welche die Charakteristik des Individuums (Phänotyp) kodiert. Die Evolution läuft in mehreren Zyklen ab, den so genannten Generationen. Ziel ist die selektive Reproduktion und zufällige Veränderung der Gene um die Leistungsfähigkeit des einzelnen Individuums zu erhöhen. Zur Bewertung des Phänotyps wird ein Leistungskriterium durch die sogenannte Fitness-Funktion und das Individuum evaluiert. Wird beispielsweise die Evolution auf eine Parametrisierung einer mathematischen Funktion angewendet, deren Fehler bezüglich einer vorgegebenen Menge von Zielwerten gering werden soll, dann ist die Fitness des Individuums umso

höher, je näher die Funktion die vordefinierten Werte annähert. Ein einfacher generischer Algorithmus beginnt mit einer zufälligen Belegung aller Gene. Jede neue Generation entsteht durch Anwendung der evolutionären Regeln: selektive Reproduktion, Mutation und Kreuzung.

1. Die *selektive Reproduktion* beinhaltet das Auswählen der besten Individuen einer Generation, welche unverändert in die nächste Generation übernommen werden.
2. Unter der *Mutation* wird die zufällige Veränderung der Gene eines Individuums verstanden.
3. Die *Kreuzung* führt dazu, dass die Gene von zwei Individuen kombiniert werden und daraus ein neues Individuum entsteht.

Durch Parameter wie die Mutationswahrscheinlichkeit des Individuums oder des Gens im Individuum sowie die Anzahl an zu übernehmenden und zu kreuzenden Individuen kann der Verlauf der Evolution beeinflusst werden. Bei idealer Parametrisierung ist davon auszugehen, dass bei hinreichend vielen Wiederholungen die Fitness des besten Individuums nicht mehr zunimmt und somit ein lokales Maximum der Leistungsfähigkeit gefunden wurde.

1.5. Gliederung der Arbeit

In der vorliegenden Arbeit werden alle Aspekte thematisiert, die zur Realisierung der Richtungsbestimmung von Schallquellen notwendig sind. Im Kapitel 2 *Beschreibung und Evaluation der Hardware* wird die Zielplattform des humanoiden Roboters vorgestellt und der physikalische Aufbau des zur Untersuchung verwendeten Experimentalsystems beschrieben. Es wird gezeigt, dass der Roboter die physikalischen, sensorischen und elektrischen Voraussetzungen erfüllt, um die benötigten richtungsabhängigen Schallinformationen digitalisieren zu können. Dazu zählt die Betrachtung der äußeren Form des Kopfes und des künstlichen Außenohres, das für die Modulation richtungsabhängiger Schallinformationen von Bedeutung ist. Um die Audiosignale möglichst verlustfrei zu digitalisieren, müssen geeignete Sensoren und elektronische Komponenten ausgewählt werden. Dies wurde teilweise bereits in einer früheren Evaluation [Wol08] untersucht und wird hier noch einmal zusammenfassend dargestellt. Eine Neuerung in dieser Arbeit ist die Verwendung einer

Platine zur Analog-Digital-Wandlung der Stereosignale, die für den Einsatz im Roboterkopf entwickelt wurde. Im Anschluss an die Beschreibung der Hardware werden in Kapitel 3 *Experimentaldesign und Reizgenerierung* Annahmen bezüglich des späteren Anwendungsszenarios getroffen und Anforderungen für die Schallquellen-Lokalisation abgeleitet. Daraus werden ein Experimentalaufbau sowie ein statisches Schallsignal entworfen, welches zur Richtungsbestimmung genutzt wird. Das Experimentaldesign ist für die Generierung eines Testdatensatzes erforderlich, mit dessen Hilfe die Konzeption und Untersuchung des Modells erfolgt. Im Kapitel 4 *Konzeption der Modelle und Verfahren* werden aufbauend auf dem entworfenen Schallsignal Modelle und Verfahren zur Onset-Detektion und Bestimmung der horizontalen und vertikalen Richtung entwickelt. Die Richtungsanalyse wird im Weiteren getrennt betrachtet, da die Verarbeitung unterschiedlicher Signaleigenschaften im Vordergrund steht. In Kapitel 5 *Untersuchung und Auswertung* wird nachgewiesen, dass das entwickelte Verfahren grundsätzlich funktioniert. Zusätzlich wird die Genauigkeit des Verfahrens bei der Anwendung auf den Testdatensatz des vorliegenden Roboterkopfes untersucht und in ausgewählten Fällen die Implementierung erläutert sowie konkrete Belegungen der Modellparameter genannt. Kapitel 6 stellt die Zusammenfassung der Ergebnisse und den Ausblick auf weiterführende Arbeiten dar.

2. Beschreibung und Evaluation der Hardware

Im ersten Abschnitt dieses Kapitels wird die Plattform des humanoiden Roboters vorgestellt, auf der die Schallquellen-Lokalisation implementiert werden soll. Zum Experimentieren und zur Untersuchung dient jedoch ein Ersatzsystem, welches die Entwicklung und Analyse vereinfacht. Es besteht aus einem Prototyp des Roboterkopfes, der eine realistische Schallaufnahme vergleichbar mit der des finalen Systems erlaubt. Des Weiteren dient ein separater Lautsprecher zur Schallerzeugung sowie ein PC, der zur Signalgenerierung, -verarbeitung und Untersuchung der Lokalisationsverfahren verwendet wird. Die ausführliche Beschreibung sowie die Evaluation des Roboterkopfes befindet sich im zweiten Unterkapitel *Experimentalsystem*.

2.1. Humanoider Roboter

Als Zielplattform für die Realisierung der Schallquellen-Lokalisation dient ein humanoider Roboter (Abbildung 2.1). Der Roboterkopf ist mit einem binauralen Audiosystem zur Schallquellen-Lokalisation ausgestattet. Ein binaurales System ist im einfachsten Fall ein Sensorsystem aus zwei Mikrofonen, das sich auf einer horizontalen Achse befindet. Handelt es sich jedoch um ein humanoides System, wird es um zwei künstliche, nach menschlichem Vorbild geformte *Pinnae*¹ erweitert. Zudem befinden sich die Kunstohren nicht frei schwebend im Raum, sondern sind seitwärts in einer ovalen Schale eingelassen, welche die Proportionen des menschlichen Kopfes aufweisen (Abbildung 2.2). Als *Pinnae* dienen zwei industriell gefertigte Silikon-Ohrmuscheln. Miniaturmikrofonkapseln führen das einfallende Signal zur eingebetteten elektronischen Hardware, welche die Schallquellen-Lokalisation ausführen soll. Diese ist identisch mit der elektronischen Hardware des Roboterkopfes im Experimentalsystem und wird im Abschnitt 4.2.2 *Prototyp des Roboterkopfes* genauer vor-

¹*Pinna* (Plural *Pinnae*) ist die lateinische Bezeichnung für Ohrmuschel.



Abbildung 2.1. – Darstellung des 1,25 m großen autonomen humanoiden Roboters, welcher im Rahmen des Projektes ALE-AR entwickelt wurde.

gestellt. Zusätzlich verfügt der Kopf über drei Freiheitsgrade² zum seitlichen Neigen und horizontalen und vertikalen Ausrichten des Kopfes auf ein Objekt, zum Beispiel eine Schallquelle. Jedes Gelenk besitzt jedoch nur einen begrenzten Aktionswinkel, welcher der menschlichen Bewegungsfreiheit nachempfunden wurde. Die Rotation des Kopfes um die vertikale Achse ist beispielsweise auf den Bereich zwischen $+90,0$ und $-90,0$ Grad beschränkt. Folglich muss für eine komplette 180-Grad-Drehung zusätzlich die Rotation des gesamten Roboters oder des Torsos durchgeführt werden, wie es ebenfalls beim Menschen notwendig ist.

²Ein Freiheitsgrad stellt eine Dimension dar, in der sich das physikalische System frei bewegen kann. Es handelt sich um die Rotation um eine Achse oder die Translation in einer vorgegebenen Richtung.

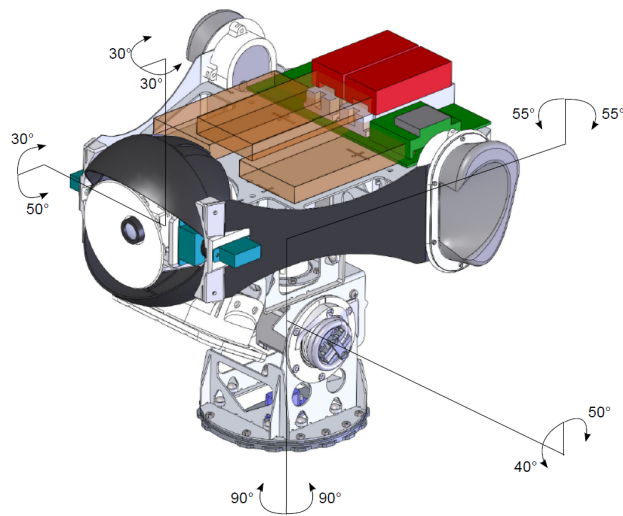


Abbildung 2.2. – Darstellung des humanoiden Roboterkopfes ohne Verschalung. Ausgestattet ist er unter anderem mit einem Kamerasystem zur Bildverarbeitung, einem Lautsprecher zur Schallerzeugung sowie einem binauralen Audiosystem zur Schallquellen-Lokalisation. Das System verfügt über drei Gelenke im Hals zur Rotation des Kopfes sowie zwei weiteren im Auge, welche jeweils die eingezeichneten Öffnungswinkel besitzen.

Die vorliegende Arbeit befasst sich ausschließlich mit der Lokalisation der Schallquelle, nicht aber mit der anschließenden Steuerung zur Ausrichtung des Roboters. Methoden zur Ausrichtung eines mobilen autonomen Roboters auf eine Schallquelle werden beispielsweise in [AHSK04], [Hil07] und [KKOO08] vorgestellt.

2.2. Experimentalsystem

Für die Konzeption und Untersuchung von Verfahren zur Schallquellen-Lokalisation wird ein experimentell erzeugter Datensatz von richtungsabhängigen Stereoaufnahmen benötigt. Zur Aufnahme dient wie bereits erwähnt ein *Prototyp des Roboterkopfes*, der im nachfolgenden Abschnitt vorgestellt wird. Es werden die externen Merkmale des Kopfes vorgestellt und die Ergebnisse einer Voruntersuchung zur Eignung des Kopfes zur Schallquellen-Lokalisation zusammenfassend dargestellt. Zusätzlich wird die eingebettete, elektronische Hardware beschrieben, die im Experiment zur Aufnahme der Audiodaten verwendet wird und im finalen Roboterkopf der Analyse der Schallsignale und Richtungsbestimmung dienen soll. Anschließend wird im

zweiten Abschnitt *Wiedergabesystem und zusätzliche Messtechnik* die Hardware vorgestellt, die zur Erzeugung und Wiedergabe von Schallsignalen und zur Festlegung physikalischer Umgebungsparameter während der Messung dient.

2.2.1. Prototyp des Roboterkopfes

Im Experiment dient ein Prototyp des Roboterkopfes als Messsystem, der speziell für die Untersuchung der Schallquellen-Lokalisation entwickelt wurde. Dieser ist im Gegensatz zum finalen Roboterkopfsystem auf einem Ständer fixiert und weist eine vereinfachte, ovale Kopfform auf. Der Kunstkopf sowie der schematische Aufbau der integrierten Elektronik werden in Abbildung 2.3 dargestellt. Nachfolgend wird der Systemaufbau näher beschrieben und anschließend die Funktionsfähigkeit des Kopfes durch die Evaluation der Bestandteile nachgewiesen.

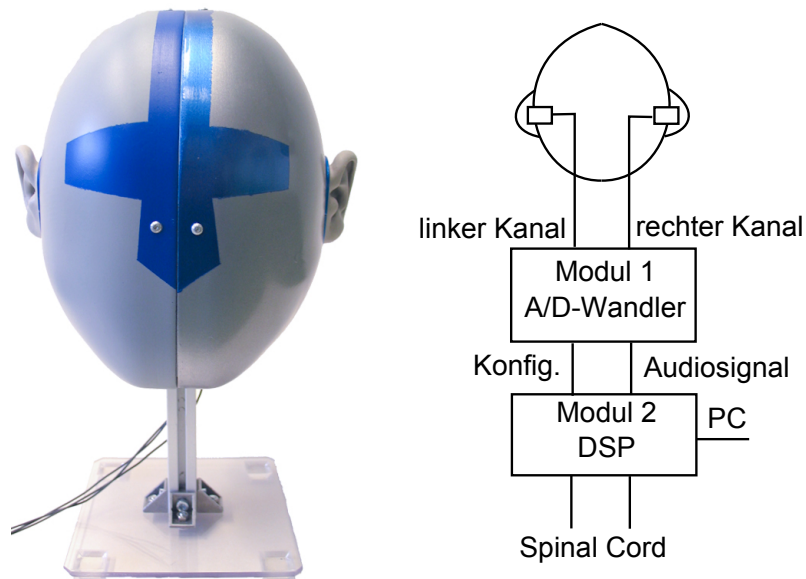


Abbildung 2.3. – Links: Der Prototyp des humanoiden Roboterkopfes, der aus einer ovalen Schale, zwei Silikon-Ohren und darin integrierten Miniaturmikrofonkapseln besteht. Im Inneren befindet sich die Elektronik, welche zur Aufnahme beziehungsweise Signalverarbeitung dient. Rechts: Schematische Darstellung der modularen elektronischen Hardware des Roboterkopfes.

Beschreibung des Systems

Die Schale des Kopfes besteht aus Kunststoff und weist eine ovale Form auf. Zwei nach menschlichem Vorbild geformte Silikon-Ohren befinden sich seitwärts des Kopfes auf Augenhöhe. In ihnen befindet sich je eine Mikrofonkapsel des Typs WM61A von Panasonic. Der Abstand der Mikrofonkapseln beträgt 18 cm. Diese sind über circa 30 cm lange Kabel mit der elektronischen Hardware verbunden, welche modular in zwei Platinen gegliedert ist. Das erste Modul ist mit einem *Analog-Digital-Wandler* (A/D-Wandler) ausgestattet und dient somit der Digitalisierung der analogen Audiosignale. Der Schaltplan der Platine wird im Anhang in Abbildung A.1 dargestellt. Neben der Entwicklung des Layouts wurden unter anderem die Bestückung der Platine, der Test und die Inbetriebnahme im Rahmen dieser Arbeit durchgeführt. Eine fehlerfreie Funktionsweise wurde für eine Samplingrate von 46 875 Hz³ bei einer Signalaufösung von 16 bit nachgewiesen, welche im Weiteren genutzt wird.

Das zweite Modul ist mit einem Mikrocontroller ausgestattet und dient somit der digitalen Signalverarbeitung. Es handelt sich um einen ARM-basierten 32-Bit-Prozessor der Firma STMicroelectronics mit der vollständigen Bezeichnung STM32-F103CBT6. Dieser ist mit einem 128 kB-Flash-Speicher und einem Arbeitsspeicher von 20 kB ausgestattet. Die maximale Taktfrequenz beträgt 72 MHz. Die vergleichsweise geringe Rechenleistung sowie der sehr beschränkte Arbeitsspeicher müssen wie bereits erwähnt bei der Konzeption des Analyseverfahrens zur Schallquellen-Lokalisation berücksichtigt werden, da auf dieser Hardware die Berechnung im finalen System ablaufen soll. Beide Platinen sind über eine I2C-Schnittstelle zur Konfiguration des A/D-Wandlers und eine I2S-Schnittstelle zur Übertragung der 16-Bit Audiodaten miteinander verbunden. Die Signalverarbeitungsplatine verfügt zusätzlich über eine serielle Schnittstelle, welche zur Datenübertragung zum PC genutzt wird. Diese Schnittstelle ist nur im Entwicklungssystem verfügbar. Wird die elektronische Hardware im finalen Roboterkopf eingesetzt, findet die Kommunikation über einen Datenbus, den so genannten *Spinal Cord*, statt, der die Prozessoren des verteilten Systems des Roboters miteinander verbindet. In diesem Fall findet die Richtungsanalyse des Roboters autonom auf dem eingebetteten System statt. Das Resultat der Schallquellen-Lokalisation wird über den Spinal Cord verbreitet, sodass eine andere Routine die Ausrichtung des Roboters auf die Schallquelle ausführen kann.

³Die Standard-Samplingrate von 44 100 Hz, war nicht verfügbar.

Evaluation des Roboterkopfes

In der selbst durchgeführten Voruntersuchung *Machbarkeitsstudie zur Realisierung der Phonotaxis am humanoiden Roboter mit Hilfe eines binauralen Audiosystems* [Wol08] wurden das Sensorsystem sowie die äußeren Merkmale des Roboterkopfes untersucht, um zu prüfen, ob die für die Schallquellen-Lokalisation benötigten Eigenschaften vorliegen.

Der erste Teil der Untersuchung beschäftigte sich mit der Evaluation eines geeigneten Mikrofonkapseltyps für den Einsatz im humanoiden Roboterkopf. Dazu wurden 14 handelsübliche miniaturisierte Mikrofonkapseln ausgewählt und hinsichtlich der klassischen Kriterien Sensitivität, Signal-Rausch-Verhältnis (SNR), lineare und nicht-lineare Verzerrungen untersucht. Zusätzlich wurden die Kapseln auf die Einstreuung von Vibrationen über die Kapselwand untersucht. Dies sollte als Indikator für die Störanfälligkeit durch den körpereigenen Schall dienen, der durch die Motoren in den Gelenken verursacht wird. Da die von den Servomotoren ausgehenden Vibrationen direkt über die Roboterkonstruktion zur Mikrofonkapsel über-

Typ	Sensitivität	SNR	Lineare Verzerrung	Nichtlin. Verzerrung	Kapselwand	Gesamt
EM-1050	+	+	-	○	+	+2
EM-6022BC	○	+	-	-	○	-1
EM-6027	-	○	--	-	-	-5
EM-6035B	+	+	-	○	+	+2
EM-6050U	-	○	○	○	+	0
EMY-6018RBC	○	○	--	+	+	0
EMY-62NL101	-	+	--	-	○	-3
POM-2735PR	+	+	-	○	+	+2
POM-2746LR	○	+	○	+	○	+2
POM-2746PR	-	+	-	○	○	-1
PUM-3546LR	-	-	○	○	○	-2
WM-61A	+	+	○	+	+	+4
WM-64MNT	+	○	++	+	○	+4
WM-64PNT	○	○	○	+	+	+2

Tabelle 2.4. – Zusammenfassung der Mikrofonkapsel-Untersuchung aus der Voruntersuchung. Die verwendeten Symbole besitzen die Wertigkeiten: +2 (++), +1 (+), 0 (○), -1 (-), -2 (--). Die Gesamtbewertung jedes Kapseltyps in der letzten Spalte ergibt sich aus der Summe der Bewertungen in den fünf Disziplinen. Sie erstreckt sich von -5 bis +4 und dient als Indikator für die Eignung des Kapseltyps für den Einsatz in einem humanoiden Roboter.

tragen werden, kann dies die Amplitude des externen Schalls deutlich übersteigen. Können die Signale nicht ausreichend gedämpft werden, ist während der Bewegung des Roboters keine Schallquellen-Lokalisation möglich oder es muss mit zusätzlichen Filtertechniken zur Signalverbesserung gearbeitet werden. Die Ergebnisse der Kapseluntersuchung werden in Tabelle 2.4 zusammengefasst. Die Bewertung Null entspricht dem Durchschnittswert aller Messungen, sodass ein Wert größer Null für eine überdurchschnittliche und ein Wert kleiner Null für eine unterdurchschnittliche Bewertung steht. Die Gesamtbewertung eines Typs ergibt sich aus der Summe der Einzelbewertungen. Die genaue Klassifizierung der Einzelbewertungen ist in der Voruntersuchung nachzulesen.

Eine Kapsel mit überdurchschnittlicher Bewertung ist der Typ WM-61A von Panasonic. Es handelt sich um eine Kondensator-Mikrofonkapsel mit einem Durchmesser von 6 mm, omnidirektionaler Richtcharakteristik und Vibrationsresistenz. Der Kapseltyp hebt sich von anderen unter anderem auf Grund seines hervorragenden Signal-Rausch-Verhältnisses von 62 dB und der hohen Sensitivität von -35 dB ab. Der lineare Frequenzbereich liegt zwischen 20 Hz und 2 kHz. In der Untersuchung konnte zusätzlich festgestellt werden, dass der Anteil an nicht-linearer Verzerrung sowie die Anfälligkeit auf Störeinstrahlung durch die Kapselwand vergleichsweise niedrig sind. Da dieser Kapseltyp im relativen Vergleich in der Gesamtbewertung eine der beiden höchstbewerteten war, wird er für die Realisierung der Schallquellen-Lokalisation eingesetzt. Er wurde dem gleichbewerteten Kapseltyp WM-64MNT vorgezogen, da er sich löten und sich somit leichter verbauen lässt.

Im zweiten Teil der Untersuchung wurde die Eignung der äußeren Form des Roboterkopfes in Kombination mit dem Kunststoff analysiert. Diese ist für die richtungsabhängige Frequenzveränderung des Schalls verantwortlich, auf deren Basis die vertikale Schallquellen-Lokalisation durchgeführt wird [Beg94]. Dazu wurden monaurale *head related transfer functions* (HRTFs) gemessen, analysiert und mit einem Referenzsystem verglichen. Unter einer HRT-Funktion wird eine richtungsabhängige Übertragungsfunktion verstanden, welche im Ohr gemessen wird. Die Frequenzanteile des Signals werden in Abhängigkeit der Einfallsrichtung auf unterschiedliche Weise durch die äußere Form des Kopfes und insbesondere der Ohren so beeinflusst, dass der Schall einer linearen richtungsabhängigen Filterung unterliegt. Mit geeigneten Verfahren kann durch Analyse monauraler und binauraler Eigenschaften, die aus den HRTF extrahiert werden, auf die vertikale und teilweise auch auf die horizontale

Einfallrichtung des Direktschalls geschlossen und so die Ausrichtung der Schallquelle zum Kopf rekonstruiert werden. Nach Gierlich [Beg94] wird die HRTF durch die Pinnae im Bereich zwischen 2 bis 14 kHz richtungsabhängig beeinflusst, während der Kopf nur im tieffrequenten Bereich zwischen 500 Hz und 1,6 kHz wirkt. Ein wichtiger Aspekt der Voruntersuchung war es, in diesem einfachen Roboterkopfsystem auffällige, richtungsabhängige Frequenzunterschiede festzustellen. In einem ersten Schritt wurde ein Frequenzband identifiziert, das in allen verfügbaren Frequenzgängen des Systems eine Amplitudenverstärkung aufweist. Es befindet sich zwischen 3,5 und 4,5 kHz und ist auf die Resonanz des Cavuum Concha⁴ zurückzuführen. Dagegen sind im Frequenzband zwischen 4 und 14 kHz Unterschiede des Betragsspektrums erkennbar, die sich mit der Variation der vertikalen und horizontalen Schallquellenposition ändern. Dies wird anhand ausgewählter Frequenzgänge in Abbildung 2.5 deutlich. Die überlagerten Betragsspektren aufeinander folgender Schallquellenpo-

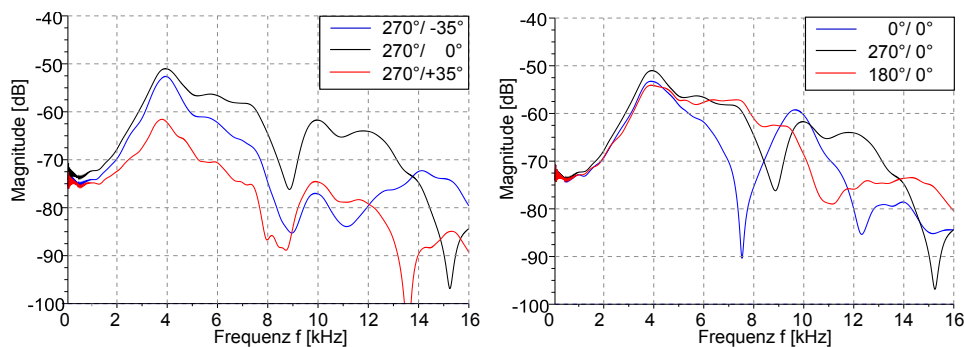


Abbildung 2.5. – Darstellung ausgewählter, monauraler Frequenzgänge, welche mit einem Referenzmikrofon im rechten Ohr des Prototyps des Roboterkopfes gemessen wurden. Es ist deutlich erkennbar, dass sowohl bei horizontaler (links) als auch bei vertikaler (rechts) Variation der Schallquelle deutliche Unterschiede im Pegel und Verlauf der Frequenzgänge auftreten.

sitionen lassen erahnen, dass die Übergänge zwischen den Spektren teilweise linear verlaufen, was für die Konzeption des Verfahrens zur vertikalen Richtungsanalyse von Bedeutung ist.

Als Nächstes wurden Vergleiche zwischen diesem System und einem konkreten Referenzsystem mit der Bezeichnung *Knowles Electronic Manikin for Acoustic Research* (KEMAR) angestellt. Es handelt sich bei dem Vergleichsobjekt um einen standardisierten, nach menschlichem Vorbild geformten Kunstkopf, welcher mit hochprä-

⁴*Cavuum Concha* ist die lateinische Bezeichnung für Ohrkanal.

ziser Audiomesstechnik ausgestattet ist. Mit diesem ist es möglich, die akustische Wahrnehmung des Menschen realistisch abzubilden. Der Vergleich der beiden Systeme wurde auf Basis eines umfangreichen HRTF-Datensatzes⁵ durchgeführt, den Bill Gardner und Martin Keith vom Massachusetts Institute of Technology (MIT) zu Forschungszwecken frei zur Verfügung stellten [GK94]. Ziel des Vergleichs war es, ähnliche richtungsabhängige Eigenschaften der Betragsspektren des Referenzsystems ebenfalls im vorliegenden System zu identifizieren. Dazu wurde unter anderem die spektrale Position eines für die vertikale Ausrichtung dominierenden Merkmals verglichen. Es handelt sich um eine schmalbandige Dämpfung im Frequenzspektrum, die alternativ als Tal beziehungsweise Dip bezeichnet wird. Es wurden jeweils Spektren beider Systeme mit gleicher Schallquellenposition verglichen. Tatsächlich befindet sich der Dip in beiden Messungen in ähnlichen Frequenzbändern. Um zu

Elevation [°]	Prototyp		KEMAR	
	Frequenzband der Dips [Hz]	Breite [Hz]	Frequenzband der Dips [Hz]	Breite [Hz]
-40	7200–8600	600	6200–7000	800
-20	7900–8500	600	6600–7800	1200
0	9000–9100	100	8200–9200	1000
+20	10300–10700	400	9400–11200	1800
+40	10400–13400	3000	10800–11800	1000

Tabelle 2.6. – Vergleich der Messungen des Prototypen mit denen des Referenzsystems KEMAR. Untersucht wird die Position der Frequenzbänder eines markanten, vom Erhebungswinkel abhängigen Dips.

zeigen, dass der Dip eine überwiegend vertikale Bedeutung hat, werden die Positionen des Dips von mehreren Messungen der gleichen vertikalen, aber verschiedenen horizontalen Richtungen zu einem Frequenzband zusammengefasst (siehe Tabelle 2.6). Da sich die Bänder mit wachsender Höhe der Schallquelle in beiden Systemen auf ähnliche Weise verschieben, kann geschlossen werden, dass ein vergleichbares Merkmal beobachtet wurde und es sich um eine Eigenschaft handelt, die vom Erhebungswinkel abhängig ist. Der Frequenzbereich des Dips befindet sich beim vor-

⁵Grundlage für die Messungen war das Kunstkopf-Modell DB-4004, der sich auf einem Torso befand. Als künstliches Pinna kam das Modell DB-061 und das Mikrofon Etymotic ER-11 zum Einsatz.

liegenden System zwischen 7,2 und 13,4 kHz, während er sich beim KEMAR von 6,2 bis 11,8 kHz erstreckt. Die leichte Abweichung der Frequenzbereiche kann auf die unterschiedliche äußere Gestalt der Systeme zurückgeführt werden. Bei genauerer Betrachtung ist die Breite der Bänder beim vorliegenden Prototyp ungleichmäßiger verteilt als beim Referenzsystem. Auch überschneiden sich aufeinander folgende Bänder stärker. Somit wäre eine Lokalisation allein auf diesem Merkmal grundsätzlich möglich, unterläge jedoch einer starken Ungenauigkeit. Aus dieser Analyse und einigen weiteren theoretischen Überlegungen konnte darauf geschlossen werden, dass dieses binaurale System im Hinblick auf die monauralen Filtereigenschaften bei unterschiedlichen Schallquellen-Richtungen mit dem KEMAR-Kopf und somit dem menschlichen Kopf vergleichbar ist. Es ist daher geeignet, in der weiteren Konzeption und Untersuchung der Schallquellen-Lokalisation verwendet zu werden.

2.2.2. Wiedergabesystem und zusätzliche Messtechnik

Als Schallquelle kommt der aktive Lautsprecher HS50M von Yamaha zum Einsatz. Er besitzt einen regelbaren Leistungsverstärker und einen separaten Tief- und Hochtoner. Die Verstärkung wird auf +4 dB voreingestellt. Der lineare Frequenzgang des Lautsprechers liegt zwischen 50 Hz und 20 kHz. Die Generierung und Wiedergabe des Signals wird über einen PC gesteuert. Als externes Audio-Interface dient das EMU 0404-USB von Creative Labs, welches mit einem konfigurierbaren D/A-Wandler ausgestattet ist. Das Signal wird mit einer Abtastrate von 44 100 Hz und einer Signalaufösung von 16 bit generiert. Daher ist ein maximales Signal-Rausch-Verhältnis von 90 dB möglich, die höchste erzeugbare Frequenz liegt bei 22 050 Hz.

Zur Messung des physikalischen Schalldruckpegels, alternativ *sound pressure level* (SPL) genannt, wurde das SoundLevelMeter 5055 von Peak-Tech verwendet. Die Angabe der Pegel erfolgt A-gewichtet, sodass die Amplitudenbewertung vergleichbar mit der menschlichen Wahrnehmung ist.

3. Experimentaldesign und Reizgenerierung

Im vorherigen Kapitel wurde die verfügbare Hardware zur Schallquellen-Lokalisierung vorgestellt. Auf dieser Basis wird im vorliegenden Kapitel ein Experimentalaufbau entwickelt und mit Hilfe eines geeigneten Signals ein Audiodatensatz aufgenommen. Im ersten Abschnitt *Experimentalanforderung* werden alle Anforderungen aufgezählt, welche sich aus dem möglichen Anwendungsszenario der Schallquellen-Lokalisation ergeben. Daraus werden vereinfachende Annahmen bezüglich des vorliegenden Experimentalaufbaus getroffen. Der Entwurf eines akustischen Signals wird im zweiten Abschnitt *Design des akustischen Reizes* beschrieben. Aufgrund des begrenzten Arbeitsspeichers ist es notwendig, ein spezielles Signal zu konzipieren, das sowohl eine geringe Länge aufweist, um vollständig im Speicher des STM32 gepuffert zu werden als auch die benötigten Informationen zur Richtungsbestimmung enthält. Auf Basis des Testsignals werden im letzten Abschnitt *Experimentaldesign und Messung* ein Experimentalaufbau aus den Schlussfolgerungen der Analyse der Experimentalanforderung entwickelt sowie die Durchführung der Audiomessungen geschildert.

3.1. Experimentalanforderung

Die Aufgabe, ein Verfahren zur Schallquellen-Lokalisation zu entwickeln, welches unter allen Bedingungen verlässlich funktioniert, ist zu umfangreich, um in dieser Arbeit vollständig bewältigt zu werden. Daher werden die zu erwartenden Bedingungen der möglichen Anwendungsdomäne der Schallquellen-Lokalisation am Beispiel der Sprachexperimente im Rahmen des Projekts ALEAR analysiert. Folgende Annahmen bezüglich des realistischen Experimentalaufbaus und der Umwelt werden getroffen:

1. Das Experiment findet in einem nichthallfreien, geschlossenen Raum mit mehreren autonomen Robotern und verschiedenen unbewegten Objekten statt. Jeder Roboter soll in der Lage sein, die anderen Roboter anhand ihrer Audiosignale zu orten.
2. Die Roboter kommunizieren geräuschlos über eine Funkschnittstelle. Die akustischen Signale können von den Robotern gezielt genutzt werden, um beispielsweise auf natürliche Weise die Aufmerksamkeit der anderen Agenten auf sich zu ziehen.
3. Die zu ortende Schallquelle kann sich theoretisch an einem beliebigen Ort im Raum (auch in beliebiger Höhe) befinden. Am wahrscheinlichsten ist jedoch, dass sich die Schallquelle auf gleicher Höhe wie das binaurale System eines stehenden, ortenden Roboters befindet.
4. Der Roboter bewegt sich im Allgemeinen nicht permanent, sondern nur temporär zur Exploration des Raums, zur Reizzuwendung oder um bestimmte Aktionen auszuführen wie Laufen, Stabilisieren, Greifen, Zeigen, Gestikulieren. Andere akustische Signale, die als Störgeräusche interpretiert werden könnten, entstehen unter anderem durch die Motoren und Gelenke bei der Bewegung der Roboter.
5. Der normale Ruhepegel im Raum hat ein konstantes Niveau. Es wird hauptsächlich durch die Motorgeräusche im Haltezustand geprägt, die den Roboter aufrecht halten.

Aus diesen Bedingungen werden Schlussfolgerungen für den Experimentalaufbau der Audiomessungen in der vorliegenden Arbeit gezogen. Die Wahl des Experimentalaufbaus hat auch Folgen auf das Schallsignal¹, welches zur Ortung benötigt wird, und das Lokalisationsverfahren zur Berechnung der Richtung:

Aus Punkt 1 folgt für den Messaufbau, dass die Messungen nicht wie bei Audioaufnahmen üblich in einem hallfreien Messlabor durchgeführt werden sollten. Stattdessen wird ein geschlossener, durchschnittlich verhallter Raum wie ein Konferenzraum oder ein Büro verwendet. Beim Lokalisationsverfahren muss also das Problem der Schallreflexionen im Raum berücksichtigt werden. Die grundlegenden

¹Das Schallsignal, welches die Grundlage für die Lokalisation darstellt, wird im Weiteren auch *Ortungssignal* genannt.

Zusammenhänge wurden im Abschnitt *Schall und Schallreflexion* des Kapitels 1.3 *Einführung grundlegender Begriffe* vorgestellt.

Entsprechend Punkt 2 läuft der Informationsaustausch nicht akustisch ab, sondern geräuschlos per Funk. Dies erlaubt es, die Audiosignale, welche jeder Roboter mit einem Lautsprecher von sich gibt, gezielt für die Lokalisation zu verwenden. Der Messaufbau muss daher aus einem Lautsprecher, welcher die Schallquelle beziehungsweise den zu ortenden Roboter repräsentiert, und einem binauralen System bestehen, welches das Aufnahmesystem des ortenden Roboters ersetzt. Die Entwicklung des Lokalisationsverfahrens kann dann auf einem gewöhnlichen PC ablaufen.

Auch kann aus Punkt 2 geschlussfolgert werden, dass die Ortung verschiedener Schallquellen sequenziell ablaufen kann, da zu einer Zeit nur ein Roboter im Fokus der Aufmerksamkeit der anderen Roboter steht. Daher kann das Ortungsproblem so genannter konkurrierender Schallquellen² vernachlässigt werden. In dieser Arbeit wird von einem ungestörten Ortungssignal ausgegangen.

Aus Punkt 3 wird geschlossen, dass sich das binaurale Messsystem in einer Höhe eines stehenden Roboters befinden soll. Auch der Lautsprecher des zu ortenden Roboters befindet sich im Normalfall auf gleicher Höhe. Von diesem Aufbau soll in der vorliegenden Arbeit ausgegangen werden. Die horizontale Schallquellen-Lokalisation sollte eine Genauigkeit von im Mittel 5 bis 10 Grad aufweisen, um eine genaue horizontale Ausrichtung auf die Schallquelle zu ermöglichen. Im Falle der vertikalen Richtung wird mindestens eine grobe räumliche Zuordnung in drei Bereiche gefordert. Es soll unterscheidbar sein, ob Schall von oben, aus gleicher Höhe oder von unten kommt.

Entsprechend Punkt 4 muss zusätzlich nicht die Überlagerung durch Störgeräusche, die beispielsweise durch Motoren verursacht werden, berücksichtigt werden. Da diese Störgeräusche nur temporär auftreten, kann im Falle einer detektierten Störung, die Messung zu einem späteren Zeitpunkt erneut durchgeführt werden, wenn alle Roboter zur Ruhe gekommen sind. Dazu kann ein Kommando per Funk an alle Roboter übertragen werden, welches dazu führt, dass alle Körperaktionen für eine Weile eingestellt werden. Somit wird der Lokisationsalgorithmus auf eine ungestörte Audioaufnahme angewendet.

²Mehrere Roboter erzeugen gleichzeitig Schallsignale, welche sich gegenseitig überlagern. Die Aufgabe, ein Signal aus einer Menge überlagerter Signale herauszufiltern und zu lokalisieren, ist auch als Cocktail-Party-Effekt oder Blind-Source-Separation (BSS) bekannt.

Aus Punkt 5 können weitere Schlussfolgerungen für die Schallerzeugung gezogen werden. Der Schallpegel des zu ortenden Signals sollte deutlich über dem Umgebungsrauschen liegen, um ein möglichst gutes Signal-Rausch-Verhältnis zu erhalten. Der maximale Pegel des Signals wird durch die Ausgangsleistung des Lautsprechers des schallerzeugenden Roboters sowie durch den Abstand zwischen Schallquelle und Empfänger beschränkt. Daher kann kein absoluter Pegel vorgegeben werden. Der Rauschpegel im Raum kann in Abhängigkeit des gewählten Aufnahmeortes variieren. Während der Audiomessungen sollen sowohl Wiedergabe- als auch Rauschpegel konstant sein. Die Abstände zwischen Sender und Empfänger sind begrenzt.

Bevor nun das endgültige Experimentaldesign abgeleitet und angewendet werden soll, welches den Audiomessungen dieser Arbeit zu Grunde liegt, wird im folgenden Abschnitt ein geeignetes Ortungssignal für die Messungen entwickelt.

3.2. Design des Audiosignals zur Lokalisation

Unter einem *Reiz*, beziehungsweise einem *Stimulus*, wird ein physiologisch wirksamer innerer oder äußerer Faktor verstanden, der den Organismus zu einer Zustandsänderung (beispielsweise Empfindung) veranlasst [Sch94]. Diese Definition lässt sich analog auf die Robotik übertragen, wobei der Organismus durch den Roboter ersetzt wird. Der Reiz ist im vorliegenden Fall ein akustisches Signal, dessen Amplitude deutlich über dem normalen Rauschpegel der Umgebung liegt und somit durch Signalverarbeitungsmethoden erkannt und verarbeitet werden kann. Als Reaktion auf diesen Reiz ist es dem Roboter beispielsweise möglich, sich der Quelle zuzuwenden. Dieses Verhalten wird in der Biologie als *Phonotaxis* bezeichnet.

Ziel ist es, einen speziellen akustischen Reiz zu generieren, welcher unter Berücksichtigung der Einschränkungen der elektronischen Hardware des Roboterkopfes zur Analyse des Onsets sowie der horizontalen und vertikalen Lokalisation der Schallquelle geeignet ist. Alle Anforderungen werden im ersten Abschnitt *Analyse der Anforderungen* analysiert und daraus Bedingungen für die Signalstruktur abgeleitet. Anschließend wird im Abschnitt *Synthese des Audiosignals* das Signal konstruiert.

3.2.1. Analyse der Anforderungen

Ein Schallsignal zur zweidimensionalen Lokalisation der Schallquelle muss bestimmte Grundvoraussetzungen erfüllen. Diese werden nachfolgend aus verschiedenen Anforder-

derungen verschiedener Domänen hergeleitet. Die Experimentalanforderungen fließen dabei ebenfalls ein:

1. Anforderungen der Hardware

Aufgrund des beschränkten Arbeitsspeichers des Prozessors, der im Roboterkopf zur Audioaufnahme und -verarbeitung dient, darf das Schallsignal eine bestimmte Dauer nicht überschreiten. Der Puffer zur Aufzeichnung des 16 bit Stereosignals, soll maximal 4000 Samples (dies entspricht circa 8 kB) umfassen. Somit wird das Signal vollständig im Arbeitsspeicher des Signalprozessors gepuffert. Zusätzlich bleibt freier Speicher für die Verarbeitung. Daraus ergibt sich bei einer Abtastrate von 46 875 Hz eine maximale Dauer des Signals von 43 ms.

2. Anforderung der Onset-Detektion

Der Definition im Abschnitt *Signalstruktur und Onset* des Kapitels 1.3 kann entnommen werden, dass der Onset der früheste Zeitpunkt ist, an dem ein Signal erstmals vom Umgebungsrauschen unterscheidbar ist. Dies entspricht also der *time of arrival* (TOA) des Direktschalls in beiden Ohren. Damit die Detektion des Onsets mit einfachen Verfahren im Zeitbereich erfolgen kann, soll das Signal mit einer einfachen Sinusschwingung beginnen, die trotz Reflexionen und Umgebungsrauschen fehlerfrei identifiziert werden kann.

3. Anforderung der horizontalen Schallquellen-Lokalisation

Die horizontale Schallquellen-Lokalisation basiert grundsätzlich auf der Analyse der Laufzeitdifferenz des Direktschalls zwischen beiden Ohren, auch *interaurale time difference* (ITD) genannt [Bla69]. Diese kann beispielsweise durch Korrelation der beiden Kanäle im Zeitbereich bestimmt werden [BC78]. Dazu wird ein Signalvektor zeitlich zum anderen verschoben und die Ähnlichkeit der beiden Vektoren bestimmt. Die ITD ist die zeitliche Verschiebung mit der größten Übereinstimmung. Unterliegt das Ortungssignal, welches in beiden Ohren gemessen wird, bei der Übertragung denselben Filtereinflüssen, liefert die Korrelation ein verlässliches Ergebnis. Dies kann aufgrund der Symmetrie des Kopfes und der Ohren ausschließlich beim Einfall des Direktsignals aus der Medianebene garantiert werden. Die Laufzeitdifferenz und die Differenz der Signalspektren des Direktschalls ist in diesem Fall null. Zusätzlich

wird der Direktschall durch Reflexionen im Raum überlagert, die zu Mehrdeutigkeiten bei der Korrelation im Zeitbereich führen. Bekannte Verfahren weichen daher in den Frequenzbereich aus, wo die Korrelation separat in den einzelnen Bändern durchgeführt und anschließend die Verschiebung mit der größten Häufigkeit bestimmt [CNNU03], [CKN⁺08] wird. Diese Methode ist zwar verlässlicher, jedoch soll aus Komplexitätsgründen auf Verfahren im Frequenzbereich in dieser Arbeit verzichtet werden. Ein viel einfacherer Ansatz, ist die Berechnung der Laufzeitdifferenz auf Basis der TOA im Zeitbereich. Diese wird mittels der Onset-Detektion in beiden Kanälen bestimmt. Die zeitliche Differenz beider Onsets entspricht der gesuchten Laufzeitdifferenz. Es wird also eine hinreichend genaue Bestimmung der Onsets gefordert.

4. Anforderung der vertikalen Schallquellen-Lokalisierung

Wie schon in Kapitel 4.2.2 *Prototyp des Roboterkopfes* festgestellt wurde, werden für die Bestimmung der vertikalen Richtung einer Schallquelle monaurale oder binaurale Eigenschaften analysiert, um Rückschlüsse auf die Elevation der Schallquelle zu ziehen. Diese richtungsabhängigen Informationen werden in bestimmten Frequenzbändern des Schalls transportiert. Nach Gierlich [Beg94] erstrecken sich diese Bänder beim Menschen von 2 bis 14 kHz. Bei circa 3 kHz wird das Spektrum durch die Resonanzfrequenz des Ohrkanals überlagert, die eine richtungsunabhängige Anhebung des Pegels verursacht. Die Resonanzfrequenz kann in Abhängigkeit von der individuellen Länge des Ohrkanals variieren. Die Analyse der HRTFs des vorliegenden Roboterkopfes ergab eine Resonanz bei circa 4 kHz. Richtungsabhängige Frequenzen wurden anhand des markanten Dips bis 13,4 kHz gemessen. Es wird daher zum einen gefordert, dass das Frequenzspektrum mindestens 1 kHz vor der Resonanzfrequenz beginnt und bis 14 kHz verläuft, und zum anderen, dass die Frequenzanteile ohne aufwendige Transformation im Zeitbereich analysierbar sind.

Im nächsten Abschnitt wird ein geeignetes Audiosignal synthetisiert, das diesen Anforderungen genügt.

3.2.2. Synthese des Audiosignals

Als Grundlage für das Audiosignal wird ein so genannter *Sinussweep* verwendet. Es handelt sich um eine periodische Schwingungsfolge mit konstantem Pegel und linear

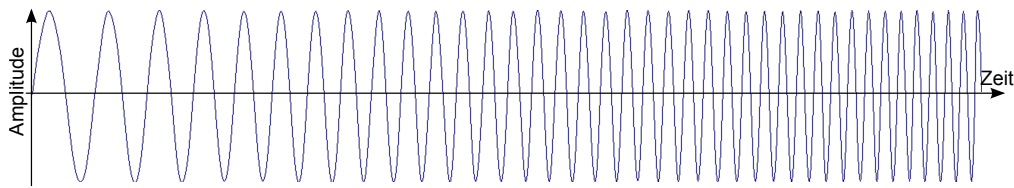


Abbildung 3.1. – Schematische Darstellung eines Sinussweeps mit einer linear steigenden Frequenz von 3 bis 14 kHz. Der zugehörige Audiovektor besteht aus 38 Schwingungen und hat eine Gesamtlänge von 4,5 ms.

oder logarithmisch steigender Frequenz. Die Frequenz des Sinussweeps soll linear von 3 auf 14 kHz steigen (siehe Abbildung 3.1). Die Signallänge beträgt 4,5 ms. Die geringe Dauer stellt einen Kompromiss dar. Je kürzer das Signal ist, desto weniger wird es durch störende Reflexionen überlagert. Berücksichtigt wurden die zu erwartenden Reflexionszeiten des Schalls (siehe Tabelle 5.7 auf Seite 69). Je länger das Signal ist, desto mehr Frequenzen können im Sinussweep dargestellt werden. In diesem Fall enthält das Signal 38 Schwingungen bei einer Frequenzauflösung von circa 280 Hz. Dies führt bei einer Abtastrate von 44 100 Hz zu einem Audiovektor von 200 Samples.

Mit diesem Signal werden die vier genannten Anforderungen erfüllt:

- Auf dem Aufnahmesystem beträgt die Signallänge 213 Samples, da nicht dieselbe Abtastrate auswählbar war, wie beim Wiedergabesystem. Gewählt wurde die Frequenz 46 875 Hz. Das 16-bit-Signal kann bei einem Umfang von circa 6,7 kB vollständig im Speicher des STM32 verarbeitet werden.
- Das Signal beginnt mit einer Sinusschwingung, welche trotz linearer Filterung eine deterministische Wellenstruktur aufweist. Der Onset beider Kanäle lässt sich somit im Zeitbereich analysieren.
- Aus den Onsets lässt sich die ITD bestimmen. Sie bilden somit die Grundlage zur Berechnung der horizontalen Richtung.
- Der Frequenzgang des Signals lässt sich im Zeitbereich analysieren. Das Spektrum umfasst den gewünschten Frequenzbereich von 3 bis 14 kHz. Anhand der Information über die Verhältnisse der Frequenzpegel können Rückschlüsse auf die vertikale Schallquellen-Richtung gezogen werden.

Die Analyse eines solchen Signals im Zeitbereich funktioniert wie bereits erwähnt nur unter der Voraussetzung, dass keine Störsignale auftreten. Sollte das Ortungssignal durch ein anderes überlagert werden oder das SNR zu gering sein, sind Fehlinterpretationen bei der Analyse möglich. Dies wird durch die Anforderungen, die an den Experimentalaufbau gestellt werden, ausgeschlossen. Nachfolgend wird der endgültige Aufbau des Audioexperiments entwickelt und die Generierung des Audiodatensatzes auf Basis des Ortungssignals beschrieben.

3.3. Experimentaldesign und Messdurchführung

Im vorliegenden Unterkapitel wird mit Hilfe der bereits vorgestellten Hardware ein Experimentalaufbau entwickelt. Dazu wird in Abschnitt 3.3.1 *Versuchsaufbau* ein vereinfachter Messaufbau aus den Experimentalanforderungen abgeleitet. Anschließend wird eine Gruppe von verschiedenen Experimentalkonfigurationen bestimmt. In Abschnitt 3.3.2 *Messdurchführung* wird die Vorgehensweise bei der Erzeugung des Testdatensatzes beschrieben. Ziel der Messung ist es, einen repräsentativen Audiodatensatz zu erzeugen, mit dessen Hilfe die Verfahren zur Onset-Detektion und Schallquellen-Lokalisation entwickelt und untersucht werden können. Der zu erzeugende Audiodatensatz hat drei wichtige Funktionen:

- In der Onset-Detektion handelt es sich um einen einfachen Testdatensatz, der zur Erprobung und Parametrisierung des Verfahrens dient.
- Bei der Bestimmung des Azimuts wird er zur empirischen Modellbildung und Richtungsberechnung benötigt.
- Im Rahmen der vertikalen Richtungsbestimmung bildet ein Teil der Daten die Grundlage zur Evolution geeigneter Parameter, die zur Berechnung der Richtung benötigt werden. Der andere Teil dient als Testdaten zur Bestimmung der Genauigkeit.

Der fehlerfreien Erstellung des Datensatzes wird daher eine hohe Bedeutung beigemessen, weshalb der Versuchsaufbau und die Messung nachfolgend ausführlich beschrieben werden.

3.3.1. Versuchsaufbau

In Abschnitt 3.1 *Experimentalanforderung* wurde festgelegt, dass die Messungen unter ähnlichen Bedingungen statt finden, wie sie in der späteren Anwendungsdomäne des Roboters auftreten können. Es wird daher kein schallisoliertes, hallfreies Audiolabor für die Messungen verwendet, sondern ein klassischer Konferenzraum mit einigen abseits stehenden Objekten. Dieser ist circa 15 m lang und 8 m breit. Der Lautsprecher und der Roboterkopf befinden sich in der Raummitte. Der mit Filzteppich ausgelegte Boden und die Decke in einer Höhe von circa 3 m bilden die ersten Reflexionsflächen. Der Abstand zu den Wänden beträgt mindestens 2,5 m. Alle anderen Gegenstände im Raum besitzen mindestens einen Abstand von 2 m zum Messaufbau und stellen keine nennenswerten Reflexionsflächen dar. Der Schalldruckpegel des Umgebungsrauschens am Ort des Roboterkopfes beträgt 37 dB/ μ Pa. Der maximale Pegel des Schallsignals liegt am gleichen Ort bei 68 dB/ μ Pa.

Die verschiedenen Anordnungen von Kopf und Lautsprecher werden schematisch in Abbildung 3.2 dargestellt. Zwischen dem Roboterkopf und dem Lautsprecher befinden sich keine den Direktschall verdeckenden Objekte. Eine Ausnahme bildet der Kopf selbst, welcher eine abschattende Wirkung auf das zu messende Signal hat, wenn die Schallquelle sich auf der dem Ohr abgewandten Seite des Kopfes befindet.

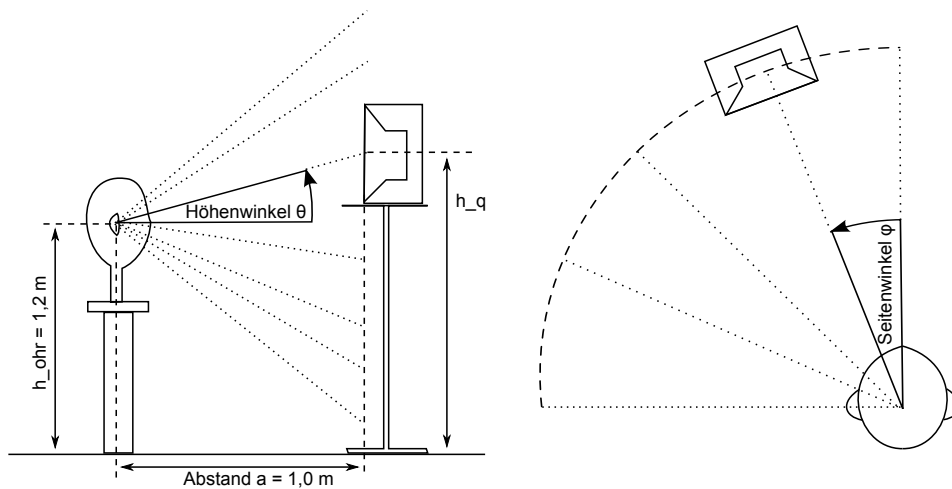


Abbildung 3.2. – Darstellung der Anordnungen zwischen dem Roboterkopf und dem 1 m entfernten Lautsprecher. Variiert wird die Höhe der Schallquelle h_q beziehungsweise der Höhenwinkel ϑ (links) und der Seitenwinkel φ (rechts). Es ergeben sich daraus 35 verschiedene Anordnungen.

Der Lautsprecher befindet sich auf einem höhenverstellbaren Ständer. Der feste, horizontale Abstand zum Roboterkopf beträgt circa 1 m. Der Lautsprecher zeigt eben zum Kopf. Das binaurale System bleibt stets nach vorne ausgerichtet und befindet sich auf einer festen Höhe von 1,2 m, die der eines stehenden Roboters entspricht. Zur Aufnahme des Audio-Datensatzes wird die Signalquelle in horizontaler und vertikaler Richtung zum Roboterkopf variiert.

Auf Basis der Erfahrungen der Voruntersuchung wurden 35 repräsentative Richtungen abgeleitet, die den Raum möglicher Schallquellenpositionen charakterisieren (siehe Tabelle 3.3). Es wurden in der horizontalen Ebene fünf Richtungen ausgewählt, welche den Raum zwischen 0 und 90 Grad mit einer durchschnittlichen Schrittweite von 22 Grad abdecken. In der vertikalen Ebene wurden sieben Positionen zwischen -48 und $+42$ Grad gewählt. Die Schrittweite beträgt hier durchschnittlich 15 Grad. Die Höhenwinkel leiten sich aus den folgenden einstellbaren Höhen des Lautsprechers ab: 0,1 m, 0,6 m, 0,85 m, 1,2 m, 1,45 m, 1,7 m und 2,1 m. Aufgrund der Symmetrie des Roboterkopfes zur Medianebene werden für den Audiodatensatz Seitenwinkel auf der linken Raumhälfte berücksichtigt. Messungen auf der gegenüberliegenden Seite ergeben bei symmetrischer Anordnung den gleichen Datensatz mit vertauschter Kanalzuordnung. Geringfügige Abweichungen der Datensätze können jedoch durch Varianzen zwischen den Übertragungseigenschaften der Mikrofonkapseln entstehen. Diese werden bei der Modellbildung und Unter-

ϑ	φ				
	0°	20°	45°	70°	90°
-48°	1	2	3	4	5
-31°	6	7	8	9	10
-19°	11	12	13	14	15
0°	16	17	18	19	20
$+14^\circ$	21	22	23	24	25
$+27^\circ$	26	27	28	29	30
$+42^\circ$	31	32	33	34	35

Tabelle 3.3. – Auflistung der 35 Konfigurationen zwischen Roboterkopf und Lautsprecher. Jede Anordnung wird durch ein Paar aus Höhenwinkel ϑ und Seitenwinkel φ beschrieben. Die zugehörigen Messungen bilden den Testdatensatz, welcher durch eine eindeutige ID gekennzeichnet ist.

suchung in der vorliegenden Arbeit vernachlässigt. Im realen Einsatz der Modelle muss jedoch eine Abstraktion von den spezifischen Kapseleigenschaften erfolgen, sofern signifikante Varianzen in einer Baureihe auftreten sollten. Insbesondere bei der vertikalen Richtungsbestimmung kann dies sonst Einfluss auf die Genauigkeit der Richtungsbestimmung haben.

3.3.2. Messdurchführung

Im vorhergehenden Abschnitt wurden der Versuchsaufbau sowie die verwendeten Experimentanordnungen beschrieben. Folgende drei Schritte sind für die Durchführung einer Messung erforderlich:

1. Signalwiedergabe und -aufnahme

Die Wiedergabe des Audiosignals wird vom PC aus gesteuert. Die Aufnahme des Audiosignals erfolgt mittels eines Programms auf dem Mikrocontroller des Roboterkopfes. Das zu ortende Schallsignal (Nutzsignal) wird in einem Messzyklus automatisch durch die Überschreitung eines Schwellwertes detektiert und aufgenommen. Im Rahmen der vorliegenden Arbeit wurde ein kleines Programm zur Aufnahme für den STM32 geschrieben. Folgender Programmablauf findet statt: Der A/D-Wandler des Modul 1 (siehe Abbildung 2.3 liefert entsprechend seiner Abtastfrequenz fortlaufend einen seriellen 32-Bit-Wert (Stereosample). Dieser wird durch eine zeitgesteuerte *Interrupt Service Routine* (ISR) bitweise eingelesen und als zwei Integerwerte in einem Ringpuffer abgelegt. Nach vollständigem Einlesen beider Samples wird ein Vergleich mit einem Schwellwert durchgeführt, welcher weit über dem Umgebungsrauschen liegt. Bei erstmaliger Überschreitung in einem Messzyklus wird angenommen, dass ein analysierbares Signal anliegt (unter anderem der Direktschall des Ortungssignals). Der Messzyklus wird abgeschlossen, indem eine feste Anzahl weiterer Samples eingelesen wird. Der resultierende Signalvektor enthält den Direktschall des Ortungssignals und einige Reflexionen.

2. Datenübertragung

Nach der Signalaufzeichnung wird der Audiovektor automatisch über die serielle Schnittstelle zum PC übertragen, wo die Audiodaten beispielsweise mit Hilfe eines Terminal-Programms erfasst und in einer Datei zwischengespeichert werden. Die Aufzeichnung auf dem eingebetteten System beginnt nach

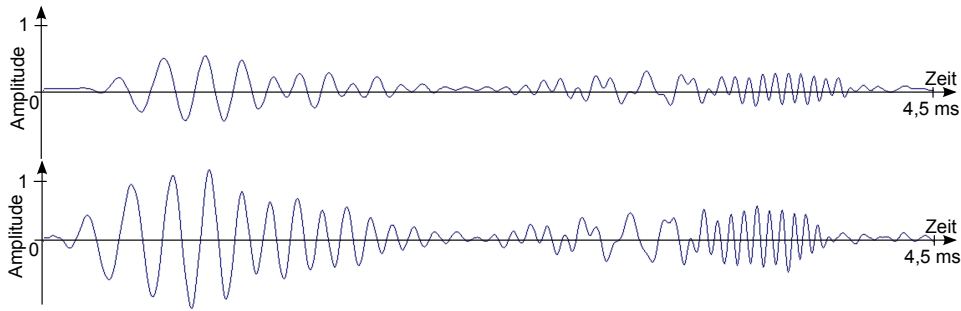


Abbildung 3.4. – Exemplarische Darstellung einer Stereoaufnahme (linker Kanal oben, rechter Kanal unten) des Sinussweeps. Als Messanordnung dient Konfiguration 12 ($20^\circ / -19^\circ$). Es ist deutlich die frequenzabhängige Amplitude sowie die zeitliche Verschiebung beider Direktsignale zueinander zu erkennen. Im letzten Drittel beider Vektoren treten Überlagerungen des Direktschalls mit schwachen Reflexionen auf. Dies ist an den ungleichmäßig alternierenden Wellenbergen und -tälern zu erkennen.

einer definierten Pause, die größer als die Abklingzeit des Ortungssignals ist, erneut. Die Aufnahme wird fünf Mal für eine Messanordnung wiederholt. Die Daten einer Messkonfiguration werden aufeinander folgend in derselben Datei abgespeichert. So können bei späterer Verarbeitung Fehler im Signalvektor identifiziert werden. Zusätzlich können die Verfahren zur Signalanalyse auf die Robustheit bezüglich natürlicher Varianzen bei der Übertragung überprüft werden.

3. Datenformatierung

In einem weiteren Schritt werden die Rohdaten mittels eines Audioeditors formatiert und in eine 16-Bit Stereo-Wave-Datei exportiert. Im Dateinamen wird die horizontale und vertikale Anordnung der Aufnahme vermerkt.

Auf diese Weise entstand ein Audiodatensatz aus 35 Dateien mit je fünf aneinander gereihten Stereovektoren, der für die Offline-Analyse und Konzeption der Verfahren genutzt wurde. Exemplarisch ist in Abbildung 3.4 der Ausschnitt des Ortungssignals eines aufgenommenen Signalvektors abgebildet.

4. Konzeption der Modelle und Verfahren

Die theoretische Entwicklung der Teilverfahren und -modelle, die zur Schallquellen-Lokalisation nötig sind, wird in den nachfolgenden Unterkapiteln beschrieben. Dazu werden die drei Schwerpunkte Onset-Detektion, Bestimmung des Azimuts und Bestimmung der Elevation separat betrachtet. Es bestehen Abhängigkeiten zwischen den Verfahren, auf die explizit hingewiesen wird.

4.1. Verfahren zur Detektion des Onsets

Aus vielen Untersuchungen, wie denen von Blauert, Cobben [BC78] und Hartmann [Har83], [Har97] geht hervor, dass die Schallquellen-Lokalisation trotz verschiedener Einflüsse wie Reflexionen, Rauschen und Überlagerung mehrerer Quellen auf Basis des ungestörten Direktschalls am erfolgreichsten ist. Auch der Mensch bestimmt die horizontale Richtung einer Schallquelle anhand des Direktschalls, welcher vor den Reflexionen eintrifft. Früh eintreffende Reflexionen werden nicht als separates Schallereignis wahrgenommen, sodass deren Richtungsinformationen ignoriert werden und die Richtungsbestimmung auf Basis des Direktschalls unbeeinträchtigt bleibt. Dieser psychoakustische Effekt wird als *Präzedenzeffekt* [WNR49] beziehungsweise *Gesetz der Ersten Wellenfront* bezeichnet [Bla97]. Haas zeigte zudem, dass dieser Effekt selbst dann noch gilt, wenn das verzögerte Signal bis zu 10 dB über dem Pegel der ersten Wellenfront liegt. Daher ist ebenfalls die Bezeichnung *Haas-Effekt* [Haa51] verbreitet. Aus diesem Grund spielt die Detektion des ungestörten Direktschalls eine große Bedeutung für die Bestimmung der horizontalen Richtung der Schallquelle. Wie bereits bei der Begriffseinführung im Abschnitt *Signalstruktur und Onset* des Kapitels 1.3 ausgeführt wurde, wird der Beginn eines Signals und somit der Beginn des ungestörten Direktschalls als Onset bezeichnet. Die Detektion des Onsets bildet eine Voraussetzung zur horizontalen Schallquellen-Lokalisation.

Zur Detektion des Direktschalls wurde 1987 von Zurek erstmals eine Methode auf Basis eines Onset-Detektors vorgestellt [Zur87]. Die Onset-Detektion wurde bis

heute durch Überführung der Analyse in den Frequenzbereich weiter verfeinert und verbessert [BDDS04], [Ric05], [Pek06]. Eine aktuelle Methode ist beispielsweise das Modell der interauralen Kohärenz (IC) von Faller und Marimaa [FM04], welches ebenfalls bei der Überlagerung konkurrierender Schallquellen in verhallten Räumen erlaubt, die Signalabschnitte zu detektieren, in denen die Lokalisation zuverlässig funktioniert.

Eine sehr einfache Methode im Zeitbereich kann jedoch angewendet werden, wenn die Signalstruktur deterministisch ist. Unter normalen Umständen ist dies nicht gegeben, da sich natürliche Signale meist aus verschiedenen Frequenzen zusammensetzen. In Kapitel 3.2 *Synthese des Audiosignals* wird das Signal so konstruiert, dass es mit einer reinen Sinusschwingung beginnt, welche sich leicht detektieren lässt. Die Filterung des Signals ist bei der Übertragung zwischen Schallquelle und -empfänger vollständig linear, sodass sich zwar Phase und Amplitude, jedoch nicht die Frequenz verändern. Sehr früh eintreffende Reflexionen können die Form der Schwingung durch die Superposition verändern, was dann der Fall ist, wenn der Direktschall sehr nah an einer Reflexionsfläche vorbei führt. Dieser Spezialfall wird bei der Parametrisierung der Implementierung in Kapitel 5.1 berücksichtigt.

Da die Frequenz der ersten Schwingung bekannt ist, kann durch Analyse der ersten Periode eines noch unbekanntes Signals mit begrenzter Sicherheit festgestellt werden, ob es sich um das für die Lokalisation gedachte Signal handelt. Somit ist eine Erkennung des Signals in einem Vektor, welcher außerhalb des Ortungssignals Rauschen enthält, auf einfachem Weg, aber mit begrenzter Zuverlässigkeit möglich. Das Verfahren zur Onset-Detektion besteht in der schrittweisen Analyse eines Audiovektors, der entsprechend der Messdurchführung in Kapitel 3.3.2 *Messdurchführung* aufgenommen wurde. Das gesuchte Ortungssignal befindet sich in diesem Vektor und kann anhand des Signalpegels detektiert werden. Dazu wird der Vektor sampleweise aufsteigend mit dem Schwellwert $C_{\min\text{Level}}$ verglichen, bis dieser erstmalig überschritten wird. Anschließend findet das Analyseverfahren zur Bestimmung des möglichen Onsets Anwendung, welches in Abbildung 4.1 anhand eines Audiovektors schematisch dargestellt ist. Die Detektion gliedert sich in folgende drei Schritte:

1. Detektion des ersten lokalen Maximums

Der Signalvektor wird nach dem Sample des ersten positiven, lokalen Maximums durchsucht. Da es sich bei dem Ausgangssignal um eine Schwingung handelt, markiert dieses die Mitte des ersten Wellenberges der Schwingung.

2. Detektion des ersten lokalen Minimums

Da eine Wellenperiode immer einen positiven und einen negativen Extremalwert besitzt, wird die Detektion beider als Bedingung vorausgesetzt, um zu validieren, dass es sich um den Beginn des Schallsignals handelt.

3. Festlegung des Onsets

Als Onset kann theoretisch jeder beliebige Punkt der ersten Periode gewählt werden. Entscheidend ist, dass er sich in allen Messungen an derselben Stelle der Periode befindet, damit auf dessen Basis die Laufzeitdifferenz zwischen zwei Kanälen bestimmt werden kann. Theoretisch eignet sich zum Beispiel der Zeitpunkt eines Nulldurchgangs oder eines Extremalwertes, da er sich leicht bestimmen lässt.

Die Genauigkeit der Onset-Bestimmung wird durch die zeitliche Quantisierung des analogen Schallsignals begrenzt, welche von der Abtastrate abhängig ist. Aus diesem Grund kann der Zeitpunkt der maximalen Amplitude einer Schwingung oder eines Nulldurchgangs nicht ohne weiteres exakt bestimmt werden. Die gemessene Zeitpunkt des Onsets weicht maximal um einen halben Sample vom tatsächlichen Zeitpunkt ab. Durch geometrische Interpolation oder aufwendigere, aber robustere Verfahren ist es möglich, die Genauigkeit zu erhöhen [SR99]. Im Weiteren soll jedoch darauf verzichtet werden, da die sample-genaue Bestimmung des Onsets hinreichend für die Lokalisation ist.

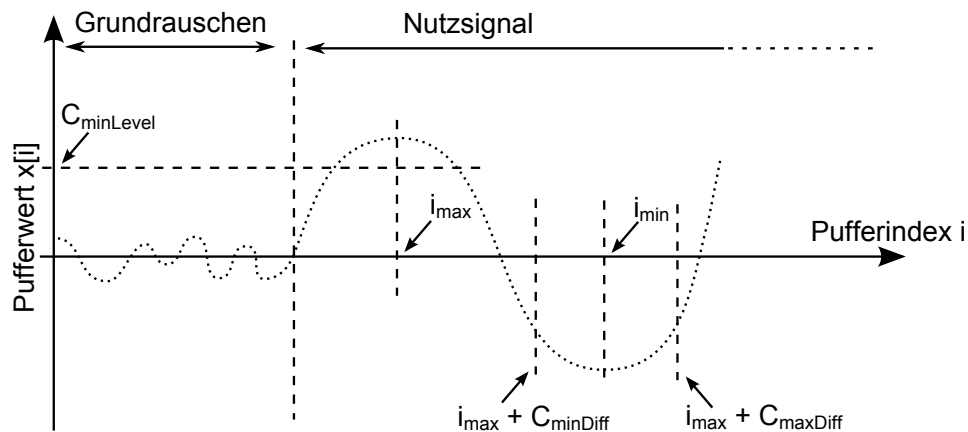


Abbildung 4.1. – Schematische Darstellung des Onset-Detektors. Durch Analyse des ersten Wellenberges und -tales im Zeitbereich wird der Onset des Signals bestimmt und der Signaltyp identifiziert.

Folgende mathematischen Zusammenhänge und Bedingungen werden aus den oben genannten Schritten abgeleitet, welche in Form eines Algorithmus auf den Vektor angewendet werden können:

1. Maximum-Detektion

$$\begin{aligned} \text{Bedingung: } & x(i) > C_{\text{minLevel}} \\ & \wedge x(i) > x(i-1) \\ & \wedge x(i) > x(i+1) \end{aligned}$$

Falls wahr, weise $i_{\text{max}} = i$ zu und fahre mit 2. fort, ansonsten erhöhe i .

2. Minimum-Detektion

$$\begin{aligned} \text{Bedingung: } & x(i) < -C_{\text{minLevel}} \\ & \wedge x(i) < x(i-1) \\ & \wedge x(i) < x(i+1) \\ & \wedge i - i_{\text{max}} \leq C_{\text{maxDiff}} \\ & \wedge i - i_{\text{max}} \geq C_{\text{minDiff}} \end{aligned}$$

Falls wahr, weise $i_{\text{min}} = i$ zu und fahre mit 3. fort, ansonsten erhöhe i .

3. Onset-Bestimmung

Weise $i_{\text{onset}} = i_{\text{min}}$ zu und beende.

Als Onset wird in diesem Beispiel das Minimum der Schwingung gewählt. Der Signalvektor x beinhaltet alle gepufferten Samples eines Audiokanals. Der Pufferindex i liegt im Bereich 0 bis $C_{\text{maxLength}}$. Die Konstanten C_{minDiff} und C_{maxDiff} bezeichnen einen zeitlichen Toleranzbereich, in dem das Minimum bezogen auf das Maximum der Schwingung liegen muss. Aus diesen mathematischen Beziehungen wird ein entsprechender Algorithmus synthetisiert, der im Kapitel 5.1 anhand des Audiodatensatzes untersucht wird.

4.2. Geometrisches Modell zur Bestimmung des Azimuts

Die Analyse der horizontalen Richtung setzt im Allgemeinen voraus, dass die Signallaufzeitdifferenz zwischen beiden Ohren bekannt ist [Bla69]. Diese wird in Stereosystemen *time difference of arrival* (TDOA) beziehungsweise in binauralen Systemen *interaural time difference* (ITD) genannt. Dazu werden im linken und rechten Kanal des binauralen Systems Signalabschnitte gesucht, die demselben Schallereignis zugeordnet werden können. Durch die Bestimmung der TOA in beiden Kanälen, welche

beispielsweise durch die Onset-Detektion erfolgt, lässt sich die ITD aus der Differenz der beiden Sample-IDs berechnen.

In diesem Unterkapitel soll ein Verfahren zur horizontalen Richtungsbestimmung auf Basis der Laufzeitdifferenz schrittweise hergeleitet werden. Dazu wird vom minimalistischen Modell von Hornborstel und Wertheimer ausgegangen, das bereits 1920 eingeführt wurde [Bla74]. Dieses wird im ersten Abschnitt *Minimales Modell des Stereosystems* beschrieben und anschließend schrittweise erweitert. Um dieses Modell anwenden zu können, sind einige Vereinfachungen des Lokalisationsproblems notwendig:

1. Der Abstand zwischen Schallquelle und Mikrofonsystem ist sehr viel größer als der halbe Abstand der Mikrofone.
2. Das Aufnahmesystem ist ein Stereosystem aus zwei omnidirektionalen Mikrofonen, die mit einem festen Abstand in der horizontalen Ebene in gleicher Höhe angeordnet sind.
3. Die Schallquelle befindet sich auf derselben Höhe wie die Mikrofone.

Im darauf folgenden Abschnitt *1D-Modell des Roboterkopfes* wird das Modell um Punkt 2 der Aufzählung erweitert. Es handelt sich beim Messobjekt nicht um ein Stereosystem mit omnidirektionalen Mikrofonen, sondern um ein binaurales System mit einem Kunstkopf. Dies führt zum Modell von de Boer [Bla74]. Im letzten Abschnitt *2D-Modell des Roboterkopfes* wird Punkt 3 der Vereinfachung aufgehoben. Da Schallquelle und -empfänger in der vertikalen Ebene variieren können, führt dies zu Mehrdeutigkeiten bei der Interpretation der ITD. Um den korrekten Seitenwinkel zu bestimmen, wird ein eigenes Modell unter Zuhilfenahme des Höhenwinkels bestimmt.

4.2.1. Minimales Modell eines Stereosystems

Im Abschnitt *Schall und Schallreflexion* der Einführung grundlegender Begriffe wurde bereits beschrieben, dass der Schall einer punktförmigen Schallquelle bei hinreichend großer Entfernung als lineare Welle betrachtet werden kann. Die Schallwellen bewegen sich parallel auf omnidirektionale Mikrofone zu. Da sich Send- und Empfangssystem auf einer horizontalen Ebene befinden, kann das Problem in den zweidimensionalen, kartesischen Raum abgebildet werden. Unter diesen Bedingungen

kommt das *Sinusgesetz* zur Anwendung, welches in Abbildung 4.2 veranschaulicht wird.

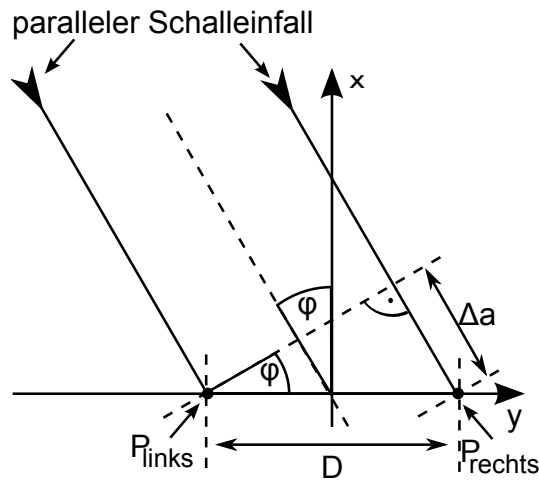


Abbildung 4.2. – Grafische Darstellung des Modells von Hornborstel und Wertheimer zur Berechnung des Seitenwinkels φ . Als Grundlage dient das *Sinusgesetz*, welches von parallel einfallendem Schall auf zwei Mikrofone im freien Feld ausgeht. Die Mikrofone werden durch die zwei Punkte P_{links} und P_{rechts} dargestellt.

Die horizontale Richtung wird durch den Seitenwinkel φ dargestellt. Ausgehend von dem Sinusgesetz im rechtwinkligen Dreieck

$$\Delta a = D \cdot \sin(\varphi) \quad (4.1)$$

wird der Winkel φ zwischen Ankathete und Hypotenuse wie folgt berechnet:

$$\varphi = \arcsin\left(\frac{\Delta a}{D}\right). \quad (4.2)$$

Der Dreieckswinkel φ an Punkt P_{links} ist aufgrund der orthogonalen Beziehung zwischen der x- und y-Achse auch gleichzeitig der gesuchte Seitenwinkel des Direktschalls (siehe Abbildung 4.2). Da jedoch anstelle von Δa die Laufzeitdifferenz l_a gegeben ist, muss die Formel zur Vollständigkeit durch die Gleichung

$$\Delta a = \frac{l_a \cdot c}{f} \quad (4.3)$$

erweitert werden. Sie lautet dann:

$$\varphi = \arcsin\left(\frac{l_a \cdot c}{D \cdot f}\right). \quad (4.4)$$

Es werden zur Berechnung die Konstanten Mikrofonabstand D , die Schallgeschwindigkeit c und die Samplingrate f benötigt.

Die Genauigkeit dieses Modells zur Schallquellen-Lokalisation wird durch die sample-genaue Onset-Bestimmung beschränkt und ändert sich mit dem Seitenwinkel. Die maximale Winkelabweichung $\Delta\varphi_l$ vom tatsächlichen Winkel entspricht etwa der Hälfte der Schrittweite zwischen zwei aufeinander folgenden diskreten Winkeln φ_l :

$$\Delta\varphi_l = \frac{1}{2} (\varphi_l - \varphi_{l+1}). \quad (4.5)$$

Bei Anwendung von Formel 4.4 auf ein Stereomikrofonsystem mit den Mikrofonabständen des Roboterkopfes ($D = 0,18$ m, $f = 46\,875$ Hz, $c = 340$ m/s) treten Laufzeitdifferenzen i von 0 ($\varphi = 0$ Grad) bis 24 Samples ($\varphi = 90,0$ Grad) auf. Die maximale Winkelabweichung wächst mit steigendem l exponentiell. Im Frontalbereich beträgt die Winkelauflösung $\Delta\varphi$ 1,2 Grad, beim maximalen Seitenwinkel (maximale Laufzeitdifferenz) dagegen nur noch 8,0 Grad. Eine vollständige Liste der diskreten Winkel und zugehörigen maximalen Winkelabweichungen befindet sich im Anhang in Tabelle A.2.

Dieses Modell ist nur eine Näherung für zwei Mikrofone im freien Feld und kann daher nicht auf den Roboterkopf angewendet werden. Das nachfolgend beschriebene erweiterte Modell berücksichtigt die Ausdehnung des Roboterkopfes.

4.2.2. 1D-Modell des Roboterkopfes

Als Messsystem kommt in dieser Arbeit ein binaurales System mit einem Kunstkopf zum Einsatz. Somit ist der direkte Schallweg Δa durch die Außenschale des Kopfes verdeckt. Der tatsächliche Pfad des Schalls ist der um die Kopfschale gekrümmte Weg Δs (siehe Abbildung 4.3). Der Fehler des minimalen Modells wächst mit zunehmendem Seitenwinkel, da die gemessene Laufzeitdifferenz und der berechnete Seitenwinkel desto stärker auseinander laufen, je größer die Wegdifferenz zwischen Δa und Δs ist. Eine einfache Substitution lautet:

$$\Delta a = \Delta s \cdot \frac{D}{\Delta s_{\max}}. \quad (4.6)$$

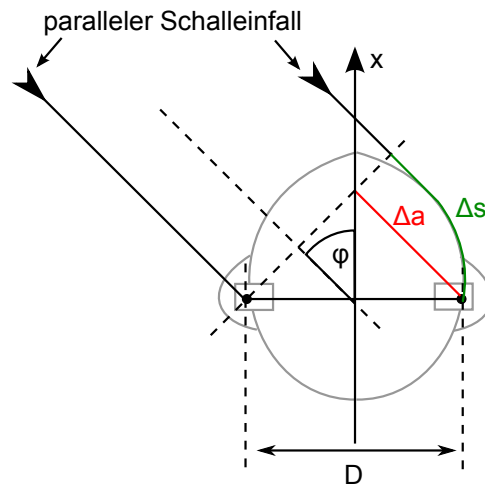


Abbildung 4.3. – Erweiterung des einfachen Modells zweier Aufpunkte um die Existenz eines runden Kopfes. Es ergibt sich eine Verlängerung des Schallweges S verglichen zur ungehinderten Strecke Δa durch die abschattende Wirkung des Kopfes.

Die Gleichung 4.1 lautet mit substituiertem Δa :

$$\Delta s \cdot \frac{D}{\Delta s_{\max}} = D \cdot \sin(\varphi) \quad (4.7)$$

beziehungsweise nach Umstellung:

$$\Delta s = \Delta s_{\max} \cdot \sin(\varphi). \quad (4.8)$$

Es handelt sich bei der Korrektur um eine grobe Approximation, da kein linearer Zusammenhang zwischen Δa und Δs besteht. In [Bla74] wird die aus empirischen Ergebnissen abgeleitete alternative Formel von De Boer genannt, welche zu einer genaueren Näherung führen soll:

$$\Delta s = \Delta s_{\max} \cdot (\varphi + \sin(\varphi)) \quad (4.9)$$

Beide Näherungsmodelle 4.6 und 4.9 werden in Abbildung 4.4 durch die kontinuierlichen Graphen dargestellt. Auf der Abszisse wurde die normierte Laufzeitdifferenz und auf der Ordinate der resultierende Seitenwinkel dargestellt. Beim direkten Ver-

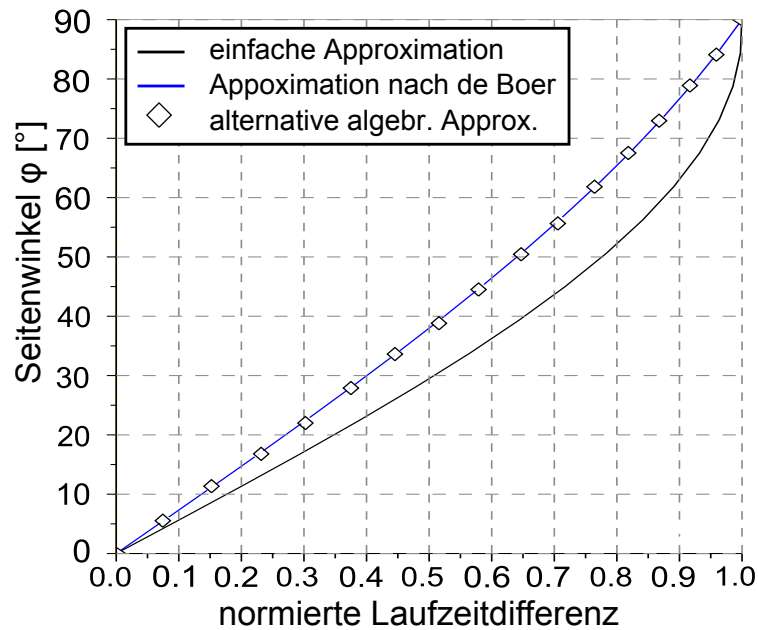


Abbildung 4.4. – Graphische Gegenüberstellung der Näherungsmodelle der einfachen Approximation (Formel 4.8), der Approximation nach De Boer (Formel 4.9) und der algebraischen Alternative zum Modell von de Boer (Formel 4.10).

gleich beider Modelle ergeben sich gravierende Unterschiede bei den resultierenden Winkeln. Die berechneten Winkel besitzen bei gleichen Laufzeitdifferenzen eine Abweichung von bis zu 12 Grad. Es wird daher das Modell nach de Boer bevorzugt. Die zugehörige Gleichung ist jedoch nicht nach der Laufzeitdifferenz l auflösbar, sodass eine alternative algebraische Näherung benötigt wird. So wurde in der vorliegenden Arbeit eine Näherung bestimmt, die minimal von den Resultaten von De Boer abweicht. Die Gleichung lautet:

$$\Delta s = \Delta s_{\max} \cdot x_S \cdot \sin\left(\frac{\varphi}{x_\varphi}\right) \quad (4.10)$$

mit den Konstantenbelegungen $x_\varphi = 1,50$ und $x_S = 1,14$. Der zugehörige Graph überlagert die blaue Kurve und wurde daher diskret durch Rauten dargestellt, um eine Unterscheidung zu ermöglichen (siehe Abbildung 4.4). Entsprechend lautet die

Formel zur Azimutberechnung:

$$\varphi = x_\varphi \cdot \arcsin\left(\frac{\Delta s}{\Delta s_{\max} \cdot x_S}\right). \quad (4.11)$$

Durch die Substitution von Δs mit $\Delta s = l \cdot c / f$ und die analoge Ersetzung von Δs_{\max} durch $\Delta s_{\max} = L \cdot c / f$, wobei L die empirisch bestimmte, maximale Laufzeitdifferenz ist, lautet die endgültige Formel zur Berechnung des Azimuts:

$$\varphi = x_\varphi \cdot \arcsin\left(\frac{l}{L \cdot x_S}\right). \quad (4.12)$$

Aufgrund der Verlängerung des Schallweges durch die Krümmung um den Kopf wird die maximale Laufzeitdifferenz des minimalen Stereosystems von 24 Samples auf circa 37 Samples erhöht. Dies hat positive Folgen für die Winkelgenauigkeit. Im Frontalbereich beträgt die maximale Winkelabweichung 1,0 statt 1,2 Grad und im Seitenbereich (Azimut = 90 Grad) 2,1 statt 8,0 Grad (siehe Tabelle A.2 im Anhang).

Dieses Modell kann bereits für den Roboterkopf angewendet werden. Es unterliegt jedoch weiterhin einer Einschränkung. Der mit diesem Modell berechnete Azimut ist ausschließlich für Schallquellen korrekt, die sich auf der gleichen vertikalen Ebene wie der Roboterkopf befinden. Für höher oder niedriger angeordnete Schallquellen nimmt der Fehler mit wachsender Höhe und mit wachsendem Seitenwinkel zu. Dieser Zusammenhang wird im folgenden und letzten Abschnitt erläutert. Durch Einführung der vertikalen Dimension als Eingangsparameter wird das Modell des Roboterkopfes zur Azimutbestimmung vervollständigt.

4.2.3. 2D-Modell des Roboterkopfes

Die Grundlage für das Verfahren zur Berechnung des Seitenwinkels ist die Laufzeitdifferenz zwischen beiden Ohren. Die ITD ist jedoch unter bestimmten Umständen keine ausreichende Basis für die Berechnung des Azimuts einer Schallquelle. Der Einfachheit halber wird noch einmal auf das anfängliche Modell von Hornbostel und Wertheimer Bezug genommen, also die Ausdehnung des Roboterkopfes vernachlässigt. Es lässt sich jedoch analog auf das erweiterte Modell des Roboterkopfes anwenden.

Alle Positionen der Schallquelle mit gleicher Laufzeitdifferenz l befinden sich auf einer Hyperbel (siehe Abbildung 4.5). Entsprechend den Beschränkungen des Ver-

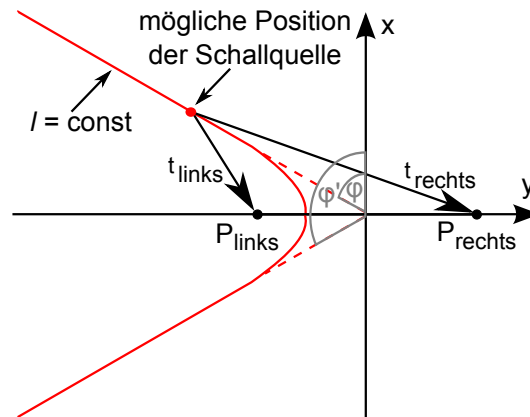


Abbildung 4.5. – Die rote Linie stellt den geometrischen Ort aller Punkte dar, welche die gleiche Laufzeitdifferenz l zwischen linkem und rechtem Mikrofon (P_{links} und P_{rechts}) in der Horizontalebene aufweisen. Ab einer bestimmten Entfernung befinden sich alle Positionen auf zwei Geraden, die das Zentrum des Stereosystems mit dem Winkel φ beziehungsweise φ' durchstoßen.

suchsaufbaus werden nur Positionen in größerer Entfernung ($a \gg \frac{1}{2}D$) betrachtet. Es kann vereinfachend davon ausgegangen werden, dass sich alle Positionen auf zwei Strahlen befinden, deren Schnittpunkt im Zentrum des Systems liegt. Die Strahlen weisen die Seitenwinkel φ und φ' auf, wobei zwischen beiden die Beziehung $\varphi' = 180^\circ - \varphi$ gilt. Auf welchem der beiden Strahlen die Schallquelle liegt, ist mit Hilfe der Laufzeitdifferenz nicht unterscheidbar. Dies wird auch als *Front-Back-Ambivalenz* bezeichnet. Ein Lösungsansatz für die Reduzierung dieser Mehrdeutigkeit der ITD basiert auf dem Modell eines rotierenden Kopfes. Aus der positiven oder negativen Veränderung der Laufzeitdifferenz kann auf eine der beiden Positionen geschlossen werden. Dieses Modell, welches erstmalig von van Söst 1929 vorgestellt wurde, ist unter anderem in [Bla74], [ISK03] und [AHSK04] auf seine Genauigkeit hin untersucht worden. Alternativ ist es ebenfalls mit einem starren Kopf möglich, dieses Problem auf Basis der monauralen Eigenschaften zu lösen, wie es Blauert 1969 in der Medianebene demonstrierte [Bla69],[OFS08]. In dieser Arbeit kann das Problem vernachlässigt werden, da alle Seitenwinkel zwischen 0 und 90 Grad liegen (siehe 3.3 *Experimentaldesign und Messung*). Eine mögliche Erweiterung des entwickelten Modells wird in Kapitel 6 *Zusammenfassung und Ausblick* genannt.

Weitere Mehrdeutigkeiten der Laufzeitdifferenz treten auf, wenn der vertikalen Winkel der Schallquelle auch ungleich Null sein kann [Bla74]. In Abbildung 4.6 sind

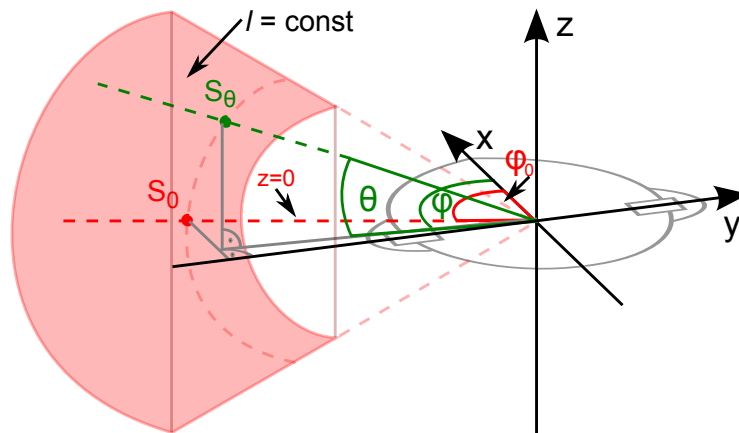


Abbildung 4.6. – Der rot markierte Mantel eines seitlich liegenden, halbierten Kegelstumpfes stellt den geometrischen Ort aller Positionen dar, welche die gleiche Laufzeitdifferenz zwischen linkem und rechtem Ohr im Raum aufweisen und vor der Frontalebene liegen. Die exemplarischen Schallquellenpositionen S_0 und S_ϑ liegen in gleicher x - z -Ebene auf dem Kegelmantel. S_0 besitzt die Elevation 0 Grad und den kleineren Seitenwinkel φ_0 . S_ϑ ist der Seitenwinkel φ und Elevation ϑ zugehörig.

alle Positionen der Schallquelle, die dieselbe Laufzeitdifferenz und einen Seitenwinkel unter 90 Grad aufweisen, durch eine rote Fläche dargestellt. Es handelt sich um den Mantel eines auf der Seite liegenden, vertikal halbierten Kegelstumpfes. Dieses geometrische Gebilde kann auch durch eine unendliche Menge von Strahlen gebildet werden, die sich im Zentrum des Kopfes schneiden und alle einen separaten Seiten- und Höhenwinkel besitzen. Zwei dieser Strahlen sind in der Abbildung exemplarisch mit dem Höhenwinkel 0 Grad und dem beliebigen Winkel ϑ eingezeichnet, auf der je eine Schallquellenposition in derselben z -Ebene liegt. Wird mit der Formel 4.12 der Azimut der Schallquellenposition S_ϑ bestimmt, so lautet die Lösung φ_0 anstelle von φ . Dies liegt daran, dass das zugrunde liegende 1D-Modell die Laufzeitdifferenz so interpretiert, als läge die Schallquelle in der x - y -Ebene ($z = 0$). Folglich kann nur der Seitenwinkel des Strahls berechnet werden, der den Punkt S_0 schneidet. Je stärker die Elevation der Schallquelle von dieser Ebene ($z = 0$) abweicht, desto größer ist der Fehler des berechneten Seitenwinkels.

Für eine Erweiterung des Modells wird also ein zusätzlicher Eingangsparameter benötigt. Dazu bietet sich der Höhenwinkel ϑ an, welcher im nächsten Kapitel 4.3 *Verfahren zur Bestimmung der Elevation* bestimmt wird. So wird in Formel 4.12 ein Korrekturterm $g(\vartheta)$ eingeführt, der vom Höhenwinkel abhängig ist und den Fehler

des Seitenwinkels ausgleicht:

$$\varphi = x_\varphi \cdot \arcsin\left(\frac{l}{L \cdot x_S}\right) + g(\vartheta). \quad (4.13)$$

Mathematische Modelle für $g(\vartheta)$ benötigen neben dem Elevationswinkel weitere Parameter, wie beispielsweise die Entfernung der Schallquelle. Mit dem binauralen System ist es jedoch nicht möglich, diese Information zur Verfügung zu stellen. Daher wird die Approximationsfunktion empirisch auf Basis des Audiodatensatzes bestimmt (siehe Kapitel 5.2 *Bestimmung des Azimuts*).

4.3. Verfahren zur Bestimmung der Elevation

Im ersten Abschnitt dieses Kapitels werden allgemeine Ansätze zur vertikalen Richtungsbestimmung diskutiert und benötigte Grundkonzepte ausgewählt. Darauf aufbauend werden im nächsten Abschnitt das Verfahren synthetisiert und die Einzelschritte erläutert. Die Implementierung und Untersuchung des Verfahrens auf Basis des Testdatensatzes findet im Kapitel 5.3 statt, wo das Verfahren angewendet und auf seine Genauigkeit hin untersucht wird.

4.3.1. Grundkonzepte des Verfahrens

Nachfolgend werden einige grundlegende Konzepte dargestellt, die für das Verfahren zur Bestimmung der Elevation benötigt werden. Dies betrifft zum einen allgemeine Konzepte zur Richtungsbestimmung, aber auch spezielle Modelle zur Bestimmung der Richtung. Es wird auch argumentiert, warum diese in dem vorliegenden Verfahren Anwendung finden.

Allgemeine Konzepte zur vertikalen Richtungsbestimmung

Verbreitete Methoden zur vertikalen Richtungsbestimmung basieren auf den Konzepten der monauralen beziehungsweise binauralen spektralen Eigenschaften (MSC¹, BSC²) [WB06], [Beg94]. Die MSC sind spektrale Eigenschaften des einfallenden Schalls eines einzelnen Ohres. Die Analyse ist im Allgemeinen nur für Signale möglich, deren Frequenzzusammensetzung bekannt ist. So können Abweichungen des Spektrums auf Filtereinflüsse bei der Signalübertragung zurückgeführt werden, was

¹MSC ist die englische Abkürzung für *monaural spectral cues*.

²BSC ist die englische Abkürzung für *binaural spectral cues*.

Rückschlüsse auf die Einfallsrichtung zulässt. Bei nicht deterministischen Signalen wie beispielsweise der menschlichen Sprache werden stattdessen binaurale spektrale Eigenschaften analysiert, die über einen längeren Zeitraum gemittelt werden. Die BSC berechnen sich aus dem Differenzspektrum beider Ohren, welches von der veränderlichen Frequenzzusammensetzung des zu ortenden Signals weitestgehend unabhängig ist. Am Beispiel des Modells vom Chisaki et al. [CNNU03] wird aus der so genannten interauralen Amplitudendifferenz (ILD³) und der interauralen Phasendifferenz (IPD⁴) ein spektraler Fingerabdruck des anliegenden Signals berechnet und mit einer Datenbank von Differenzen verglichen, für die die zugehörigen Richtungen bekannt sind. Dieses Verfahren wurde im Rahmen einer Voruntersuchung implementiert und mit Hilfe des bereits erwähnten KEMAR-Datensatzes [GK94] untersucht. Anhand einfacher Sprachbeispiele konnte nachgewiesen werden, dass das Verfahren für isolierte Schallquellen einwandfrei funktioniert und sogar für konkurrierende Schallquellen teilweise zufriedenstellende Ergebnisse liefert [CKN⁺08], [WLH⁺09]. Allerdings ist dieses Verfahren sehr speicher- und rechenaufwendig, sodass es für die Realisierung in dem vorliegenden Roboterkopfsystem nicht in Frage kommt. Ein Nachteil der BSC ist die Tatsache, dass sich die Elevation einer Schallquelle nicht in der Medianebene bestimmen lässt. Die spektralen Differenzen sind für alle Elevationen null, sodass diese sich auf diese Weise nicht unterscheiden lassen.

Vorgehen zur Extraktion monauraler spektraler Eigenschaften

Das Problem der Medianebene soll nicht extra behandelt werden, daher wird die vertikale Schallquellen-Lokalisation in dieser Arbeit auf Basis der monauralen spektralen Eigenschaften (MSC) bestimmt. Die geforderte Bedingung, dass die Frequenzzusammensetzung des verwendeten Schallsignals bekannt ist, wird erfüllt, da es sich um einen linearen Sinussweep mit deterministischen Frequenzverteilung von 3 bis 14 kHz handelt (siehe Kapitel 3.2). Dies erlaubt die Extraktion der Frequenzzusammensetzung mit einfachen Mitteln aus der Hüllkurve des aufgenommenen Audiosignals. Somit kann auf die Transformation des Signals mittels der *Fast-Fourier-Transformation* (FFT) oder der Wavelet-Transformation verzichtet werden [MH04]. Die genaue Vorgehensweise bei der gewählten Variante wird im nächsten Abschnitt *Synthese des Verfahrens* dargestellt.

³ILD ist die englische Abkürzung für *interaural level difference*.

⁴IPD ist die englische Abkürzung für *interaural phase difference*.

Motivation zur Reduktion spektraler Informationen

Ein weiterer Baustein des vorliegenden Verfahrens zur Bestimmung der Elevation folgt aus der Arbeit von Blauert [Bla69]. Er fand anhand von Experimenten in der Medianebene heraus, dass der Mensch bereits die grobe vertikale Richtung (vorn, hinten, oben) einer Schallquelle korrekt zuordnen kann, wenn die Energie des Signals in bestimmten Frequenzbändern dominiert. Diese Bänder werden häufig auch *Blauert'sche Bänder* genannt. Daraus kann geschlussfolgert werden, dass zur Schallquellen-Lokalisation nicht zwangsläufig einzelne Frequenzanteile ausgewertet werden müssen, sondern eine Reduktion des Spektrums in Frequenzbändern erfolgen kann.

Beim vorliegenden Roboterkopf wurde in der Untersuchung in Kapitel 2.2.1 *Prototyp des Roboterkopfes* nachgewiesen, dass Zusammenhänge zwischen der vertikalen Richtung und Frequenzbändern nicht nur in der Medianebene erkennbar sind. Die Diagramme in Abbildung 2.5 und Tabelle 2.6 veranschaulichen diese These exemplarisch. Daher soll es Bestandteil dieses Verfahrens sein, das Spektrum des Audiosignals geeignet in Bänder zu untergliedern und die Amplitude innerhalb eines Bandes zu mitteln. So wird eine Datenreduktion erreicht. Es ist auch davon auszugehen, dass die Robustheit der Schallquellen-Lokalisation bezüglich Störungen im Spektrum erhöht werden kann, was die Auswertung in Kapitel 5.3.4 zeigen wird.

Vorgehen zur Reduktion spektraler Informationen

Für die geeignete Einteilung des Spektrums in Bänder gibt es keine allgemeine Festlegung. Viele biologisch inspirierte Verfahren zur Schallquellen-Lokalisation benutzen eine Gammaton-Filterbank zur Reduktion des Spektrums in wenige, logarithmisch steigende Bänder, wobei sich die Anzahl und Breite der Bänder in verschiedenen Verfahren unterscheidet [Beg94] [WB06]. Andere Modelle benutzen das gesamte Spektrum als Datenquelle wie [CNNU03] und [SR99]. In dem vorliegenden Verfahren soll die geeignete Einteilung der Bänder genetisch bestimmt werden. Es wird eine künstliche Evolution eingesetzt, um optimale Parameter für die Anordnung und Breite der Bänder zu suchen. Dies geschieht im Rahmen eines überwachten maschinellen Lernverfahrens, welches auf Lerndaten basiert, die aus einem Satz von Audiovektoren und den zugehörigen vertikalen Richtungen bestehen, angewendet wird. Beide Konzepte wurden bereits in der Einleitung vorgestellt (siehe Kapitel 1.4 *Allgemeine Verfahren und Konzepte*).

Vorgehen zur Berechnung der Elevation

Die gesuchten vertikalen Richtungsinformationen können aus den spektralen Eigenschaften abgeleitet werden, die aus einem Audiovektor eines Ohres extrahiert werden. Es wird vorausgesetzt, dass es sich um das Ohr handelt, welches der Schallquelle zugewandt ist. Anhand der Laufzeitunterschiede zwischen beiden Audiokanälen ist dieser leicht zu detektieren.

Zur Berechnung der Elevation wird die *Principal Component Analysis* verwendet, die in Kapitel 1.4 *Allgemeine Verfahren und Konzepte* vorgestellt wurde. Nachfolgend wird das Konzept in angepasster Form zur Ermittlung der Elevation genutzt: Den Ausgangspunkt bildet die Formel 1.5 zur Berechnung der N dekorrelierten Zufallsvektoren \mathbf{g}_ν , welche nachfolgend noch einmal kurz aufgeführt wird:

$$g_{\nu\mu} = \mathbf{V}^T(\lambda_\nu) \cdot [f_{1\mu} \ f_{2\mu} \ \dots \ f_{N\mu}]^T. \quad (4.14)$$

Die Eingangsgrößen sind N korrelierte Zufallsvektoren oder Messreihen \mathbf{f}_ν mit jeweils M Elementen $f_{\nu\mu}$. Der Index ν (mit $\nu = 1..N$) bezeichnet ein Element der Menge der Variablen beziehungsweise eine Dimension der Datenbasis, wohingegen μ (mit $\mu = 1..M$) ein Element der Menge der Messungen angibt. Die Ausgangsgrößen werden analog mit $g_{\nu\mu}$ bezeichnet. Der Vektor $[f_{1\mu} \ f_{2\mu} \ \dots \ f_{N\mu}]$ entspricht den extrahierten Frequenzinformationen eines Audiovektors. Der Eigenvektor $\mathbf{V}(\lambda_\nu)$ mit der Dimension $N \times 1$ gibt die Transformationsvorschrift für die Hauptkomponente ν an.

Gesucht wird der Eigenvektor \mathbf{V}_1 , welcher die Eingangsvektoren $[f_{1\mu} \ f_{2\mu} \ \dots \ f_{N\mu}]$ auf eine Hauptkomponente g_1 , nämlich die zugehörige vertikale Richtung $g_{1\mu}$ abbildet. Daher kann die Dimension der Ausgangsgrößen auf $R = 1$ reduziert werden.

Gegeben sind sowohl die Eingangsgrößen als auch die Ausgangsgrößen, die den Elevationen in Grad der Messanordnungen der Lerndaten entsprechen. Die Transformationsmatrix enthält demnach nur den Eigenvektor $\mathbf{V}(\lambda_1)$. Die zu berechnende Formel lautet:

$$g_{1\mu} = \mathbf{V}^T(\lambda_1) \cdot [f_{1\mu} \ f_{2\mu} \ \dots \ f_{N\mu}]^T. \quad (4.15)$$

Der Index ν kann für g_ν im Weiteren entfallen, da er für die Ausgangsgröße konstant ist, analog wird statt $\mathbf{V}(\lambda_1)$ die Bezeichnung \mathbf{V} verwendet. Da die Transformation \mathbf{V} für alle Messungen μ gelten soll, lautet die Formel:

$$\mathbf{g} = \mathbf{V}^T \cdot \mathbf{F}. \quad (4.16)$$

Die Matrix \mathbf{F} setzt sich aus den Elementen $f_{\nu\mu}$ zusammen und hat die Dimension $N \times M$. Es wird anschließend nach dem gesuchten Eigenvektor umgestellt:

$$\mathbf{V}^T = \mathbf{g} \cdot \mathbf{F}^{-1}. \quad (4.17)$$

Bei der Matrix \mathbf{F}^{-1} handelt es sich um die Pseudo-Inverse von \mathbf{F} , auch *Moore-Penrose-Inverse* genannt [Sto02]. In [MH04] wurde festgestellt, dass die Hauptachsentransformation „problem- oder prozessabhängig ist, da die Transformationsmatrix an die statistischen Eigenschaften des signalgenerierenden Prozesses angepasst wird“. Wurden also die Lerndaten repräsentativ für die statistischen Eigenschaften des gesamten Datenraums ausgewählt, dann lässt sich die Transformationsmatrix auch auf andere Vektoren anwenden. Ziel ist es, die vertikale Richtung unbekannter Vektoren des Datenraums mit guter Näherung zu bestimmen.

Im nächsten Abschnitt werden die beschriebenen Konzepte geordnet und zu einem Verfahren zur Richtungsbestimmung vereinigt.

4.3.2. Synthese des Verfahrens

Das Verfahren zur Bestimmung der vertikalen Richtung gliedert sich in die *Lernphase* und die *Testphase*. Letztere dient zur Untersuchung der Genauigkeit des Verfahrens, aber auch zur Auswahl der optimalen Konfiguration für unbekannte Audiovektoren. Hinsichtlich des erwarteten Speicher- und Rechenbedarfs des nachfolgend beschriebenen Verfahrens, soll an dieser Stelle auf folgendes hingewiesen werden: Wenn das Verfahren zur Schallquellen-Lokalisation auf einem Roboter Anwendung findet, dann läuft nur ein Teil, nämlich der Berechnungsalgorithmus zur Bestimmung der vertikalen Richtung, auf der eingebetteten Hardware des Roboters ab. Der Großteil, bestehend aus der Lernphase und einem Teil der Testphase, wird einmalig im Vorfeld offline zur Kalibrierung der benötigten Parameter auf einem PC ausgeführt und unterliegt daher nicht denselben hohen Anforderungen der eingebetteten Hardware wie der Berechnungsalgorithmus. Die Einzelschritte beider Phasen sind in Abbildung 4.7 dargestellt. Die Daten der Lernphase bestehen aus einer Menge des Typs (\mathbf{a}_i, el_i) , welcher je einen Audiovektor \mathbf{a}_i sowie die zugehörige tatsächliche Elevation el_i beinhaltet. Die Gesamtzahl der Elemente wird mit A bezeichnet. Die Lerndaten sind eine Teilmenge des Audiodatensatzes aus Kapitel 3.3. Analog werden die *Testdaten* aus den übrigen Audiovektoren des Datensatzes und ihren Elevationen

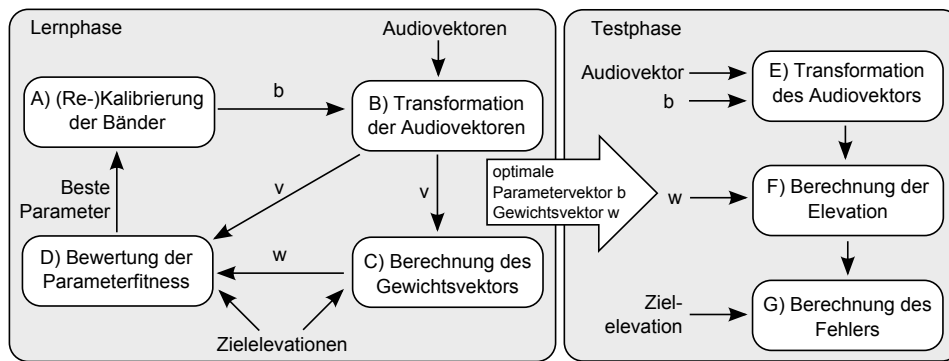


Abbildung 4.7. – Die vertikale Schallquellen-Lokalisation ist in die Trainings- (links) und die Testphase (rechts) gegliedert. Sie setzen sich aus Einzelschritten zusammen, die nacheinander angewendet werden. Die Pfeile symbolisieren den Datenfluss. Die Daten lauten: b (Bandparametervektor), v (reduzierte Audiovektoren), w (Gewichtsvektor).

gebildet.

Die Lernphase wird durch die künstliche Evolution gesteuert. Die Individuen einer Population werden mit dem Index j identifiziert. Die Gene umfassen den Parametervektor \mathbf{b}_j , der im **Schritt A** genauer definiert wird. Die Schritte der Lernphase werden nachfolgend erläutert:

A) (Re-)Kalibrierung der Bänder

Wie bereits im vorhergehenden Abschnitt dieses Kapitels beschrieben wurde, wird ein reduzierter spektraler Vektor aus einem gegebenen Audiovektor extrahiert. Dazu ist die Einteilung des Vektors in *Bänder* nötig. Die Konfiguration der Bänder wird im Parametervektor \mathbf{b}_j gespeichert. Mit der Bezeichnung *Bänder* sind in Anlehnung an die Frequenzbänder eines Spektrums zusammenhängende Abschnitte der Hüllkurve gemeint. Die Analogie wurde gewählt, da die Hüllkurve eine Art Amplitudenspektrum des zugrunde liegenden Sinussweeps darstellt.

Folgende Regeln sollen bei der Einteilung des Audiovektors gelten: Die Anzahl B der Bänder ist innerhalb einer Evolution konstant. Die Bänder werden so angeordnet, dass sie den Audiovektor vollständig abdecken und sich nicht überlappen. Das erste Band beginnt mit dem ersten Element des Vektors, das letzte endet mit dem letzten Index. Der Parametervektor beinhaltet die $B - 1$ Sample-IDs, welche die dazwischen liegenden Grenzen der Bänder darstellen. Sie sind aufsteigend angeordnet. Jedes Band darf eine festgelegte Mindestbreite nicht unterschreiten. Zu Beginn der

Evolution wird der Parametervektor zufällig mit gültigen Indizes belegt. Vor jedem neuen Evolutionszyklus werden die Parameter der neuen Vektoren entsprechend evolutionärer Regeln aus der vorhergehenden Generation berechnet (siehe Listing A.1). Wie diese Bänder zur Extraktion der spektralen Eigenschaften auf den Audiovektor anzuwenden sind, wird im nächsten Schritt beschrieben.

B) Extraktion eines komprimierten, spektralen Vektors

Gegeben sei ein Vektor von Bandparametern \mathbf{b}_j sowie ein Audiovektor \mathbf{a}_i . Mit der Methode „extract“ soll die Bestimmung des reduzierten, spektralen Vektors \mathbf{v}_i aus dem Audiovektors \mathbf{a}_i unter Verwendung des Parametervektors \mathbf{b}_j zusammengefasst werden:

$$\mathbf{v}_i = \text{extract}(\mathbf{a}_i, \mathbf{b}_j). \tag{4.18}$$

Im ersten Schritt wird die Hüllkurve des Audiovektors bestimmt, welche eine Fre-

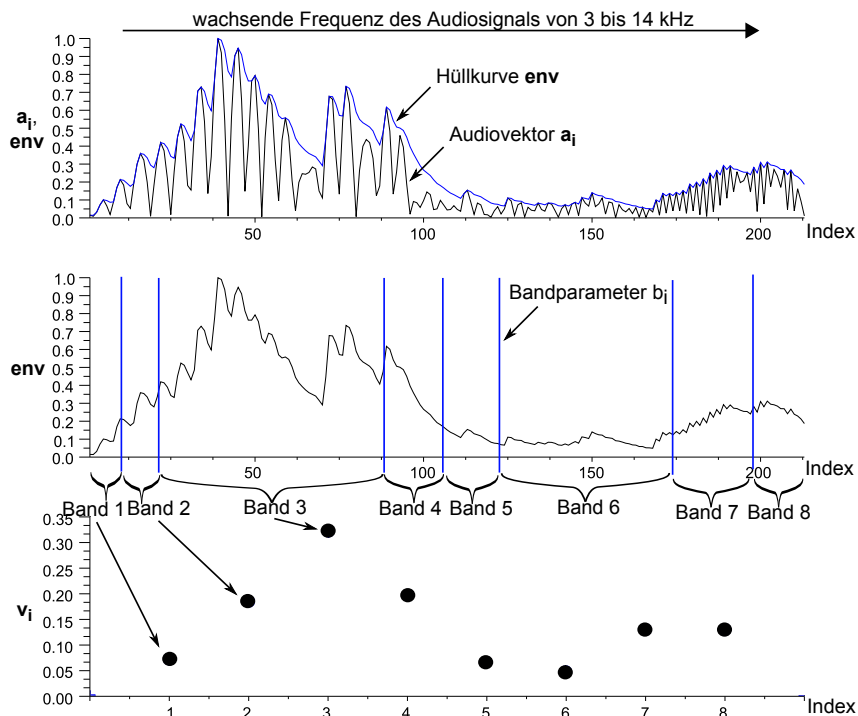


Abbildung 4.8. – Schematische Darstellung der stufenweisen Extraktion des reduzierten spektralen Vektors \mathbf{v}_i . Im oberen Diagramm wird die Hüllkurve \mathbf{env} des Audiovektors \mathbf{a}_i extrahiert. Das mittlere Diagramm zeigt die Untergliederung der Hüllkurve in Bänder. Der resultierende Vektor \mathbf{v}_i berechnet sich aus den Mittelwerten der Elemente eines Bandes.

quenzabbildung des Audiovektors darstellt (siehe Diagramm 1 in Abbildung 4.8). Durch die aufsteigende Anordnung isolierter Schwingungen des Audiosignals kann theoretisch zu jeder Schwingung der mittlere oder maximale Pegel bestimmt werden. Die Hüllkurve bildet somit den groben Charakter des Frequenzspektrums des Signals nach. Allerdings können die Signalreflexionen, welche den Direktschall überlagern, die spektrale Abbildung verfälschen. Es wird davon ausgegangen, dass die Reflexionen im Allgemeinen nur schwach sind und der Einfluss durch die weitere Verarbeitung des Signals vernachlässigt werden kann. Der zweite Teil der Extraktion eines spektralen Vektors besteht in der Reduktion des Hüllkurvenvektors. Dazu wird der Vektor in Abhängigkeit der Bandparameter \mathbf{b}_j aus **Schritt A** vollständig in Abschnitte gegliedert und diese anschließend durch Mittelwertbildung zu je einem Wert zusammengefasst (siehe Diagramm 2 in Abbildung 4.8). Diese Werte bilden in der Reihenfolge der Bänder die Elemente des Vektors \mathbf{v}_i (siehe Diagramm 3 in Abbildung 4.8).

C) Bestimmung des Gewichtsvektors

Die Grundlage für die Berechnung des Gewichtsvektors \mathbf{w} bildet die Gleichung 4.17. Diese wurde im vorhergehenden Abschnitt aus der PCA hergeleitet. Unter Berücksichtigung der in diesem Kapitel verwendeten Bezeichner lautet sie:

$$\mathbf{w} = \mathbf{e}l \cdot \mathbf{F}^{-1}. \quad (4.19)$$

Der Gewichtsvektor \mathbf{w} entspricht dem Eigenvektor \mathbf{V} . Dieser berechnet sich aus der Eingangsmatrix \mathbf{F} , welche sich aus den Zeilenvektoren \mathbf{v}_i zusammensetzt, und den Ausgangsgrößen, die durch den Vektor $\mathbf{e}l$ zusammengesetzt werden. Die Anzahl B der Bänder entspricht der Dimension der Vektoren N und die Anzahl der Audiovektoren A der Lerndaten der Anzahl der Messungen M .

Nach erfolgreicher Bestimmung des Gewichtsvektors \mathbf{w} kann für einen beliebigen Eingangsvektor \mathbf{v}_i die Elevation $estEl_i$ interpoliert werden:

$$estEl_i = \mathbf{w} \cdot \mathbf{v}_i^T. \quad (4.20)$$

Die Gleichung 4.20 wird für die Umsetzung der *Schritte D und F* benötigt. Die Genauigkeit der Interpolation unbekannter Vektoren wird in Kapitel 5.3.3 *Analyse der Richtungsgenauigkeit* untersucht.

D) Bewertung der Fitness

Als letzter Schritt der Evolution wird die *Fitness* der Parametervektoren bestimmt, um die besten Vektoren für den nächsten Zyklus auszuwählen. Die Fitness fit eines Individuums j berechnet sich aus der durchschnittlichen Abweichung $\overline{\Delta el}$ der berechneten vertikalen Richtungen $estEl$ von den tatsächlichen Elevationen el :

$$\overline{\Delta el} = \frac{\sum_{A}^{i=1} |el_i - estEl_i|}{A}. \quad (4.21)$$

Der Gewichtsvektor \mathbf{w} wird auf alle Eingangsvektoren \mathbf{v}_i angewendet und die zugehörige Richtung el_{est_i} bestimmt. Die Fitnessfunktion eines Individuums lautet:

$$fit = \frac{\sum_{A}^{i=1} |el_i - \mathbf{w} \cdot \mathbf{v}_i^T|}{A}. \quad (4.22)$$

Die Individuen mit der besten Fitness werden in die nächste Generation übernommen und die übrigen Individuen durch Mutation und Kreuzung neu generiert. Die Evolutionsparameter Mutationsrate, Kreuzungsrate sowie die Dauer der Evolution werden in der Untersuchung geeignet festgelegt.

Nach Abschluss der Evolution werden der Parametervektor \mathbf{b}_{opt} sowie der zugehörige Gewichtsvektor \mathbf{w} des besten Individuums in die Testphase übernommen. Diese werden auf unbekannte Testdaten angewendet und die zugehörige Richtung interpoliert. Damit lässt sich eine Aussage über die Genauigkeit des Verfahrens treffen. Die Testphase setzt sich aus den folgenden drei Schritten zusammen, die denen der Lernphase sehr ähnlich sind:

E) Transformation der Audiovektoren

Die Transformation läuft für alle Audiovektoren \mathbf{a}_i analog zum **Schritt B** der Lernphase mit der Methode *extract* ab.

F) Berechnung der Elevation

Die Berechnung der Elevation jedes Testvektors läuft entsprechend der Gleichung 4.20 aus **Schritt C** ab.

G) Berechnung des Fehlers

Die durchschnittliche Abweichung wird entsprechend der Formel 4.21 aus **Schritt C** berechnet.

Zusammenfassung

Die Evolution der Bandparameter in **Schritt A** hat zum Ziel, dass nach der Vektortransformation in **Schritt B** der Informationsgehalt des Ausgangsvektors für die Richtungsbestimmung maximal wird und Reflexionen herausgemittelt werden. Anschließend wird im **Schritt C** auf Grundlage der PCA ein Gewichtsvektor bestimmt, der sowohl für die Lerndaten zum kleinsten Fehler führt (siehe Fitnessberechnung in **Schritt D**), als auch für unbekannte Daten eine Richtungsschätzung mit hinreichender Genauigkeit zulässt (siehe **Schritt E** bis **G**). Im Kapitel 5.3 Untersuchung und Analyse der *Bestimmung der Elevation* wird dieses Verfahren implementiert und mit Hilfe des Testdatensatzes untersucht.

5. Untersuchung und Auswertung

In den folgenden drei Unterkapiteln werden die im vorhergehenden Kapitel hergeleiteten und entwickelten Modelle und Verfahren implementiert und auf Basis des Testdatensatzes parametrisiert und untersucht. Es wird die korrekte Funktionsweise der Implementierung nachgewiesen und die Genauigkeit der Verfahren bestimmt. Begonnen wird mit der Untersuchung der Onset-Detektion. Anschließend wird die Untersuchung der horizontalen Richtungsbestimmung durchgeführt, welche eine korrekte Onset-Detektion voraussetzt. Abschließend wird die Bestimmung der Elevation untersucht und ausgewertet.

5.1. Detektion des Onsets

Ziel der Untersuchung des Modells der Onset-Detektion ist es zu zeigen, dass das Modell geeignet ist, den Onset der Signalvektoren des Testdatensatzes zu detektieren. Dazu wird im ersten Abschnitt der implementierte Code der Onset-Detektion kurz vorgestellt und die Parametrisierung erläutert. Anschließend wird das Modell auf alle Vektoren des Testdatensatzes angewendet und das Modell auf seine Genauigkeit hin untersucht. Anschließend wird das Modell der Onset-Detektion auf den fehlerfreien Datensatz angewendet, um eine allgemeine Aussage über die Genauigkeit der Onset-Detektion treffen zu können.

5.1.1. Anpassung der Modellparameter

Als Grundlage für die Untersuchung dient der Audiodatensatz bestehend aus je fünf nacheinander aufgenommenen Stereo-Signalvektoren der 35 Messanordnungen mit je einem Onsetpaar pro Einzelmessung. Somit sind insgesamt 350 Onsets beziehungsweise 175 Onsetpaare zu bestimmen. Ein Ausschnitt der Implementierung der Onset-Detektion wird nachfolgend aufgeführt:

```

1 for c = 1:2                                // left/right channel
2   for s = 1:BUFFERLENGTH-4                // samples
3     if(buf(s,c) > MINLEVEL & buf(s,c) > buf(s-1,c)
4     & buf(s,c) >= buf(s+1,c) & buf(s,c) > buf(s+2,c)
5     & buf(s,c) > buf(s+3,c) & buf(s,c) > buf(s+4,c))
6       for n = MINDIFF:MAXDIFF
7         if(buf(s+n,c) < -MINLEVEL
8         & buf(s+n,c) < buf(s+n-1,c)
9         & buf(s+n,c) <= buf(s+n+1,c))
10          onset(channel) = s + n;          // i_onset = i_min
11          break;                          // finish searching min
12        end
13      end
14      if(onset(channel) ~= 0) // onset found
15        break;                          // finish parsing buffer
16      end
17    end
18  end
19 end
20 if(onset(1) ~= 0 & onset(2) ~= 0)
21   ITD = onset(2) - onset(1)
22 end

```

Listing 5.1 – Scilab-Code der Onset-Detektion

Die empirisch bestimmten Parameterbelegungen werden in Tabelle 5.1 aufgelistet. Die Konstante $C_{\max\text{BufferLevel}}$ bezeichnet die maximale Amplitude, die in dem jeweiligen Signalvektor gemessen wurde. Damit wird die Mindestamplitude des Onsets

$C_{\text{bufferLength}}$	400
C_{minLevel}	$0,045 \cdot C_{\max\text{BufferLevel}}$
C_{minDiff}	4
C_{maxDiff}	15

Tabelle 5.1. – Deklaration der Konstanten des Verfahrens zur Onset-Detektion und ihre Belegungen in dieser Implementierung.

$C_{\min\text{Level}}$ relativ zum Maximalpegel des Audiovektors bestimmt. Eine geringfügige Erweiterung ist im Vergleich zum entworfenen Modell bezüglich der Bedingung des Maximums notwendig, wie anhand des Quellcodes ersichtlich ist. Zudem musste die Konstante $C_{\max\text{Diff}}$ sehr tolerant eingestellt werden. Grund dafür ist, dass bei einer extremen Lautsprecheranordnung nah am Boden die erste Reflexion sehr früh auftritt, die die Wellenform der ersten Schwingung leicht verzerrt. Mit diesen Parametern wird im nachfolgenden Abschnitt die Untersuchung der Onset-Detektion durchgeführt.

5.1.2. Untersuchung

Durch Anwendung des Verfahrens zur Onset-Detektion auf den Test-Datensatz wird der Onset für jeden Signalvektor bestimmt. Da das zeitliche Auftreten des Nutzsignals im Signalvektor variieren kann, müsste der tatsächliche Onset für jeden der 350 Vektoren manuell festgelegt werden, um die Genauigkeit der Detektion bestimmen zu können. Dies ist aus zeitlichen Gründen nicht möglich. Stattdessen wird die Tatsache ausgenutzt, dass das zeitliche Auftreten des Nutzsignals in den beiden Kanälen einer Stereoaufnahme in einem festen zeitlichen Verhältnis stehen muss. Eine einfache Methode zur Analyse der korrekten Onset-Detektion sieht dann wie folgt aus:

In einem ersten Schritt wird für jeden der Stereosignale der Onset sowohl für den linken als auch für den rechten Kanal detektiert. Daraus ergeben sich 175 Onsetpaare. Die fünf Messungen einer Messkonfiguration werden zu einer Gruppe zusammengefasst und innerhalb dieser verglichen. Für jedes Stereovektorpaar kann auf Basis der detektierten Onset-Samples die Laufzeitdifferenz des Signals berechnet werden. Diese sollte innerhalb einer Messkonfigurationsgruppe um maximal einen Sample variieren. Dies entspricht der doppelten Ungenauigkeit, welche bei der Bestimmung der Zeitpunkt eines einzelnen Onsets aufgrund des Quantisierungsfehlers theoretisch auftreten kann wie bereits in der im Kapitel 4.1 erläutert wurde. Maximal kann die Abweichung allerdings zwei Samples betragen, wenn Faktoren wie Signalrauschen bei der Übertragung oder Messungenauigkeiten einen Einfluss haben. Eine vollständige Auflistung aller Laufzeitdifferenzen und der maximalen Variation innerhalb einer Gruppe befindet sich in Tabelle 5.2. Anhand der Tabelle ist ersichtlich, dass fast alle berechneten Onset-Differenzen innerhalb der erwarteten Parameter liegen. In der letzten Zeile *Abw.* wird die maximale Abweichung der größten und kleinsten

Laufzeitdifferenz innerhalb einer Messkonfiguration dargestellt. In den Anordnungsgruppen 4 und 14 gibt es jeweils einen anomalen Wert, welcher bei genauer Betrachtung der Audiovektoren auf einen Signalfehler zurückzuführen ist. Dieser entsteht durch eine Wiedergabestörung des Lautsprechers, die selten auftritt und akustisch wahrnehmbar ist.

ID	1	2	3	4	5	6	7	8	9	10	11	12
M_1	2	-6	-14	-20	-23	1	-8	-17	-26	-29	2	-9
M_2	2	-6	-14	-21	-23	0	-9	-17	-25	-29	1	-8
M_3	2	-6	-14	-21	-24	1	-9	-17	-24	-28	1	-9
M_4	2	-6	-15	-27	-24	0	-9	-17	-26	-29	1	-9
M_5	2	-6	-14	-20	-23	1	-9	-17	-25	-28	1	-9
<i>Abw.</i>	0	0	1	7	1	1	1	0	2	1	1	1
ID	13	14	15	16	17	18	19	20	21	22	23	24
M_1	-19	-28	-34	-1	-8	-20	-28	-37	0	-8	-19	-27
M_2	-19	-28	-33	-1	-8	-19	-28	-36	0	-8	-19	-27
M_3	-19	-29	-34	-2	-8	-20	-28	-36	0	-7	-19	-27
M_4	-19	-35	-33	-2	-8	-20	-28	-37	1	-9	-18	-27
M_5	-19	-28	-33	-2	-8	-20	-29	-37	0	-8	-19	-27
<i>Abw.</i>	0	7	1	1	0	1	1	1	1	2	1	0
ID	25	26	27	28	29	30	31	32	33	34	35	
M_1	-33	-1	-6	-16	-26	-32	-2	-7	-15	-21	-26	
M_2	-33	-1	-7	-16	-25	-32	-2	-7	-15	-22	-27	
M_3	-33	-1	-6	-16	-25	-32	-2	-7	-16	-21	-26	
M_4	-33	-1	-6	-16	-26	-32	-1	-6	-16	-21	-27	
M_5	-33	-1	-6	-16	-25	-32	-2	-7	-15	-21	-26	
<i>Abw.</i>	0	0	1	0	1	0	1	1	1	1	1	

Tabelle 5.2. – Übersicht über die spaltenweise gruppierten Laufzeitdifferenzen in Samples der fünf Messungen M_i jeder Konfigurationsgruppe. In der letzten Zeile *Abw.* wird die maximale Abweichung zwischen der kleinsten und größten Laufzeitdifferenz je Anordnungsgruppe in Samples angegeben.

5.1.3. Auswertung

Über die Genauigkeit der Onset-Bestimmung kann nur eine relative Aussage getroffen werden, indem die Laufzeitdifferenzen der mehrfach wiederholten Stereomessungen verglichen werden. In der Untersuchung wurde gezeigt, dass die Onset-Detektion mit dem gewählten Verfahren und einer geeigneten Parametrisierung erfolgreich auf die Testdaten angewendet werden kann. Die Genauigkeit liegt innerhalb der erwarteten Toleranz von einem Sample (in Ausnahmefällen auch zwei Samples). Die zwei anomalen Messergebnisse sind auf Störungen der Signalwiedergabe zurückzuführen, die trotz sorgfältiger Messdurchführung auftreten können. Für die weitere Verwendung der Onset-Differenzen werden diese vernachlässigt. Der Test-Datensatz deckt ein großes Spektrum an Schallquellenpositionen ab. Zudem ist die Detektionsmethode durch eine zusätzliche Anpassung robust bezüglich des Spezialfalls von früh eintreffenden Reflexionen. Dies lässt den Analogieschluss zu, dass die Onset-Detektion unter vergleichbaren Bedingungen für unbekannte, aber sich in die Messanordnung einreihende Positionen der Schallquelle mit vergleichbarer Genauigkeit ausgeführt werden kann.

Im nächsten Kapitel wird überprüft, ob die berechneten Laufzeitdifferenzen zu den erwarteten Seitenwinkeln führen. Als Grundlage werden die gemittelten Laufzeitdifferenzen einer Messanordnung verwendet (siehe Spalte 4 in Tabelle A.3).

5.2. Bestimmung des Azimuts

Ziel dieser Untersuchung ist es, unter anderem das entwickelte 2D-Modell eines Roboterkopfes, welches als endgültiges Modell in Abschnitt 4.2 bestimmt wurde, auf Grundlage der empirischen Untersuchung zu vervollständigen. Dazu wird der Fehler des zuvor bestimmten 1D-Modells des Roboterkopfes auf Grundlage des Testdatensatzes analysiert, um daraus Verallgemeinerungen für die Ausprägung des Approximationsterms im 2D-Modell abzuleiten. Abschließend wird eine Aussage über die Genauigkeit des Modells bei Anwendung auf den Testdatensatz getroffen.

Die Berechnung des Seitenwinkels erfolgt auf Grundlage der durchschnittlichen Laufzeitdifferenzen auf Basis der Onset-Bestimmung (siehe Tabelle A.3 im Anhang). Zur Richtungsberechnung und zur Fehlerbestimmung werden die tatsächlichen Seiten- und Höhenwinkel des Messaufbaus benötigt, die im Weiteren mit φ_{soll} und ϑ_{soll} bezeichnet werden.

5.2.1. Untersuchung und Vervollständigung des Modells

Um eine geeignete Erweiterung des 2D-Modells zu bestimmen, wird der Seitenwinkel φ_0 des 1D-Modells auf Basis der Gleichung 4.12 berechnet. In Abbildung 5.3 wird die Abweichung des Seitenwinkels $\Delta\varphi_0$ zum originalen Seitenwinkel des Messaufbaus dargestellt, welche sich wie folgt berechnet: $\Delta\varphi_0 = \varphi_{\text{soll}} - \varphi_0$. Die Werte wurden entsprechend ihres ursprünglichen Höhenwinkels ϑ_{soll} (siehe Bezeichnung in der Legende) durch zusammenhängende Graphen dargestellt. Es fällt leichter, Abhängigkeiten vom Höhenwinkel zu erkennen. Zusätzlich wurde die gemittelte Abweichung je Seitenwinkel diskret eingezeichnet um eine mögliche Verallgemeinerung bezüglich des Seitenwinkels treffen zu können.

Es wird jetzt versucht, aus den Graphen Schlussfolgerungen für die Ausprägung des Korrekturterms $g(\vartheta)$ in Gleichung 4.13 des Seitenwinkels herzuleiten. Mit Hilfe der Korrektur soll es möglich sein, den Fehler des 2D-Modells des Roboterkopfes auf einem gleichmäßigen niedrigen Niveau zu halten. Folgende Aussagen können aufgrund der graphischen Auswertung getroffen werden: Zum einen ist eine Abhän-

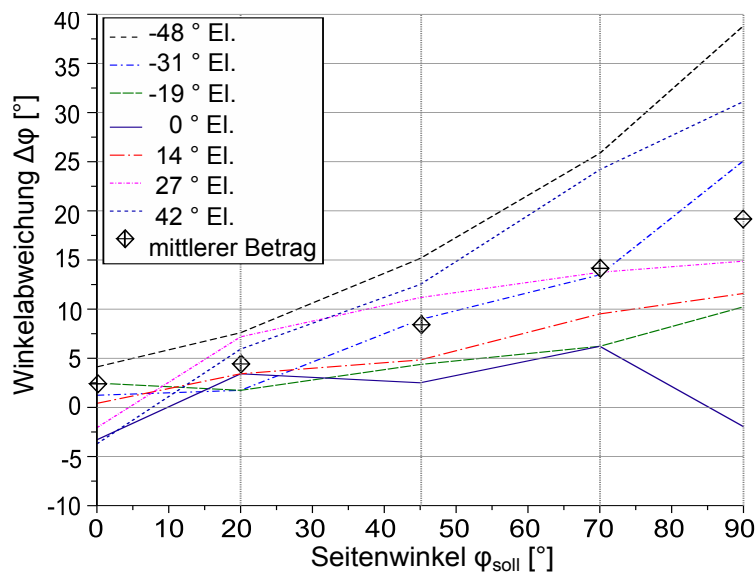


Abbildung 5.3. – Darstellung der Differenz $\Delta\varphi_0$ des Seitenwinkels vom tatsächlichen Winkel φ_{soll} , die nach Formel 4.12 berechnet wurde. Jeder Graph repräsentiert die Ergebnisse der Messanordnungen mit einer festen Elevation ϑ_{soll} . Zusätzlich wurde die mittlere Abweichung mit Rauten eingezeichnet.

gigkeit vom Seitenwinkel selbst zu erkennen, da fast alle Graphen mit wachsendem Seitenwinkel linear ansteigen. Bei einer Elevation von 0 Grad, scheint sich die Winkelabweichung dagegen im Mittel aufzuheben. Der Azimut sollte also einfach in den Korrekturterm einfließen. Zum anderen ist festzustellen, dass der Anstieg desto größer ist, je höher der Betrag der zugehörigen Elevation ist. Auch ist auffällig, dass die Graphen mit einer niedrigen Elevation von +14 und -19 Grad sehr flach ansteigen (das Maximum beider liegt bei circa 10 Grad) und die Graphen mit einer mittleren Elevation +27 und -31 Grad nur etwas steiler sind (das Maximum beider liegt im Mittel bei 20 Grad). Dagegen ist der Anstieg der Graphen +42 und -48 Grad Elevation deutlich größer (das Maximum beider liegt im Mittel bei 35 Grad). Betrachtet man dazu die Verhältnisse der zugrunde liegenden Elevationen, so kann daraus geschlossen werden, dass die Elevation quadratisch in den Korrekturterm einfließen sollte, um den Fehler auszugleichen. Zusammenfassend lautet die Korrekturfunktion $g(\vartheta)$ wie folgt:

$$g(\vartheta) = \varphi_0 \cdot \vartheta^2 \cdot x_{\text{norm}}. \quad (5.1)$$

Der Normierungsfaktor x_{norm} wird mit Wert $x_{\text{norm}} = 2,2/8100 = 2,72 \cdot 10^{-4}$ eingeführt. Der Divisor ergibt sich aus dem Quadrat des maximalen Höhenwinkels. Nachdem alle Parameter fest standen, wurde der Dividend durch manuelle Minimierung des Gesamtfehlers bestimmt. Die im Entwurf noch unvollständige Formel 4.13 des 2D-Modells zur Berechnung des Seitenwinkels kann durch die konkrete Approximationsfunktion 5.1 vervollständigt werden und lautet:

$$\varphi = (1 + \vartheta^2 \cdot x_{\text{norm}}) \cdot x_{\varphi} \cdot \arcsin\left(\frac{l}{L \cdot x_S}\right). \quad (5.2)$$

Die Konstante L wird auf 36,6 festgelegt. Dies wird aus der durchschnittlichen Laufzeitdifferenz des maximalen Seitenwinkels in der Transversalebene der Kopfes abgeleitet (entsprechend Konfiguration 20 in Tabelle A.3 im Anhang). Nachdem nun das 2D-Modell des Roboterkopfes vervollständigt wurde, kann die Gleichung 5.2 noch einmal auf die Testdaten angewendet werden um die Winkelabweichung zu bestimmen. Diese wird analog zu Abbildung 5.3 in Abbildung 5.4 veranschaulicht. Es ist eine deutliche Verbesserung zu sehen. Die genaue Auswertung des Fehlers erfolgt im nächsten Abschnitt.

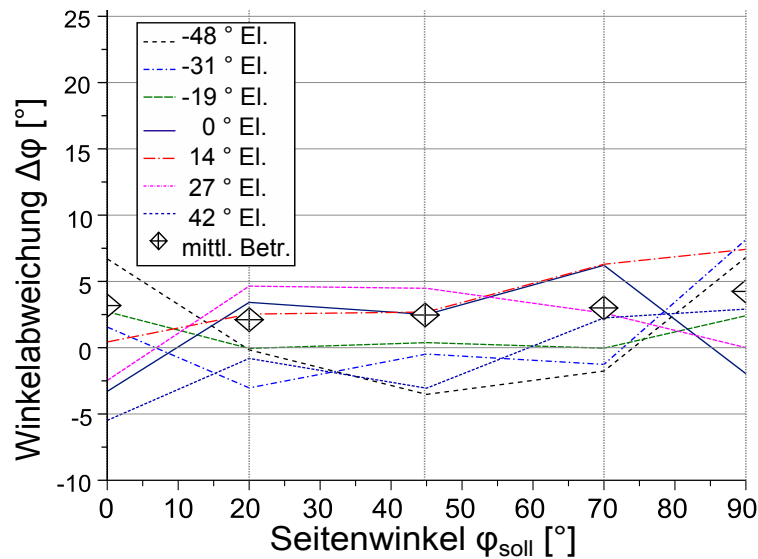


Abbildung 5.4. – Darstellung der Differenz des Seitenwinkels vom tatsächlichen Winkel φ_{soll} nach Formel 5.2 berechnet. Jeder Graph repräsentiert die Ergebnisse einer Messanordnung mit einem festen Höhenwinkel ϑ_{soll} .

5.2.2. Auswertung

Durch die Erweiterung des 1D-Modells zum 2D-Modell kann für Positionen mit einem Höhenwinkel ungleich Null eine deutliche Verbesserung der Berechnung des Seitenwinkels festgestellt werden (siehe Tabelle A.3 im Anhang). Die durchschnittliche Abweichung des Winkels beträgt für die Testdaten im 1D-Modell 9,8 Grad. Im endgültigen 2D-Modell sind es dagegen nur noch 3,0 Grad, was weniger als ein Drittel des vorhergehenden Fehlers ist. Zudem konnte die maximale Abweichung von 39,8 auf 8,4 Grad gesenkt werden, was etwa einem Fünftel entspricht.

Die benötigte Approximationsfunktion, welche die Elevation und den ursprünglichen Seitenwinkel des 1D-Modells verwendet, wurde mit Hilfe der Testdaten aus der Winkelabweichung des 1D-Modells abgeleitet. Da das Modell teilweise auf einem beschränkten Datensatz basiert und zwei empirisch bestimmte Konstanten benötigt, ist diese Ausprägung des 2D-Modells nur auf den hier verwendeten Roboterkopf anwendbar. Eine Fehlergenauigkeit von durchschnittlich 3,0 und maximal 8,4 Grad für die Testdaten ist sehr gut. Allerdings wird der reale Fehler etwas höher sein, da auch die zugrundeliegenden, gemessenen Ausgangspositionen der Schallquelle φ_{soll} und ϑ_{soll} einem systematischen Fehler unterliegen, der durch die begrenzte zur Ver-

fügung stehende Messtechnik begründet ist.

Zudem wurde in dieser Untersuchung anstelle des im nächsten Kapitel zu berechnenden Höhenwinkels der gemessene Winkel des Messaufbaus verwendet. Im späteren Einsatz steht dieser jedoch nicht zur Verfügung. Die reale Genauigkeit der Methode hängt also zusätzlich davon ab, mit welcher Genauigkeit der vertikale Winkel bestimmt wird. Aus der endgültigen Formel 5.2 geht hervor, dass der Einfluss des Elevationsfehlers aufgrund der quadratischen Abhängigkeit mit wachsender Elevation zunimmt. Dies wird aufgrund der Multiplikation bei kleinen Ausprägungen des Seitenwinkels φ_0 gedämpft und bei großen verstärkt. Bei einem theoretischen Offset der Elevation um $+5,0$ Grad erhöht sich der durchschnittliche Fehler des berechneten Seitenwinkels des Testdatensatzes um nur $0,6$ auf $3,6$ Grad. Wird beispielsweise von einem theoretischen Offset der Elevation um $+10,0$ Grad ausgegangen, beträgt der mittlere Fehler des Seitenwinkels $5,2$ Grad und liegt unter dem durchschnittlichen Fehler des 1D-Modells (siehe Tabelle A.3 im Anhang).

Im nächsten Kapitel wird die Elevation auf Grundlage eines entwickelten Verfahrens bestimmt und untersucht. Nach der Bestimmung der Winkelgenauigkeit der Elevation wird noch einmal kurz Bezug auf die Auswirkungen auf den Horizontalfehler genommen.

5.3. Bestimmung der Elevation

Die Untersuchung und Auswertung des Verfahrens zur vertikalen Richtungsanalyse gliedert sich in vier Teile:

Es beginnt mit einer kurzen Beschreibung und Erläuterung ausgewählter Bestandteile der Implementierung des Verfahrens aus Kapitel 4.3. Dazu wird insbesondere auf den Hüllkurvenoperator und die Evolution eingegangen. Anschließend wird die korrekte Funktionsweise der künstlichen Evolution nachgewiesen. Somit kann das Resultat der Parameter-Analyse als optimale Lösung angesehen werden. Dies ist wichtig, um die Resultate der Untersuchung hinsichtlich der Genauigkeit des Verfahrens korrekt interpretieren zu können. Im letzten Schritt wird die Winkelgenauigkeit des Verfahrens anhand zweier Varianten der Testdatenaufteilung untersucht. In der Auswertung wird eine Verallgemeinerung der zu erwartenden Richtungsgenauigkeit auf Grundlage der Untersuchungsergebnisse durchgeführt.

5.3.1. Erläuterung der Implementierung

Nachfolgend wird die Umsetzung der speziellen Bestandteile des Verfahrens zur vertikalen Richtungsanalyse vorgestellt. Dazu gehört die Extraktion der Audiovektoren (Lern-, Testdaten) aus dem Testdatensatz, die Bildung der Hüllkurve des Audiovektors (**Schritt B**, siehe Abbildung 4.7) sowie die Parametrisierung der Evolution (**Schritt A, D**).

Für die Umsetzung der mathematischen Formeln (**Schritt B, C, E – G**), die in Kapitel 4.3 vorgestellt wurden, werden Standard-Funktionen von Scilab verwendet. Daher wird hier auf eine genaue Betrachtung der Implementierung verzichtet.

Extraktion der Audiovektoren

Die benötigten Audiovektoren \mathbf{a}_i bestehen aus einer zusammenhängenden Samplefolge, die aus der ersten Aufnahme jeder Richtungskonfiguration des Testdatensatzes entnommen werden. Für die monaurale Analyse wird nur ein Kanal benötigt. Es wird der Kanal des Ohres verwendet, welcher der Schallquelle zugewandt ist. Der Audiovektor \mathbf{a} ist der Ausschnitt der Aufnahme, der 10 Samples vor dem detektierten Onset (Detektion siehe Kapitel 5.1) beginnt und 213 Samples lang ist. Die Vektorlänge entspricht der Dauer des 200-sample-langen Ausgangssignals, das in Kapitel 3.2 festgelegt wurde unter Berücksichtigung der Abtastfrequenz des Aufnahmesystems ($f = 46\,875\text{ Hz}$). Eine grafische Übersicht aller extrahierten Audiovektoren befindet sich in Abbildung A.4 im Anhang. Zur Untersuchung des Verfahrens im nächsten Kapitel werden dieselben 35 Vektoren des Audiodatensatzes auf verschiedene Weisen in Lern- und Testdaten aufgeteilt (siehe Abbildung 5.7). Im Weiteren wird zur Unterscheidung von Variante 1 und 2 gesprochen.

Bildung der Hüllkurve

Zur Bestimmung der Hüllkurve des Audiovektors, die zur Extraktion der Frequenzeigenschaften des Signals benötigt wird, dient der Algorithmus im Codeausschnitt 6.2. Es wird der Absolutbetrag des Vektors \mathbf{a} gebildet und dieser anschließend normiert. Die Normierung führt dazu, dass die weitere Analyse vom ursprünglichen Audiopegel weitgehend entkoppelt wird. Folglich ist die Bestimmung der vertikalen Richtung vom Ausgangspegel bei der Signalerzeugung sowie der Entfernung zwischen Quelle und Empfänger relativ unabhängig.

```

23 f = 1/8;          // weighting factor f
24 a = abs(a);      // absolute of audiovector a
25 a = a/max(a);    // normalization of a
26 env = zeros(length(a),1); // define envelope vector env
27 env(1) = a(1);
28 for i = 2 : length(a)
29     if(a(i) > env(i-1))
30         env(i) = a(i);
31 else
32     env(i) = env(i - 1) * (1 - f) + a(i) * f;
33 end

```

Listing 5.2 – Scilab-Code der Hüllkurvenbestimmung.

Die Hüllkurve wird nach folgender einfachen Regel gebildet: Über alle Elemente des Eingangsvektors \mathbf{a} wird geprüft, ob der Wert des aktuellen Samples $a(i)$ größer ist als der letzte Wert der bereits extrahierten Hüllkurve $env(i-1)$. In diesem Fall wird der Wert an die Stelle i der Hüllkurve übernommen. Anderenfalls wird der aktuelle Hüllkurvenwert mit der Formel eines einfachen IIR-Filters erster Ordnung berechnet (Beispiel siehe Abbildung 4.7 auf Seite 52).

Konfiguration der Evolution

Die bereits in Kapitel 4.3 beschriebene Evolution der Parametervektoren wurde mit den in Tabelle 5.5 dargestellten Konstanten konfiguriert. Der evolvierte Bandpara-

Anzahl der Generationen je Evolution	50
Anzahl der Wiederholung der Evolution	15
Anzahl der Individuen	20
Anzahl der selektierten Individuen	8
Mutationsfaktor pro Generation	0,5
Mutationsfaktor pro Individuum	0,5
Mindestbreite des ersten Bandes [Samples]	10
Mindestbreite der übrigen Bänder [Samples]	3
Anzahl der Bänder	8 bis 20

Tabelle 5.5. – Zusammenfassung der Parametrisierung der Evolution in der Lernphase. Durchgeführt wird die Lernphase für unterschiedliche aber innerhalb der Evolution feste Anzahl an Bändern (letzte Zeile).

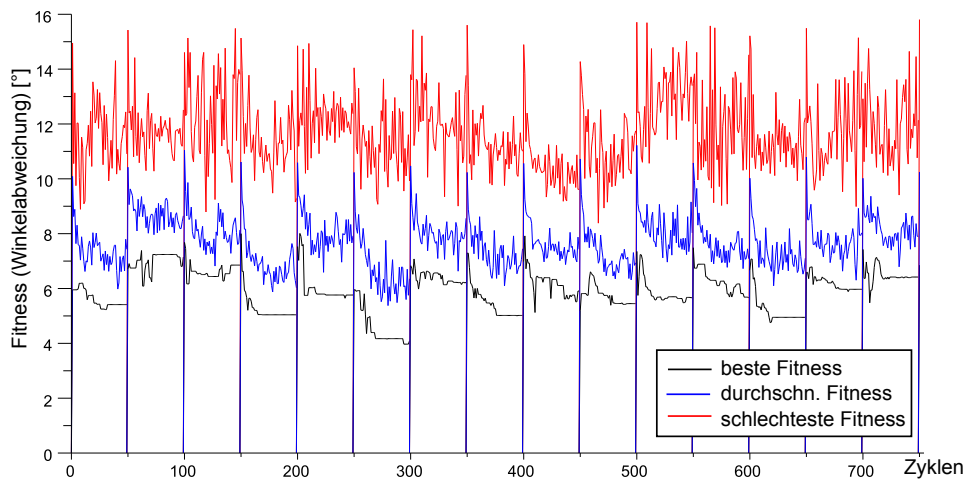


Abbildung 5.6. – Grafische Darstellung der Entwicklung der Fitness in 15 Wiederholungen einer Evolution mit 50 Generationen. Die Graphen repräsentieren die Individuen mit der besten (schwarz), schlechtesten (rot) Fitness sowie die durchschnittliche Fitness (blau) aller Individuen.

metervektor wird für eine feste Anzahl an Bändern gebildet, in dem nach 15-maliger Wiederholung der Evolution das Individuum mit der besten Fitness aus allen Durchläufen ausgewählt wird. Der zugehörige Bandparametervektor und Gewichtsvektor wird in der Testphase verwendet, um die Genauigkeit der Richtungsbestimmung der Testdaten zu bestimmen. Die Scilab-Implementierungen der wichtigsten Funktionen der Evolution werden in Listing A.1 dargestellt. Exemplarisch wird in Abbildung 5.6 der typische Verlauf der Fitness ausgewählter Individuen während der Evolution dargestellt.

5.3.2. Nachweis der korrekten Funktionsweise

Im vorhergehenden Abschnitt wurden die wichtigsten Abschnitte der Implementierung vorgestellt und die korrekte Funktionsweise der Evolution veranschaulicht. So kann nun das Verfahren bezüglich seiner Richtungsgenauigkeit anhand zweier Beispiele untersucht werden. Dazu wurden zwei Varianten ausgewählt, welche den Testdatensatz in unterschiedlicher Weise in Lern- und Testdaten gliedern.

In beiden Fällen werden aus dem Testdatensatz 30 Audiovektoren verschiedener Konfigurationen als Lern- und fünf als Testdaten ausgewählt (siehe Abbildung 5.7). Es kommen dieselben Vektoren in anderer Kombination zum Einsatz. In der ersten Variante soll geprüft werden, wie gut das Verfahren vertikale Richtungen unbekann-

ϑ	φ					ϑ	φ				
	0°	20°	45°	70°	90°		0°	20°	45°	70°	90°
-48°	1	2	3	4	5	-48°	1	2	3	4	5
-31°	6	7	8	9	10	-31°	6	7	8	9	10
-19°	11	12	13	14	15	-19°	11	12	13	14	15
0°	16	17	18	19	20	0°	16	17	18	19	20
+14°	21	22	23	24	25	+14°	21	22	23	24	25
+27°	26	27	28	29	30	+27°	26	27	28	29	30
+42°	31	32	33	34	35	+42°	31	32	33	34	35

Tabelle 5.7. – Darstellung der Aufteilung zwischen Lern- und Testdaten (letztere wurden dick gedruckt). Variante 1 (links): Aus fünf der sieben Elevationsgruppen (Zeilen) wird je eine Konfiguration desselben Seitenwinkels als Testdaten entnommen. Variante 2 (rechts): Alle fünf Konfigurationen einer Elevation (27 Grad) werden als Testdaten verwendet.

ter Vektoren interpolieren kann, die einem Seitenwinkel angehören, der nur mit zwei Vektoren in den Vektoren vertreten ist. In Variante 2 werden alle fünf Vektoren einer Elevationsgruppe als Testdaten ausgewählt. Es soll geprüft werden, wie gut das Verfahren eine Richtung interpolieren kann, die nicht durch Audiovektoren in den Lerndaten vertreten ist.

Analyse der Gewichtsvektoren

Der Analyse der Gewichtsvektoren geht voraus, dass die Lernphase für die zwei Lerndaten-Varianten ausgeführt wurde. Ergebnis dieser Phase sind 15 Parametervektoren \mathbf{b} und zugehörige Gewichtsvektoren \mathbf{w} , welche pro Anzahl an Bändern als Gewinner der mehrfach wiederholten Evolution hervorgehen.

Anhand der Gewichtung der Audiovektoren lassen sich weitere Eigenschaften des Verfahrens zur Bestimmung der vertikalen Richtung analysieren und diskutieren. Es können zusätzlich Aussagen über den Einfluss der Anzahl der Bänder getroffen werden und der in Kapitel 3.2 entwickelte Sinussweep hinsichtlich seiner Eignung zur vertikalen Richtungsbestimmung untersucht werden.

Aus den Einträgen des Gewichtsvektors geht die Relevanz der Bänder für die Richtungsbestimmung hervor. Allgemein ist zu sagen: Je höher der Betrag des Gewichts ist, desto relevanter ist das zugehörige Band für die Richtungsbestimmung. Die Gewichte lassen sich jedoch schlecht zwischen verschiedenen Vektoren vergleichen, da die Anordnung der Bänder sehr verschieden sein kann. Zur besseren Vergleichbarkeit und Verallgemeinerbarkeit sollen sie auf die einzelnen Samples des Eingangsvektors

aufgeschlüsselt werden. Dazu wird der so genannte *erweiterte Gewichtsvektor* eingeführt. Auf Grundlage des Bandparametervektors wird ein 213-dimensionalen Vektor rekonstruiert, dessen Samplewerte das Gewicht des zugehörigen Bandes beinhalten.

Um eine allgemeine Tendenz der Gewichtung des Vektors der unterschiedlichen Anzahl der Bänder vergleichen zu können, werden die erweiterten Gewichtsvektoren der Gewinner der 15 Wiederholungen einer Evolution gemittelt und in Abbildung 5.8 visualisiert. Jedes Diagramm repräsentiert eine andere Anzahl an Bändern. Die Ergebnisse der beiden Varianten werden durch die schwarzen (Variante 1) und blauen Graphen (Variante 2) gegenübergestellt. Im achten Diagramm wird nicht das Resultat einer Evolution, sondern der Gewichtungsvektor dargestellt, welcher berechnet wird, wenn anstelle der reduzierten Vektoren \mathbf{v}_i die 213-dimensionalen Hüllkurvenvektoren¹ verwendet werden. Aus der Abbildung der Gewichtsvektoren lassen sich folgende Tendenzen feststellen:

Bei genauerer Betrachtung der beiden Graphen innerhalb eines Diagramms, fällt auf, dass sich beide sehr ähneln. Dies kann auf die starke Übereinstimmung der Lern-daten zurückgeführt werden, welche 26 gleiche und nur vier unterschiedliche Vektoren aufweisen. Folglich werden ähnliche Gewichte berechnet. In den ersten sieben Diagrammen führt die Mittelwertbildung der Gewichte zusätzlich zur gegenseitigen Angleichung der Kurven.

Nachfolgend werden die schwarzen Graphen der ersten sieben Diagramme verglichen. Es wird deutlich, dass unabhängig von der Anzahl der Bänder ein ähnlicher Gewichtungsverlauf berechnet wird. Die Ähnlichkeit zeigt, dass das Evolutionsverfahren im Mittel ein reproduzierbares Ergebnis liefert, das relativ unabhängig von der Anzahl der Bänder ist. Je mehr Bänder aber zur Verfügung stehen, desto genauer kann der Gewichtsvektor angepasst werden und desto mehr Details enthält der gemittelte Vektor. Die gemittelten Gewichtsvektoren weisen einen ähnlichen Verlauf zur Gewichtung des Originalvektors im letzten Diagramm auf.

Eine Gemeinsamkeit aller Kurven sind die besondere Dynamik und die vergleichsweise hohen Gewichte in den ersten 50 Samples. Allgemein kann daraus geschlossen werden, dass diese Samples eine höhere Relevanz für die Berechnung der vertikalen Richtung aufweisen als die späteren. Nachfolgend wird auf mögliche Gründe und Folgen eingegangen. Analog zu den Gewichtsvektoren werden die Audiovektoren

¹Im Weiteren wird der 213-dimensionalen Hüllkurvenvektoren auch als Originalhüllkurvenvektor beziehungsweise Originalvektor bezeichnet.

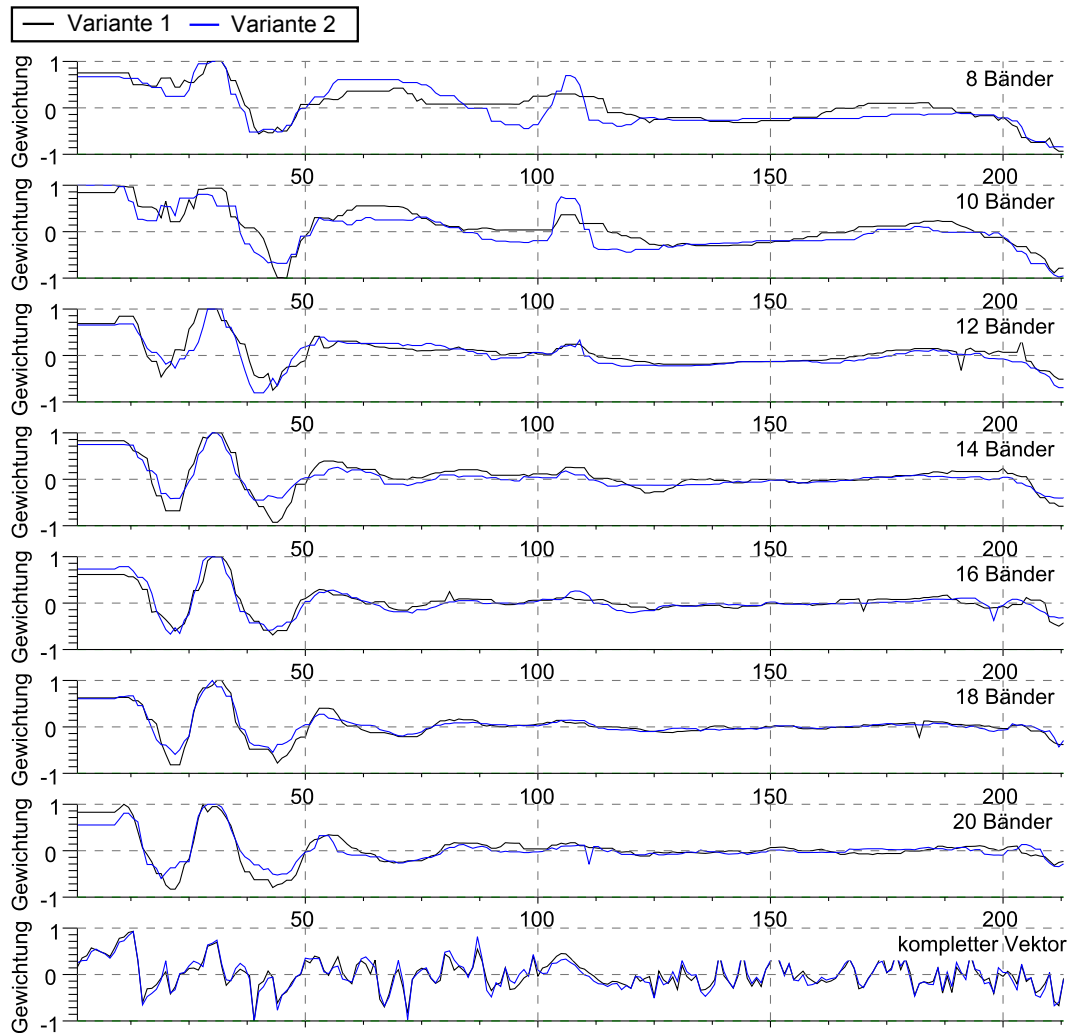


Abbildung 5.8. – Übersicht der mittleren Gewichtung der Hüllkurve beider Varianten. Gemittelt wurde über die besten Ergebnisse von je 15 Evolutionen (Diagramm 1–7). Zum Vergleich (Diagramm 8): Der Gewichtsvektor nach Anwendung der PCA auf die unkomprimierte Hüllkurve.

betrachtet, welche die Grundlage der Gewichtsbestimmung darstellen. Es ist erkennbar, dass in über 80 Prozent der Fälle die maximale Amplitude des Vektors bei circa 40 Samples liegt (siehe Abbildung A.4 im Anhang). Wie bereits in der Voruntersuchung festgestellt wurde, kann dies auf die richtungsunabhängige Resonanzfrequenz des Ohrkanals zurückgeführt werden. Diese erstreckt sich im Frequenzband von 3,5 bis 4,5 kHz und im Audiovektor circa von 30 bis 50 Samples. Da diese Resonanz richtungsunabhängig auftritt, wird das Signal an dieser Stelle in fast allen Vektoren verstärkt. Folglich werden auch die bereits auf den Direktschall durch den Kopf und die Pinnae aufmodulierte Richtungsinformationen in diesem Frequenzbereich verstärkt. Tatsache ist, dass die Anhebung auch zu einer Hervorhebung der Richtungsinformationen führt, die in diesen Frequenzen enthalten sind. Möglicherweise befinden sich in diesem Frequenzbereich aber nur wenige bis keine nützlichen Richtungsinformationen, sodass anstelle dessen Störungen und Rauschen verstärkt werden. Wäre dieser Bereich jedoch unbedeutend für die Richtungsbestimmung, dann hätte dies im Evolutionsprozess durch die Entwicklung eines breiten Bandes an dieser Stelle gedämpft und somit neutralisiert werden müssen. Dies ist jedoch nicht geschehen. Stattdessen scheint die Bedeutung im Vergleich zum Gewichtsvektor des Originalvektors eher noch gestiegen zu sein. Die Verstärkung dieses frühen Abschnitts des Direktschalls hat noch einen weiteren positiven Effekt. Dazu wird das zeitliche Auftreten der ersten Reflexion des Schalls betrachtet, die im Allgemeinen störend für die Interpretation der Hüllkurve als Abbild des Frequenzspektrums ist. Grundsätzlich nimmt mit wachsender Dauer des Direktschalls die Wahrscheinlichkeit zu, dass dieser durch Reflexionen überlagert wird. In Tabelle 5.9 werden berechnete Verzögerungszeiten der ersten Reflexion in Samples für verschiedene Höhen der Schallquelle zwischen 0,2 und 1,2 m und Entfernungen zwischen 0,5 und 4,0 m dargestellt. Aus der Tabelle geht hervor, dass beispielsweise bei einem Abstand von 1 m, welcher im Testaufbau verwendet wurde, Reflexionen innerhalb der ersten 50 Samples selten auftreten. Bei einer Schallquellhöhe unter 0,25 m oder bei größerem Abstand treten die Verzögerungen früher auf. Extrem niedrige oder hohe Höhen der Schallquelle gehören jedoch nicht zum Normalfall des Anwendungsszenarios. Das erste Viertel des Direktschallvektors wird potentiell am wenigsten durch Reflexionen beeinträchtigt. Es ist somit vorteilhaft, dass die ersten 50 Samples des Signals durch die Resonanzfrequenz des Ohres verstärkt werden, da sie die am wenigsten gestörten Richtungsinformationen enthalten.

h [m]	a [m]			
	0,5	1,0	2,0	4,0
0,2	51	42	28	16
0,4	101	84	56	32
0,6	150	123	83	47
0,8	196	160	109	62
1,0	237	193	133	77
1,2	269	221	155	92

Tabelle 5.9. – Auflistung der Verzögerungszeiten der ersten Reflexion in Samples, wenn sich das binaurale System in einer Höhe von 1,20 m befindet und der Boden die erste Reflexionsfläche darstellt. Berechnet werden diese für verschiedene Höhe h der Schallquelle (Zeilen) und verschiedenen Abständen a zum Empfänger (Spalten).

In Abbildung 5.10 sind die endgültigen „erweiterten“ Gewichtsvektoren je Bänderanzahl dargestellt, die in allen Evolutionen die beste Fitnessbewertung erhalten haben. Hier bestätigt sich die anfänglich starke und mit der zunehmenden Dauer des Direktschalls fallende Bedeutung der Bänder für die Richtungsbestimmung. Eine Auflistung der zugehörigen Parameter- und Gewichtsvektoren befindet sich im Anhang A.5, welche die Grundlage für die Berechnungen im nächsten Abschnitt bilden.

5.3.3. Analyse der Richtungsgenauigkeit

In Abbildung 5.11 werden die Resultate der Lern- und Testphase für beide Varianten dargestellt. Auf der Abszisse ist die Anzahl der verwendeten Bänder abgetragen, auf der Ordinatenachse die Abweichung vom Sollwert in Grad. Die verschiedenen Graphen stellen die durchschnittliche (schwarz) und maximale (blau) Winkelabweichung der Lerndaten sowie die durchschnittliche (grün) und maximale (rot) Abweichung der Testdaten dar. Für die durchschnittliche Abweichung der Testdaten wurde zusätzlich die Standardabweichung eingezeichnet (schwarz). Zur Berechnung werden die evolvierten Parameter- und Gewichtsvektoren verwendet (siehe Tabelle A.5 im Anhang). Der letzte Wert jedes Graphen (Bänderanzahl 213) stellt zum Vergleich die Winkelgenauigkeit der Richtungsbestimmung auf Grundlage der Originalvektoren dar. In einem separaten Diagramm werden die mittleren Winkelabweichungen zusätzlicher Varianten der Einteilung in Lern- und Testdaten dargestellt (siehe Abbildung 5.12). Diese sollen die Ergebnisse der ersten beiden Varianten bestätigen

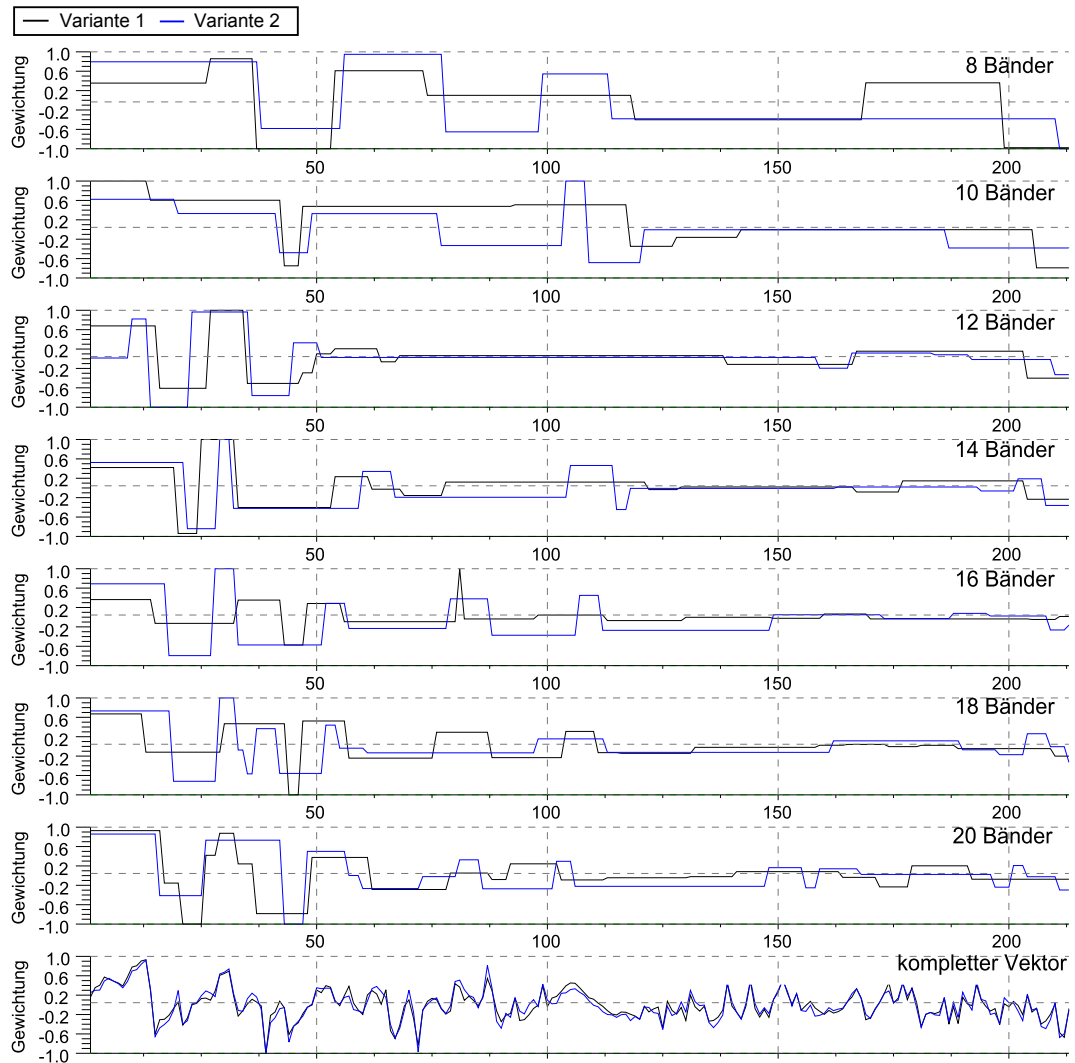


Abbildung 5.10. – Übersicht der optimalen „erweiterten“ Gewichtsvektoren beider Varianten. Zum Vergleich (Diagramm 8): Die Gewichte der ungefilterten Hüllkurve als Eingangsvektor der PCA.

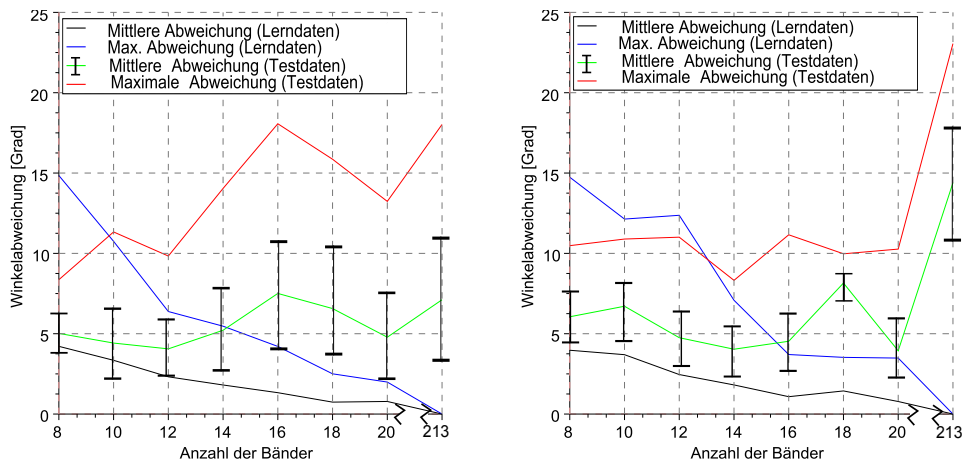


Abbildung 5.11. – Vergleich spezieller Winkelabweichungen der Evolution mit einer fixen Anzahl an Bändern von Variante 1 (links) und Variante 2 (rechts). Die verschiedenen Graphen stellen die durchschnittliche (schwarz) und maximale (blau) Winkelabweichung der Lerndaten sowie die durchschnittliche (grün) und maximale (rot) Abweichung der Testdaten dar. Zum Vergleich werden durch den letzten Wert jedes Vektors die Abweichungen für den Originalvektor visualisiert.

und weitere Schlussfolgerungen erlauben.

Nachfolgend werden die ersten beiden Diagramme analysiert und erste Aussagen getroffen: In einem ersten Schritt wird die Winkelabweichung der Lerndaten betrachtet. Es bestätigt sich die Behauptung, welche im Abschnitt 5.3 aufgestellt wurde, dass die Genauigkeit mit wachsender Anzahl an Bändern zunimmt. Je mehr Dimensionen im Eingangsvektor enthalten sind, desto besser findet die Anpassung auf die Lerndaten statt. Die durchschnittliche Winkelabweichung sinkt in beiden Varianten von 4 Grad (bei 8 Bändern) auf unter 1 Grad (bei 20 Bändern), analog fällt die maximale Abweichung der Lerndaten von 15 auf unter 2 beziehungsweise 3 Grad. Im Fall des Originalvektors geht die Abweichung sogar gegen 0 Grad, weil mehr Dimensionen zur Verfügung stehen als Richtungen berechnet werden müssen. Dieses überdimensionierte lineare Gleichungssystem erlaubt eine optimale Anpassung an die abzubildenden Richtungen der Lerndaten. Dies kann sich auch nachteilig auf die Testdaten auswirken, wie weiter unten gezeigt wird.

Wichtiger für eine spätere Verallgemeinerung der Richtungsgenauigkeit sind die Resultate der Testdaten. In beiden Varianten werden je fünf Testvektoren ausgewählt, die zwei Extremfälle darstellen. Im ersten Fall stammen die Vektoren aus verschiedenen Elevationen mit gleichem Seitenwinkel. Im zweiten Fall ist es umge-

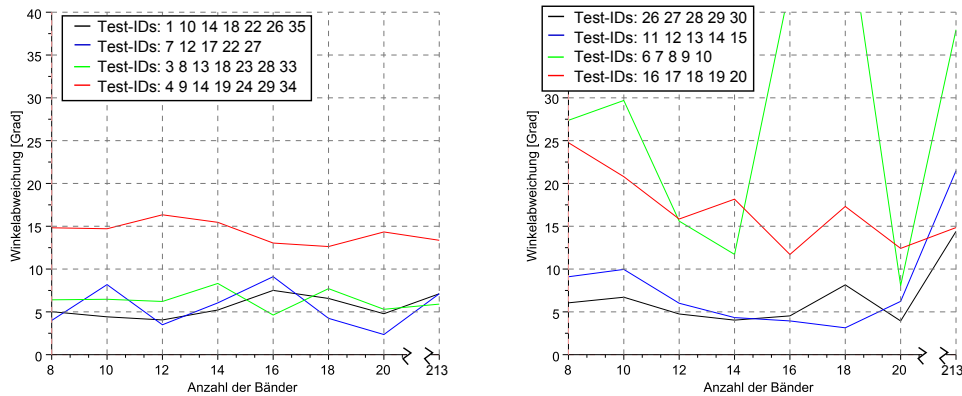


Abbildung 5.12. – Darstellung der mittleren Abweichung verschiedener Varianten der Einteilung des Datensatzes in Lern- und Testdaten. Zum Vergleich wurden mit dem letzten Wert jedes Vektors die Abweichungen für den Originalvektor (Bänderanzahl = 213) visualisiert.

kehrt. Entsprechend fehlen diese Vektoren in den Lerndaten, sodass Lücken in der Winkelauflösung der Lerndaten entstehen.

Folgende Aussagen können aus den Diagrammen abgeleitet werden:

1. Es wurde ein einfaches Verfahren entwickelt, welches die Richtungsbestimmung von unbekanntem Audiovektoren erlaubt. Für die zwei Varianten der Dateneinteilung gibt es je eine Lösung, die eine Genauigkeit von durchschnittlich weniger als 5,0 Grad aufweist. In Variante 1 beträgt die geringste mittlere Abweichung bei 12 Bändern 4,1 Grad (maximale Abweichung 9,8 Grad, Standardabweichung 3,7 Grad). Bei Variante 2 sind es bei 14 Bändern durchschnittlich 4,0 Grad (maximale Abweichung 8,3 Grad, Standardabweichung 3,2 Grad).

2. Es kann aus den beiden Resultaten keine allgemeine Aussage darüber abgeleitet werden, welche Anzahl der Bänder optimal für die Richtungsbestimmung unbekannter Vektoren ist. Die Einbeziehung der Ergebnisse in Abbildung 5.12 zeigt, dass bei unterschiedlichen Varianten der Einteilung in Lern- und Testdaten der durchschnittliche Fehler nicht bei denselben Bänderzahlen minimal wird. Um hier eine Tendenz erkennen zu können, müssen zusätzliche Testvektoren verwendet werden. Diese stehen jedoch in dieser Untersuchung nicht zur Verfügung, da der Umfang des Lerndatensatzes nicht weiter eingeschränkt werden soll.

3. In Abbildung 5.12 wird die mittlere Abweichung zusätzlicher Varianten dargestellt. Diese werden analog zu Variante 1 und 2 (ebenfalls im Diagramm abgetragen) in Lern- und Testdaten untergliedert. Bei der Betrachtung des Verlaufs der mittleren

Abweichung im linken Diagramm scheint es kaum einen Unterschied zu geben, ob eine Datenreduktion in Bänder erfolgt ist, oder nicht. Im rechten Diagramm variiert die Kurve in Abhängigkeit der gewählten Anzahl an Bändern dagegen deutlich. Im Falle des Originalvektors (Anzahl der Bänder = 213) findet keine Datenreduktion der Hüllkurven vor der Bestimmung der Gewichte statt. Es existiert in beiden Diagrammen mindestens eine Bänderanzahl pro Variante, die für die jeweiligen Testvektoren im Mittel eine genauere Lösung darstellt als die Gewichte auf Basis der Originalvektoren. Die Ergebnisse insbesondere im rechten Diagramm sind daher ein Hinweis darauf, dass durch die Datenreduktion eine Verbesserung der Richtungsgenauigkeit hervorgerufen wird.

4. Der Vergleich der Kurven der mittleren Abweichung lässt eine Aussage über die Auswahl der Testdaten zu. Einige Varianten weisen einen höheren mittleren Fehler auf als andere. Beispiele für eine geringe resultierende Positionsgenauigkeit sind die Varianten der roten Graphen in beiden Diagrammen sowie der grüne im rechten Diagramm von Abbildung 5.12. Dies lässt darauf schließen, dass die Messanordnungen der Lerndaten in diesen Fällen ungünstig gewählt wurden, um die dazwischen liegenden vertikalen Richtungen der Testdaten hinreichend gut interpolieren zu können. Im rechten Diagramm wird deutlich, dass dies durch die Dimensionsreduktion der Eingangsdaten teilweise kompensiert werden kann, sodass in allen Varianten eine Lösung mit einer mittleren Genauigkeit von unter 15 Grad existiert.

5. Die Genauigkeit der berechneten Elevation hat im 2D-Modell der Azimut-Bestimmung einen Einfluss auf die Genauigkeit des Azimut-Winkels. Darauf wurde bereits im Kapitel 5.2 hingewiesen. Der mittlere Fehler liegt in beiden Varianten der vertikalen Richtungsberechnung unter 5,0 Grad. Folglich liegt der horizontale Fehler unter 3,6 Grad (vorletzte Spalte in Abbildung A.3).

5.3.4. Auswertung

Im ersten Teil dieses Unterkapitels wurden relevante Abschnitte der Implementierung vorgestellt, erläutert und ihre korrekte Funktionalität nachgewiesen. Im dritten Teil fand die Analyse der Richtungsgenauigkeit auf Grundlage zuvor evaluierter Parameter- und Gewichtsvektoren statt. Grundsätzlich hat sich gezeigt, dass das Verfahren für verschiedene Varianten der Testdaten eine mittlere Genauigkeit aufweist, welche im Rahmen der vertikalen Schallquellen-Lokalisation hinreichend ist, wenn sie durch weitere Testdaten bestätigt werden kann. In Kapitel 3.1 *Experimen-*

talanfordering wurde festgelegt, dass eine grobe Klassifizierung der vertikalen Richtung zur Ausrichtung des Kopfes in Folge der Lokalisation ausreicht. Dies wird im Allgemeinen von allen Varianten erfüllt, da mindestens eine Bandparametrisierung existiert, die eine mittlere Winkelgenauigkeit unter 15 Grad ergab. Die maximale Abweichung kann jedoch deutlich größer sein, sodass in Einzelfällen eine Zuordnung zu einer falschen Klasse möglich ist.

Die Verwendung der vertikalen Richtung im 2D-Modell des Roboterkopfes zur Bestimmung des Azimuts ist bei einer großen Abweichung nachteilig. Die Genauigkeit der Elevation sollte im Mittel zwischen 5 bis 10 Grad liegen, da sonst die Beeinträchtigung der horizontalen Richtung unter Umständen größer ist, als die resultierende Korrektur auf Grundlage der Elevation. Durch die Erhöhung der Anzahl der Messanordnungen und der resultierenden Verbesserung der Winkelauflösung der Lerndaten, kann die benötigte Genauigkeit bei unbekanntem Vektoren erreicht werden. Um dies zu bestätigen müssen derartige Untersuchungen auf einer größeren Datenbasis durchgeführt werden.

6. Zusammenfassung und Ausblick

6.1. Zusammenfassung

Das Ziel dieser Arbeit war es, die Schallquellen-Lokalisation für humanoide Roboter unter Verwendung eines binauralen Audiosystems zu realisieren. Nachfolgend werden die Resultate der Entwicklung und Untersuchung zusammenfassend dargestellt:

Im ersten Schritt wurde ein Prototyp eines humanoiden Roboterkopfes hinsichtlich seiner Eignung zur Schallquellen-Lokalisation evaluiert und mit der Elektronik zur Stereosignalverarbeitung erweitert. In Audioexperimenten wurde nachgewiesen, dass es mit einer einfachen Kopfform sowie den nach menschlichem Vorbild geformten Silikonohren möglich ist, von der vertikalen Richtung abhängige Frequenzunterschiede auf den einfallenden Schall zu modulieren. Die elektronische Hardware des binauralen Systems – bestehend aus Miniaturmikrofonkapseln, einem Modul zur A/D-Wandlung und einem Modul zur Verarbeitung – stellte eine geeignete Grundlage zur störungsarmen Digitalisierung des Audiosignals dar. Sie erlaubt die Ausführung der Schallquellen-Lokalisation mit einfachen Algorithmen. Aufgrund der elektronischen Beschränkung des vorliegenden eingebetteten Systems hinsichtlich des Speichers und der Rechenleistung, war die Entwicklung eines im Zeitbereich ausführbaren Verfahrens zur zweidimensionalen Schallquellen-Lokalisation nötig. Zur Verringerung der Komplexität der Lokalisationsaufgabe wurden die Anforderungen des Anwendungsszenarios analysiert und einige vereinfachende Annahmen für den Experimentalaufbau abgeleitet. Zu berücksichtigen waren: Schallreflexionen im nicht-hallfreien Raum, die gezielte Nutzung der vom Roboter erzeugten Schallsignale zur Lokalisation und eine ungestörte Übertragung des zu ortenden Schalls mit ausreichendem Signal-Rausch-Abstand. Diese Annahmen führten auch zur Entwicklung eines angepassten Audiosignals, mit dessen Hilfe die Schallquellen-Lokalisation im Zeitbereich unter den genannten Beschränkungen möglich war. Der Sinussweep erfüllte die Anforderungen der Hardware und erlaubte aufgrund seiner deterministischen Ausgangsstruktur die Analyse und Berechnung der Richtung in beiden Raumdimensionen.

Auf der Grundlage des entwickelten Experimentalaufbaus und des speziellen Audiosignals wurde ein Audiodatensatz für die weitere Untersuchung aufgenommen. Das Verfahren zur Schallquellen-Lokalisation gliedert sich in drei Bestandteilen: Den ersten Teil bildet der Onset-Detektor, welcher den Beginn des Direktschalls in einem Audiovektor auf Basis des mathematischen Modells einer Sinusschwingung bestimmt. Die Implementierung wurde mit Modellparametern auf den Testdatensatz angepasst, sodass eine robuste und hinreichend genaue Detektion der Onsets möglich war. Durch den relativen Vergleich der ITDs verschiedener Messungen konnte in 94 Prozent der Testfälle eine Genauigkeit von einem Sample pro Stereopaar festgestellt werden, was der technisch maximal realisierbaren Genauigkeit entspricht. In Ausnahmefällen wurde eine Genauigkeit von 2 Samples erreicht, was auf zusätzliches Signalrauschen zurückzuführen war. Somit war die hinreichend genaue Bestimmung der Laufzeitdifferenz des Direktschalls zwischen beiden Ohren möglich. Anschließend wurde geschlussfolgert, dass diese Genauigkeit auch bei Schallvektoren anderer Richtung erreicht werden kann, da der Testdatensatz repräsentativ gewählt wurde.

Den zweiten Bestandteil stellt das zweidimensionale Modell des Roboterkopfes zur Bestimmung des Azimuts dar. Es setzt die Bestimmung der Laufzeitdifferenz und der Elevation voraus. Das Verfahren bestand aus zwei Teilen: Zum einen basiert es auf dem erweiterten Sinussatz zur groben Bestimmung des Azimut auf Basis der ITD. Zum anderen wurde eine Approximationsfunktion zur Korrektur des Seitenwinkels eingeführt, welche die Elevation der Schallquelle berücksichtigt. Die Bestimmung der Modellparameter geschah empirisch mit Hilfe des Audiodatensatzes. In der Untersuchung wurde im Falle einer fehlerfreien Bestimmung der vertikalen Richtung eine Genauigkeit des Seitenwinkels von durchschnittlich 3,0 Grad (Maximalabweichung: 8,4 Grad) erreicht. In Abhängigkeit des Fehlers der Elevation verringert sich die Genauigkeit allerdings. Der Einfluss des Elevationsfehlers nimmt linear zur Laufzeitdifferenz und quadratisch zum Betrag der Elevation zu. Für zwei theoretische Fälle wurde eine statistische Berechnung des Fehlers durchgeführt, die darauf hinweist, dass die Beeinträchtigung des Seitenwinkels durch die Abweichung der Elevation im Mittel nur wenige Grad beträgt (siehe Tabelle A.3). Im Extremfall kann jedoch eine starke Abweichung der Elevation zu einer starken Verfälschung des Azimuts führen. Bedingt durch die Abhängigkeit des Seitenwinkels von der Elevation und der Laufzeitdifferenz ist die Genauigkeit im Frontalbereich höher als im Seitenbereich. Die Winkelauflösung eines Samples ist hier doppelt so hoch (1,0 Grad) wie

im Seitenbereich (2,1 Grad) (siehe Tabelle A.3). Damit besteht eine Analogie zur menschlichen Fähigkeit der Schallquellen-Ortung, welche ebenfalls im Frontalbereich am genauesten ist und zu den Seiten und nach oben und unten hin abnimmt [Beg94].

Der dritte Bestandteil des Verfahrens zur Schallquellen-Lokalisation ist die Bestimmung der Elevation. Das entwickelte Verfahren setzt eine einmalige Lernphase zur Offline-Kalibrierung der Parametervektoren (Band- und Gewichtsparameter) voraus. Über die Genauigkeit des Verfahrens wurde folgendes festgestellt: Es konnte eine Abhängigkeit der Genauigkeit von der Wahl des Lerndatensatzes festgestellt werden. Bei Verwendung der gleichen Anzahl an Lernvektoren, variierte die mittlere Richtungsgenauigkeit der Testvektoren für unterschiedliche Varianten der Einteilung in Lern- und Testdaten (siehe Abbildung 5.12). Daraus wurde geschlussfolgert, dass eine angemessene Anzahl repräsentativer Schallquellenpositionen als Grundlage für die Lerndaten benötigt werden, um die dazwischenliegenden Elevationen unbekannter Audiovektoren mit hinreichender Genauigkeit interpolieren zu können. Die Untersuchung zeigte auch, dass bereits mit 28 beziehungsweise 30 Schallquellenanordnungen der Lerndaten gute Ergebnisse in der Richtungsberechnung von unbekanntem Testvektoren erzielt werden können. Die (geringste) mittlere Genauigkeit lag bei den exemplarischen Varianten 1 und 2 unter 5 Grad und die maximale Abweichung bei circa 10 Grad. Die Methode der Dimensionsreduktion der Eingangsvektoren durch das Bänder-Konzept erhöhte die Robustheit und Genauigkeit der Richtungsberechnung in einigen Fällen. Es war allerdings keine abschließende Verallgemeinerung auf eine bestimmte Anzahl an Dimensionen (Bändern) und die Richtungsgenauigkeit im realen Einsatz möglich, da eine zu geringe Zahl von Vektoren (5 beziehungsweise 7 Testvektoren) für die Testphase zur Verfügung stand.

Die Realisierung der Schallquellen-Lokalisation mit dem gewählten binauralen System, dem dreistufigen Verfahren auf Basis eines speziellen Signals wurde erfolgreich durchgeführt. Das entwickelte Verfahren lässt sich mit einfachen Algorithmen im Zeitbereich, also ohne Transformation des Audiovektors in den Frequenzbereich, realisieren, sodass die Richtungsbestimmung auf einem eingebetteten System mit begrenzten Ressourcen eines Roboters ausführbar ist. Das Verfahren setzt die Adaption der Modellparameter auf das jeweilige binaurale System voraus, was auf Grundlage eines Testdatensatzes empirisch geschehen muss.

Die möglichen Anwendungsgebiete der realisierten Schallquellen-Lokalisation sind vielfältig. Denkbare Anwendungen sind:

- In Kombination mit einer autonomen Steuerung ist ein Roboter in der Lage, seinen Körper auf eine Schallquelle, in diesem Fall einen anderen Roboter, horizontal und den Kopf zusätzlich vertikal auszurichten (Phonotaxis), um einen anderen Agenten zu fokussieren. Die Ausrichtung auf eine Schallquelle könnte auch auf Menschen aufgeweitet werden, was eine denkbare Anwendung für Serviceroboter ist.
- Ein großes Anwendungsgebiet ergibt sich aus der Sensordatenfusion insbesondere in Kombination mit visuellen Daten. So kann ein internes Weltbild erstellt werden, bei dem nicht nur Objekte im Raum, sondern auch ihre Schalleigenschaften abgespeichert werden. Es können somit Objekte außerhalb des Blickfeldes akustisch geortet und ihre Position im internen Weltbild aktualisiert werden. Eine praktische Anwendung könnte zum Beispiel im Roboterfußball liegen, um die dynamische Position aktiver Mitspieler und gegebenenfalls auch der Gegner trotz fehlendes Sichtkontaktes zu aktualisieren.
- Visuelle und akustische Sensordaten können, sofern sie sich auf dasselbe Objekte im Sichtfeld beziehen, zur gegenseitigen Korrektur der abgeleiteten Richtungsinformation verwendet werden. Die Parameter des Lokalisationsverfahrens könnten durch adaptive Lernmechanismen korrigiert werden, wenn eindeutige Fehler in der Schallquellen-Ortung erkannt wurden. Somit wäre auch die Anpassung an neue Geräusche oder Veränderungen der Umgebung möglich.

6.2. Ausblick

Die Untersuchung des vorliegenden Verfahren lieferte für den Testdatensatz hinreichend gute Ergebnisse für die zweidimensionale Schallquellen-Lokalisation. Die Evaluierung am echten Roboter steht noch aus. Es wurden zur Reduzierung der Komplexität einige Vereinfachungen des Experimentalaufbaus durchgeführt. Daraus ergeben sich mögliche Erweiterungen, die in weiterführenden Arbeiten untersucht werden können: Eine Voraussetzung ist die Erweiterung des Experimentalaufbaus und des Testdatensatzes. Um den Azimut rundum bestimmen zu können, ist der Testdatensatz um Messungen zwischen 90 und 180 Grad zu erweitern. Die in Kapitel 4.2.3 *2D-Modell des Roboterkopfes* erwähnte Front-Back-Ambivalenz bei der Bestimmung des Seitenwinkels muss nun in dem Verfahren berücksichtigt werden.

Dies ist möglich, wenn das zweidimensionale Modell des Roboterkopfes neben der Laufzeitdifferenz und der Elevation um einen dritten (binären) Eingangsparameter ergänzt wird. Der zusätzliche Parameter wird durch die Frequenzanalyse des Schalls berechnet [Bla74], sodass das Verfahren zur Berechnung der Elevation angewendet werden kann. Die Bestimmung des Gewichtsvektors erfolgt analog. Allerdings findet die Abbildung nicht auf die Elevation, sondern auf ein zweiwertiges Skalar (zur Unterscheidung in vorne und hinten) statt. Mit dieser zusätzlichen Information kann der Seitenwinkel eindeutig zugeordnet werden.

Ein weites Spektrum für zukünftige Arbeiten liegt in der ausführlicheren Untersuchung und Verbesserung des Verfahrens zur Bestimmung der Elevation: Für eine allgemeine Aussage zur Genauigkeit sind umfassendere Messdaten nötig. Es ist zu untersuchen wie dicht das Netz der Schallquellenpositionen der Lerndaten gewählt werden muss, um eine gleichbleibend hohe Positionsgenauigkeit für beliebige dazwischen liegende Test-Positionen zu erreichen. Zudem steht eine abschließende Aussage über eine ideale Anzahl von Dimensionen der Datenreduktion für die vertikale Richtungsbestimmung aus.

Eine andere offene Frage ist, ob die Schallquellen-Lokalisierung auf verschiedene Entfernungen abstrahiert werden kann. Insbesondere auf die vertikale Richtungsbestimmung kann die Veränderung der Entfernung einen Einfluss haben, da sich die Reflexionszeiten verändern, was trotz gleicher Richtung auf Grund der Superposition mit den Reflexionen zu einer Veränderung des Signalvektors führen kann. Ein weiterer Aspekt für eine Untersuchung stellt die Orientierung der Schallquelle dar. Diese war in den Messungen stets horizontal auf das messende System ausgerichtet. Im Allgemeinen kann der Roboterkopf und damit der Lautsprecher in eine beliebige Richtung im Raum orientiert sein. Es gilt also auch zu untersuchen, welche Veränderungen sich für die Schallquellen Lokalisation ergeben, wenn die Schallquelle beispielsweise eine vollständig abgewandte Blickrichtung aufweist oder der Direkt-schall durch Hindernisse gedämpft wird.

Die Einsatzmöglichkeiten des vorliegenden Verfahrens außerhalb des Anwendungsszenarios oder ähnlichen Bedingungen sind begrenzt, da es auf einem speziell entwickelten Schallsignal für die Analyse im Zeitbereich aufbaut. Es wird beispielsweise nicht möglich sein, eine robuste Lokalisation unter zusätzlichem Störeinfluss durchzuführen oder die Richtungsinformationen zweier überlagerter Schallsignale zu unterscheiden.

Robotern stehen allgemein aufgrund der benötigten Autonomie und Mobilität nur begrenzte Ressourcen zur Speicherung und Berechnung zur Verfügung. Daher gibt es einen Bedarf an einfachen und effizient umsetzbaren Modellen und Verfahren zur Schallquellen-Lokalisation. Diese Tatsache sollte ein Anreiz für die Wissenschaft sein, um bekannte Modelle unter den genannten Gesichtspunkten neu zu bewerten und weitere innovative Methoden zu entwickeln, die auf beschränkten Systemen ausführbar sind. Diese Arbeit liefert dazu einen Beitrag, da gezeigt werden konnte, dass für spezielle Probleme der Schallquellen-Lokalisation in der Robotik einfache Verfahren genügen, um eine hinreichende Richtungsgenauigkeit zu erzielen.

i [Sampl.]	Stereomodell		Roboterkopf	
	φ [°]	$\Delta\varphi_i$ [°]	φ [°]	$\Delta\varphi_i$ [°]
0	0	1,2	0,0	1,0
1	2,4	1,2	2,0	1,0
2	4,8	1,2	4,1	1,0
3	7,2	1,2	6,1	1,0
4	9,6	1,2	8,2	1,0
5	12,1	1,2	10,2	1,0
6	14,5	1,2	12,3	1,0
7	17,0	1,3	14,3	1,0
8	19,5	1,3	16,4	1,0
9	22,1	1,3	18,5	1,0
10	24,7	1,3	20,6	1,1
11	27,4	1,4	22,7	1,1
12	30,1	1,4	24,8	1,1
13	32,9	1,5	26,9	1,1
14	35,8	1,5	29,1	1,1
15	38,8	1,6	31,2	1,1
16	42,0	1,7	33,4	1,1
17	45,3	1,8	35,7	1,1
18	48,8	1,9	37,9	1,1
19	52,6	2,1	40,2	1,2
20	56,7	2,4	42,5	1,2
21	61,4	2,8	44,8	1,2
22	66,9	3,1	47,2	1,2
23	74,1	8,0	49,6	1,2
24	90,0	8,0	52,0	1,2
25			54,5	1,3
26			57,1	1,3
27			59,7	1,3
28			62,4	1,4
29			65,2	1,4
30			68,0	1,5
31			71,0	1,5
32			74,0	1,6
33			77,2	1,7
34			80,6	1,8
35			84,1	1,9
36			87,9	2,1
37			92,0	2,1

Tabelle A.2. – Mathematisch bestimmte Seitenwinkel φ und resultierende maximale Winkelabweichung $\Delta\varphi_i$ in Abhängigkeit der Laufzeitdifferenz i des minimalen Stereomodells (Formel 4.1) und des Roboterkopfmodells (Formel 4.12) bei gleicher Parametrisierung.

Konf.- ID	φ_{soll} [°]	ϑ_{soll} [°]	i [Samp.]	φ_0 [°]	$\Delta\varphi_0$ [°]	φ_ϑ [°]	$\Delta\varphi_\vartheta$ [°]	$\Delta\varphi_{\vartheta+5^\circ}$ [°]	$\Delta\varphi_{\vartheta+10^\circ}$ [°]
1	0	-48	2,0	-4,1	4,1	-6,7	6,7	6,2	5,7
2	20	-48	-6,0	12,4	7,6	20,2	-0,2	1,4	2,7
3	45	-48	-14,2	29,8	15,2	48,5	-3,5	0,2	3,4
4	70	-48	-20,5	44,1	25,9	71,8	-1,8	3,7	8,5
5	90	-48	-23,4	51,2	38,8	83,2	6,8	13,1	18,8
6	0	-31	0,6	-1,2	1,2	-1,6	1,6	1,5	1,4
7	20	-31	-8,8	18,3	1,7	23,0	-3,0	-1,6	-0,5
8	45	-31	-17,0	36,1	8,9	45,5	-0,5	2,3	4,6
9	70	-31	-25,5	56,5	13,5	71,3	-1,3	3,1	6,7
10	90	-31	-28,6	64,9	25,1	81,8	8,2	13,2	17,3
11	0	-19	1,2	-2,5	2,5	-2,7	2,7	2,6	2,5
12	20	-19	-8,8	18,3	1,7	20,1	-0,1	0,8	1,3
13	45	-19	-19,0	40,6	4,4	44,6	0,4	2,2	3,5
14	70	-19	-28,2	63,8	6,2	70,0	-0,0	2,8	4,8
15	90	-19	-33,4	79,8	10,2	87,6	2,4	6,0	8,5
16	0	0	-1,6	3,3	-3,3	3,3	-3,3	-3,3	-3,4
17	20	0	-8,0	16,6	3,4	16,6	3,4	3,3	3,0
18	45	0	-19,8	42,5	2,5	42,5	2,5	2,2	1,4
19	70	0	-28,2	63,8	6,2	63,8	6,2	5,8	4,5
20	90	0	-36,6	92,0	-2,0	92,0	-2,0	-2,6	-4,5
21	0	14	0,2	-0,4	0,4	-0,4	0,4	0,5	0,5
22	20	14	-8,0	16,6	3,4	17,5	2,5	1,8	0,8
23	45	14	-18,8	40,2	4,8	42,3	2,7	0,9	-1,5
24	70	14	-27,0	60,5	9,5	63,7	6,3	3,6	0,1
25	90	14	-33,0	78,4	11,6	82,6	7,4	3,9	-0,7
26	0	27	-1,0	2,1	-2,1	2,5	-2,5	-2,6	-2,8
27	20	27	-6,2	12,8	7,2	15,4	4,6	3,6	2,4
28	45	27	-16,0	33,8	11,2	40,5	4,5	1,8	-1,4
29	70	27	-25,4	56,3	13,7	67,4	2,6	-1,9	-7,2
30	90	27	-32,0	75,1	14,9	90,0	0,0	-6,0	-13,1
31	0	42	-1,8	3,7	-3,7	5,5	-5,5	-5,9	-6,4
32	20	42	-6,8	14,1	5,9	20,8	-0,8	-2,5	-4,4
33	45	42	-15,4	32,5	12,5	48,1	-3,1	-7,0	-11,4
34	70	42	-21,2	45,8	24,2	67,8	2,2	-3,3	-9,4
35	90	42	-26,4	58,9	31,1	87,1	2,9	-4,2	-12,1
$ \overline{\Delta\varphi_x} $					9,7		3,0	3,6	5,2

Tabelle A.3. – Übersicht der berechneten Seitenwinkel ohne (φ_0 , Formeln 4.12) und mit (φ , Formel 5.2) Berücksichtigung des Höhenwinkels ϑ_{soll} . Zusätzlich wurden die Winkelabweichungen vom tatsächlichen Seitenwinkel φ_{soll} berechnet. Zeile $|\overline{\Delta\varphi_x}|$ bezeichnet den Durchschnitt der Beträge der jeweiligen Spalte.

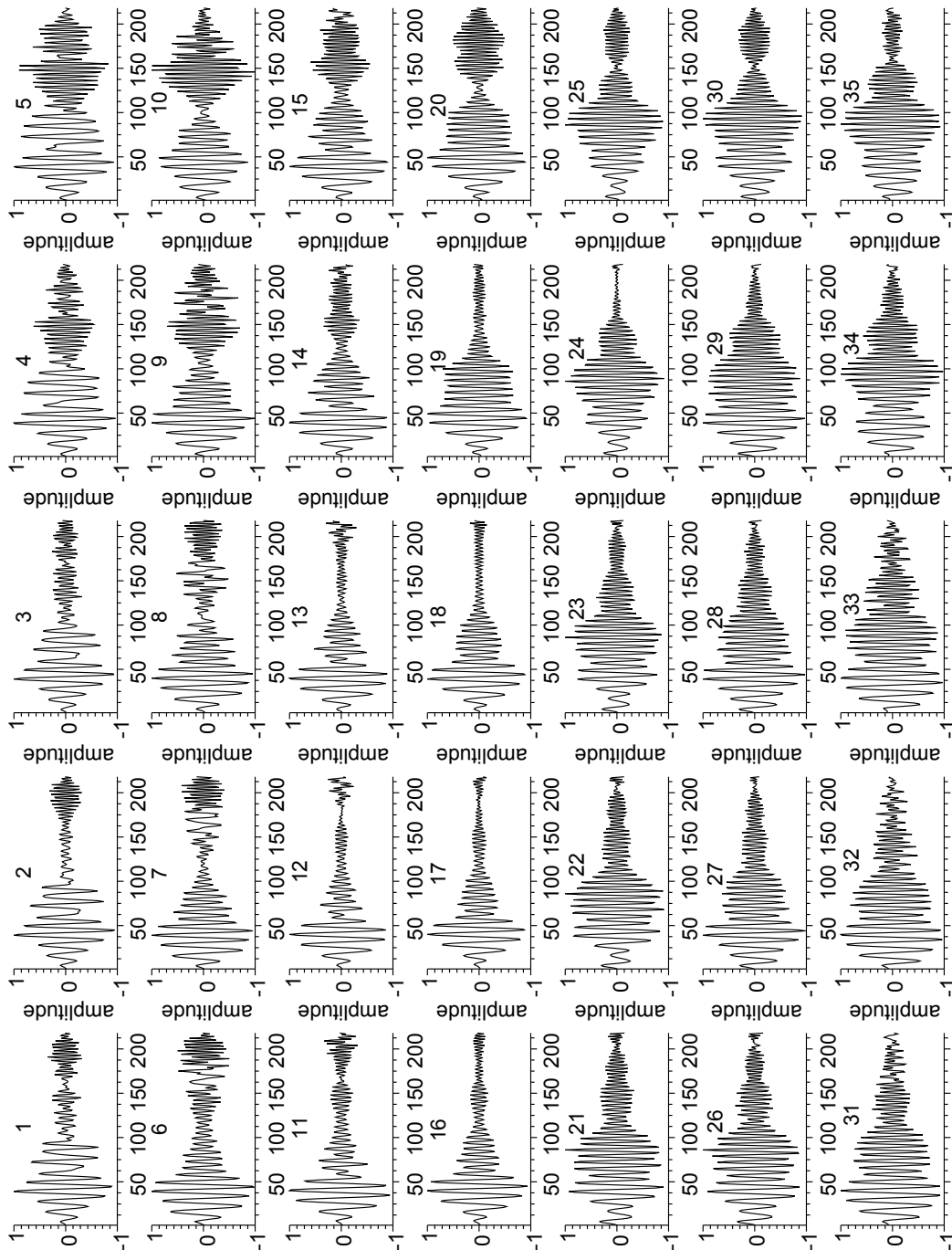


Abbildung A.4. – Abbildung aller 35 Audiovektoren \mathbf{a}_i , welche aus dem Testdatensatz extrahiert wurden und als Basis für die vertikale Richtungsbestimmung dienen. Angeordnet wurden sie in einer Matrix analog zur Tabelle 3.3 auf Seite 32.

Bdr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Variante 1																				
8	26	36	53	73	118	168	198													
10	13	42	46	92	92	117	127	-393	205											
12	15	26	228	181	-1063	193	-131	-61	-1	-299	203	-267								
14	19	24	32	53	61	68	77	121	128	128	166	176	203							
16	14	32	42	47	55	80	81	97	112	129	148	159	169	204	210					
18	12	29	43	46	56	75	87	103	110	115	131	158	164	173	180	188	209			
20	16	20	25	28	32	36	48	61	78	87	91	102	112	130	140	163	171	178	191	
723	-120	-779	326	680	188	-611	292	-223	41	-63	189	-69	-34	-16	64	-28	-181	157	-58	
Variante 2																				
8	37	55	77	98	113	210	214													
10	19	41	48	76	103	108	120	186	213											
12	9	13	22	35	44	50	58	165	183	191	209	-225								
14	21	28	31	59	66	104	114	117	162	193	201	207	214							
16	17	27	32	51	56	78	87	106	111	148	172	187	195	208	212					
18	18	28	32	34	36	41	51	54	60	97	112	161	189	197	203	208	212			
20	15	25	42	47	56	59	72	80	85	101	105	147	155	158	167	196	200	203	210	
624	-300	-727	364	1	-194	-15	237	-197	214	-160	119	-183	102	17	-173	151	-16	-215		

Tabelle A.5. – Zeilenweise werden abwechselnd der Bandparametervektor \mathbf{b} und zugehörige Gewichtsvektor \mathbf{w} aufgelistet, welche zu den Individuen mit der besten Fitness der jeweiligen Anzahl an Bändern (erste Spalte) gehören. Diese Vektoren bilden die Grundlage für die Testphase und werden in Abbildung 5.10 als „erweiterte“ Gewichtsvektoren auf Seite 74 dargestellt.

```

34 // ** Funktion **
35 // Initialisierung der ersten Population einer Evolution
36 // ** Ausgangsparameter **
37 // P0: Vektor der Gene der initialen Generation (IDs der Bänder,
38 //     in welche die Hüllkurve zergliedert wird)
39 function P0 = initialPopulation()
40     rawP0 = zeros(curNoOfFilters-1, noOfIndividuals);
41     for individual = 1: noOfIndividuals
42         for gen = changeGenIDs
43             temp = int((rand(1, 'uniform')*spectralRange)+minSpectralLine-1)
44             while(min(abs(temp -rawP0(:, individual)))) < minBW)
45                 temp = int((rand(1, 'uniform')*spectralRange)+minSpectralLine-1)
46             end
47             rawP0(gen, individual) = temp;
48         end
49     end
50     P0 = gsort(rawP0, 'r', 'i');
51 endfunction
52
53 // ** Funktion **
54 // Selektion der besten Individuen der letzten Generation und
55 // Übernahme in die aktuelle Generation
56 // ** Eingangs-/Ausgangsparameter **
57 // population: Matrix mit den Genen aller Individuen der aktuellen Gener.
58 // noOfBest: Angabe, wie viele Individuen übernommen werden sollen
59 // sortedFitness: sortierter Vektor der Fitness aller Individuen einer Gener.
60 // selectedIndividuals: Gene des ausgewählten Individuums
61 function selectedIndividuals = selection(population, noOfBest, sortedFitness)
62     for individualID = 1:noOfBest
63         selectedIndividuals(:, individualID) =
64             population(:, sortedFitness(individualID, 2));
65     end
66 endfunction
67
68 // ** Funktion **
69 // Generierung eines neuen Individuums durch zufällige Kreuzung der
70 // Gene zweier zufällig gewählter Individuen aus der Liste der besten
71 // Individuen der vorherigen Generation
72 // ** Eingangs-/Ausgangsparameter **
73 // population: Matrix mit den Genen aller Individuen der vorherigen Gener.
74 // noOfBest: Angabe, wie viele Individuen zu den besten Individuen gehören
75 // crossedIndividuals: resultierende Gene des neuen Individuums

```

```

76 function crossedIndividuals = crossing(population, noOfBest)
77     crossedIndividuals = zeros(curNoOfFilters-1, noOfIndividuals-noOfBest)
78     for individualID = 1 : noOfIndividuals - noOfBest
79         ID1 = round(rand(1, 'uniform') * (noOfBest-1)+ 1); //1<=IDs<=noOfBest
80         ID2 = round(rand(1, 'uniform') * (noOfBest-1)+ 1);
81         while (ID1 == ID2)
82             ID2 = round(rand(1, 'uniform') * (noOfBest-1)+ 1);
83         end
84         genCut = round(rand(1, 'uniform') * (curNoOfFilters - 2) + 1);
85         crossedIndividuals(1:genCut, individualID)=population(1:genCut, ID1);
86         crossedIndividuals(genCut + 1 : curNoOfFilters - 1, individualID) =
87             population(genCut + 1 : curNoOfFilters - 1, ID2);
88     end
89     crossedIndividuals = gsort(crossedIndividuals, 'r', 'i');
90 endfunction
91
92 // ** Funktion **
93 // Generierung einer neuen Generation durch beschränkte, aber
94 // zufällige Mutation der Gene der Individuen einer Generation
95 // ** Eingangs-/Ausgangsparameter **
96 // P: Matrix der Gene aller Individuen der aktuellen Generation
97 // mutationFactorIndividual: Mutationswahrscheinlichkeit eines Individuums
98 // mutationFactorgenID: Mutationswahrscheinlichkeit eines Gens
99 // noOfBest: Angabe, wie viele Individuen zu den besten Individuen gehören
100 function P = mutation (P, mutationFactorIndividual, mutationFactorgenID)
101     for individualID = 1:noOfIndividuals
102         if rand(1, 'uniform') <= mutationFactorIndividual
103             for genID = changeGenIDs
104                 if rand(1, 'uniform') <= mutationFactorgenID
105                     temp = 0;
106                     oldV = P(genID, individualID);
107                     P(genID, individualID) = 0
108                     iter = 0;
109                     while((temp <= minSpectralLine) | (temp >= maxSpectralLine)
110                         | (min(abs(temp-P(:, individualID))) < minBW & iter <100) )
111                         temp = oldV + rand(1, 'uniform') * genVariance;
112                         iter = iter +1;
113                     end
114                     P(genID, individualID) = int(temp);
115                 end
116             end
117         P(:, individualID) = gsort(P(:, individualID), 'r', 'i');

```

```

118     end
119 end
120 endfunction
121
122 // ** Funktion **
123 // Bestimmung der Fitness eines Individuums
124 // ** Eingangs-/Ausgangsparameter **
125 // W: zugehöriger Gewichtsvektor eines Individuums
126 // Xin: Matrix aus den extrahierten Frequenzvektoren der Audiovektors
127 // Y: Vektor der zugehörigen (Soll)-Elevationen
128 // fitness: resultierende Fitness des Individuums
129 function fitness = getFitness(W,Xin,Y)
130     error = zeros(length(Y),1);
131     for i = 1 : length(Y)
132         error(i) = abs(Y(i) - W * Xin(:,i));
133     end
134     fitness = mean(error);
135 endfunction
136
137 // ** Funktion **
138 // Auswahl des besten Individuums aus der Menge aller Individuen einer Pop.
139 // ** Eingangs-/Ausgangsparameter **
140 // population: Matrix mit den Genen aller Individuen der aktuellen Gener.
141 // fitness: Vektor der Fitness der Individuen
142 // noOfCurGeneration: ID der aktuellen Generation
143 // bestIndividualGens: Vektor der Gene des ausgewählten Individuums
144 // bestIndividualFitness: Fitness des ausgewählten Individuums
145 // generOfBestIndivid: Generationsid des ausgewählten Individuums
146 function [bestIndividualGens, bestIndividualFitness, generOfBestIndivid] =
147     saveBestIndividual(population, fitness, noOfCurGeneration)
148     [minV, minI] = min(fitness);
149     if (minV < bestIndividualFitness)
150         bestIndividualGens = population(:, minI);
151         bestIndividualFitness = minV;
152         generOfBestIndivid = noOfCurGeneration;
153     end
154 endfunction

```

Listing A.1 – Scilab-Implementierung ausgewählter Funktionen der Evolution.

Literaturverzeichnis

- [AF01] ANGELL, J.R. ; FITE, W.: The monaural localization of sound. In: *Psychological Review* 8 (1901), S. 225–243
- [AHSK04] ANDERSSON, Sean B. ; HANDZEL, Amir A. ; SHAH, Vinay ; KRISHNAPRASAD, P. S.: Robot phonotaxis with dynamic sound-source localization. In: *Proceedings of the 21st IEEE International Conference on Robotics and Automation – IEEE ICRA 2004* Bd. 6. Piscataway, NJ, USA : IEEE, 2004, S. 4833–4838
- [AP80] AXENFELD ; PAU: *Lehrbuch und Atlas der Augenheilkunde*. Stuttgart, Germany : Gustav Fischer Verlag, 1980
- [ARA+06] ASFOUR, T. ; REGENSTEIN, K. ; AZAD, P. ; SCHRÖDER, J. ; BIERBAUM, A. ; VAHRENKAMP, N. ; DILLMANN, R.: ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control. (2006), S. 7
- [BC78] BLAUERT, Jens ; COBBEN, W.: In: *Some consideration of binaural cross correlation analysis* Bd. 39. 1978, S. 96–104
- [BDA+05] BELLO, Juan P. ; DAUDET, Laurent ; ABDALLAH, Samer ; DUXBURY, Chris ; DAVIES, Mike ; SANDLER, Mark B.: A Tutorial on Onset Detection in Music Signals. In: *Speech and Audio Processing, IEEE Transactions on* 13 (2005), S. 1035–1047
- [BDDS04] BELLO, Juan P. ; DUXBURY, Chris ; DAVIES, Mike ; SANDLER, Mark B.: On the use of phase and energy for musical onset detection in the complex domain. In: *Signal Processing Letters, IEEE* 11 (2004), S. 553–556
- [Beg94] BEGAULT, Durand R.: *3D-Sound - for virtual reality and multimedia*. Cambridge, MA, USA : AP Professional, 1994

- [Ben10] BENCKENDORFF, Christian: *Technische Realisierung multimodaler Sensorik für humanoide Roboter*. Berlin, Germany : Humboldt Universität zu Berlin, 2010. – Diplomarbeit
- [Bla69] BLAUERT, Jens: *Untersuchung zum Richtungshören in der Medianebene bei fixiertem Kopf*. Aachen, Germany, Technischen Hochschule Aachen, Fakultät für Elektrotechnik, Diss., 1969
- [Bla74] In: BLAUERT, Jens: *Räumliches Hören*. Stuttgart, Germany : S. Hirzel Verlag Stuttgart, 1974, S. 25ff.
- [Bla85] In: BLAUERT, Jens: *Räumliches Hören: neue Ergebnisse und Trends seit 1972*. 2. Auflage. Stuttgart, Germany : S. Hirzel Verlag Stuttgart, 1985, S. 24ff.
- [Bla97] In: BLAUERT, Jens: *Räumliches Hören: neue Ergebnisse und Trends seit 1982*. 2. Auflage. Stuttgart, Germany : S. Hirzel Verlag Stuttgart, 1997
- [BLP+06] BEIRA, R. ; LOPES, M. ; PRAÇA, M. ; SANTOS-VICTOR, J. ; BERNARDINO, A. ; METTA, G. ; BECCHI, F. ; SALTARÉN, R.: Design of the Robot-Cub (iCub) Head. In: *Proceedings of IEEE International Conference on Robotics and Automation, ICRA (2006)*
- [CHM+06] CHENG, G. ; HYON, S.-H. ; MORIMOTO, J. ; UDE, A. ; JACOBSEN, S. C.: Cb: A humanoid research platform for exploring neuroscience. In: *IEEERAS/RSJ International Conference on Humanoid Robots (Humanoids 2006)*. 2006, S. 182–187
- [CKN+08] CHISAKI, Yoshifumi ; KAWANO, Sho ; NAGATA, Kyoko ; MATSUO, Kotaro ; NAKASHIMA, Hidetoshi ; USAGAWA, Tsuyoshi: Azimuthal and elevation localization of two sound sources using interaural phase and level differences. In: *Acoust. Sci & Tech.* 29 (2008), Nr. 2, S. 139–148
- [CNK01] CHENG, Gordon ; NAGAKUBO, Akihiko ; KUNYOSHI, Yasuo: *Continuous humanoid interaction: An integrated perspective – gaining adaptivity, redundancy, flexibility – in one*. Tsukuba, Ibaraki 305-8568, Japan, 2001

- [CNNU03] CHISAKI, Yoshifumi ; NAKANISHI, Takashi ; NAKASHIMA, Hidetoshi ; USAGAWA, Tsuyoshi: Concurrent Speechsignal Separation Based On Frequency Domain Binaural Model. In: *Proceedings of the International Workshop on Acoustic Echo and Noise Control–IWAENC2003*. Kyoto, Japan, 2003, S. 255–258
- [FM04] FALLER, Christof ; MERIMAA, Juha: Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. In: *J. Acoustic Soc. Am.* 116 (2004), S. 3075–3089
- [GK94] GARDNER, Bill ; KEITH, Martin: *HRTF Measurements of a KEMAR Dummy-Head Microphone*. Cambridge, MA, USA: The MIT Press, 1994. – <http://sound.media.mit.edu/KEMAR.html>
- [Haa51] HAAS, Helmut: Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache. In: *Acustica* 1 (1951), S. 49–58
- [Har83] HARTMANN, William M.: Localization of Sound in Rooms. In: *J. Acoustic Soc. Am.* 74 (1983), S. 1380–1391
- [Har97] HARTMANN, William M.: Listening in a room and the precedence effect. In: *Binaural and Spatial Hearing in Real and Virtual Environments* (1997), S. 349–376
- [Hil07] HILD, Manfred: *Neurodynamische Module zur Bewegungssteuerung autonomer mobiler Roboter*. Berlin, Germany, Humboldt-Universität zu Berlin, Institut für Informatik, Diss., 2007
- [ISK03] IWAYA, Yukio ; SUZUKI, Yōiti ; KIMURA, Daisuke: Effects of head movement on front-back error in sound localization. In: *Acoustical Science and Technology* 24 (2003), Nr. 5, S. 322–324
- [Jat99] JATHO, Martin: *Phonotaxis von Laubheuschrecken und Grillen*. 1999. – Frankfurt, Germany: J.W. Goethe Universität, Zoologisches Institut. <http://user.uni-frankfurt.de/~mjatho/phono.html>
- [KKOO08] KIM, Hyun-Don ; KOMATANI, Kazunori ; OGATA, Tetsuya ; OKUNO, Hiroshi G.: Design and Evaluation of Two-Channel-Based Sound Source

- Localization over Entire Azimuth Range for Moving Talkers. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems – IROS 2008*. Nice France : IEEE, 2008, S. 2197–2203
- [LM08] LIU, P.R. ; MENG, M.Q.-H.: Robotic sound source localisation algorithm with cues selection mechanism. In: *Electronics Letters* (2008), S. 1493–1494
- [MH04] In: MEFFERT, Beate ; HOCHMUTH, Olaf: *Werkzeuge der Signalverarbeitung*. Pearson Studium, 2004, S. 91–100
- [NF00] NOLFI, Stefano ; FLOREANO, Dario: *Evolutionary Robotics*. Cambridge, Massachusetts : MIT Press, 2000
- [Nil96] NILSSON, Nils J.: *Introduction to machine learning – an early draft of a proposed textbook*. <http://robotics.stanford.edu/~nilsson/MLDraftBook/MLBOOK.pdf>, 1996
- [OFS08] ONO, Nobutaka ; FUKAMACHI, Souichiro ; SAGAYAMA, Shigeki: Sound Source Localization with Front-Back Judgement by Two Microphones Asymmetrically Mounted on a Sphere. In: *Journal of Multimedia* 3 (2008), Nr. 3, S. 1–9
- [Pek06] PEKONEN, Jussi: *Onset Detection Methods for Musical Sounds*. 2006. – Helsinki University of Technology
- [Ric05] RICARD, Julien: *An Implementation Of Multi-band Onset Detection*. 2005. – MIREX 2005, University of Illinois
- [Sch94] SCHWACHULLA, Wolfram: *Meyers großes Handlexikon*. Bd. 17. Mannheim, Germany : Meyers Lexikonverlag, 1994
- [SR99] STROBEL, N. ; RABENSTEIN, Rudolf: Concurrent Speechsignal Separation Based On Frequency Domain Binaural Model. In: *Proceedings of the International Conference of Acoustics–ICASSP99*. Phonix, USA, 1999, S. 255–258
- [Sto02] STOER, J.: *Numerische Mathematik 1*. Berlin, Germany : Springer Verlag, 2002

- [WB06] WANG, DeLiang ; BROWN, Guy J.: *Computational Auditory Scene Analysis*. Hoboken, NJ, USA : IEEE Press, John Wiley & Sons, Inc., 2006
- [WLH⁺09] WOLFF, Robert ; LASSECK, Mario ; HILD, Manfred ; VILARROYA, Oscar ; HADZIBEGANOVIC, Tarik: Towards Human-Like Production and Binaural Localization of Speech Sounds in Humanoid Robots. In: *Proceedings of the 3rd International Conference on Bioinformatics and Biomedical Engineering-iCBBE 2009* Bd. II: Biomedical Engineering. Beijing, China : IEEE, 2009, S. 1–4
- [WNR49] WALLACH, H. ; NEWMAN, E. B. ; ROSENZWEIG, M. R.: The precedence effect in sound localization. In: *Am. J. Psychol.* 57 (1949), S. 315–336
- [Wol08] WOLFF, Robert: *Machbarkeitsstudie zur Realisierung der Phonotaxis am humanoiden Roboter mit Hilfe eines binauralen Audiosystems*. Berlin, Germany : Humboldt Universität zu Berlin, Institut für Künstliche Intelligenz, 2008. – Studienarbeit
- [ZH06] ZHANG, Peter X. ; HARTMANN, William H.: Lateralization of sine tones–interaural time vs phase. In: *Acoustical Society of America* (2006), S. 3471–3474
- [Zur87] ZUREK, Patrick M.: The precedence effect. In: *Directional Hearing* (1987), S. 85–105

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

Berlin, den 16. Juni 2010

Robert Wolff

Einverständniserklärung

Ich erkläre hiermit mein Einverständnis, dass die vorliegende Arbeit in der Bibliothek des Institutes für Informatik der Humboldt-Universität zu Berlin ausgestellt werden darf.

Berlin, den 16. Juni 2010

Robert Wolff