UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

FAKULTÄT FÜR
ELEKTROTECHNIK,
INFORMATIK UND
MATHEMATIK

Geometriekalibrierung akustischer Sensornetze

Von der Fakultät für Elektrotechnik, Informatik und Mathematik
der Universität Paderborn

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

genehmigte Dissertation

von

Dipl.-Ing. Florian Jacob

Erster Gutachter:	Prof. Dr.-Ing. Reinhold Häb-Umbach
Zweiter Gutachter:	Prof. Dr. Peter Schreier
Tag der mündlichen Prüfung:	21.12.2016

Paderborn 2017

Diss. EIM-E/329

Abstract

The recording of acoustic signals by several microphones is of great importance for many modern signal processing algorithms. Multi-channel recordings can be used to exploit spatial information and thus enable, for example, the suppression of speech signals or interfering noises impinging from certain directions. Furthermore, multi-channel recordings are prerequisite for the localization of speakers or acoustic events. The aforementioned techniques are applied for automatic speech recognition systems, hearing aids, advanced teleconferencing- and hands-free systems. Moreover, the performance of the algorithms that are used increases with the number of microphones as well as their spatial diversity. Therefore, spatially distributed sensors, that are composed to an acoustic sensor network (ASN) are preferred to a local accumulation of microphones.

The spatial configuration of theses sensors is mostly unknown, although it is mandatory for applications like acoustic localization. Hence, the task of geometry calibration algorithms is to automatically determine the geometric configuration of the sensors. So far, the existing algorithms primarily utilize special calibration signals and measure signal propagation times or time differences of arrival (TDOA), that allow for a computation of the corresponding distances. Timing measurements, however, require a sampling rate synchronization, which is not always present due to the spatial separation of the sensors.

This thesis is concerned with the development of geometry calibration algorithms for acoustic and audio-visual sensor networks, that do not require special calibration-signals and minimize the synchronization requirements. The calibration is carried out based on direction of arrival estimates (DOA-estimates), that are extracted from speech signals. Therefore, this thesis firstly addresses the development and analysis of DOA-estimators. However, the focus is on the design and examination of geometry calibration algorithms.

Due to reverberation and imperfect correlation properties of speech signals the DOA-estimates contain errors. Furthermore, outlier measurements are caused if no line-of-sight (LOS) propagation path from the source to the microphones is present. Core aspect of this thesis is the embedding of developed calibration techniques into a random sample consensus (RANSAC) framework to ensure a robust calibration. A calibration solely based on direction estimates only achieves a relative calibration whereby an unknown scaling factor remains. In order to fix the scaling ambiguity different strategies are examined. First of all, acoustic concepts are investigated, but the main objective is the development of audio-visual approaches. Finally, the performance of developed geometry calibration algorithms is evaluated by simulations as well as experiments in real environments.

Kurzfassung

Die Aufnahme akustischer Signale durch mehrere Mikrofone bildet die Grundlage für viele moderne Signalverarbeitungsalgorithmen. Mehrkanalige Aufnahmen gestatten die Ausnutzung räumlicher Informationen und ermöglichen somit bspw. die Unterdrückung von Sprachsignalen oder Störgeräuschen aus bestimmten Richtungen. Außerdem schaffen mehrkanalige Aufnahmen die Voraussetzungen für die Lokalisierung von Sprechern oder akustischen Ereignissen. Anwendung finden diese Techniken z. B. bei der Spracherkennung, in Hörgeräten und in Telekonferenz- ebenso wie in Freisprechsystemen. Die Leistungsfähigkeit der verwendeten Algorithmen steigt, sowohl mit zunehmender Anzahl der Mikrofone als auch mit wachsendem räumlichen Abstand. Daher werden anstatt kompakter Mikrofonansammlungen bevorzugt verteilte Sensoren, die gemeinsam ein sogenanntes akustisches Sensornetz (ASN) darstellen, eingesetzt.

Die räumliche Anordnung der Sensoren ist zumeist unbekannt, obwohl die Kenntnis dieser bspw. die Voraussetzung für die akustische Lokalisierung ist. Die Aufgabe der Geometriekalibrierung besteht deshalb in der automatischen Bestimmung der geometrischen Anordnung der Sensoren. Bislang existierenden Verfahren verwenden vorwiegend spezielle Kalibrierungssignale und messen Signallaufzeiten bzw. Signallaufzeitdifferenzen, die anschließend einen Rückschluss auf die zugehörigen Distanzen erlauben. Die Zeitmessung erfordert jedoch eine Abtastsynchronisation, die aufgrund der räumlichen Trennung der Sensoren häufig nicht gegeben ist.

Diese Arbeit beschäftigt sich daher mit der Entwicklung von Verfahren zur Geometriekalibrierung akustischer sowie audio-visueller Sensornetze, die keine Hilfsmittel, wie z. B. Kalibrierungssignale, erfordern und darüber hinaus die Synchronisationsanforderungen auf ein Minimum reduzieren. Zur Kalibrierung dienen Einfallswinkel (engl. *direction of arrival* (DOA)), die aus den Aufnahmen eines Sprachsignals extrahiert werden. Aufgrund dessen befasst sich diese Arbeit zunächst mit der Entwicklung und Analyse von Winkelschätzern. In Zentrum stehen aber der Entwurf und die Untersuchung von Geometriekalibrierungsalgorithmen.

Ausgelöst durch Nachhall und die nicht idealen Korrelationseigenschaften von Sprachsignalen treten bei der Winkelschätzung Störungen auf. Zudem ergeben sich Ausreißer, wenn kein direkter Signalausbreitungspfad (engl. *line-of-sight* (LOS)) von der Signalquelle zu den Mikrofonen vorliegt. Kernaspekt der Arbeit ist daher die Einbettung der entwickelten Algorithmen in einen *Random Sample Consensus* (RANSAC) um ein robustes Kalibrierungsverfahren zu entwickeln. Ferner gestattet die Kalibrierung allein durch Winkelschätzungen nur eine relative Bestimmung der Geometrie, sodass ein unbekannter Skalierungsfaktor verbleibt. Deswegen werden außerdem Ansätze zur Fixierung der Skalierung der Geometrie begutachtet. Zunächst werden rein akustische Lösungsmöglichkeiten untersucht, im Mittelpunkt stehen jedoch audio-visuelle Strategien. Zur Bewertung der Leistungsfähigkeit der entworfenen Geometriekalibrierungsalgorithmen dienen abschließend Simulationen ebenso wie Untersuchungen in realen Umgebungen.

Danksagung

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter im Fachgebiet Nachrichtentechnik der Universität Paderborn. Eine Förderung erfolgte im Rahmen des Projektes „Unüberwachte audio-visuelle Geometriekalibrierung von verteilten Mikrofonfeldern“ durch die Deutsche Forschungsgemeinschaft (DFG).

An dieser Stelle möchte ich mich insbesondere bei Herrn Prof. Dr.-Ing. Reinhold Häb-Umbach für die ausgezeichnete Betreuung und umfassende Unterstützung bedanken. Die zahlreichen Anregungen, die sehr gute Arbeitsatmosphäre sowie die Rahmenbedingungen im Fachgebiet Nachrichtentechnik haben entscheidend zu dieser Arbeit beigetragen. Weiterhin gilt mein Dank Herrn Prof. Dr. Peter Schreier für die Übernahme des Korreferates.

Ich danke allen wissenschaftlichen Kollegen des Fachgebietes Nachrichtentechnik für das freundschaftliche Arbeitsklima und die gemeinsame Zeit. Insbesondere danke ich Dipl.-Ing. Aleksej Chinaev für die vielfältigen Diskussionen zu fachlichen Themen ebenso wie zu gesellschaftlichen und politischen Fragestellungen. Dipl.-Ing. Oliver Walter danke ich dafür, dass er jederzeit zum Ideenaustausch bereit war und M. Sc. Lukas Drude für die detailverliebten Gespräche über die ideale Realisierung von Softwarekomponenten, aber auch für die konstruktiven Anmerkungen zu dieser Arbeit. Desweiteren danke ich Dr.-Ing. Jörg Schmalenströer, der mit seinen kritischen Fragen und Denkanstößen immer wieder zur Problemlösung beigetragen hat. Darüber hinaus danke ich auch M. Sc. Thomas Glarner, M. Sc. Jahn Heymann und M. Sc. Prerna Arora und den ehemaligen Kollegen Dr.-Ing. Volker Leutnant, Dipl.-Ing. Dang Hai Tran Vu und Dr.-Ing. Manh Kha Hoang.

Weiterhin gilt mein Dank auch allen Studierenden, deren Abschlussarbeiten ich in den letzten Jahren betreuen durfte.

Für die willkommene Abwechslung und den notwendigen Ausgleich während der Erstellung dieser Arbeit möchte ich mich bei meinen Freunden Dipl.-Ing. Tobias Tangemann, Dipl.-Ing. Jan Christoph Müller und Julian Abraham bedanken.

Meiner Schwester M. A. Isabelle danke ich für das sorgfältige und kritische Korrekturlesen. Abschließend gilt ein großer Dank meinen Eltern, aber auch meinen Großeltern, die mich immer wieder motiviert und gefördert haben und mir dadurch die idealen Möglichkeiten für meine wissenschaftliche Laufbahn eröffnet haben.

Inhaltsverzeichnis

1 Einleitung

Die zunehmende Verbreitung technischer Systeme und ihre Vernetzung ist, insbesondere im Bereich der Kommunikation, allgegenwärtiger Bestandteil der modernen Gesellschaft. Computer, Laptops, Tablets und Smartphones ermöglichen es, praktisch zu jeder Zeit, an beliebigen Orten, in Echtzeit mit dem gewünschten Kommunikationspartner in Verbindung zu treten. Neben der schriftlichen Form der Telekommunikation, z. B. durch E-Mail, SMS oder *Instant Messaging*, ist die mündliche Kommunikation die am weitesten verbreitete Form, da sie das natürlichste Mittel des Informationsaustausches darstellt. Eine deutliche Präferenz zur mündlichen Kommunikation zeigt sich speziell bei Smartphones, die neben der eigentlichen Telefonie schon von ca. 25 % der Anwender genutzt werden, um Textnachrichten mithilfe automatischer Spracherkennung zu diktieren, anstatt diese einzutippen [Bun14].

Grundvoraussetzung für eine Kommunikation mittels gesprochener Sprache ist die Aufnahme, Übertragung und Verarbeitung der akustischen Signale. Zur Aufnahme verfügen moderne Smartphones über mehr als ein Mikrofon und erlauben damit sowohl eine Störgeräuschunterdrückung als auch eine Fokussierung auf den Zielsprecher. Diese als *Beamforming* bezeichnete räumliche Filterung nutzt aus, dass sich aufgrund der örtlichen Trennung der Mikrofone unterschiedliche Übertragungseigenschaften ausbilden. Die Kombination der verschiedenen Übertragungswege führt letztendlich zur Steigerung der Signalqualität.

Um einen Gewinn durch die Filterung zu erzielen, ist eine ausreichende räumliche Diversität der Mikrofone notwendig, die jedoch im Widerspruch zu einer möglichst kompakten Bauform der Smartphones steht. Andererseits gestattet eine Vernetzung verschiedener Geräte einen Informationsaustausch und damit eine kooperative Signalverarbeitung in Form eines akustischen Sensornetzes (ASN). Ein einzelnes Mikrofon bzw. eine kompakte Ansammlung mehrerer Mikrofone besitzt möglicherweise eine große Entfernung zur gewünschten Signalquelle und eignet sich vorrangig zur Aufnahme des lokalen Schallfeldes. Der Einsatz eines akustischen Sensornetzes, das aus räumlich verteilten Mikrofonen besteht, ermöglicht hingegen die Abdeckung größerer Bereiche. Ein ASN besitzt dadurch eine höhere Wahrscheinlichkeit, dass sich ein oder sogar mehrere Mikrofone in der Nähe der Quelle befinden.

In einem Freisprechszenario könnte somit bspw. die Aufnahme von den Mikrofonen des Smartphones unmittelbar an ein benachbartes Laptop verlagert werden, sofern sich dieses näher am Zielsprecher befindet. Alternativ dazu erlauben *Beamforming*-Verfahren die Kombination aller Mikrofonsignale des ASN und erzielen damit eine Steigerung der Empfindlichkeit in Richtung des Zielsprechers bei gleichzeitiger Unterdrückung von Störgeräuschen, wie z. B. dem Rauschen eines Laptoplüfters.

Neben dem skizzierten Freisprechszenario profitieren auch zahlreiche weitere Anwendungsbereiche vom Einsatz eines ASN oder werden sogar erst dadurch möglich. Ein weiterer Anwendungsbereich eines akustischen Sensornetzes ist bspw. ein Telekonferenzsystem in einem Besprechungsraum, das aus mehreren Mikrofonen besteht, um bestmögliche Aufnahmebedingungen für alle Teilnehmer zu bieten [WC97; SH10]. Darüber hinaus finden ASN u. a. bei der Lokalisierung akustischer Ereignisse [GH10], der Sprecherlokalisation [PF14c; WHP04], in Hörgeräten [BM09; Doc+09], bei der automatischen Spracherkennung [KMR12], aber auch in Systemen zur Realisierung ambienter Intelligenz [PST07] sowie zur Überwachung Anwendung [ZBS09].

Angesichts der vielfältigen Einsatzmöglichkeiten akustischer Sensornetze sind sie essentieller Bestandteil zahlreicher Forschungsvorhaben. Wenig Berücksichtigung findet dabei der Aspekt, dass die räumliche Anordnung der Mikrofone in der Regel unbekannt ist, obwohl die Kenntnis dieser bspw. die Grundvoraussetzung für die Lokalisation von Ereignissen bzw. Personen darstellt [BAS95]. Sofern ein ASN aus mobilen Endgeräten, wie z. B. Smartphones, besteht, ist die Position der Mikrofone innerhalb des Sensornetzes möglicherweise sogar zeitveränderlich.

Andererseits lässt sich durch die Kenntnis der Sensorkonfiguration eine höhere Leistungsfähigkeit erzielen [TTH14]. Letztendlich erzeugen die akustischen Signalverarbeitungsverfahren, die die Kenntnis der geometrischen Anordnung der Mikrofone voraussetzen bzw. davon profitieren, die Nachfrage nach Verfahren zur automatischen Bestimmung der räumlichen Mikrofonanordnung. Der Vorgang, diese zu ermitteln, wird als *Geometriekalibrierung* bezeichnet und kann sowohl durch manuelle als auch durch automatische Verfahren erfolgen. Allerdings erfordert schon das manuelle Einmessen der Sensorpositionen von kleinen Netzen, die aus wenigen Sensoren bestehen, einen erheblichen Zeitaufwand. Außerdem kann abhängig von der Umgebung, in der das Sensornetz zum Einsatz kommt, die manuelle Kalibrierung eine enorme Herausforderung bedeuten. Eine automatische Kalibrierung ist daher erstrebenswert, um die Kalibrierung der Sensornetze zu erleichtern und die Fehleranfälligkeit zu reduzieren. Weiterhin eröffnet sie die Möglichkeit, dynamisch veränderliche Sensornetze einzusetzen, die bspw. aus Smartphones bestehen oder die Nutzung größerer Installationen, bis hin zu riesigen Mikrofonarrays (engl. *huge microphone array* (HMA)), die z. T. mehrere hundert Mikrofone besitzen [SSP05].

Die vielfältigen Einsatzmöglichkeiten akustischer Sensornetze und die damit verbundene Nachfrage nach automatischen Geometriekalibrierungsverfahren hat bereits zur Entwicklung von verschiedenen Ansätzen zur automatischen Geometriekalibrierung geführt. Während erste Varianten lediglich darauf abzielten, das manuelle Einmessen der Sensorkonfiguration zu erleichtern [BS05], besteht das Ziel moderner Verfahren in der vollständigen Automatisierung. Um dieses Ziel zu erreichen, kommen z. B. eigens angefertigte Lautsprecherkonstruktionen [SSP05], spezielle Kalibrierungssignale [RD04; Red+09] oder aktive Sensorknoten, die sowohl über Mikrofone als auch Lautsprecher verfügen [HF11; PMH11], zum Einsatz. Im Gegensatz dazu beschäftigt sich die vorliegende Arbeit mit der Entwicklung von Geometriekalibrierungsverfahren, die keine Lautsprecher, künstlichen Kalibrierungssignale oder andere Hilfsmittel erfordern, sondern die zur Kalibrierung eines ASN notwendigen Informationen aus einem Sprachsignal extrahieren. Der Fokus liegt dabei auf der Kalibrierung von Sensornetzen, die zur Lokalisation von Ereignissen und/oder Sprechern sowie zum Einsatz in Telekonferenzsystemen geeignet

sind. Da dort vorwiegend kompakte Sensorknoten, die aus wenigen Mikrofonen bestehen, Verwendung finden, sollen die zu entwickelnden Algorithmen für ein solches Szenario konzipiert werden.

Das nachfolgende Kapitel 2 erläutert die dieser Arbeit zugrunde liegenden Szenarien inklusive der zugehörigen Rahmenbedingungen. Dabei gilt es insbesondere festzulegen, welche Parameter eines Sensornetzes durch die automatische Kalibrierung zu ermitteln sind. Daran anschließend erfolgt ein Überblick über existierende Geometriekalibrierungslösungen. Dieser beinhaltet neben einer ausführlichen Darstellung akustischer Kalibrierungsvarianten auch einen kurzen Exkurs in nicht-akustische Anwendungsbereiche. Ausgehend davon werden schließlich die wissenschaftlichen Ziele dieser Arbeit definiert und die erforderliche Genauigkeit der Geometriekalibrierung festgelegt.

Zumal akustische Signale die Informationsquelle für die Geometriekalibrierung darstellen, bilden sie einen wichtigen Bestandteil dieser Arbeit. Daher gibt Kapitel 3 zunächst einen kurzen Überblick über die Konzepte zur Beschreibung der Schallausbreitung in Räumen. Darauf aufbauend wird die Simulation der Schallausbreitung betrachtet, da simulierte Audiosignale sowohl zur Entwicklung als auch Evaluierung der Geometriekalibrierungsalgorithmen genutzt werden.

Angesichts der im Verlauf von Kapitel 2 dargelegten Entscheidung, die Geometriekalibrierung mithilfe von Einfallswinkeln durchzuführen, befasst sich Kapitel 4 mit der Auswahl eines Verfahrens zur Schätzung der Einfallswinkel aus den Mikrofonsignalen. Dabei werden sowohl existierende Ansätze als auch ein ebenfalls im Rahmen dieser Arbeit konzipierter Winkelschätzer berücksichtigt. Die Untersuchungen, die die Grundlage für die Auswahl eines Winkelschätzers bilden, weisen zudem auf einen erheblichen systematischen Fehler bei der Winkelschätzung hin. Daher werden außerdem die Ursachen dieses Fehlers analysiert, um eine verlässliche Winkelschätzung für die Geometriekalibrierung zu erzielen. Gleichzeitig dienen die gewonnenen Erkenntnisse zur Entwicklung eines Modells zur statistischen Beschreibung des Fehlers der Einfallswinkelschätzung.

Kapitel 5 beginnt mit der Analyse eines vielversprechenden Kalibrierungsalgorithmus. Dieser bildet die Grundlage für die eigenen Entwicklungen eines akustischen Geometriekalibrierungsverfahrens. Außerdem soll durch Simulationen einerseits die Wirksamkeit der eigenen Entwicklungen geprüft und andererseits die Leistungsfähigkeit des Algorithmus im Hinblick auf den realen Einsatz bewertet werden.

Aufgrund der im Rahmen dieser Untersuchungen festgestellten Reduktion der Leistungsfähigkeit durch Messausreißer, beschäftigt sich Kapitel 6 mit der Steigerung der Robustheit des in Kapitel 5 entwickelten Verfahrens gegenüber Messausreißern. Kern dieser Betrachtungen ist die Nutzung eines *Random Sample Consensus* (RANSAC), um damit den Einfluss der Ausreißer zu reduzieren. Darüber hinaus dienen Erweiterungen des RANSAC-Konzeptes dazu, dieses an die speziellen Herausforderungen, die sich aus dem zur Kalibrierung genutzten Einfallswinkelverfahren ergeben, anzupassen und gleichzeitig zu einer recheneffizienteren Anwendung des Konzeptes beizutragen.

Die Kalibrierung durch das Einfallswinkelverfahren und den RANSAC gestattet zunächst nur eine relative Angabe der Sensorpositionen, sodass ein unbekannter Skalierungsfaktor zwischen dem Kalibrierungsergebnis und der tatsächlichen Sensorkonfiguration verbleibt. Daher werden in Kapitel 7 zwei Varianten zur Gewinnung der erforderlichen Skalierung entwickelt. Diese beschränken sich auf die Ausnutzung der Informationen eines akustischen Sensornetzes.

Da im Umfeld von ASN häufig nicht nur Mikrofone, sondern auch Kameras zur Verfügung stehen, werden in Kapitel 8 zusätzlich audio-visuelle Ansätze zur Lösung der Skalierungsproblematik konzipiert und begutachtet. Durch die Kombination der Informationen aus beiden Modalitäten resultiert außerdem eine Beschreibung aller Sensoren in einem gemeinsamen Koordinatensystem, die wiederum die Grundlage für die Ausnutzung von Synergien bildet.

Die entwickelten Geometriekalibrierungslösungen, die in den durchgeführten Untersuchungen die besten Resultate erzielten, werden abschließend in Kapitel 9 in zwei realen Szenarien angewendet, um ihre Praxistauglichkeit zu bewerten. Schließlich fasst Kapitel 10 die Ergebnisse dieser Arbeit zusammen.

2 Grundlagen der Geometriekalibrierung

Akustische Sensornetze bilden die Grundlage für die Realisierung zahlreicher Signalverarbeitungsaufgaben. Abschnitt 2.1 beginnt deshalb mit einer kurzen Beschreibung ausgewählter Einsatzmöglichkeiten, bei denen die Kenntnis der geometrischen Anordnung der Mikrofone Voraussetzung für den jeweiligen Einsatzzweck ist. Trotz der unterschiedlichen Verwendungsmöglichkeiten, lassen sich jedoch gewisse Basiskonfigurationen identifizieren. Darüber hinaus erläutert Abschnitt 2.1 auch, welche Parameter im Rahmen des Kalibrierungsprozesses zu bestimmen sind. Somit dient dieser Abschnitt zur Konkretisierung der Ausgangssituation.

Im Anschluss daran erfolgt in Abschnitt 2.2 ein Überblick über den bisherigen Stand der Forschungen auf dem Gebiet der Geometriekalibrierung. Um eine systematische Übersicht über die Vielzahl der akustischen Ansätze zu gestatten, werden zuerst Kategorien definiert und danach ausgewählte Verfahren aus den jeweiligen Kategorien vorgestellt. Der Fokus dieses Überblicks liegt auf Algorithmen zur Kalibrierung akustischer Sensornetze. Allerdings dient ein kurzer Exkurs in nicht-akustische Anwendungsbereiche, zu denen u. a. die Kalibrierung von Kameranetzen oder Funksensornetzen gehören, dazu, die dort verwendeten Techniken zu skizzieren. Aufbauend auf den erläuterten Eigenschaften der existierenden Geometriekalibrierungsverfahren werden in Abschnitt 2.3 die wissenschaftlichen Ziele dieser Arbeit definiert. Abschnitt 2.4 befasst sich schließlich mit der Kalibrierungsgenauigkeit der zu entwickelnden Geometriekalibrierungsalgorithmen.

2.1 Szenario

Angesichts der zahlreichen Einsatzmöglichkeiten für akustische Sensornetze existieren diese in unterschiedlichsten Ausprägungen. Grundsätzlich besteht ein ASN jedoch aus mehreren, räumlich verteilten Sensorknoten. Die Anforderungen, die sich aus dem jeweiligen Einsatzzweck ergeben, variieren und beeinflussen den Aufbau einzelner Sensorknoten, deren räumliche Positionierung sowie die verwendete Anzahl. Im einfachsten Fall besitzt ein Sensorknoten lediglich ein einzelnes Mikrofon. Meistens verfügt er hingegen über mehrere Mikrofone, die eine kompakte Gruppe (Array) bilden. Da die Anzahl der Mikrofone eines Sensorknotens nur eine untergeordnete Rolle spielt, wird die Bezeichnung *Sensor* bzw. *Sensorknoten* im Rahmen dieser Arbeit sowohl für ein einzelnes Mikrofon als auch für kompakte Mikrofonarrays verwendet und kennzeichnet lediglich, dass die Mikrofone zu ein und demselben Gerät gehören.

Zu den bekanntesten Einsatzbereichen akustischer Sensornetze gehört die Lokalisation und Verfolgung von Ereignissen und/oder Sprechern [WLW03]. Dazu wird in [WLW03] ein Sensornetz verwendet, bei dem jeder der an den Wänden des Raumes montierten Sensoren lediglich über ein einzelnes Mikrofon verfügt. Mithilfe von Messungen der Signallaufzeitdifferenzen und eines Partikel-Filters erfolgt anschließend die Verfolgung des Sprechers. Im Gegensatz dazu sind in [MP10] mehrere Sensoren an einem mobilen Serviceroboter befestigt, damit dieser Personen in seinem Umfeld lokalisieren kann, um anschließend mit ihnen zu interagieren.

In intelligenten Umgebungen (engl. *smart home*) ermöglichen akustische Sensornetze in Kombination mit *Beamforming*-Verfahren die Fokussierung der Mikrofone auf die Zielquelle, bei gleichzeitiger Unterdrückung von Störquellen [MGC13]. Bevorzugt sind auch hier die Sensoren an den Wänden des Raumes montiert. Sofern eine schwenkbare Kamera vorhanden ist, lässt sich diese, basierend auf der akustischen Lokalisierung, die z. B. durch eine Triangulation erfolgt [WHP04], in Richtung des detektierten Ereignisses ausrichten [WC97; HBE00].

Weiterhin erlauben akustische Sensornetze die Realisierung sogenannter akustischer Kameras. Dabei wird, vergleichbar zu einer visuellen Kamera, der Raum mithilfe mehrerer Mikrofone akustisch abgetastet, sodass eine visuelle Darstellung des Schallfeldes (Geräuschkarte) entsteht [Red+09]. Ferner sind akustische Sensornetze Bestandteil von Systemen zur Sprecheridentifikation und Extraktion von Kontextinformationen [SH10]. Sehr eng verwandt mit diesem Themenbereich ist auch die automatische Annotation von audio-visuellen Daten [Kuh+07], bei der ebenfalls auf Sensornetze zurückgegriffen wird. Aber auch die automatische Spracherkennung profitiert von der Verwendung akustischer Sensornetze, sofern kein Nahbereichsmikrofon vorhanden ist [KMR12].

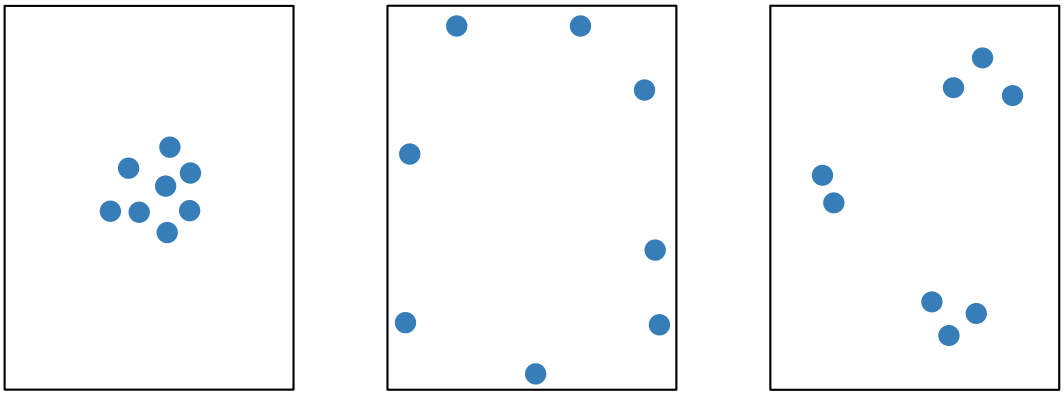
Im Rahmen dieser Arbeit soll ein multimedialer Besprechungsraum zur audio-visuellen Kommunikation (engl. *telepresence*) als Anwendungsszenario für ein ASN dienen. Schon heute liefern vornehmlich Firmen aus der Telekommunikationsbranche, aber auch Netzwerkausrüster, Lösungen, die meist aus einem Konferenztelefon, Kameras und Lautsprechern sowie Anzeigegeräten bestehen. Ziel dabei ist es, den Nutzern des Systems den Anschein zu vermitteln, dass sich die zugeschalteten Teilnehmer im gleichen Raum befinden, um dadurch den Eindruck einer persönlichen Besprechung zu gewährleisten.

Zur Steigerung der Sprachqualität verfügen heutige Konferenztelefone über zusätzliche Mikrofone, die bei Bedarf durch den Nutzer aktiviert werden können. In zukünftigen Systemen könnten sich angesichts der voranschreitenden Vernetzung auch die Mikrofone der Smartphones und Laptops der Teilnehmer nahtlos in das akustische Sensornetz integrieren. Dadurch lässt sich eine größere räumliche Diversität der Signale erzielen, die in Kombination mit modernen Störgeräuschunterdrückungstechniken und *Beamforming*-Verfahren zu einer bestmöglichen Sprachqualität führt. Außerdem kann eine schwenkbare Kamera, die sich basierend auf einer akustischen Positionsschätzung auf den gerade aktiven Besprechungsteilnehmer ausrichtet, dazu beitragen, den Anschein zu erwecken, als würde man den Teilnehmer direkt ansehen. Somit lässt sich der Eindruck einer persönlichen Besprechung weiter verstärken. Zusätzlich ist durch die Integration eines Spracherkenners auch eine automatische Transkription des Gesprächsverlaufs möglich, die gespeist durch ein System zur Extraktion von Kontextinformationen, gleichzeitig Information darüber enthalten kann, wann bspw. eine Person den Raum verlässt. Angesichts der Vielzahl der Anwendungsmöglichkeiten akustischer Sensornetze, die im

Kontext eines multimedialen Besprechungsraumes auftreten können und gleichzeitig von der Kenntnis der Sensoranordnung profitieren, stellt ein Besprechungsraum ein geeignetes Szenario zur Betrachtung von Geometriekalibrierungsverfahren dar.

Aus dem geschilderten Kommunikationsszenario und den weiteren Verwendungsmöglichkeiten eines akustischen Sensornetzes lassen sich drei grundsätzliche Sensorkonstellation ableiten. Schematisch sind diese in Abb. 2.1 skizziert. Die begrenzenden Rechtecke kennzeichnen dabei die Wände des Raumes. Ferner gilt es zu beachten, dass ein Sensorknoten, dargestellt durch einen blauen Punkt, sowohl ein einzelnes Mikrofon, aber auch ein Array, repräsentieren kann.

Sofern das Sensornetz aus Mikrofonen besteht, die bspw. auf einem Konferenztisch stehen, konzentrieren sich die Sensoren relativ kompakt in der Mitte des Raumes (vgl. Abb. 2.1a). In intelligenten Umgebungen, in denen eine möglichst unauffällige Integration der Sensoren in die Umgebung im Vordergrund steht, werden die Sensoren in der Nähe der Wände platziert. Somit entsteht eine Konstellation vergleichbar zu Abb. 2.1b. Auch bei der Lokalisierung und Verfolgung von Personen findet eine solche Anordnung häufig Verwendung. Insbesondere bei Systemen zur Lokalisierung von Ereignissen treten aber auch räumlich verteilte Gruppen von Sensoren auf (vgl. Abb. 2.1c).



(a) Konstellation mit innenliegenden Sensoren.

(b) Konstellation mit außenliegen Sensoren.

(c) Konstellation mit räumlich verteilten Sensorgruppen.

Abbildung 2.1: Schematische Darstellung verschiedener Konstellationen akustischer Sensornetze. Die Rechtecke kennzeichnen dabei die Wände des Raumes, während die blauen Punkte einen Sensorknoten darstellen.

Unabhängig von der vorliegenden Sensorkonstellation besteht das Ziel der Geometriekalibrierung in der Bestimmung der Positionen der Sensorknoten. Sofern ein Sensorknoten nur ein einzelnes Mikrofon besitzt, das keine omnidirektionale Richtcharakteristik [Gau+14] hat oder er sich aus mehreren Mikrofonen zusammensetzt, ist die Orientierung ebenfalls zu kalibrieren. Im Fall eines einzelnen Mikrofons beschreibt die Orientierung die Ausrichtung der Richtcharakteristik. Bei einem Sensorknoten, der aus mehreren Mikrofonen besteht, ist hingegen die Orientierung des Knotens gemeint, die dazu dient zusammen mit der Position und dem Aufbau des Sensorknotens die Positionen der einzelnen Mikrofone festzulegen. Außerdem umfasst die Geometriekalibrierung

nicht nur die Kalibrierung der Sensorpositionen und optionalen Orientierungen, sondern auch die Kalibrierung der Mikrofonpositionen innerhalb eines Sensorknotens.

Voraussetzung für die Angabe der Positionen und der optionalen Orientierungen ist ein Bezugspunkt bzw. ein Referenzkoordinatensystem. In der Regel gestatten die betrachteten Kalibrierungsverfahren allerdings keine Kalibrierung zu einem absoluten Bezugspunkt, wie z. B. einer Ecke des Raumes, sondern lediglich eine relative Beschreibung der Lage der Sensoren zueinander.

Weitere Eigenschaften der Mikrofone, wie z. B. ihre Richtcharakteristik oder auch die Mikrofonverstärkung, sind zwar z. T. für die spätere Nutzung des akustischen Sensornetzes notwendig, aber nicht Bestandteil der Geometriekalibrierung. Dennoch existieren Ansätze, die die Kalibrierung von Geometrie und Mikrofonverstärkung in einem gemeinsamen Verfahren vereinigen [SSP05; GKH14].

2.2 Existierende Kalibrierungsverfahren

Das Problem der Geometriekalibrierung ist thematisch eng mit dem der Ereignislokalisation verknüpft. Einerseits erfolgt die Lokalisierung von Ereignissen häufig durch Triangulation bzw. Trilateration[1] und setzt daher die Kenntnis der Sensoranordnung voraus. Andererseits lässt sich die Geometriekalibrierung als inverse Problemstellung auffassen, bei der die Sensorpositionen ausgehend von den Ereignispositionen bestimmt werden sollen. Während Verfahren wie z. B. [SSP05] eine Lautsprecherkonstruktion mit bekannten Abständen verwenden, damit die Positionen der Ereignisse vorliegen, sind diese in der Regel jedoch unbekannt.

Um einen systematischen Überblick über die unterschiedlichen Algorithmen zur akustischen Geometriekalibrierung zu gestatten, werden zunächst Merkmale zur Klassifikation dieser Algorithmen definiert. Die verwendeten Merkmale umfassen erstens die räumliche Anordnung der Mikrofone und Sensorknoten sowie zweitens die zur Kalibrierung verwendeten Informationen (siehe Abb. 2.2). Aufbauend auf den im Folgenden näher erläuterten Kategorien, präsentieren die sich daran anschließenden Abschnitte ausgewählte Algorithmen, um einen Einblick in die grundlegenden Konzepte der Geometriekalibrierung zu gewähren.

Bei der räumlichen Anordnung der Mikrofone und Sensorknoten lassen sich drei Klassen identifizieren. Sofern alle betrachteten Mikrofone Bestandteil eines Sensors sind und ein kompaktes Array mit begrenzten Ausmaßen bilden, bei dem lediglich die Positionen der Mikrofone innerhalb des Sensorknotens bzw. Arrays kalibriert werden sollen, wird dieser Vorgang als *Intra-Array-Kalibrierung* bezeichnet. Allerdings bieten nur wenige Ansätze die erforderliche Genauigkeit, da der Abstand der Mikrofone meist im Bereich von 0,10 m [Con+12] oder sogar noch darunter [MLH08] liegt und der Kalibrierungsfehler dementsprechend deutlich kleiner ausfallen muss.

Der Übergang zur Kategorie *Inter-Mikrofon-Kalibrierung* ist fließend, da es sich hier ebenfalls um die Kalibrierung von Konstellationen aus einzelnen Mikrofonen handelt, die sich jedoch in größeren räumlichen Anordnungen befinden und damit

[1] Der Begriff Trilateration oder auch Lateration beschreibt Verfahren zur Positionsbestimmung basierend auf (drei) Abstands- bzw. Entfernungmessungen. Triangulationverfahren verwenden stattdessen (drei) Winkelmessungen.

nicht notwendigerweise zum selben Sensorknoten gehören. Handelt es sich bei den Sensorknoten hingegen um Mikrofonarrays, deren interne Geometrie bekannt ist oder durch ein Intra-Array-Kalibrierungsverfahren ermittelt wurde, wird die Kalibrierung der Sensorpositionen und Orientierungen als *Inter-Array-Kalibrierung* bezeichnet.

Neben der Einteilung, die sich an der räumlichen Anordnung der Mikrofone bzw. dem Aufbau der Sensorknoten orientiert, lassen sich die Geometriekalibrierungsalgorithmen auch anhand der Informationen, die die Grundlage für die Durchführung des Kalibrierungsprozesses bilden, gruppieren. Die erste und zugleich größte Gruppe verwendet *Distanzen*, die häufig durch Signallaufzeitmessungen von der Signalquelle zu den Sensoren gewonnen werden. Die zweite Gruppe nutzt Messungen der *Zeitdifferenzen* zwischen der Detektion eines Ereignisses an den verschiedenen Sensoren. Die Verfahren aus diesen beiden Gruppen eignen sich vornehmlich für die Intra-Array-Kalibrierung und Inter-Mikrofon-Kalibrierung, ermöglichen aber auch die Inter-Array-Kalibrierung.

Für die dritte Gruppe ist es hingegen notwendig, dass jeder Sensorknoten über mehrere Mikrofone verfügt. Daher gestatten diese Verfahren ausschließlich eine Inter-Array-Kalibrierung. Durch den Einsatz eines Mikrofonarrays pro Sensorknoten wird eine Bestimmung der *Positionen* der Ereignisse im Koordinatensystem des jeweiligen Sensorknotens möglich. Zudem lässt sich durch die Kombination der Positionsschätzungen von mehreren Ereignissen anschließend eine Koordinatentransformation bestimmen aus der sich die Sensorpositionen ergeben. Auf die Methoden wie die erforderlichen Positionen bestimmt werden, wird jedoch erst in Abschnitt 2.2.3 näher eingegangen.

Abb. 2.2 stellt die Überschneidungen der zuvor eingeführten Kategorien für akustische Geometriekalibrierungsverfahren noch einmal in grafischer Form dar. Um dabei hervorzuheben, dass der Übergang zwischen den Kategorien Intra-Array-Kalibrierung und Inter-Mikrofon-Kalibrierung fließend ist, wird an dieser Stelle eine gestrichelte Linie verwendet, während die harte Abgrenzung zwischen Inter-Mikrofon-Kalibrierung und Inter-Array-Kalibrierung als durchgezogene Linie darstellt wird.

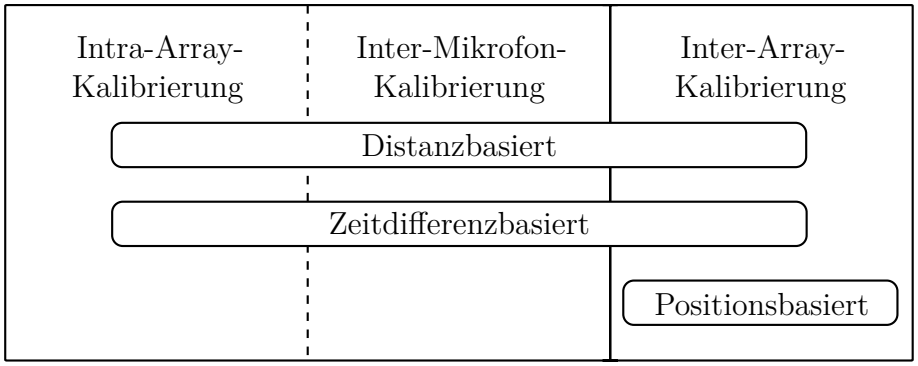


Abbildung 2.2: Einteilung akustischer Geometriekalibrierungsverfahren.

In den Abschnitten 2.2.1 bis 2.2.3 werden jeweils exemplarisch einige der Verfahren für distanz-, zeitdifferenz- und positionsbasierte Kalibrierungsalgorithmen vorgestellt. Im Anschluss daran gibt Abschnitt 2.2.4 einen Ausblick auf Lösungsansätze des Geometriekalibrierungsproblems für Sensornetze, die statt der Mikrofone über andere Sensoren verfügen. Dieser Exkurs dient dazu, die Strategien in anderen Themenbereichen herauszustellen, um diese ggf. auch in der akustischen Kalibrierung aufgreifen zu können.

2.2.1 Distanzbasierte Kalibrierung

Die Basis für zahlreiche Ansätze zur Kalibrierung der räumlichen Anordnung von Mikrofonen ist die Multidimensionale Skalierung (MDS) [Bir03]. Die Durchführung der MDS erfordert die Kenntnis der Distanzen zwischen sämtlichen Sensorpaaren. In [Bir03] wird von einer manuellen Messung, z. B. mit einem Bandmaß, ausgegangen. Aus der Matrix, die die Quadrate aller paarweisen Distanzmessungen beinhaltet, lässt sich anschließend die relative Position der Mikrofone zueinander zurückgewinnen. Dabei nutzt die MDS aus, dass die Eigenvektoren der quadrierten Distanzmatrix den Unterraum aufspannen, der die Positionen der Mikrofone beinhaltet. Eine ebenfalls in [Bir03] vorgestellte Erweiterung ermöglicht es, auch dann die Positionen der Mikrofone zu ermitteln, wenn einige der Distanzen unbekannt sind. Grundlage dafür ist, dass die vollständige Distanzmatrix redundante Informationen enthält. Die als Multidimensionale Skalierung mit Basispunkten (BMDS) bezeichnete Verallgemeinerung dieses Konzeptes bestätigt, dass anstatt einer vollständigen Matrix auch Distanzmessungen zu einer Gruppe von Basispunkten ausreichen [BS05].

Das Ziel von akustischen Geometriekalibrierungsverfahren besteht jedoch darin, den Kalibrierungsprozess durch die Gewinnung der erforderlichen Informationen aus den zur Verfügung stehenden Mikrofonsignalen zu automatisieren und auf manuell zu ermittelnde Informationen zu verzichten. Allerdings greifen viele Verfahren auf die MDS bzw. BMDS zurück und unterscheiden sich hauptsächlich in den zur Distanzschätzung aus den Audiosignalen eingesetzten Methoden.

Eine Möglichkeit zur Schätzung des Abstandes zwischen den Sensoren bietet die Analyse der Kohärenzfunktion in Umgebungen mit diffusem Rauschen. Dazu findet ein Abgleich zwischen dem theoretischen Modell der Kohärenzfunktion von diffusem Rauschen mit der Schätzung dieser Funktion aus den Aufnahmen eines Mikrofonpaares statt [MLH08]. Gemäß dieses Modells ist die Kohärenz eine Funktion des Mikrofonabstands und liefert somit die für die Anwendung der MDS erforderlichen Distanzen. Die Präzision dieser Distanzschätzung lässt sich durch die Kombination mit *Generalized Cross Correlation with Phase Transform* (GCCPhat) steigern [Vel+15]. Da die Annahme eines diffusen Schallfeldes jedoch geringe Mikrofonabstände voraussetzt, eignen sich diese Verfahren hauptsächlich zur Intra-Array-Kalibrierung. Durch die in [Tag+15] präsentierte Erweiterung lässt sich das Kohärenzmodell auch dann nutzen, wenn ein einzelnes Mikrofon einen größeren Abstand besitzt.

Ein alternatives Vorgehen zur Distanzschätzung wird in [PHM13] beschrieben. Die Zeitdifferenz zwischen dem Eintreffen des zu einem akustischen Ereignis gehörenden Signals an zwei verschiedenen Mikrofonen (engl. *time difference of arrival* (TDOA)), ist abhängig vom Abstand der Mikrofone und der Quellposition. Die größtmögliche Zeitdifferenz (engl. *maximum time difference of arrival* (MTDOA)) entsteht, wenn sich ein Ereignis auf der Verbindungsgeraden der Mikrofone, jedoch nicht zwischen diesen befindet. Sofern für jedes Mikrofonpaar ein Ereignis vorliegt, welches diese Bedingung erfüllt, ergeben sich aus den maximalen Zeitdifferenzen multipliziert mit der Schallgeschwindigkeit die paarweisen Abstände der Mikrofone. Diese wiederum erlauben die Anwendung der MDS. Angesichts der erzielten Genauigkeiten eignen sich MTDOA-Verfahren vorwiegend zur Inter-Mikrofon-Kalibrierung [PPH14] oder ggf. auch zur Inter-Array-Kalibrierung.

Im Gegensatz zu den Verfahren die auf die MDS zurückgreifen, erfolgt die Inter-Mikrofon-Kalibrierung in [Con+12] durch eine Triangulation der Mikrofone. Dazu kommt ein Lautsprecher zum Einsatz, der ein spezielles Kalibrierungssignal von mehr als hundert verschiedenen Gitterpunkten wiedergibt. Die Zeit, die das Signal vom Lautsprecher bis zum Mikrofon benötigt (engl. *time of flight* (TOF)), wird aus der Übertragungsfunktion errechnet, die sich aus dem ausgesandten Kalibrierungssignal und dem empfangenen Mikrofonsignal ergibt. Unter der Annahme, dass ein direkter Ausbreitungspfad zwischen Lautsprecher und Mikrofon existiert (engl. *line-of-sight* (LOS)), liefert die Position des ersten Maximums der Übertragungsfunktion die TOF. Diese beschreibt die möglichen Mikrofonpositionen, die auf einem um den Lautsprecher verlaufenden Kreis liegen. Aus den Schnittpunkten der Kreise von allen Lautsprecherpositionen resultiert anschließend die geschätzte Mikrofonposition.

Eine Alternative zur Positionierung eines Lautsprechers an zahlreichen Punkten bietet [SSP05]. Ein spezielles Gerät aus fünf Lautsprechern dient zur Wiedergabe eines Kalibrierungssignals. Die Messung der TOF ermöglicht die Schätzung der Distanz zu allen Lautsprechern, sodass die gleiche Ausgangssituation wie auch bei BMDS vorliegt. Zur Bestimmung der Mikrofonposition wird allerdings eine Simplex-Methode genutzt.

Die Messung der TOF erfordert indes eine Zeitsynchronisation zwischen Lautsprechern und Mikrofonen. Die Messung des Zeitpunktes, zu dem ein Signal an den Mikrofonen eintrifft (engl. *time of arrival* (TOA)), benötigt hingegen nur eine Abtastsynchronisation der Mikrofone und kommt deshalb ohne eine Synchronisation mit der Signalquelle aus. Damit jedoch eine Schätzung der TOA möglich wird, sind z. B. impulshafte Signale [GKH13] oder bekannte Kalibrierungssignale [GKH14] notwendig.

Die TOA, die sich aus dem Zeitpunkt, zu dem das Ereignis aufgetreten ist (engl. *onset*), und der TOF zusammensetzt, bildet die Grundlage für eine weitere Gruppe von Kalibrierungsverfahren. In [GKH13] wird eine Rang-Approximation genutzt, um basierend auf einer TOA-Messung zunächst die TOF und die *Onsets* zu ermitteln sowie eine ggf. nicht perfekte Abtastsynchronisation der Mikrofone auszugleichen. Anschließend gestattet eine weitere Rang-Approximation, die ursprünglich zur signallaufzeitdifferenzbasierten Kalibrierung diente [Thr05] (siehe Abschnitt 2.2.2) und in [CDM12] weiterentwickelt wurde, die Bestimmung der Mikrofonpositionen aus der TOF.

Die zweite Rang-Approximation stellt dabei ein Verfahren zur Minimierung der Zielfunktion dar, die die gemessenen Distanzen und die sich aufgrund der Mikrofon- und Ereignispositionen ergebenden Abstände in Verbindung setzt. Da diese Zielfunktion genauso viele Freiheitsgrade besitzt, wie es Mikrofon- und Ereignispositionen gibt, gestaltet sich, insbesondere in Anwesenheit von Rauschen, eine Optimierung, z. B. mit einem Gradientenverfahren, problematisch und erhöht die Gefahr, nur ein lokales Optimum zu erreichen [Cro+12]. Durch die Verwendung der Rang-Approximation kann ein Teilproblem geschlossen gelöst werden, während im zweiten Teil lediglich die Lösung eines nichtlinearen Optimierungsproblems mit neun Parametern unabhängig von der Anzahl der Ereignisse oder Mikrofone verbleibt.

Die Machbarkeitsstudie [PN08] zeigt zudem eine Möglichkeit, die TOA aus der TDOA zu gewinnen und eine Positionsbestimmung durchzuführen. Die Verwendung der so ermittelten TOA in dem von [Kua+13] präsentierten Kalibrierungsansatz erfolgt durch [KA13]. Allerdings erfordert die Anwendung der beschriebenen Verfahrensweise eine manuelle Annotation der zur Verfügung stehenden Audiodaten.

Aktive Sensorknoten, die aus Mikrofonen und Lautsprechern bestehen, ermöglichen die Bestimmung der TOA aus der Übertragungsfunktion zwischen verschiedenen Sensorknoten [PMH11]. Sofern jeder Sensorknoten neben einem Lautsprecher über mehrere Mikrofone verfügt, lässt sich die TOA auch ohne eine Abtastsynchronisation der Sensorknoten ermitteln. Zur Berechnung der Sensorpositionen wird anschließend ebenfalls die MDS eingesetzt. In einem weiteren Schritt dienen die Mikrofone eines Knotens zur Schätzung des Signaleinfallswinkels (engl. *direction of arrival* (DOA)). Aus den Signaleinfallswinkeln und den Positionen der Sensorknoten wird abschließend die Orientierung des Sensorknotens gewonnen.

Ein Konzept, dass sich von allen bislang beschriebenen Kalibrierungsverfahren sehr deutlich unterscheidet, wird in [Iva14] vorgestellt. Zur Kalibrierung dient dort eine Signalquelle, die impulshafte Geräusche emittiert. Die Besonderheit des Algorithmus besteht darin, dass in die Kalibrierung nicht nur die TOA-Messung des direkten Signalpfades einfließt, sondern auch die TOA-Messungen der auftretenden Reflexionen. Somit liefert jedes Ereignisse mehrere Informationen und der in [Iva14] erläuterte Algorithmus ist dadurch in der Lage sowohl die Geometrie des Raumes als auch die absolute Position der Mikrofone im Raum zu gewinnen.

2.2.2 Zeitdifferenzbasierte Kalibrierung

Zahlreiche der bisher betrachteten Ansätze stützen sich vornehmlich auf bekannte Kalibrierungssignale, um daraus die TOF bzw. TOA zu ermitteln. Wenn das Signal jedoch unbekannt ist oder keine Zeitsynchronisation zwischen Signalquelle und den Sensoren vorliegt, scheiden diese Verfahren aus. Eine Alternative bilden daher Ansätze, die mit Zeitdifferenzmessungen (TDOA) arbeiten und durch den Einsatz von Korrelationsverfahren, wie z. B. GCCPhat [KC76], auch ohne Kenntnis des Quellsignals funktionieren.

Die schon im vorherigen Abschnitt thematisierte Rang-Approximation stammt ursprünglich aus [Thr05] und ermöglicht dort die Minimierung einer TDOA-Zielfunktion, da eine direkte Minimierung in realen Anwendungen aufgrund zahlreicher Freiheitsgrade und lokaler Minima keine ausreichende Präzision liefert. Die in [Thr05] getroffene Näherung, dass das zu einem Ereignis korrespondierende Signal alle Sensoren unter demselben Einfallswinkel erreicht, erlaubt eine starke Vereinfachung der Zielfunktion. Dadurch lässt sich schlussfolgern, dass der Rang der Matrix, die die TDOA beinhaltet, dem Rang der Matrix der Sensor- bzw. Ereignispositionen entspricht. Somit ist eine Lösung des Kalibrierungsproblems durch eine Singulärwertzerlegung (engl. *singular value decomposition* (SVD)) und eine anschließende Optimierung mit neun Parametern möglich. Wenn die tatsächliche Anordnung mit der verwendeten Näherung übereinstimmt, also die Ereignisse sehr weit von den Mikrofonen entfernt sind (vgl. Abb. 2.1a), liefert bereits die Näherung ein zufriedenstellendes Ergebnis. Anderenfalls dient die Näherungslösung als Startwert für die Minimierung der ursprünglichen Zielfunktion. Bemerkenswert ist insbesondere, dass das Verfahren auch in Situationen, in denen die zur Approximation genutzte Annahme keinesfalls erfüllt ist (siehe Abb. 2.1b), z. T. noch gute Ergebnisse erzielt, wie eigene Untersuchungen gezeigt haben.

Anstatt die Lösung des Optimierungsproblems zunächst durch eine Rang-Approximation anzunähern, nutzt [PF14b] neben TDOA-Messungen zusätzlich Signaleinfallswinkel.

Voraussetzung dafür ist allerdings, dass ein Sensorknoten über mehr als ein Mikrofon verfügt. Somit eignet sich dieser Ansatz ausschließlich zur Inter-Array-Kalibrierung. Zur Optimierung der Zielfunktion dient anschließend ein iteratives Verfahren. Dazu wird basierend auf den Sensorpositionen und -orientierungen aus dem vorangegangenen Iterationsschritt sowie den Einfallswinkeln die Position der Ereignisse mittels Triangulation errechnet. Danach gestatten die Ereignispositionen und die TDOA-Messungen ein Update der Sensorpositionen bzw. -orientierungen.

Allerdings erfordert die Schätzung der TDOA zwischen den Mikrofonen eine Abtastsynchronisation der Sensorknoten. Diese ist u. a. bei der Verwendung von mehreren Smartphones nicht gegeben. Um dennoch mit TDOA-Schätzungen arbeiten zu können, ist zusätzlich die Bestimmung des zeitlichen Versatzes zwischen den verschiedenen Sensorknoten notwendig. Bei der Nutzung von Smartphones liefert bspw. der integrierte Kompass eine Orientierung, die zusammen mit dem bekannten Abstand zwischen Mikrofon und Lautsprecher des jeweiligen Smartphones die benötigten Zusatzinformationen bereitstellt, um neben den Sensorpositionen auch die zeitliche Anpassung zu ermitteln [HF11]. Alternativ dazu gestattet auch die Kombination von MDS und einer TDOA-Zielfunktion die Kalibrierung der Mikrofonpositionen, ohne dabei die Kenntnis der Orientierungen vorauszusetzen, dafür werden allerdings erneut aktive Sensorknoten benötigt [RKL05].

2.2.3 Positionsbasierte Kalibrierung

Sofern die verwendeten Sensorknoten aus mehreren Mikrofonen bestehen, wird außerdem eine individuelle Lokalisation der Ereignisse durch jeden Sensor möglich. Die Positionsschätzungen wiederum bilden die Grundlage für eine weitere Klasse von Kalibrierungsalgorithmen, die sich jedoch aufgrund der Voraussetzung, dass ein Sensorknoten aus einem Array bestehen muss, nur zur Inter-Array-Kalibrierung eigenen.

Eine Möglichkeit zur Lokalisierung der Signalquellen ist die Abtastung des Raumes durch jeden Sensorknoten mittels *Steered Response Power* (SRP) [DBA07]. Sobald die Positionen von mehreren Ereignissen vorliegen, kann durch die Bestimmung einer Koordinatentransformation zur Abbildung der Ereignispositionen aus dem Koordinatensystem eines Sensors auf die Ereignispositionen eines weiteren Sensors die Lage der Sensoren zueinander ermittelt werden [Val+10a]. Eine Alternative zur Lokalisierung mittels SRP-Techniken bildet die Messung der TDOA zwischen den Mikrofonen innerhalb eines Sensors und die anschließende Lösung eines *spherical least squares*-Problems [Val+10b]. Zusätzlich erläutert [Val+10b] auch eine Erweiterung der Schätzung von Koordinatentransformationen, die zur Berücksichtigung des anisotropen[2] Fehlers bei der Lokalisierung führt.

Die geometrische Anordnung der Mikrofone innerhalb der jeweiligen Arrays wurde bei den bisher dargestellten positionsbasierten Algorithmen als bekannt vorausgesetzt. [Hen+09] verwendet zunächst die bereits erläuterte Kohärenzfunktion von diffusem Rauschen in Kombination mit MDS [MLH08] zur Kalibrierung der Mikrofonpositionen innerhalb der Sensorknoten. Anschließend dient erneut ein SRP-Verfahren zur Bestimmung der Ereignispositionen in kartesischen Koordinaten. Diese wiederum werden

[2]Das Wort *anisotrop* kommt aus dem griechischen [Bro72a] und bezeichnet hier, dass der Fehler der Lokalisierung eine Vorzugsrichtung besitzt.

zur Schätzung einer Koordinatentransformation eingesetzt, aus der anschließend die Positionen der Mikrofongruppen hervorgehen.

Weiterhin lassen sich die Mikrofonarrays als akustische Kameras nutzen. Die Abtastung eines Raumes durch eine akustische Kamera liefert eine visuelle Darstellung der akustischen Signalquellen. Die visuelle Darstellung schafft die Voraussetzung für die Anwendungen von Ansätzen zur Kalibrierung von Kameranetzen. Dort dient bspw. die Epipolargeometrie zur Herstellung von geometrischen Beziehungen zwischen den Bildern verschiedener Kameras, die das gleiche Objekt zeigen [HZ04]. Die Realisierung akustischer Kameras durch einen *Delay-and-Sum Beamformer* (DSB) erlaubt somit die Wiederverwendung dieser Techniken zur Kalibrierung akustischer Sensornetze [Red+09].

2.2.4 Nicht-akustische Anwendungsbereiche

Die bisherige Übersicht der Geometriekalibrierungsalgorithmen beschränkt sich auf Verfahren die akustische Signale als Informationsquelle verwenden. Allerdings zeigt das letzte Beispiel, dass ursprünglich für andere Sensortypen entwickelte Techniken das Potenzial haben auch im akustischen Umfeld genutzt zu werden. Aufgrund der sehr unterschiedlichen Eigenschaften der verschiedenen Sensortypen gestaltet sich eine Übertragung der Strategien in eine andere Domäne insbesondere dann problematisch, wenn zur Kalibrierung explizit spezielle Eigenschaften des jeweiligen Sensortyps Berücksichtigung finden. Außerdem kommen im Unterschied zum akustischen Umfeld in anderen Anwendungsbereichen sehr viel häufiger aktive Sensoren zum Einsatz.

So dient in [BD10] eine aktive Kamera zur Erstellung von Panoramaaufnahmen. Anschließend wird durch Drehung, Neigung und Zoom der gemeinsame Bildausschnitt der Kameras maximiert, bevor abschließend die eigentliche Kalibrierung durch eine Bestimmung der Koordinatentransformationsparameter erfolgt. Obwohl aktive Kameras den Eingriff in den Kalibrierungsprozess gestatten, wird auch im visuellen Bereich auf Hilfsmittel, wie etwa spezielle LED-Konstruktionen zurückgegriffen [KLB08], um zuverlässigere Basispunkte für die Ermittlung der Koordinatentransformation zu erhalten.

Ein großer Anwendungsbereich von Geometriekalibrierungsverfahren ist die Positionsbestimmung von Funksensoren [Pat+05]. Dort verfügen die Sensorknoten meist nicht nur über einen Empfänger, sondern zusätzlich über einen Sender. Die Kenntnis der Sendeleistung sowie die Berücksichtigung eines Pfadverlustmodells erlaubt es, aus der gemessenen Empfangssignalstärke (engl. *received signal strength* (RSS)) die Entfernung zu schätzen [WY11]. Ebenfalls weit verbreitet ist der Einsatz von TOF-Messungen. Allerdings führen Abschattungen der Signale zu Situationen in denen keine direkte Signalkomponente (engl. *non-line-of-sight* (NLOS)) vorliegt, sodass Maßnahmen erforderlich sind, um die daraus resultierenden Messfehler explizit zu behandeln [Che99].

Eine Strategie, die im Bereich der akustischen Sensornetze wenig Anwendung findet, ist die Kalibrierung basierend auf Signaleinfallswinkeln. Bei der Kalibrierung sowohl von Funksensornetzen [SZW14] als auch von Infrarotsensoren [KL08] spielen sie jedoch eine wichtigere Rolle. Die Messung der DOA zu einer beweglichen Infrarotquelle, z. B. einer Person, ist ausreichend, um die Anordnung von Infrarotsensoren zu bestimmen [KWL08]. Bei der Kalibrierung von Funksensoren werden neben den Einfallswinkelmessungen zusätzlich Ankernknoten benötigt [SZW14]. Basierend auf den bekannten Positionen der Ankernknoten lassen sich anschließend die Positionen weiterer Sensoren ermitteln.

2.3 Ziele der Arbeit

Das Ziel dieser Arbeit ist die Entwicklung und Evaluierung von Algorithmen zur automatischen Geometriekalibrierung akustischer Sensornetze. Der vorangegangene Überblick über die bisher konzipierten Verfahren zeigt das breite Spektrum unterschiedlicher Ansätze zur Lösung des Kalibrierungsproblems. Diese Ansätze tragen schon jetzt zur Erleichterung der Inbetriebnahme von kompakten Arrays mit nur wenigen Sensoren [MLH08] bis hin zu sehr großen Installationen aus mehr als 1000 Mikrofonen [Wei+07] bei.

Um präzise Ergebnisse zu erzielen, kommen vorwiegend TOF- und TOA-Messungen zum Einsatz, die allerdings eine Zeitsynchronisation zwischen der Signalquelle und den Mikrofonen voraussetzen bzw. Signale mit speziellen Korrelationseigenschaften benötigten. Dank der Omnipräsenz von Smartphones kann zur Wiedergabe der Kalibrierungssignale auf diese zurückgegriffen werden [GKH14], sodass keine dedizierten Lautsprecher mehr erforderlich sind [Con+12]. Auch wenn durch den Einsatz von geeigneten Kalibrierungssignalen oder durch die Nutzung von TDOA-Messungen die Anforderung einer Zeitsynchronisation zwischen der Signalquelle und den Sensoren entfällt, ist dennoch eine Abtastsynchronisation der Mikrofone aller Sensorknoten notwendig.

Bei einem kompakten Array ist die Abtastsynchronisation zumeist unmittelbar gegeben, da alle Mikrofonsignale gemeinsam abgetastet werden. Allerdings ist eine Intra-Array-Kalibrierung in vielen Fällen gar nicht notwendig, da es sich bei einem kompakten Array meist um ein Werkstück handelt und daher die Positionen der Mikrofone schon durch den Fertigungsprozess vorliegen. Falls trotzdem eine Kalibrierung erforderlich ist, gestattet die Kohärenzfeldanalyse in Kombination mit der MDS [MLH08] sowie die darauf aufbauende Erweiterung [Vel+15], die zusätzlich GCCPhat integriert, die Kalibrierung. Mit einem maximalen Fehler von nur 0,02 m, bei moderatem Nachhall, erzielt die automatische Kalibrierung hier bereits eine ausreichende Präzision.

Diese Arbeit beschränkt sich deshalb auf räumlich verteilte Sensorknoten, die im Hinblick auf die potenziellen Anwendungsbereiche räumliche Filterung, Sprecherlokalisation und Telekonferenzsysteme die entscheidende Rolle spielen. Ferner wird nur der Teilbereich der Inter-Array-Kalibrierung berücksichtigt, da sich ein Sensornetz in den zuvor erwähnten Einsatzbereichen bspw. aus Laptops und/oder Tablets zusammensetzt, die schon heute zumeist ein Array aus zwei Mikrofonen besitzen. Auch modere Smartphones haben häufig ein zweites Mikrofon, dieses ist jedoch vorzugsweise zur Unterdrückung von Störgeräuschen geeignet, gleichwohl bietet ein Smartphone genügend Platz, um in zukünftigen Modellen auch ein Mikrofonarray zu integrieren.

Weiterhin kann aufgrund der räumlichen Trennung der Sensorknoten bei der Inter-Array-Kalibrierung eine Abtastsynchronisation zwischen den verschiedenen Knoten nicht mehr vorausgesetzt werden. Eine große Gruppe der etablierten Verfahren [HF11; GKH14; PMH11] verwendet deshalb aktive Sensorknoten und kombiniert Kalibrierung und Synchronisation in einem gemeinsamen Prozess. Im Gegensatz dazu bieten positionsbasierte Kalibrierungsalgorithmen [Val+10a; Hen+09; DBA07] den Vorteil, dass lediglich eine Abtastsynchronisation innerhalb der Sensorknoten benötigt wird und zwischen den Knoten des Netzes eine ungefähre Synchronisation ausreicht. Damit diese Verfahren verlässliche Positionsschätzungen aus den pro Sensorknoten durch-

geführten TDOA-Messungen erzielen, werden Arrays mit mehr als zehn Mikrofonen verwendet [Val+10a] oder Mikrofonabstände größer als 0,1 m benötigt [Hen+09]. Diese Anforderungen stehen jedoch im Widerspruch zu einer Miniaturisierung und möglichst kostengünstigen Realisierung der Sensorknoten.

Bei der Kalibrierung im Umfeld von nicht-akustischen Sensornetzen existieren außerdem Ansätze, die mit Einfallswinkeln (DOA) arbeiten und daher ebenfalls auf eine Abtastsynchronisation zwischen den Sensorknoten verzichten können [WY11; KWL08]. Gleichzeitig bietet die Verwendung von Einfallswinkeln bei akustischen Sensornetzen den Vorteil, dass auch kompakte Arrays mit nur wenigen Mikrofonen eine verlässliche Schätzung des Einfallswinkels gestatten [DJH15].

Das vorrangige Ziel dieser Arbeit besteht daher in der Entwicklung eines akustischen Geometriekalibrierungsverfahrens, das mithilfe von Einfallswinkelschätzungen arbeitet, die von kompakten Arrays mit möglichst wenigen Mikrofonen stammen. Durch den Einsatz von Einfallswinkeln soll, wie auch bei den positionsbasierten Ansätzen, eine ungefähre Synchronisation zwischen den Sensoren ausreichen, um somit auf die wesentlich schwieriger zu erlangende Abtastsynchronisation, die bspw. für eine TOA-Messung Voraussetzung ist, verzichten zu können.

Den Ausgangspunkt für die eigenen Entwicklungen bildet das bereits in Abschnitt 2.2.4 erwähnte Kalibrierungsverfahren [KWL08]. Die Ausgangssituation dieses zur Kalibrierung von Infrarotsensornetzen entwickelten Algorithmus zeigt große Parallelen zu den im Rahmen dieser Arbeit betrachteten akustischen Sensornetzen und stellt damit eine fundierte Basis für weitergehende Untersuchungen dar.

Darüber hinaus soll der Aufwand für die Durchführung der Kalibrierung ebenso wie die Anforderungen an die eingesetzten Sensorknoten möglichst gering ausfallen. Daher muss die Kalibrierung ohne zusätzliche Hilfsmittel, wie z. B. aktive Sensorknoten oder spezielle Kalibrierungssignale, erfolgen. Stattdessen soll die zu entwickelnde Prozedur die zur Kalibrierung benötigten Informationen aus einem Sprachsignal extrahieren. Die Nutzung von Sprachsignalen ermöglicht außerdem eine Kalibrierung des Sensornetzes während des eigentlichen Betriebs, sodass keine zusätzlichen Schritte durch einen Benutzer erforderlich werden. Allerdings besitzen Sprachsignale im Vergleich zu dedizierten Kalibrierungssignalen deutlich schlechtere Korrelationseigenschaften. Dementsprechend entsteht eine ungünstigere Ausgangssituation, insbesondere wenn Nachhall vorhanden ist. Zusätzlich treten in bestimmten Szenarien (vgl. Abb. 2.1b) Situationen auf, in den sich die Sensoren z. B. hinter dem Sprecher befinden, sodass nur reflektierte Anteile des Sprachsignals die betreffenden Sensoren erreichen. Da die Einfallswinkelschätzung jedoch die dominierende Richtung identifiziert sind somit Ausreißer zu erwarten. Deshalb muss die zu entwickelnde Vorgehensweise möglichst robust gegenüber Störungen der Einfallswinkelschätzungen sein.

Die durchgeführten Analysen werden jedoch ebenfalls zeigen, dass eine Kalibrierung durch Einfallswinkel häufig nur eine skalierungsunabhängige Bestimmung der Sensoranordnung gestattet und daher ein unbekannter Skalierungsfaktor zwischen dem Kalibrierungsergebnis und der tatsächlichen Anordnung verbleibt. Dadurch entsteht unmittelbar die Notwendigkeit zur Entwicklung zusätzlicher Techniken, um auch die Skalierung eindeutig festzulegen. Zunächst sollen ausschließlich akustische Lösungen betrachtet werden. Im Zentrum stehen allerdings modalitätsübergreifende Kalibrierungsverfahren, da im Umfeld von akustischen Sensornetzen meist auch Kameras vorhanden

sind. Als exemplarisches Beispiel dient hier erneut ein Telekonferenzsystem, das neben verteilten Mikrofonarrays zusätzlich über mehrere Kameras verfügt. Da im visuellen Bereich bereits eine sehr präzise Kalibrierung der Positionen möglich ist [KLB08], sollen die Positionen der Kameras als bekannt vorausgesetzt werden, um somit die notwendigen Zusatzinformationen zur Vollendung der akustischen Kalibrierung zu liefern. Weiterhin soll die Kombination von akustischen und visuellen Sensornetzen dazu beitragen, Synergien, wie etwa die Ausrichtung einer Kamera basierend auf einer akustischen Positionsschätzung, auszunutzen.

2.4 Erforderliche Kalibrierungsgenauigkeit

Im letzten Abschnitt wurden ausführlich die wissenschaftlichen Ziele dieser Arbeit erläutert. Die Frage wie groß der Kalibrierungsfehler ausfallen darf bzw. welche Genauigkeit bei der automatischen Geometriekalibrierung erforderlich ist, blieb dabei jedoch unbeantwortet. Zumal die Geometriekalibrierung eine Vorstufe für alle Verfahren darstellt, die die Kenntnis der Geometrie zur Erfüllung ihrer Aufgaben benötigen, wird die erforderliche Genauigkeit ebenfalls von diesen Algorithmen vorgegeben. Allgemein können Fehler bei Kalibrierung solange toleriert werden bis sie die Leistungsfähigkeit des darauf aufbauenden Algorithmus einschränken. Ggf. kann auch eine Reduktion der Leistungsfähigkeit akzeptiert werden, sofern das Resultat der Algorithmen, die die geometrische Anordnung verwenden, weiterhin den Anforderungen genügt.

Im Fokus dieser Arbeit steht die Entwicklung von Geometriekalibrierungsalgorithmen für räumlich verteilte Sensorknoten, die aus kompakten Mikrofonarrays bestehen (vgl. Abschnitt 2.3). Zu den potenziellen Anwendungsbereichen gehört insbesondere die Lokalisation von Sprechern und/oder akustischen Ereignissen zur Steuerung einer Kamera in einem Telekonferenzsystem. Daher sollen die Anforderungen, die sich aus diesem Szenario ergeben, als Basis für die Festlegung der erforderlichen Kalibrierungsgenauigkeit dienen.

Bei der Erfassung eines Sprechers durch eine Kamera ist ein ausreichend großer Bildausschnitt notwendig, um sicherzustellen, dass auch Gesten oder spontane Bewegungen innerhalb des Kamerabildes liegen. Sofern hier ein Toleranzbereich, der in etwa der Breite eines menschlichen Kopfes entspricht, zugrunde gelegt wird, sollte eine akustische Lokalisierung mit einem Fehler von bis zu 0,25 m ausreichen.

Anhand des akzeptierten Lokalisierungsfehlers soll nun geklärt werden, wie groß der Kalibrierungsfehler ausfallen darf. Grundlage für diese Untersuchung bilden zufällig erzeugte Sensorkonstellationen. Die Lokalisierung des Sprechers erfolgt durch die später in Abschnitt 6.2.2 erläuterte Triangulationsmethode. Ferner werden die zur Triangulation erforderlichen Einfallswinkel mithilfe eines Modells (siehe Abschnitt 5.6) generiert, dass den Fehler des im Rahmen dieser Arbeit verwendeten Winkelschätzers (siehe Abschnitt 4.4) in Abhängigkeit der Nachhallzeit (T_{60}) beschreibt. Darüber hinaus erfolgt die Modellierung eines Kalibrierungsfehlers durch die Überlagerung der tatsächlichen Sensorpositionen und -orientierungen mit mittelwertfreien, normalverteilten Störungen.

Eine graphische Darstellung des resultierenden Lokalisierungsfehlers in Abhängigkeit verschiedener Nachhallzeiten und Kalibrierungsfehler zeigt Abb. 2.3. Dabei kennzeichnen die Balken den Mittelwert des Lokalisierungsfehlers und die Antennen markieren die

Standardabweichung. Zur intuitiveren Interpretation des Positionierungsfehlers wird dieser nicht durch die Standardabweichung der normalverteilten Störungen, sondern durch die mittlere Entfernung des Sensors zur tatsächlichen Position beschrieben (ε_P). Beim Orientierungsfehler ist hingegen die Standardabweichung σ_Ori der Normalverteilung angeben.

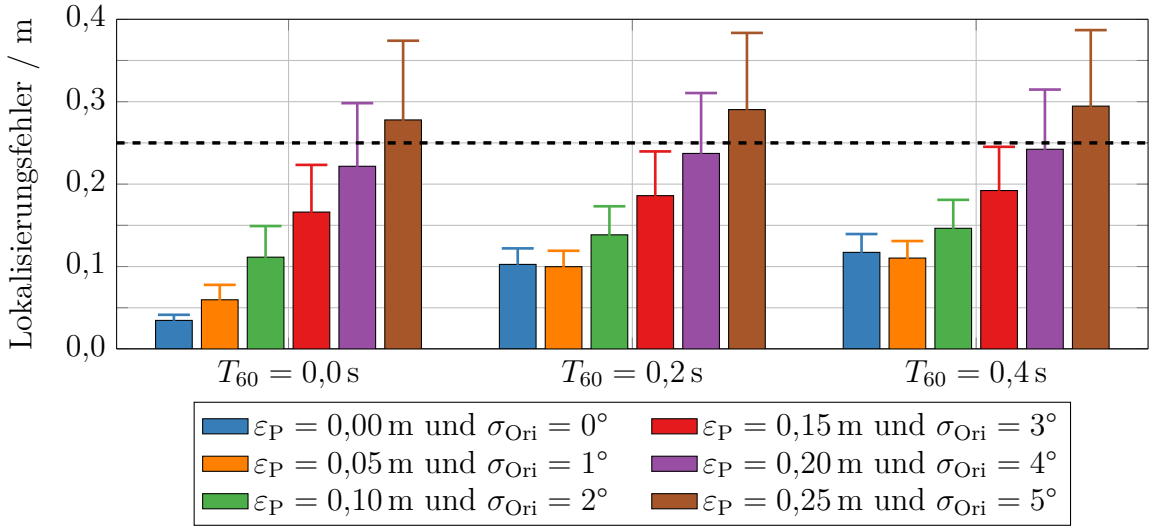


Abbildung 2.3: Auswirkung eines Positionierungs- und Orientierungsfehlers der Sensoren auf die akustische Lokalisation eines Sprechers. Die gestrichelte Linie kennzeichnet den maximal tolerierten Lokalisierungsfehler, wenn die Positionsschätzungen bspw. zur Ausrichtung einer Kamera dienen sollen.

Die Betrachtung der Ergebnisse aus Abb. 2.3 deutet auf eine deutliche Zunahme des Lokalisierungsfehlers mit ansteigendem Kalibrierungsfehler hin. Andererseits sorgt der Nachhall für Fehler bei der Einfallswinkelschätzung, die wiederum dazu führen, dass trotz der exakten Kenntnis der Sensorpositionen und -orientierungen ($\varepsilon_\mathrm{P} = 0{,}00\,\mathrm{m}$ und $\sigma_\mathrm{Ori} = 0°$) ein Lokalisierungsfehler von ca. 0,10 m auftritt. Sofern der Kalibrierungsfehler klein genug ausfällt ($\varepsilon_\mathrm{P} = 0{,}05\,\mathrm{m}$ und $\sigma_\mathrm{Ori} = 1°$) kann somit der Einfluss des Kalibrierungsfehlers gegenüber dem Lokalisierungsfehler vernachlässigt werden. Mit zunehmenden Kalibrierungsfehlern entsteht allerdings auch ein Anstieg des Lokalisierungsfehlers. Gemäß der vorangegangenen Ausführungen kann bei der Nutzung des Sensornetzes zur Ausrichtung einer Kamera ein Lokalisierungsfehler von bis zu 0,25 m toleriert werden. Ein Vergleich der erzielten Ergebnisse mit dem maximal akzeptierten Lokalisierungsfehler, der in Abb. 2.3 durch eine gestrichelte Linie dargestellt wird, dokumentiert, dass selbst bei einem Kalibrierungsfehler von 0,20 m und 4° die vorgegebene Grenze im Mittel unterschritten wird. Dementsprechend kann zur Steuerung von Kameras basierend auf einer akustischen Sprecherlokalisation ein Geometriekalibrierungsfehler von bis zu 0,20 m und 4° toleriert werden.

3 Grundlagen der Raumakustik

Bevor die Kalibrierung eines akustischen Sensornetzes beginnen kann, müssen zunächst die zur Kalibrierung notwendigen Informationen akquiriert werden. Die in Kapitel 4 erläuterte Auswahl eines geeigneten Verfahrens zur Einfallswinkelschätzung aus akustischen Signalen stellt dementsprechend eine zentrale Teilaufgabe bei der Entwicklung eines Systems zur automatischen Geometriekalibrierung akustischer Sensornetze dar.

Voraussetzung für die Auswahl eines Winkelschätzers und die spätere Evaluierung der Geometriekalibrierung sind entsprechende Audiosignale. Diese Signale wiederum sollten aus Räumen mit unterschiedlichen Reflexionseigenschaften stammen und möglichst viele verschiedene Konfigurationen aus Sprecher- und Arrayposition beinhalten, damit am Ende aussagekräftige Ergebnisse erzielt werden können. Verfügbare Datenbanken wie AV16.3 [LOG05], WSJCam0 [Fra+94], AIR [JSV09] oder MIRD [Had+14] bieten zwar mehrkanalige Aufnahmen bzw. Raumimpulsantworten zur Generierung dieser, allerdings spiegeln sie nur wenige Konfigurationen wider. Daher sind Simulationen der Aufnahmen für ein breites Spektrum verschiedener Szenarien unerlässlich.

Die Nutzung von simulierten Signalen zur Entwicklung und Evaluierung der Algorithmen sorgt jedoch auch dafür, dass die verwendete Simulationsmethode Auswirkungen auf die resultierenden Ergebnisse hat. Da außerdem einige Realisierungsaspekte der verwendeten Simulationstechniken signifikanten Einfluss auf die erzeugten Signale haben, erfolgt in Abschnitt 3.1 ein Exkurs in die Schallausbreitung in Räumen. Anschließend befasst sich Abschnitt 3.2 mit den daraus abgeleiteten Implementierungsdetails des zur Simulation von mehrkanaligen Sprachsignalen genutzten Algorithmus. Ziel dieser Beschreibungen ist es, die Differenzen verschiedener Realisierungen herauszuarbeiten und die Eigenschaften der genutzten Simulationsmethode umfassend zu charakterisieren. Dadurch soll ein Vergleich mit anderen Arbeiten ermöglicht werden, sofern dort die Details der genutzten Realisierung nicht im Unklaren bleiben.

3.1 Schallausbreitung in Räumen

Das von einer Quelle, bspw. einem Sprecher, ausgesandte Signal wird an den begrenzenden Wänden eines Raumes sowie den darin enthaltenen Einrichtungsgegenständen sowohl reflektiert als auch absorbiert. Das Verhältnis von absorbierter und reflektierter Energie hängt dabei vom jeweiligen Material ab. Die Überlagerung der reflektierten Signale bildet den Nachhall bzw. Hall. Sofern sich aufgrund der geometrischen Anordnung der Flächen und ihrer Reflexions- und Absorptionseigenschaften einzelne, deutlich wahrnehmbare, Reflexionen ausbilden, werden diese als Echo bezeichnet [CM03].

Für eine vollständige Beschreibung des Schallfeldes innerhalb des gesamten Raumes, inklusive der Effekte durch Beugung und Brechung, ist eine Berücksichtigung sämtlicher durch die Quelle angeregter Frequenzen und deren räumliche Ausbreitung notwendig. Eine solche Beschreibung erfolgt durch die *Acoustic Wave Equation* [FLS63]. Diese Differenzialgleichung 2. Ordnung charakterisiert die Ausbreitung des Schalldrucks als Funktion des Ortes und der Zeit. Angesichts der komplexen Geometrie typischer Räume und den zugehörigen Randbedingungen der Differenzialgleichungen sind diese meist nicht mehr analytisch lösbar [RNL09]. Deshalb erfolgt ihre Lösung vorwiegend numerisch durch die Finite-Elemente-Methode (FEM) [Ihl98]. Dabei wird der Raum in zahlreiche kleine Bereiche, die Finiten-Elemente, unterteilt, deren Wechselwirkungen durch die Parameter der Finiten-Elemente charakterisiert werden können. Allerdings ist die Rechenkomplexität der FEM durch die Vielzahl der benötigten Elemente beträchtlich. Daher eignet sich eine Beschreibung der akustischen Eigenschaften des Raumes durch Differenzialgleichungen vornehmlich zur detailgetreuen Modellierung eines einzelnen Raumes, nicht aber zur Erzeugung von vielen verschiedenen Szenarien.

Eine Alternative zur Beschreibung der Schallausbreitung mittels Differenzialgleichungen bildet die geometrische Akustik. Sie modelliert die Ausbreitung des Schalls als Strahlen und beschreibt deren Beugung, Brechung und Absorption [KSS68]. Dieser als *Ray-Traycing* bezeichnete Vorgang kommt auch im visuellen Bereich zur Beschreibung von Licht und Schatten in computergenerierten Grafiken zum Einsatz [RKM07]. Allerdings ist auch hier der Berechnungsaufwand für die Erzeugung von möglichst vielen verschiedenen Szenarien sehr groß.

Im Rahmen dieser Arbeit wird die Schallausbreitung deshalb mithilfe der statistischen Akustik modelliert. Ziel der statistischen Akustik ist es, eine statistische Aussage über das Schallfeld zu treffen [CM03] und somit die Berechnungen des Schallfeldes zu vereinfachen. Aus der Beschreibung der Schallausbreitung als Kugelwelle und der Modellierung in grober Näherung gleichmäßig ausgebildeter Absorptionseigenschaften des Raumes, kann die Energie einer Schallwelle mit der Anfangsenergie E_0, die sich im Raum ausbreitet, durch

$$E(t) = E_0 \cdot e^{-\frac{t}{\varsigma}} \tag{3.1}$$

dargestellt werden [CM03]. Die Dämpfungseigenschaften des Raumes ς werden dabei üblicherweise durch die Nachhallzeit ausgedrückt. Diese auch T_{60}-Zeit genannte Zeitspanne bezeichnet das Intervall, in dem der Schalldruckpegel um 60 dB abgefallen ist. Die Dämpfung ergibt sich aus der Nachhallzeit wie folgt:

$$\varsigma = -\frac{T_{60}}{6 \cdot \ln(10)}. \tag{3.2}$$

Eine experimentelle Bestätigung dieser Gesetzmäßigkeit bis zu einer gewissen Raumgröße liefert [SE94].

Große praktische Bedeutung hat jedoch die Herstellung eines Zusammenhangs zwischen Raumgeometrie, den Absorptionseigenschaften der Flächen und der daraus resultierenden Nachhallzeit. Dieser Zusammenhang ist insbesondere im folgenden Abschnitt notwendig, um bei einer Simulation für eine gegebene Raumgeometrie und Nachhallzeit die Absorption der Flächen passend einzustellen. Zur Charakterisierung der Absorptionseigenschaften einer Fläche dient der Absorptionskoeffizient, der das Verhältnis der absorbierten Energie zur Gesamtenergie beschreibt.

Es existieren zahlreiche Arbeiten, die experimentelle oder theoretische Näherungen für diese Beziehung liefern [Sab22; Eyr30; Mil32; Fit59; Ara88; NK01; LJ08]. Für den Spezialfall quaderförmiger Räume mit gleichmäßig absorbierenden Wänden, liefern alle Approximationen übereinstimmende Ergebnisse. Diese Annahme trifft aber nicht auf reale Räume zu, da bspw. ein Teppichboden den Schall stärker absorbiert als eine Fensterfront. Daher ist hauptsächlich der Fall nicht gleichmäßig absorbierender Flächen relevant. Dort liefern die Näherungen jedoch deutlich unterscheidbare Absorptionskoeffizienten. Abb. 3.1 stellt den Absorptionskoeffizienten einer ausgewählten Wand in Abhängigkeit der Nachhallzeit für einen $6{,}00 \times 8{,}00 \times 3{,}00\,\mathrm{m}^3$ großen Raum dar, bei dem Boden und Decke ca. 30 % schwächer als alle anderen Flächen reflektieren.

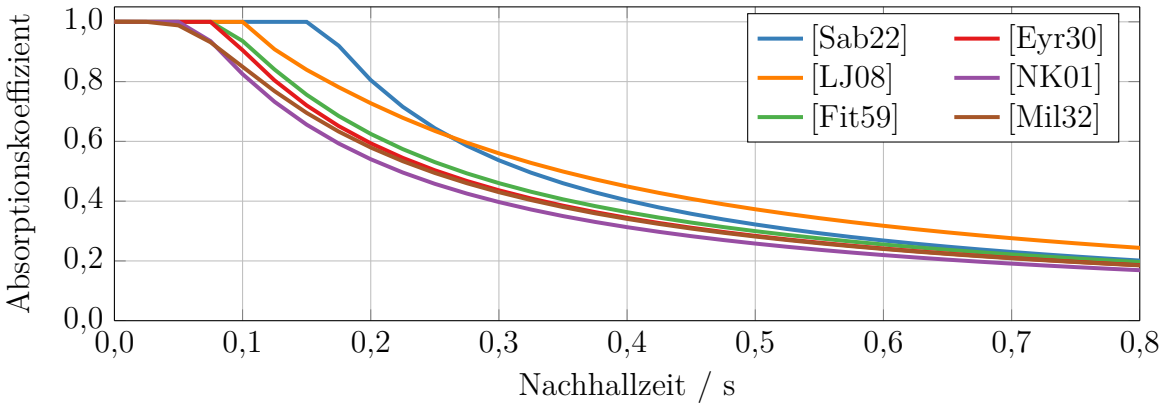


Abbildung 3.1: Vergleich verschiedener Näherungen zur Bestimmung der Absorptionskoeffizienten bei inhomogenen Reflexionseigenschaften der Wände.

Weiterhin ist in vielen Fällen keine vollständige Beschreibung des Schallfeldes im gesamten Raum nötig, sondern lediglich an dem Ort, an dem sich der Zuhörer bzw. das Mikrofon befindet. Das an diesem Ort eintreffende Signal lässt sich durch eine Raumimpulsantwort (RIA), welche spezifisch für eine Sender- und Empfängeranordnung sowie den Raum ist, vollständig beschreiben. Eine solche RIA, gemessen in einem $10{,}80 \times 10{,}90 \times 3{,}15\,\mathrm{m}^3$ großen Hörsaal ist in Abb. 3.2 dargestellt [JSV09].

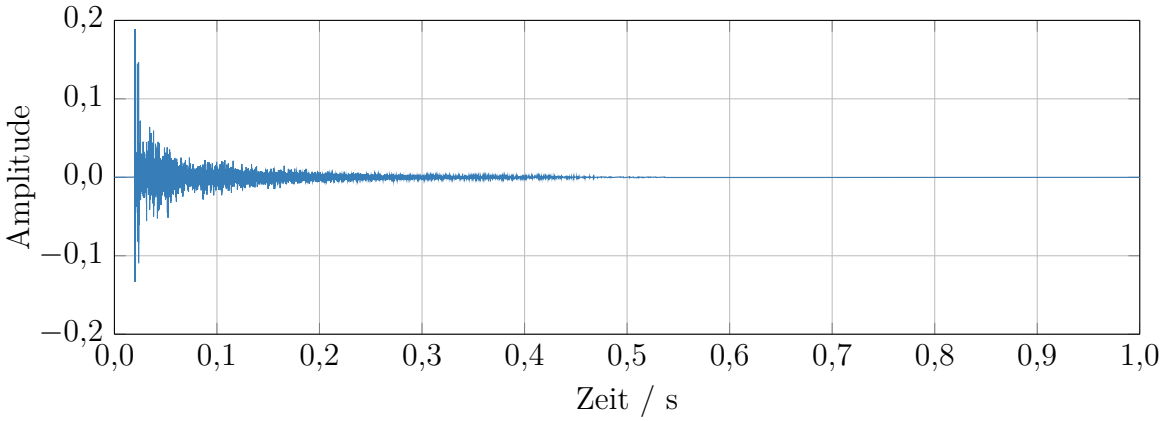


Abbildung 3.2: Beispiel einer RIA [JSV09].

Bereits in dieser Darstellung ist das in der statistischen Akustik modellierte exponentiell abklingende Verhalten (vgl. Gl. (3.1)) erkennbar. Deutlicher wird dieses Verhalten durch die Darstellung der *Energy Decay Curve* (EDC) [Sch65], die sich aus dem Verhältnis zwischen der zum Zeitpunkt t verbleibenden Energie der RIA $h(t)$ und der Gesamtenergie dieser ergibt:

$$\mathrm{EDC}(t) = 10 \cdot \log_{10}\left(\int_t^\infty h^2(t')\,\mathrm{d}t' \Big/ \int_0^\infty h^2(t')\,\mathrm{d}t'\right). \tag{3.3}$$

Die EDC der zuvor betrachteten Raumimpulsantwort ist in Abb. 3.3 visualisiert.

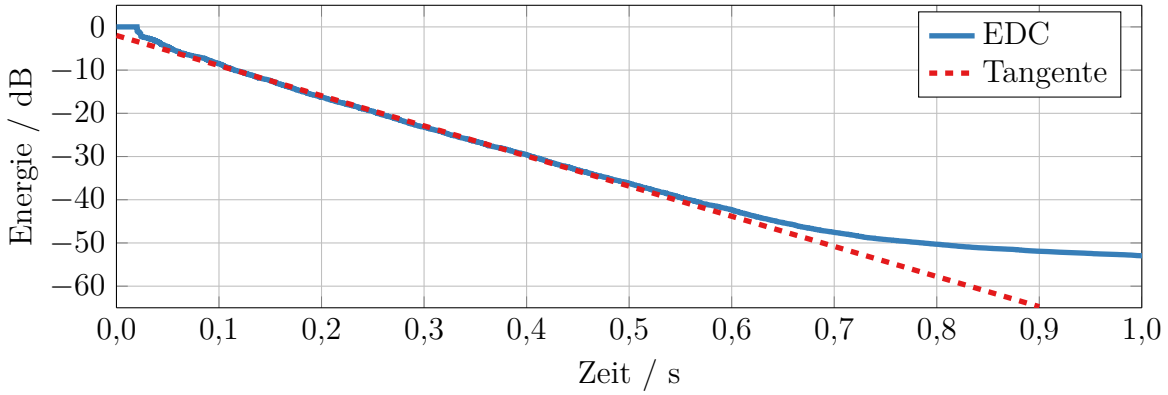


Abbildung 3.3: EDC zur RIA aus Abb. 3.2.

Der lineare Verlauf der EDC unterstreicht noch einmal das exponentielle Abklingen der Energie der RIA. Die EDC bildet auch die Grundlage zur Schätzung der Nachhallzeit mithilfe der Schroeder-Methode [Sch65]. Diese sieht die Berechnung einer Tangenten zur EDC und die anschließende Bestimmung des Punktes, an dem diese Tangente das $-60\,\mathrm{dB}$-Level erreicht, vor. Eine genauere Betrachtung der EDC aus Abb. 3.3 zeigt jedoch auch, dass eine Linearisierung durch eine Tangente erst nach dem Eintreffen der frühen Reflexionen und nicht für sehr späte Reflexionen geeignet ist. Insofern hängt das Ergebnis der Nachhallzeit-Schätzung von der Wahl des Bereiches, für den die Tangente ermittelt wird, ab. Die Tangente in Abb. 3.3 entsteht durch die in [Sch65] vorgeschlagene Linearisierung der EDC im Intervall von $[-10\,\mathrm{dB}; -40\,\mathrm{dB}]$ und liefert eine Nachhallzeit von $0{,}85\,\mathrm{s}$. Allerdings variiert der Bereich der zur Linearisierung verwendet wird [LV08], sodass sich Angaben der Nachhallzeit deutlich unterscheiden können.

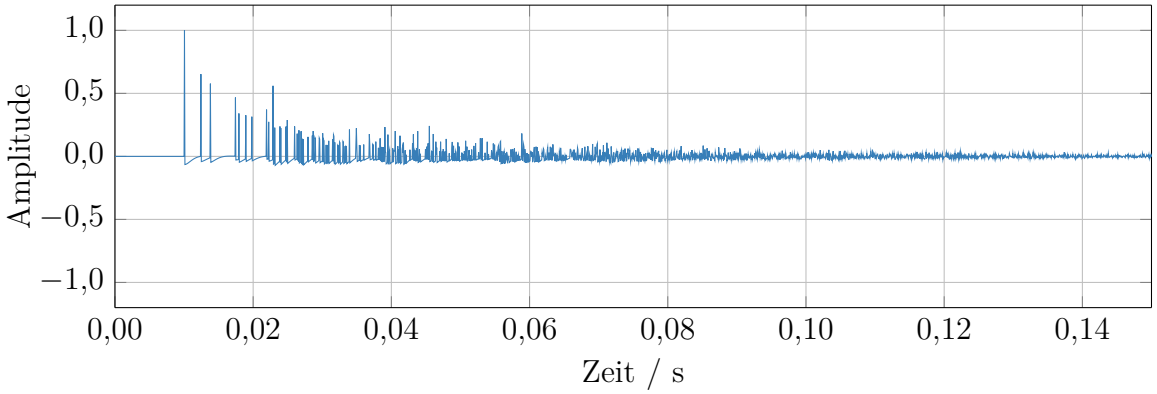
3.2 Simulation von Raumimpulsantworten

Das im zurückliegenden Abschnitt beschriebene Konzept der Raumimpulsantwort bietet ein kraftvolles Instrument zur Beschreibung der wichtigsten Eigenschaften der Schallübertragung von einer Quelle zur Senke. Zur Simulation der RIA existieren verschiedene Möglichkeiten. Wie bereits zuvor erwähnt, ist die Rechenkomplexität der FEM oder des *Ray-Traycings* sehr groß, sodass sich diese Verfahren nicht zur Generierung von Raumimpulsantworten für die Evaluierung von akustischen Signalverarbeitungsalgorithmen eignen. Daher werden zur Untersuchung von Algorithmen wie z. B. zur

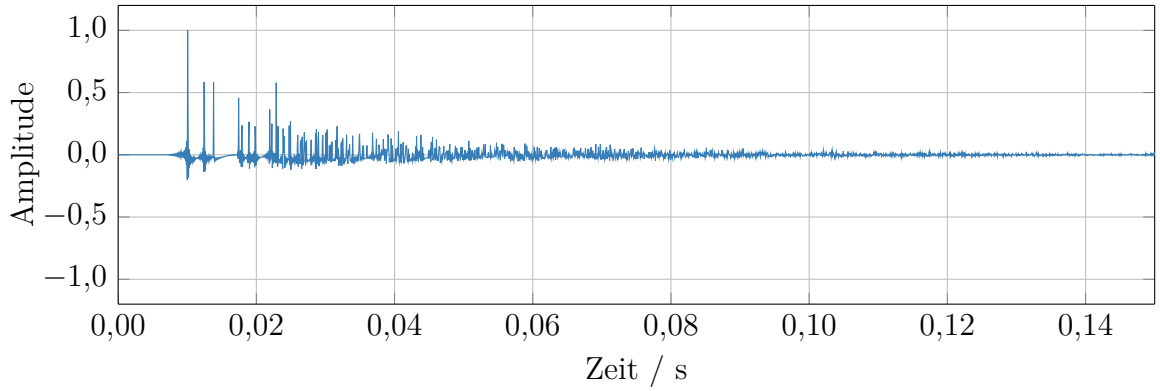
Sprecher-Lokalisation und Verfolgung [SH10], Einfallswinkelschätzung [Ara+11; DJH15], Spracherkennung [LKH14] und Quellen-Trennung [TH10], abgesehen von realen Aufnahmen fast ausschließlich mit der Spiegel-Quellen-Methode (engl. *image method* oder *image source method*) [AB79] erzeugte RIA genutzt. Da jedoch Realisierungsaspekte der Spiegel-Quellen-Methode signifikante Auswirkungen auf die resultierenden RIA haben, soll dieser Abschnitt die Unterschiede der verschiedenen Varianten beleuchten und verdeutlichen, dass ein Vergleich mit anderen Arbeiten in vielen Fällen angesichts der fehlenden Kenntnis der notwendigen Details nur eingeschränkt möglich ist.

Die Spiegel-Quellen-Methode verwendet zur Modellierung der RIA von einer Quelle zur Senke sogenannte Spiegel-Quellen. Für jede der Reflexionen an den umgebenden Wänden, die gemeinsam den Nachhall erzeugen, wird eine Spiegel-Quelle konstruiert. Die räumlichen Positionen dieser Spiegel-Quellen ergeben sich durch die Spiegelung der Quellposition an der reflektierenden Wand. Der Signalpfad von einer Spiegel-Quelle zur Senke entspricht dem Pfad, der sich aufgrund der Reflexion an der betreffenden Wand ergibt. Durch weitere Spiegelungen der Spiegel-Quellen lassen sich auch die Reflexionen an mehr als einer Wand beschreiben. Letztlich entsteht die RIA durch die Überlagerung von zahlreichen Impulsen, deren Position durch die Verzögerungen der reflektierten Signale, bezogen auf den direkten Pfad von der Quelle zur Senke, gegeben ist. Das Gewicht dieser Impulse wird aus dem Produkt der Reflexionskoeffizienten der Wände an denen das jeweilige Signal reflektiert wurde errechnet. Die dazu notwendigen Reflexionskoeffizienten beschreiben das Verhältnis der reflektierten Energie zur Gesamtenergie für die jeweiligen Wände. Dementsprechend lässt sich ein Reflexionskoeffizient durch die Subtraktion des zugehörigen Absorptionskoeffizienten vom Wert Eins berechnen.

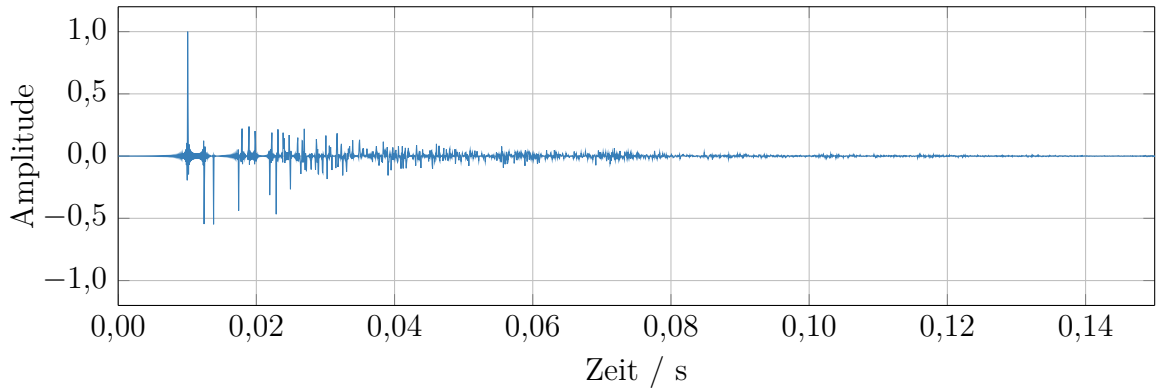
Zur Veranschaulichung der Realisierungsaspekte dient ein kurzer Vergleich ausgewählter Implementierungen. Dieser Vergleich umfasst erstens die bereits der initialen Veröffentlichung [AB79] von Allen und Berkley beiliegende Fortran-Realisierung, zweitens eine von Habets stammende C++-Implementierung mit MATLAB®-Interface [Hab10], drittens eine MATLAB®-Variante von Lehmann und Johansson [LJ10] sowie viertens eine eigene Implementierung. Um die Unterschiede hervorzuheben, wird die RIA eines $8{,}00 \times 6{,}00 \times 3{,}00\,\mathrm{m}^3$ großen Raums, mit der Quellposition $[5{,}30\ \ 3{,}35\ \ 0{,}90]^\mathrm{T}\,\mathrm{m}$, einer Senke bei $[2{,}34\ \ 1{,}78\ \ 1{,}75]^\mathrm{T}\,\mathrm{m}$ sowie einer Nachhallzeit von 0,3 s berechnet. Die RIA für die ausgewählten Implementierungen, die sich für das beschriebene Szenario ergeben, sind in Abb. 3.4 illustriert.



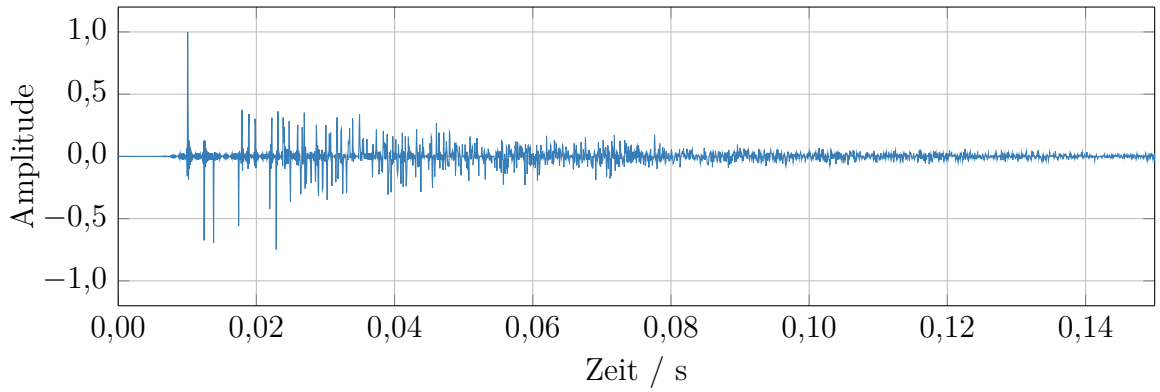
(a) Realisierung von Allen und Berkley [AB79].



(b) Realisierung von Habets [Hab10].



(c) Realisierung von Lehmann und Johansson [LJ10].



(d) Eigene Realisierung.

Abbildung 3.4: Vergleich verschiedener Implementierungen zur Simulation einer RIA.

Bei der Betrachtung der Raumimpulsantworten zeigen sich deutliche Unterschiede. Das unterschiedlich stark ausgeprägte, abklingende Verhalten, lässt sich durch die Nutzung verschiedener Approximationen zur Bestimmung der Reflexionskoeffizienten erklären (vgl. Abb. 3.1). Allerdings hat die Approximation nur Einfluss auf den Betrag bzw. das Abklingen der Amplitude und kann somit nicht allein für die Differenzen verantwortlich sein. Die Abweichungen zwischen den Realisierungen entstehen im Wesentlichen durch die unterschiedliche Modellierung der Spiegel-Quellen. Allen und Berkley verwen-

den dazu diskrete Impulse, wohingegen alle neueren Varianten sinc-Funktionen nutzen. Der Unterschied zwischen der Variante von Lehmann und Johansson bzw. der eigenen Realisierung zu den Implementierungen von Allen und Berkley sowie Habets, besteht im Auftreten von Impulsen mit negativen Gewichten. Diese Gewichte modellieren, dass es bei Auftreffen der Schallwelle auf einer reflektierenden Fläche zu einer Phaseninversion der Schallwelle kommt. Da Allen und Berkley sowie Habets diesen Effekt nicht berücksichtigen, beinhalten die Verfahren zusätzlich eine Hochpassfilterung der RIA, um den anderenfalls auftretenden, anwachsenden Gleichanteil zu unterdrücken. Die Realisierung von Lehmann und Johansson sowie die eigene Variante enthalten keinen Hochpass, weil bereits die Berücksichtigung der Phaseninversion und die daraus resultierenden positiven und negativen Impulse das Auftreten eines Gleichanteils verhindern. Der verbleibende Unterschied zwischen der eigenen Realisierung und der von Lehmann und Johansson entsteht durch die Nutzung einer Taylor-Reihen-Approximation der sinc-Funktion bei der eigenen Implementierung, zugunsten einer Reduktion der Rechenkomplexität.

Bei der akustischen Signalverarbeitung erfolgt eine Evaluierung häufig für verschiedene Nachhallzeiten, da sich der Nachhall negativ auf die Leistungsfähigkeit des Systems auswirkt. Die Darstellung der EDC in Abb. 3.5 zu den RIA aus Abb. 3.4, weist jedoch ein deutliches Missverhältnis zwischen den resultierenden Nachhallzeiten auf.

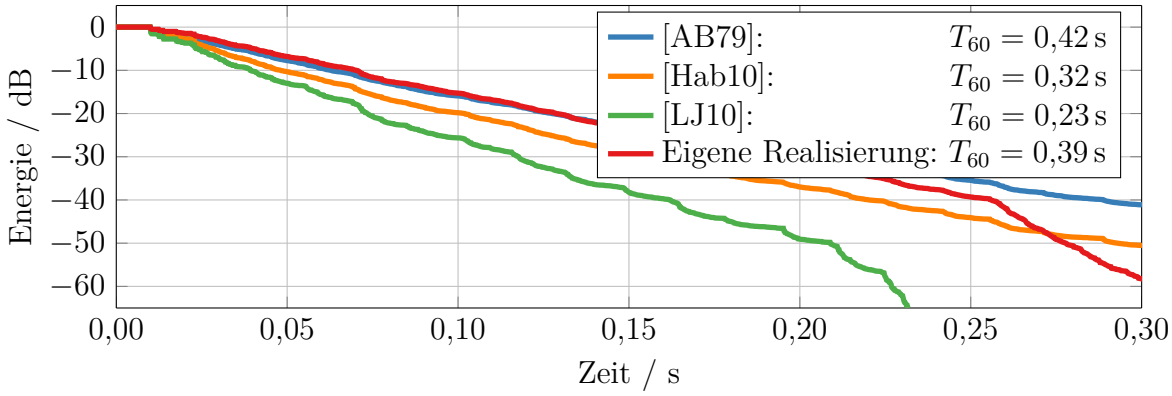


Abbildung 3.5: EDC der Raumimpulsantworten aus Abb. 3.4, inklusive der Schätzungen der Nachhallzeit unter Verwendung der Schroeder-Methode [Sch65].

Die Realisierung von Habets besitzt mit einer Nachhallzeit von 0,32 s die geringste Abweichung zu den eingestellten 0,30 s. Aber sie unterscheidet sich damit deutlich von der in der Regel zitierten Referenzimplementierung von Allen und Berkley. Eine Verwendung der Referenzimplementierung scheidet jedoch aus, da diese nur ganzzahlige Vielfache der Signalabtastrate als Verzögerungen zwischen verschiedenen Raumimpulsantworten abbilden kann. Somit ist das Verfahren zur Erzeugung von mehrkanaligen Audiosignalen, die anschließend zur Winkelschätzung dienen sollen, ungeeignet. Die eigene Variante liefert hingegen ähnliche Nachhallzeiten, wie die Referenzimplementierung. Im weiteren Verlauf soll deshalb stets die eigene Implementierung Verwendung finden, da diese der Variante von Allen und Berkley am nächsten kommt und bezüglich der anderen Varianten als *Worst-Case* betrachtet werden kann. Abschließend sei noch darauf hingewiesen, dass es sich bei Abb. 3.5 und den daraus resultierenden Werten um eine Momentaufnahme handelt, deren grundsätzliche Tendenz aber auch für andere Szenarien gilt.

3.3 Zusammenfassung

Akustische Signale stellen einen wichtigen Bestandteil dieser Arbeit dar. Daher wurden in diesem Kapitel zunächst einige Grundlagen der Schallausbreitung erläutert und Konzepte zur Modellierung dieser vorgestellt. Eine vollständige Beschreibung des Schallfeldes innerhalb eines Raumes erfordert die Nutzung von Differenzialgleichungen, deren Berechnung durch die FEM sehr aufwändig ist. Darüber hinaus ist eine vollständige Beschreibung des Schallfeldes für die im Rahmen dieser Arbeit betrachteten Fragestellungen nicht notwendig. Im Zentrum stand deshalb die Modellierung der Schallausbreitung durch eine RIA.

Die wichtigste Kenngröße zur Charakterisierung der auftretenden Reflexionen stellt dabei die Nachhallzeit (T_{60}-Zeit) dar. Eine Messung der Nachhallzeit ist mithilfe der EDC möglich, allerdings können sich durch unterschiedliche Linearisierungen deutlich variierende Angaben der Nachhallzeit ergeben.

Im zweiten Abschnitt dieses Kapitels wurde außerdem die Spiegel-Quellen-Methode zur Simulation von RIA vorgestellt. Diese bildet die Reflexionen des Signals an den Wänden durch sogenannte Spiegel-Quellen nach, deren Position durch eine Spiegelung der ursprünglichen Quelle an der betreffenden Wand entstehen. Allerdings existieren verschiedene Realisierungen, die die Spiegel-Quellen unterschiedlich implementieren. Der Vergleich von vier ausgewählten Implementierungen hat gezeigt, dass sich die Nachhallzeiten der erzeugten RIA unterscheiden. Daher hat die verwendete Realisierung der Spiegel-Quellen-Methode Einfluss auf die Audiosignale und damit auch auf das Ergebnis der Algorithmen, die mithilfe der erzeugten Signale untersucht werden. Dementsprechend ist bei einem Vergleich verschiedener Arbeiten stets auch die Berücksichtigung der verwendeten Simulationsmethode und der sich daraus ergebenden Eigenschaften notwendig.

4 Einfallswinkelschätzung

Die Realisierungen eines Algorithmus zur Geometriekalibrierung akustischer Sensornetze mithilfe von Einfallswinkeln erfordert als erstes eine Schätzung der Winkel aus den akustischen Signalen. Das primäre Ziel dieses Kapitels besteht daher in der Analyse verschiedener Winkelschätzer und der anschließenden Auswahl des für die Geometriekalibrierung am besten geeigneten Verfahrens.

Aufgrund des geplanten Anwendungsbereiches sind ausschließlich Algorithmen relevant, die die Winkelschätzung aus einem Sprachsignal beim Einsatz eines kompakten Arrays gestatten. Trotz dieser Beschränkungen existieren zahlreiche Methoden. Ein Grund für die Methodenvielfalt liegt auch darin, dass die DOA-Schätzung nicht nur im hier betrachteten akustischen Umfeld von Bedeutung ist, sondern auch bei anderen Signalklassen, wie z. B. bei den Funksignalen und die dort entwickelten Verfahren ebenfalls im akustischen Bereich Anwendung finden.

In Abschnitt 4.1 werden deshalb lediglich einige der bekanntesten Ansätze zur Einfallswinkelschätzung präsentiert, um einen Einblick in deren Funktionsweise zu geben. Zu den betrachteten Algorithmen gehören sowohl Korrelationsverfahren, wie GCC-Phat [KC76] und *Steered Response Power with Phase Transform* (SRPPhat) [DSB01], als auch moderne, modellbasierte Winkelschätzer, wie [LY10].

Im Anschluss daran befasst sich Abschnitt 4.2 mit der Entwicklung eines weiteren modellbasierten Winkelschätzers. Als Abstandsmaß zwischen den Beobachtungen und dem Modell verwendet der vorgestellte Algorithmus die komplexe Watson-Verteilung. Darüber hinaus ist der Ansatz im Gegensatz zu konventionellen Verfahren, die ausschließlich die Phasendifferenzen berücksichtigen, in der Lage, auch Amplitudendifferenzen mit in die Winkelschätzung einzubeziehen. Ferner sollen die Untersuchungen in Abschnitt 4.3 die Leistungsfähigkeit des entwickelten Algorithmus unter Beweis stellen und klären, inwieweit die Mitberücksichtigung der Amplitudendifferenzen bei der Nutzung von gerichteten Mikrofonen zu einer präziseren DOA-Schätzung beitragen kann.

Danach erfolgt in Abschnitt 4.4 ein Vergleich aller Algorithmen, um den Winkelschätzer auszuwählen, der bei der Nutzung eines kompakten Arrays die besten Ergebnisse erzielt und dementsprechend ideale Voraussetzungen für die Geometriekalibrierung bietet. Allerdings werden die Analysen bei allen untersuchten Schätzern einen systematischen Fehler bei der Bestimmung der DOA aufzeigen, sofern das eingesetzte Array nur aus zwei Mikrofonen besteht. Daher sollen weitere Untersuchungen in Abschnitt 4.5 die Ursachen dieses Phänomens ergründen. Zum Abschluss wird in Abschnitt 4.6 noch ein statistisches Modell für die Verteilung des Fehlers der Winkelschätzung entwickelt. Dieses wiederum dient im weiteren Verlauf der Arbeit schließlich zur Entwicklung eines optimalen Geometriekalibrierungsalgorithmus.

4.1 Existierende Winkelschätzer

Voraussetzung für alle Winkelschätzer ist zunächst ein Sensorknoten, der über mehr als ein Mikrofon verfügt, weil die Schätzung des Einfallswinkels mithilfe der Phasendifferenzen zwischen den verschiedenen Mikrofonsignalen erfolgt [Sch86; Bar83; DSB01; SH10; LY10; DJH15]. Die Grundlage für alle Schätzer bilden die von den Mikrofonen des Sensorknotens aufgenommenen akustischen Signale. Da die gesamte Verarbeitung der Signale zeitdiskret erfolgt, werden diese unmittelbar als Funktion des diskreten Zeitindexes l der im zeitlichen Abstand von $1/f_\mathrm{s}$ gewonnenen Abtastwerte ausgedrückt. Unter Berücksichtigung des zuvor erläuterten Konzeptes der RIA, lässt sich das vom m-ten Mikrofon empfangene Signal $x_m(l)$ durch folgendes Modell beschreiben:

$$x_m(l) = s(l) * (h_{m,\mathrm{direkt}}(l) + h_{m,\mathrm{Hall}}(l)) + n_m(l). \tag{4.1}$$

Dabei bezeichnet $s(l)$ das Quellsignal, $h_{m,\mathrm{direkt}}(l)$ beinhaltet den direkten Anteil der RIA, $h_{m,\mathrm{Hall}}(l)$ beschreibt die durch Reflexionen verursachte Mehrwegeausbreitung und $n_m(l)$ modelliert Störungen, wie etwa das Rauschen der Mikrofone.

Zur Schätzung des Einfallswinkels werden die M Signale eines Sensorknotens, $m = 1, \ldots, M$, zunächst mithilfe der Kurzzeit-Fourier-Transformation (engl. *short-time* Fourier *transform* (STFT)) im Frequenzbereich dargestellt. Dazu erfolgt eine Segmentierung des Signals $x_m(l)$ in überlappende Blöcke der Länge L. Anschließend wird jeder Block mit einer Fenster-Funktion $\omega(l)$ gewichtet und mit der diskreten Fourier-Transformation (DFT) in den Frequenzbereich transformiert [BSH07]. Somit entsteht aus dem Signal $x_m(l)$ schließlich die Repräsentation im STFT-Bereich

$$X_m(\iota, k) = \sum_{l=0}^{L-1} \omega(l) \cdot x_m(l - \iota L) \cdot e^{-\mathrm{j}\frac{2\pi}{L}lk}, \tag{4.2}$$

wobei ι den ι-ten Block und k die diskrete Frequenz $f_k = kf_\mathrm{s}/L$, $k = 0, \ldots, L/2$, kennzeichnet.

Durch die Einführung einer Vektornotation

$$\mathbf{x}(\iota, k) = \begin{bmatrix} X_1(\iota, k) & \ldots & X_M(\iota, k) \end{bmatrix}^\mathrm{T}, \tag{4.3}$$

die analog auch für alle anderen Signale aus Gl. (4.1) definiert ist, kann das Signalmodell aus Gl. (4.1) im STFT-Bereich durch eine *Multiplicative Transfer Function* (MTF) [AC07] approximiert werden:

$$\mathbf{x}(\iota, k) = \boldsymbol{h}_\mathrm{direkt}(\iota, k)\, S(\iota, k) + \tilde{\mathbf{n}}(\iota, k). \tag{4.4}$$

Die Rauschkomponente $\tilde{\mathbf{n}}(\iota, k)$ schließt dabei sowohl die durch $n_m(l)$ modellierten Störungen als auch die von der Mehrwegeausbreitung verursachten Anteile ein, da für die im weiteren Verlauf erläuterten Verfahren zur Bestimmung des Einfallswinkels nur die direkte Komponente des Signals genutzt wird.

Die Übertragungsfunktion des direkten Pfades $\boldsymbol{h}_\mathrm{direkt}(\iota, k)$ beschreibt einerseits den Phasenversatz, der proportional zu der vom Signal zurückgelegten Strecke ist und andererseits eine Dämpfung. Diese setzt sich aus dem Pfadverlust aufgrund der zurückgelegten Strecke und der Dämpfung der Mikrofone zusammen. Zur Modellierung ist es

jedoch ausreichend die relativen Verzögerungen zwischen den jeweiligen Ausbreitungspfaden, die wiederum durch die Signallaufzeitdifferenz charakterisiert werden und die relativen Pfadverluste zu verwenden. Die Betrachtung der absoluten Größen ist nicht notwendig, weil diese ebenso auch als Phase bzw. Amplitude des Quellsignals aufgefasst werden können. Dementsprechend lässt sich die direkte Komponente durch

$$\boldsymbol{h}_{\text{direkt}}(\iota, k) = \begin{bmatrix} 1 & \tilde{h}_{(2,1)} \cdot e^{-\mathrm{j}2\pi f_k \tau_{(2,1)}(\iota)} & \dots & \tilde{h}_{(M,1)} \cdot e^{-\mathrm{j}2\pi f_k \tau_{(M,1)}(\iota)} \end{bmatrix}^{\mathrm{T}} \tag{4.5}$$

darstellen. Dabei bezeichnet $\tilde{h}_{(m,1)}$ die Dämpfung auf dem Pfad zum m-ten Mikrofon im Verhältnis zum ersten Mikrofon und $\tau_{(m,1)}(\iota)$ analog dazu die TDOA zwischen dem m-ten und ersten Mikrofon im ι-ten STFT-Block. Außerdem sind die von der Mehrwegeausbreitung stammenden Anteile näherungsweise unkorreliert zu den direkten Pfaden $\boldsymbol{h}_{\text{direkt}}(\iota, k)$, sodass auch die Terme $\boldsymbol{h}_{\text{direkt}}(\iota, k)\, S(\iota, k)$ und $\tilde{\mathbf{n}}(\iota, k)$ in Gl. (4.4) als unkorreliert betrachtet werden können [LBD01].

Ausgangspunkt für alle im Folgenden betrachteten Verfahren zur Schätzung des Einfallswinkels ist stets Gl. (4.4). Eine Möglichkeit bietet das ursprünglich für Funksignale entworfene Verfahren *Multiple Signal Classification* (MUSIC) [Sch86]. Sofern omnidirektionale Mikrofone vorliegen, kann der Einfluss der Dämpfung ($\tilde{h}_{(m,1)}$) vernachlässigt werden, da sich die Dämpfungen aufgrund der von den Signalen zurückgelegten Distanzen kaum unterscheiden. Somit lässt sich der direkte Signalpfad $\boldsymbol{h}_{\text{direkt}}(\iota, k)$ auch durch den *Steering-Vector*

$$\boldsymbol{\alpha}(k, \boldsymbol{\tau}_1(\iota)) = \begin{bmatrix} 1 & e^{-\mathrm{j}2\pi f_k \tau_{(2,1)}(\iota)} & \dots & e^{-\mathrm{j}2\pi f_k \tau_{(M,1)}(\iota)} \end{bmatrix}^{\mathrm{T}} \tag{4.6}$$

approximieren, der lediglich von der Signallaufzeitdifferenz $\tau_{(m,1)}(\iota)$ abhängt. Die Signallaufzeitdifferenzen wiederum bilden den Vektor

$$\boldsymbol{\tau}_1(\iota) = \begin{bmatrix} \tau_{(2,1)}(\iota) & \dots & \tau_{(M,1)}(\iota) \end{bmatrix}^{\mathrm{T}}. \tag{4.7}$$

Die Kenntnis der geometrischen Anordnung der Mikrofone innerhalb des Sensorknotens gestattet es, zusammen mit der Fernfeldnäherung den *Steering-Vector* als Funktion des Einfallswinkels $\varphi(\iota)$ anzunähern:

$$\boldsymbol{\alpha}(k, \varphi(\iota)) \approx \boldsymbol{\alpha}(k, \boldsymbol{\tau}_1(\iota)). \tag{4.8}$$

Zur Bestimmung des Winkels aus den Mikrofonsignalen $\mathbf{x}(\iota, k)$ nutzt MUSIC das Kreuzleistungsdichtespektrum (KLDS), das sich mithilfe des Erwartungswertoperators $\mathbb{E}[\cdot]$ zu

$$\boldsymbol{\Phi}_{\mathbf{xx}}(\iota, k) = \mathbb{E}\Big[\mathbf{x}(\iota, k)\, \mathbf{x}^{\mathrm{H}}(\iota, k)\Big] \tag{4.9}$$

ergibt. Unter Verwendung der zuvor erläuterten Approximationen lässt sich das KLDS als

$$\boldsymbol{\Phi}_{\mathbf{xx}}(\iota, k) = \boldsymbol{\alpha}(k, \varphi(\iota))\, \Phi_{ss}(\iota, k)\, \boldsymbol{\alpha}^{\mathrm{H}}(k, \varphi(\iota)) + \boldsymbol{\Phi}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}}(\iota, k) \tag{4.10}$$

darstellen, wobei das Leistungsdichtespektrum (LDS) des Quellsignals $\Phi_{ss}(\iota, k)$ und das KLDS des Rauschens $\boldsymbol{\Phi}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}}(\iota, k)$ analog zu Gl. (4.9) definiert sind.

Unter der Annahme, dass die Störungen der einzelnen Kanäle unkorreliert sind, ist $\mathbf{\Phi}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}}(\iota, k)$ eine Diagonalmatrix. Damit lässt sich Gl. (4.10) als Bestimmungsgleichung für die Eigenvektoren von $\mathbf{\Phi}_{\mathbf{xx}}(\iota, k)$ auffassen. Der größte Eigenwert von $\mathbf{\Phi}_{\mathbf{xx}}(\iota, k)$ korrespondiert dabei mit der Einfallsrichtung des Quellsignals. Die in der Matrix $\mathbf{\Xi}(\iota, k)$ zusammengefassten verbleibenden $M - 1$ Eigenvektoren spannen den Unterraum in dem die Störungen liegen auf. Sofern Signal und Störungen paarweise unkorreliert sind, muss der *Steering-Vector* orthogonal zum Unterraum der Störungen sein. Der Kehrwert der Projektion eines *Steering-Vectors* für eine gegebene Richtung φ in den Unterraum der Störungen liefert das sogenannte Pseudo-Spektrum

$$P_{\mathrm{MUSIC}}(\varphi(\iota)) = \frac{1}{\sum\limits_{k=0}^{L/2} \boldsymbol{\alpha}^{\mathrm{H}}(k, \varphi(\iota))\, \mathbf{\Xi}(\iota, k)\, \mathbf{\Xi}^{\mathrm{H}}(\iota, k)\, \boldsymbol{\alpha}(k, \varphi(\iota))}. \tag{4.11}$$

Nach der Berechnung des Pseudo-Spektrums für alle möglichen Einfallswinkel, entsteht letztendlich eine Funktion, deren Maximum dem Signaleinfallswinkel entspricht.

Die bisherige Erläuterung des MUSIC-Algorithmus berücksichtigt lediglich ein einzelnes Quellsignal. Allerdings ist MUSIC auch in der Lage, den Einfallswinkel von mehreren Quellsignalen zu ermitteln. Voraussetzung dafür ist, dass die Anzahl der Mikrofone größer als die Anzahl der Quellen ist. Zur Konstruktion des Unterraums in dem die Störungen liegen ($\mathbf{\Xi}(\iota, k)$) werden genauso viele Eigenwerte verworfen wie Signale auf das Array einfallen. Die Trennschärfe der Maxima im Pseudo-Spektrum hängt jedoch von der Anzahl der Mikrofone und der Differenz zwischen den Einfallswinkeln der verschiedenen Quellen ab.

Eine Möglichkeit, die Präzision zu erhöhen und die Trennschärfe bei mehreren Quellsignalen, die aus ähnlichen Richtungen auf das Array einfallen, zu steigern, bietet die Weiterentwicklung *Root*-MUSIC (R-MUSIC) [Bar83]. Um dieses Ziel zu erreichen, wird anstelle des Pseudo-Spektrums (siehe Gl. (4.11)) eine polynomielle Darstellung aus den Eigenvektoren $\mathbf{\Xi}(\iota, k)$ entwickelt. Die Nullstellen (engl. *roots*) dieses Polynoms wiederum korrespondieren zu den gesuchten Signaleinfallsrichtungen.

Die Variante *Estimation of Signal Parameters via Rotational Invariance Techniques* (ESPRIT) [RK89] arbeitet hingegen mit einer Unterteilung des Mikrofonarrays in Teilarrays, die einen gemeinsamen Unterraum des Nutzsignalanteils aufweisen. Durch die Herstellung von Beziehungen zwischen den Teilarrays besitzt ESPRIT eine noch bessere Trennschärfe von mehreren Signalquellen und erzielt darüber hinaus präzisere Winkelschätzungen als MUSIC und R-MUSIC, bei gleichzeitiger Reduktion der Rechenkomplexität. Andererseits erfordert ESPRIT deutlich mehr Mikrofone als MUSIC bzw. R-MUSIC. Damit steht ESPRIT im Konflikt zu der Anforderung dieser Arbeit, möglichst wenig Mikrofone für die Winkelschätzung zu nutzen.

Eine Alternative zu den bisher betrachteten unterraumbasierten Verfahren MUSIC, R-MUSIC und ESPRIT stellt das für die TDOA-Messung akustischer Signale entwickelte GCCPhat [KC76] dar. Der zeitliche Versatz λ zwischen den Mikrofonen m und n wird dabei durch das Maximum der Fourier-Rücktransformierten des normierten KLDS geschätzt. Diese auch als Phat-Funktion bezeichnete Fourier-Rücktransformierte ist durch

$$\mathrm{phat}_{(n,m)}(\lambda, \iota) = \mathrm{IDFT}\left(\frac{X_n(\iota, k)\, X_m^*(\iota, k)}{|X_n(\iota, k)|\ |X_m(\iota, k)|}\right) \tag{4.12}$$

gegeben, wobei $(\cdot)^*$ die komplexe Konjugation, $|\cdot|$ den Betrag der Zahl und IDFT$(\cdot)$ die inverse diskrete FOURIER-Transformation (IDFT) bezeichnen.

Abhängig von der Abtastrate f_s ist ggf. eine Interpolation der Phat-Funktion nötig, um eine ausreichend hohe Auflösung des Parameters λ zu erzielen [TL08; LT99]. Die Multiplikation des zeitlichen Versatzes der die Phat-Funktion maximiert mit ${f_\mathrm{s}}^{-1}$ liefert schließlich die Signallaufzeitdifferenz $\tau_{(n,m)}(\iota)$ der Mikrofone m und n im ι-ten STFT-Block:

$$\tau_{(n,m)}(\iota) = \underset{\lambda}{\operatorname{argmax}} \left\{ \mathrm{phat}_{(n,m)}(\lambda, \iota) \right\} \cdot {f_\mathrm{s}}^{-1}. \tag{4.13}$$

Danach ermöglicht die Fernfeldnäherung zusammen mit der Kenntnis der Mikrofonpositionen die Berechnung des Einfallswinkels aus der Signallaufzeitdifferenz [MP10; SH10]. Für das in Abb. 4.1 dargestellte zwei-elementige Array ergibt sich mit der 2-Norm $\|\cdot\|_2$ und der Schallgeschwindigkeit c_S der Einfallswinkel zu

$$\varphi(\iota) = \arcsin\left(\frac{\tau_{(2,1)} \cdot c_\mathrm{S}}{\|\boldsymbol{m}_2 - \boldsymbol{m}_1\|_2} \right). \tag{4.14}$$

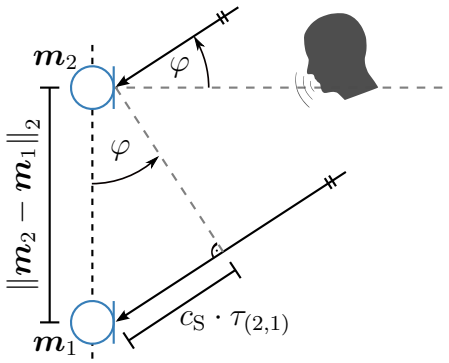


Abbildung 4.1: Geometrischer Zusammenhang zur Berechnung des Signaleinfallswinkels für ein Mikrofonpaar.

Die Verallgemeinerung des Konzeptes von GCCPhat auf mehr als zwei Mikrofone heißt SRPPhat [DSB01]. Dabei wird zunächst die Phat-Funktion jedes Mikrofonpaares berechnet. Anschließend erfolgt, vergleichbar zum Pseudo-Spektrum von MUSIC, die Bestimmung eines *Scores* für jeden möglichen Einfallswinkel. Der maximale *Score* kennzeichnet erneut den Einfallswinkel. Zur Berechnung des *Scores* wird die Signallaufzeitdifferenz $\tau_{(n,m)}(\iota)$ als Funktion des Einfallswinkels $\varphi(\iota)$ angenähert (vgl. Gl. (4.8)). Abschließend werden die Phat-Funktionen aller Mikrofonpaare für die Verzögerungen, die sich aufgrund des angenommenen Winkels $\varphi(\iota)$ ergeben, ausgewertet und summiert. Insgesamt resultiert somit der winkelabhängige *Score*

$$P_\mathrm{SRPPhat}(\varphi(\iota)) = \sum_{m=1}^{M} \sum_{n=1}^{M} \mathrm{phat}_{(n,m)}\Big(\tau_{(n,m)}(\iota) \cdot f_\mathrm{s}\Big), \tag{4.15}$$

dessen Maximum mit dem gesuchten Einfallswinkel korrespondiert.

Die Untersuchung aus [JCN04] von R-MUSIC und SRPPhat, unter Verwendung eines kompakten Arrays mit einem Mikrofonabstand von 0,04 m, dokumentiert, insbesondere bei geringem *Signal-to-Noise Ratio* (SNR) eine präzisere Winkelschätzung durch SRPPhat. Somit ist SRPPhat gegenüber R-MUSIC und damit auch gegenüber MUSIC zu bevorzugen. Außerdem belegen die Untersuchungen in [SH10], dass bei einem Array mit nur zwei Mikrofonen durch die Kombination eines *Filter-and-Sum Beamformer* und SRPPhat (FSBPhat) die Präzision einer anschließenden Triangulation steigt. Erreicht wird diese Verbesserung, indem zunächst die Filterimpulsantworten des *Filter-and-Sum Beamformer* (FSB) so berechnet werden, dass das SNR am *Beamformer*-Ausgang maximiert wird [WHP04]. Die Filterimpulsantworten des FSB dienen schließlich als Eingangssignal für die Auswertung von Gl. (4.12) anstelle der bislang verwendeten Mikrofonsignale $X_m(\iota, k)$ und $X_n(\iota, k)$.

Eine weitere Klasse von DOA-Schätzern nutzt Modelle zur Beschreibung der zu erwartenden Zusammenhänge der empfangenen Mikrofonsignale. Die Winkelschätzung erfolgt anschließend durch die Bestimmung des Abstandes zwischen den aufgrund des Modells vorhergesagten und den tatsächlichen Beobachtungen. Dieses als *Generalized State Coherrence Transform* (GSCT) bezeichnete Konzept, erlaubt einerseits die TDOA-Schätzung zu mehreren Quellen mit nur zwei Mikrofonen [NO09] und ermöglicht andererseits auch die Schätzung des Einfallswinkels [LY10]. Als Beobachtungen verwendet der von Loesch und Yang entwickelte und im Folgenden als Loesch und Yang DOA-*Estimator* (LYDE) bezeichnete Algorithmus [LY10] den Phasenversatz $\arg(X_m(\iota, k)/X_1(\iota, k))$ des m-ten Mikrofons bezogen auf den ersten Kanal. Dieser wird mit dem durch ein Modell bei Freifeldausbreitung prädizierten Phasenversatz $\arg(\boldsymbol{\alpha}(k, \varphi(\iota)))$ verglichen. Durch die Wahl des 2π-periodischen Abstandsmaßes

$$P_{\mathrm{LYDE}}(k, \varphi(\iota)) = \sum_{m=1}^{M} \cos(\arg(X_m(\iota, k)\, X_1^*(\iota, k)) - \arg(\boldsymbol{\alpha}(k, \varphi(\iota)))) \tag{4.16}$$

können auch die Frequenzen bei denen die Wellenlänge kleiner als die Hälfte des Mikrofonabstandes ist und dementsprechend räumliches Aliasing auftritt, berücksichtigt werden. Vor der Addition der Abstände $P_{\mathrm{LYDE}}(k, \varphi(\iota))$ für alle Frequenzen wird zunächst eine heuristische, nichtlineare Gewichtung des Abstandes mit der Funktion

$$\rho(\boldsymbol{x}) = 1 - \tanh\left(\sqrt{\boldsymbol{x}}\right) \tag{4.17}$$

vollzogen. Diese Gewichtung sorgt für eine schärfere Abgrenzung der Maxima in den einzelnen Frequenzbändern, sodass auch bei der Berechnung des *Scores* für alle Frequenzen

$$P_{\mathrm{LYDE}}(\varphi(\iota)) = \sum_{k=0}^{L/2} \rho(\|2 \cdot M - 2 \cdot P_{\mathrm{LYDE}}(k, \varphi(\iota))\|_2) \tag{4.18}$$

ein klares Maximum entsteht. Wie schon bei MUSIC und SRPPhat kann hier ebenfalls der Winkel zu mehreren Quellen bestimmt werden, indem so viele Maxima ermittelt werden, wie Quellen vorliegen.

4.2 Entwicklung eines Winkelschätzers

Inspiriert von der GSCT sowie dem zuvor erläuterten LYDE wurde im Rahmen der Arbeit ein weiterer Algorithmus zur Schätzung des Einfallswinkels konzipiert. Dieser nutzt ebenfalls ein Modell zur Prädiktion der zu erwartenden Beobachtungen für einen gegebenen Einfallswinkel und identifiziert anschließend das Modell, dessen Vorhersage am Besten mit den Messungen übereinstimmt. Durch ein periodisches Abstandsmaß kann räumliches Aliasing, wie auch beim LYDE, konstruktiv ausgenutzt werden.

Zur Berechnung der Wahrscheinlichkeit, ob eine beobachtete Phasendifferenz zu der Prädiktion eines Modells passt, dient die komplexe Watson-Verteilung. Sie beschreibt die Verteilung komplexer Einheitsvektoren auf der komplexen Einheitshyperkugel und wurde zunächst in der Bildverarbeitung eingesetzt [MD99]. Im Bereich der akustischen Signalverarbeitung bietet die komplexe Watson-Verteilung inzwischen, insbesondere bei der Blinden Quellentrennung (engl. *blind source separation* (BSS)), eine Alternative zur Beschreibung der Signale eines Mikrofonarrays im STFT-Bereich zu der ansonsten verwendeten komplexen Normalverteilung [JTN14; TH10; Dru+14].

Ferner gehen die in Abschnitt 4.1 präsentierten Winkelschätzer von einer omnidirektionalen Richtcharakteristik der eingesetzten Mikrofone aus und verwenden deshalb lediglich die Phasendifferenzen zur Schätzung des Winkels. Im Hinblick auf die potenziellen Einsatzgebiete der zu entwickelnden Geometriekalibrierungsverfahren können jedoch u. a. die Mikrofone von Smartphones zur Richtungsschätzung dienen. Diese besitzen meist keine omnidirektionale Richtcharakteristik, sodass neben Phasendifferenzen auch Amplitudendifferenzen vorliegen. Die Nichtberücksichtigung der Amplitudendifferenzen von direktionalen Mikrofonen führt bspw. beim *Beamforming* zu signifikanten Einbußen der Signalqualität [Gau+14]. Daher soll der im weiteren Verlauf vorgestellte Algorithmus neben den Phasendifferenzen auch die auftretenden Amplitudendifferenzen berücksichtigen.

Den Ausgangspunkt bildet weiterhin das bereits in Gl. (4.4) definierte Signalmodell. Gemäß dieses Modells wird die Länge der komplexen Vektoren $\mathbf{x}(\iota, k)$ hauptsächlich von der Leistung des Quellsignals $S(\iota, k)$ bestimmt. Die auftretenden Amplitudendifferenzen der einzelnen Elemente von $\mathbf{x}(\iota, k)$ spiegeln sowohl die Dämpfung aufgrund der zurückgelegten Strecke wieder als auch eine richtungsabhängige Sensitivität der Mikrofone. Durch die Normierung des Betrages von $\mathbf{x}(\iota, k)$ entsteht die von der Signalleistung unabhängige Darstellung des Phasenversatzes:

$$\boldsymbol{y}(\iota, k) = \frac{\mathbf{x}(\iota, k)\, X_1^*(\iota, k)}{\left\|\mathbf{x}(\iota, k)\, X_1^*(\iota, k)\right\|_2}. \tag{4.19}$$

Die resultierenden Beobachtungen $\boldsymbol{y}(\iota, k)$ beschreiben daher Punkte auf der komplexen Einheitshyperkugel ($\|\boldsymbol{y}(\iota, k)\|_2 = 1$). Dabei ist jedoch zu beachten, dass die Beobachtungen durch die Normierung zwar nicht mehr von der Amplitude des Quellsignals abhängen, die aufgrund der Dämpfung der Pfade sowie der Richtcharakteristik auftretenden relativen Amplitudendifferenzen aber weiterhin vorhanden sind.

Zumal der gesuchte Einfallswinkel mithilfe des Konzeptes der GSCT, vergleichbar zum LYDE, bestimmt werden soll, ist außerdem ein Modell zur Prädiktion der Beobachtungen für eine gegebene Signaleinfallsrichtung erforderlich. Der gesuchte Winkel korrespondiert dabei mit der Richtung der direkten Signalkomponente. Die beobachteten

Amplituden- und Phasendifferenzen, die von der direkten Komponente hervorgerufen wurden, sind ähnlich zu denen in einer nachhallfreien Umgebung. Dementsprechend gilt es, das Modell der Phasen- und Amplitudendifferenzen für den nachhallfreien Fall zu identifizieren, das bei gegebenem Einfallswinkel die größtmögliche Übereinstimmung mit den Beobachtungen erzielt.

Die erwarteten Phasendifferenzen für den Einfallswinkel φ lassen sich durch den aus Gl. (4.7) bekannten *Steering-Vector* beschreiben. Darüber hinaus soll jedoch auch die Richtcharakteristik der Mikrofone mit in die Schätzung einfließen. Diese kann für das m-te Mikrofon bspw. durch

$$\xi_m(\varphi) = c_\alpha + (1 - c_\alpha) \cdot \cos(\varphi - \omega_m) \tag{4.20}$$

modelliert werden [Hab10]. Der Winkel ω_m gibt dabei die Orientierung des m-ten Mikrofons und damit auch die Ausrichtung der Richtcharakteristik bezüglich des Koordinatensystems des Mikrofonarrays an. Die Empfindlichkeit für eine bestimmte Richtung lässt sich durch die Wahl des Parameters c_α beschreiben. Richtcharakteristiken für ausgewählte Werte von c_α sind in Abb. 4.2 visualisiert.

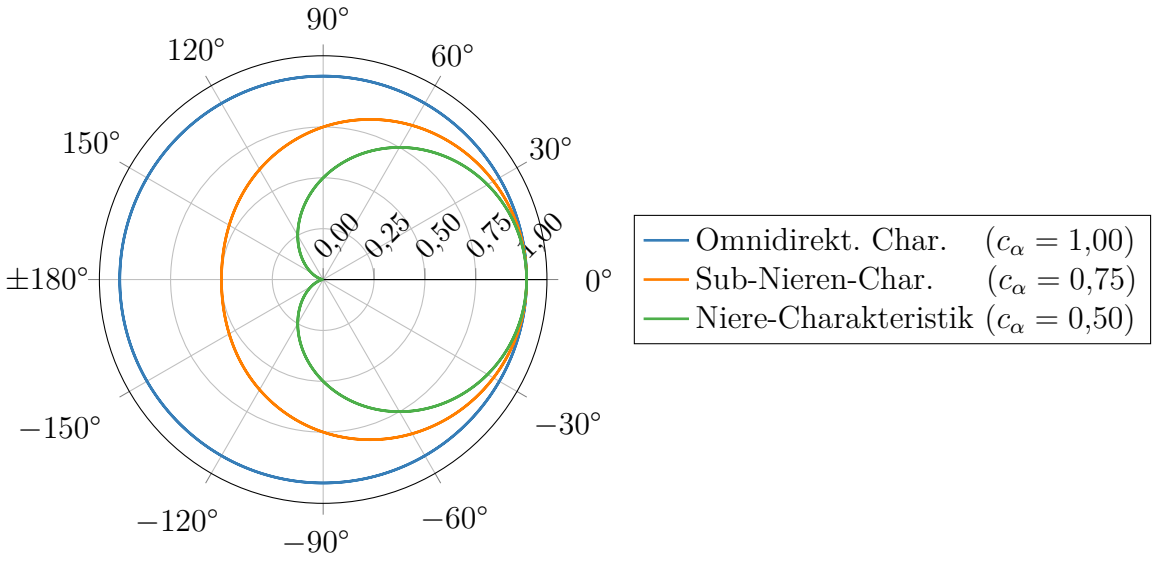


Abbildung 4.2: Beispiele für die richtungsabhängige Empfindlichkeit eines Mikrofons.

Zur besseren Anschauung beschränken sich Gl. (4.20) ebenso wie auch Abb. 4.2 auf einfache, frequenzunabhängige Richtcharakteristiken. Allerdings kann $\xi_m(\varphi)$ unmittelbar durch eine frequenzabhängige oder die für ein spezielles Mikrofon gemessene Richtcharakteristik ersetzt werden.

Die Zusammenfassung der Richtcharakteristiken für alle Mikrofone liefert den Vektor

$$\boldsymbol{\xi}(\varphi) = \begin{bmatrix} \xi_1(\varphi) & \dots & \xi_M(\varphi) \end{bmatrix}^{\mathrm{T}}. \tag{4.21}$$

Durch die Kombination von Gl. (4.8) und den Richtcharakteristiken aus Gl. (4.21) entsteht schließlich das Modell

$$\tilde{\boldsymbol{\Delta}}(k, \varphi) = \boldsymbol{\xi}(\varphi) \odot \boldsymbol{\alpha}(k, \varphi), \tag{4.22}$$

das sowohl die Phasen- als auch die Amplitudendifferenzen berücksichtigt. Der Operator $\odot$ bezeichnet dabei das Hadamard-Produkt (elementweise Multiplikation) der beteiligten Vektoren. Danach erfolgt noch die Normierung des Modells, sodass die Prädiktionen ebenfalls auf der komplexen Einheitshyperkugel liegen:

$$\boldsymbol{\Delta}(k,\varphi) = \tilde{\boldsymbol{\Delta}}(k,\varphi) / \left\| \tilde{\boldsymbol{\Delta}}(k,\varphi) \right\|_2 . \tag{4.23}$$

Da sich jetzt sowohl die beobachteten als auch die vorhergesagten Vektoren auf einer komplexen Einheitshyperkugel befinden, wird abschließend noch eine Verteilung zur Modellierung der Beobachtungen benötigt, die diese Eigenschaft berücksichtigt. Deshalb wird die komplexe Watson-Verteilung mit der Verteilungsdichtefunktion

$$\mathrm{p_W}(\boldsymbol{y}(\iota,k);\kappa,\boldsymbol{\Delta}(k,\varphi)) = \frac{1}{c_\mathrm{W}(\kappa)} \cdot \exp\Big(\kappa \cdot \left\| \boldsymbol{\Delta}^\mathrm{H}(k,\varphi)\, \boldsymbol{y}(\iota,k) \right\|_2\Big) \tag{4.24}$$

verwendet. Die Verteilung der Beobachtungen $\boldsymbol{y}(\iota,k)$ wird somit durch den Modalvektor $\boldsymbol{\Delta}(k,\varphi)$ (engl. *mode*) und den reellwertigen Konzentrationsparameter κ beschrieben. Die Normierungskonstante ist durch $c_\mathrm{W}(\kappa)$ gegeben [MD99].

Bei der Betrachtung des Exponenten in Gl. (4.24) wird außerdem ersichtlich, dass die komplexe Watson-Verteilung invariant bezüglich der Multiplikation von $\boldsymbol{y}(\iota,k)$ bzw. $\boldsymbol{\Delta}(k,\varphi)$ mit einem komplexen Skalar ist. Aufgrund dessen ist eine Darstellung der Beobachtungen unabhängig von der absoluten Phase nicht notwendig und somit entfällt die Notwendigkeit der in Gl. (4.19) durchgeführten Normalisierung der Phase. Außerdem kann jede Vektorkomponente von $\boldsymbol{y}(\iota,k)$ mit $\mathrm{e}^{\mathrm{j}2\pi}$ multipliziert werden, ohne dass sich die Wahrscheinlichkeit verändert. Daher wird auch hier räumliches Aliasing analog zur Kosinus-Distanz des LYDE (vgl. Gl. (4.16)) berücksichtigt.

Die Summation des mit der Watson-Verteilung gewichteten Abstandes über alle Frequenzen liefert, wie auch bei den bisher betrachteten Winkelschätzern, den *Score*:

$$P_\mathrm{WKM}(\varphi(\iota)) = \frac{2}{L} \cdot \sum_{k=0}^{L/2} \mathrm{p_W}(\boldsymbol{y}(\iota,k);\kappa,\boldsymbol{\Delta}(k,\varphi)). \tag{4.25}$$

Der Schätzwert für den Einfallswinkel ergibt sich schließlich ebenfalls durch die Selektion des Maximums. Zumal sich Gl. (4.25) als Kerndichteschätzer auffassen lässt [DJH15], wird der entwickelte Schätzer als Watson-Kern-Methode (WKM) bezeichnet. Auch hier sei nocheinmal darauf hingewiesen, dass sich durch die Suche mehrerer Maxima auch die Einfallswinkel zu mehr als einer Quelle ergeben.

Ein weiteres Argument für die Verwendung der komplexen Watson-Verteilung als Distanzmaß entsteht durch eine erneute Betrachtung des Exponenten der Verteilung in Gl. (4.24). Der dort berechnete Betrag des Skalarpoduktes zwischen dem Modell und den Beobachtungen entspricht der Leistung am Ausgang eines *Beamformers*, wenn $\boldsymbol{\Delta}(k,\varphi)$ die Filterimpulsantworten und $\boldsymbol{y}(\iota,k)$ das Eingangssignal darstellen. Demnach erscheint das verwendete Distanzmaß passender als die vom LYDE genutzte Variante. Außerdem kann die beim LYDE heuristisch gewählte Nichtlinearität zur Steigerung der Winkelauflösung nun durch eine probabilistisch motivierte Exponentialfunktion ersetzt werden. Zwar verbleibt auch bei der WKM die passende Wahl des Konzentrationsparameters κ, allerdings stellt die heuristische Wahl eines einzigen Parameters gegenüber einer gesamten Funktion eine deutliche Verbesserung dar.

4.3 Evaluierung des entwickelten Winkelschätzers

Nachdem der vorausgegangene Abschnitt das Konzept der WKM erläutert hat, soll nun die Leistungsfähigkeit analysiert werden. Bestandteil der Untersuchungen ist einerseits die passende Wahl des Konzentrationsparameters κ und andererseits eine Bewertung, inwieweit sich die zusätzliche Berücksichtigung der Richtcharakteristik der Mikrofone auf die Winkelschätzung auswirkt. Weitere Analysen zur Qualität der Winkelschätzungen folgen in Abschnitt 4.4 im Rahmen der Auswahl eines für die Geometriekalibrierung geeigneten Winkelschätzers.

Die Basis für die Untersuchung der WKM auf die zuvor genannten Gesichtspunkte bilden mithilfe der Spiegel-Quellen-Methode (vgl. Abschnitt 3.2) generierte mehrkanalige Audiosignale. Die dazu notwendigen Quellsignale stammen aus der TIMIT-Datenbank [G+93]. Um bei den Analysen ein möglichst breites Spektrum verschiedener Szenarien abzudecken, wurden sowohl die Raumgröße und die Anordnung des Mikrofonarrays innerhalb des Raumes, als auch die Nachhallzeit variiert. Darüber hinaus diente isotropes[1] Rauschen [HG07] zur Modellierung zusätzlicher Störungen, wie sie bspw. in Büroumgebungen auftreten. Damit trotz der Richtungsunabhängigkeit des isotropen Rauschens die Kohärenz zwischen den Mikrofonsignalen gewährleistet ist, erfolgte die Simulation des Rauschens durch [HG07].

Das zur Untersuchung verwendete Szenario besteht aus einem Mikrofonarray mit 4 Mikrofonen, die in einem Kreis mit 0,10 m Radius angeordnet sind. Weiterhin wird die Position des Mikrofonarrays innerhalb eines Raumes mit zufällig generierten Ausmaßen ebenfalls zufällig gewählt. Die Nachhallzeit variiert zwischen 0,2 s, 0,4 s und 0,6 s und das SNR beträgt 5 dB, 10 dB oder 15 dB. Außerdem wurden die zu diesen Szenarien gehörenden Audiosignale stets mit einer Blocklänge von 1024 Abtastwerten, einen Blockvorschub von 256 Abtastwerten und einem Blackman-Fenster in den STFT-Bereich überführt.

Eine Bewertung der gewonnenen Winkelschätzungen setzt weiterhin ein geeignetes Fehlermaß voraus. Eine tabellarische Angabe der Ergebnisse wie in [Ara+11] scheidet aufgrund der Menge der Experimente aus. Die Nutzung des *Root-Mean-Square Error* (RMSE) erscheint ungeeignet, da dieser von einzelnen Ausreißern dominiert wird, die aber nicht notwendigerweise ein Problem für die spätere Kalibrierung bedeuten (siehe Kapitel 6). Eine prozentuale Angabe, wie viele Schätzungen einen Fehler kleiner als ein gewählter Schwellwert aufweisen [PF13], erfordert die Wahl eines geeigneten Schwellwerts und sagt wenig über die Verteilung des Fehlers aus.

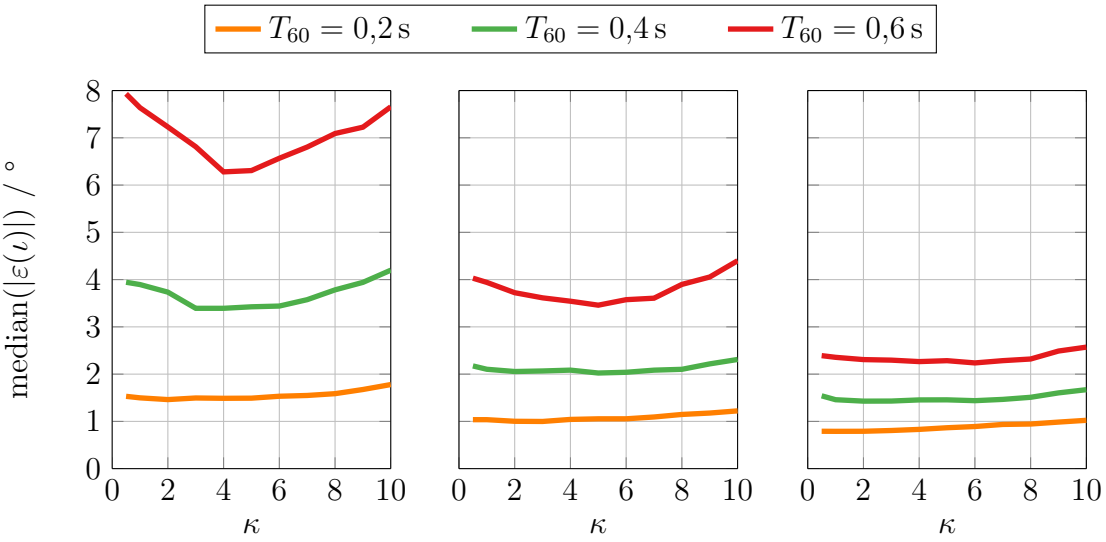
Für eine detaillierte Einschätzung, wie viel Prozent der Schätzungen einen bestimmten Fehler unterschreiten, soll deshalb das kumulative Histogramm des absoluten Fehlers genutzt werden. Dazu wird zunächst die Differenz zwischen dem Einfallswinkel $\varphi(\iota)$ und der zugehörigen Schätzung $\hat{\varphi}(\iota)$ berechnet:

$$\varepsilon(\iota) = \varphi(\iota) - \hat{\varphi}(\iota). \tag{4.26}$$

Der absolute Fehler der Einfallswinkelschätzung ergibt sich schließlich indem der Betrag von $\varepsilon(\iota)$ gebildet wird.

[1] Der Begriff isotrop bezeichnet, dass eine Eigenschaft nicht von der Richtung abhängt [Bro72b]. Isotropes Rauschen besitzt dementsprechend keine Vorzugsrichtung.

Bei der zunächst betrachteten Auswirkung des Konzentrationsparameters κ auf die Leistungsfähigkeit der Winkelschätzung ist jedoch die Darstellung eines kumulativen Histogramms für die jeweiligen Nachhallzeiten und SNR-Level zu aufwändig. Damit die Qualität der Winkelschätzung durch eine skalare Größe charakterisiert werden kann, obwohl die später präsentierten kumulativen Histogramme eine endlastige (engl. *heavy-tailed*) Verteilung des Fehlers zeigen, wird in Abb. 4.3 der Median des Betrages von $\varepsilon(\iota)$ aus 1000 Simulationen visualisiert.



(a) SNR: 5 dB (b) SNR: 10 dB (c) SNR: 15 dB

Abbildung 4.3: Auswirkung des Konzentrationsparameters κ auf den Fehler der Einfallswinkelschätzung durch die WKM für verschiedene SNR-Level und Nachhallzeiten bei der Verwendung eines zirkulären Arrays mit einem Radius vom 0,10 m und 4 Mikrofonen.

Die Betrachtung der Ergebnisse aus Abb. 4.3 belegt eindeutig, dass die Wahl des Konzentrationsparameters nur eine untergeordnete Rolle für die Qualität der erzielten Winkelschätzungen spielt. Bei sehr großem SNR (siehe Abb. 4.3c) ist der Fehler sogar nahezu unabhängig vom Konzentrationsparameter. Mit zunehmenden Störungen durch Nachhall und Rauschen (vgl. Abb. 4.3a) entsteht jedoch ein schwacher Zusammenhang. Angesichts der geringen Auswirkung des Konzentrationsparameters auf den resultierenden Fehler wird der Konzentrationsparameter für alle weiteren Untersuchungen zu $\kappa = 5$ gewählt und nicht weiter optimiert.

Wichtiger als die Wahl des Konzentrationsparameters ist hingegen die Leistungsfähigkeit der WKM, wenn gerichtete Mikrofone vorliegen. Um die erzielten Ergebnisse einschätzen zu können, soll SRPPhat als Referenzverfahren dienen. Für die Untersuchungen wurde eine Nachhallzeit von 0,4 s und ein SNR von 10 dB sowie 1000 Konstellationen aus Sensor- und Ereignispositionen verwendet. Sofern die Richtcharakteristiken der Mikrofone den bereits in Abb. 4.2 skizzierten Verläufen entsprechen, ergeben sich die in Abb. 4.4 gezeigten kumulativen Histogramme.

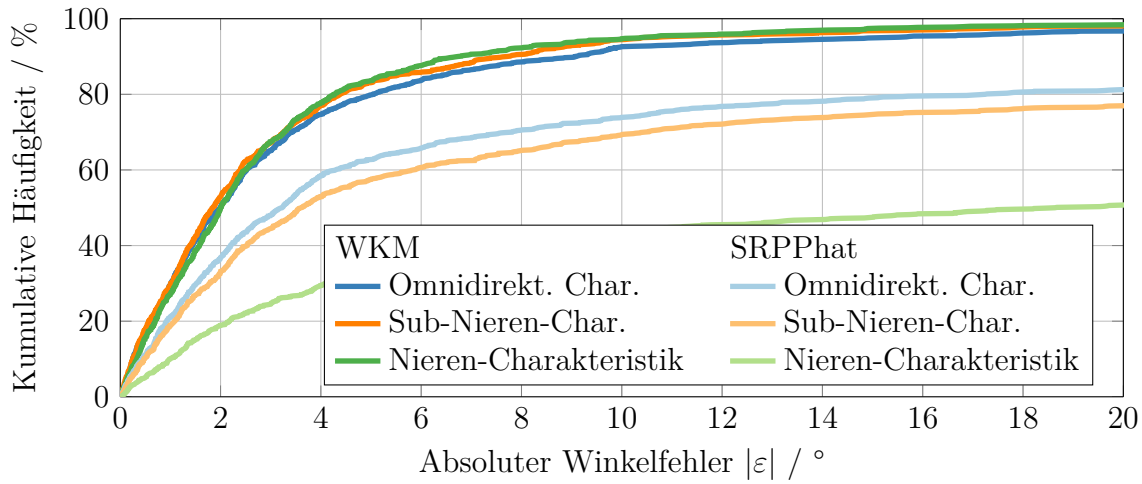


Abbildung 4.4: Kumulative Histogramme des absoluten Winkelfehlers $|\varepsilon|$ bei einer Einfallswinkelschätzung durch die WKM bzw. SRPPhat für verschiedene Richtcharakteristiken der Mikrofone bei der Verwendung eines zirkulären Arrays mit einem Radius vom 0,10 m und 4 Mikrofonen.

Ein Vergleich der Ergebnisse beim Einsatz omnidirektionaler Mikrofone zeigt zunächst, dass die entwickelte WKM gegenüber dem SRPPhat-Ansatz deutlich überlegen ist. Darüber hinaus steigt der Fehler von SRPPhat an, je größer die Abweichung zur omnidirektionalen Richtcharakteristik wird. Verantwortlich dafür sind die von den Richtcharakteristiken der Mikrofone verursachten Amplitudendifferenzen, die von SRPPhat unberücksichtigt bleiben. Damit ergibt sich eindeutig die Notwendigkeit die Richtcharakteristik der Mikrofone mit in die Winkelschätzung einzubeziehen. Die WKM ist durch die Kenntnis der Richtcharakteristik in der Lage, den bei SRPPhat auftretenden Anstieg des Fehlers zu kompensieren und kann darüber hinaus die Qualität der Winkelschätzungen gegenüber dem Szenario mit omnidirektionalen Mikrofonen z. T. sogar geringfügig verbessern.

4.4 Auswahl eines Winkelschätzers für die Geometriekalibrierung

Aufgrund der Entscheidung, die Geometrie eines akustischen Sensornetzes durch Einfallswinkelschätzungen zu bestimmen, hat die Qualität der Winkelschätzung großen Einfluss auf die spätere Genauigkeit des Kalibrierungsergebnisses. Ziel dieses Abschnittes ist es daher, die Leistungsfähigkeit der WKM mit den in Abschnitt 4.1 vorgestellten Ansätzen zu vergleichen und den Winkelschätzer auszuwählen, der beim Einsatz von kompakten Arrays mit möglichst wenigen Mikrofonen die besten Winkelschätzungen liefert.

Wie bereits in Kapitel 3 erwähnt, erfordert eine vergleichende Analyse Aufnahmen für möglichst viele Sprecher und Arrayposition in Räumen mit unterschiedlichen Reflexionseigenschaften. Zumal keine Datenbanken vorhanden sind, die eine ausreichend große Menge mehrkanaliger Aufnahmen bereitstellen, erscheint allein eine stichprobenartige

Simulation zufälliger Szenarien realisierbar. Zur Evaluierung der Algorithmen dienen daher 100 000 zufällige Konfigurationen mit Raumgrößen zwischen $4{,}00 \times 4{,}00 \times 3{,}00\,\mathrm{m}^3$ und $8{,}00 \times 8{,}00 \times 3{,}00\,\mathrm{m}^3$, einer Quellposition in $1{,}75 \pm 0{,}10\,\mathrm{m}$ Höhe und einer Entfernung zwischen Sensor und Quelle von $1{,}00 \pm 0{,}50\,\mathrm{m}$. Zunächst werden Mikrofonpaare mit einem Mikrofonabstand von 0,05 m verwendet. Die für den Vergleich erforderlichen akustischen Signale entstehen durch die Verhallung von Sprachsignalen aus der TIMIT-Datenbank [G+93] unter Verwendung der Spiegel-Quellen-Methode (vgl. Abschnitt 3.2). Vorerst wird nur die Nachhallzeit zwischen 0,0 s, 0,2 s und 0,4 s variiert und weitere Störungen, wie z. B. durch Rauschen, bleiben unberücksichtigt.

Abb. 4.5 zeigt die kumulativen Histogramme des Winkelfehlers für das zuvor erläuterte Experiment. Ohne Nachhall ($T_{60} = 0{,}0\,\mathrm{s}$) liegt der Fehler aller Winkelschätzer

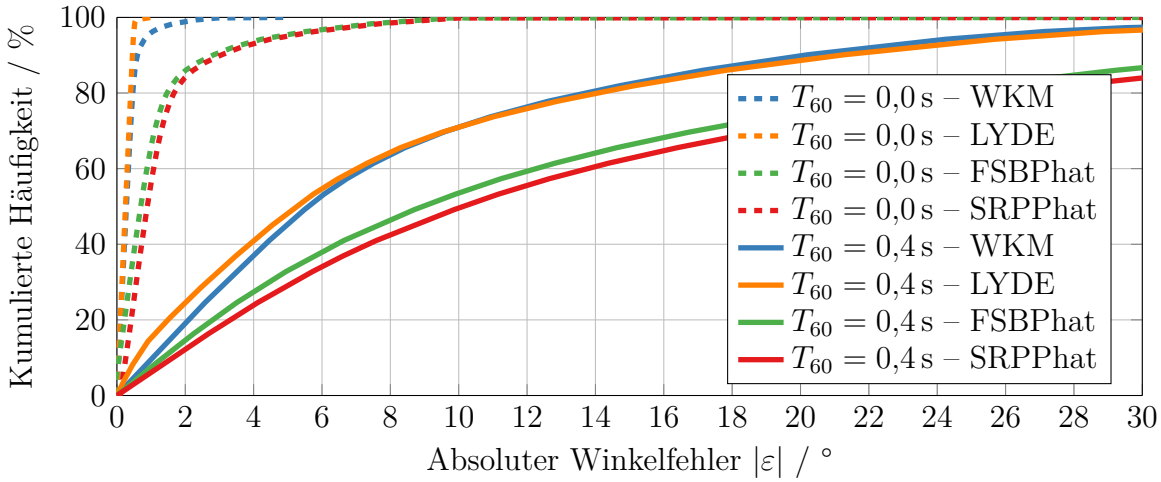


Abbildung 4.5: Kumulative Histogramme des Winkelfehlers ausgewählter Schätzer für ein zwei-elementiges Mikrofonarray mit 0,05 m Mikrofonabstand bei verschiedenen Nachhallzeiten.

größtenteils unter 2°. Der LYDE und die WKM schneiden dabei etwas besser ab als die korrelationsbasierten Konkurrenten FSBPhat und SRPPhat. Die schon bei einer Nachhallzeit von 0,0 s auftretenden Unterschiede bei der Leistungsfähigkeit der betrachteten Algorithmen bleiben auch mit steigender Nachhallzeit bestehen, allerdings kommt es zu einem signifikanten Anstieg des Fehlers insgesamt. Die Ergebnisse für eine Nachhallzeit von 0,4 s zeigen zudem, dass nur noch ein geringer Teil der Schätzungen einen Fehler im 2°-Bereich aufweist und ein großer Anteil mit Fehlern von mehr als 10° auftritt. Diese Ausgangssituation stellt eine deutliche Herausforderung für die Geometriekalibrierung dar, bzw. lässt es fraglich erscheinen, ob die vorliegenden Winkelschätzungen zu einer präzisen Kalibrierung des akustischen Sensornetzes führen. Insofern erscheint eine nähere Betrachtung des Fehlers bzw. die Identifikation der Ursachen zwingend notwendig.

Zur näheren Analyse des auftretenden Fehlers dient Abb. 4.6. Sie zeigt beispielhaft für die Anwendung von SRPPhat auf die Filterimpulsantworten des FSB (FSBPhat) den mittleren Fehler des Einfallswinkels $\bar{\varepsilon}$ (Bias) in Abhängigkeit des tatsächlichen Winkels φ. Das ausgewählte Verfahren besitzt zwar laut Abb. 4.5 einen vergleichsweise großen Fehler, aber auch für die anderen Ansätze ergeben sich qualitativ ähnliche Darstellungen. Eine vollständige Darstellung für sämtliche Algorithmen liefert Anhang A.1.

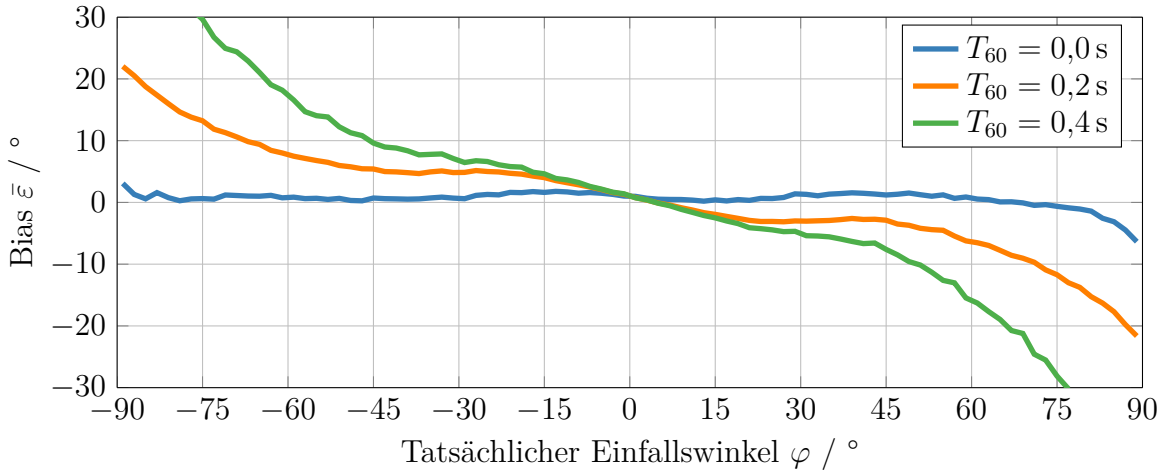


Abbildung 4.6: Richtungsabhängigkeit des Bias der Einfallswinkelschätzung eines zwei-elementigen Mikrofonarrays mit 0,05 m Mikrofonabstand bei der Anwendung von SRPPhat auf die Filterimpulsantworten eines FSB.

Die Ergebnisse aus Abb. 4.5 dokumentieren einen Anstieg des absoluten Fehlers mit zunehmender Nachhallzeit und bestätigen damit die erwartete Abhängigkeit zwischen Winkelfehler und Nachhallzeit. Abb. 4.6 weist jedoch eindeutig auf einen Bias hin, der mit zunehmendem Einfallswinkel und mit ansteigender Nachhallzeit anwächst. Ein Einfallswinkel von 0° entspricht dabei der zentralen Position vor dem Mikrofonarray, bei der der Phasenversatz zwischen den Signalen verschwindet. Die Berücksichtigung des Vorzeichens des Fehlers bestätigt die aus Laborexperimenten bekannte Beobachtung, dass die Winkelschätzungen zwei-elementiger Arrays im Bereich von ±90° tendenziell einen Bias in Richtung von 0° aufweisen. Allerdings deuten die Daten aus Abb. 4.6 auf einen mehr oder weniger ausgeprägten Bias bei fast allen Winkeln hin. Eine Erklärung für Fehler bei Einfallswinkeln, die in unmittelbarer Nähe von ±90° liegen, ergibt sich anhand der Berechnungsvorschrift des Einfallswinkels aus der TDOA (vgl. Gl. (4.14)).

Die Sensitivität der Richtungsschätzung ist gemäß [MP10] durch die Ableitung von Gl. (4.14) nach der TDOA $\tau_{(n,m)}$ gegeben. Eine bezüglich des Integrals normierte Darstellung dieser Ableitung als Funktion des Einfallswinkels φ ist in Abb. 4.7 visualisiert.

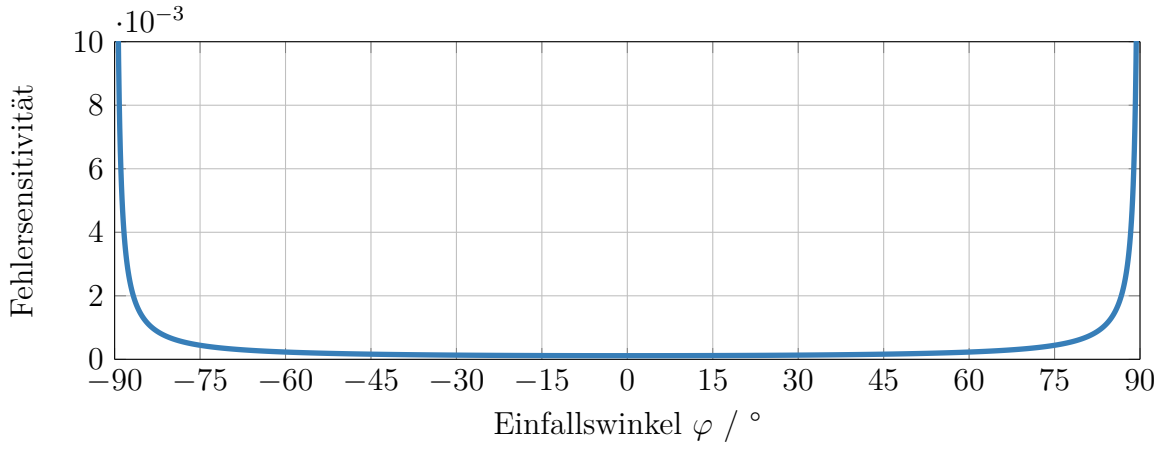


Abbildung 4.7: Sensitivität der Einfallswinkelberechnung gegenüber TDOA-Fehlern.

Anhand dieser Abbildung wird ersichtlich, dass die Sensitivität gegenüber Fehlern in der Laufzeitmessung bei ±90° am größten ist. Verantwortlich dafür ist der sehr steile Verlauf der arcsin-Funktion in diesem Bereich. Dementsprechend verursacht eine kleine Änderung der TDOA eine große Winkeländerung. Allerdings deutet eine erhöhte Sensitivität auf eine Steigerung der Varianz und nicht auf den beobachteten Bias hin, demnach müssen noch andere Ursachen vorliegen.

Eine detaillierte Untersuchung des Bias findet erst in Abschnitt 4.5 statt, deshalb soll hier stattdessen ein Vorgriff auf die dort gewonnenen Erkenntnisse erfolgen, um den Vergleich der Winkelschätzer fortführen zu können. Als ein Auslöser des Bias wird sich die lineare Anordnung der Mikrofone herausstellen. Daher sollen für den weiteren Vergleich der Winkelschätzer Sensorknoten mit mehr als zwei Mikrofonen, die allerdings nicht auf einer Linie angeordnet sein dürfen, berücksichtigt werden.

Die Ausführungen in [MP10] sehen zur Reduktion der Sensitivität eine Kombination der Winkelschätzungen mehrerer Paare, gewichtet mit dem Inversen der Sensitivität, vor. Allerdings lassen sich alle betrachteten Winkelschätzer direkt für mehr als zwei Mikrofone verwenden und gestatten so bereits während einer gemeinsamen Winkelschätzung die Ausnutzung der räumlichen Information durch zusätzliche Mikrofone. Die Sensitivität dient jedoch dazu, eine geeignete Anordnung der Mikrofone festzulegen. Damit die Anzahl der Mikrofone weiterhin möglichst klein ausfällt, wird ein Sensorknoten mit drei Mikrofonen betrachtet. Gemäß der Darstellungen in [MP10] bietet sich in diesem Fall eine Anordnung in Form eines gleichseitigen Dreiecks an.

Um zu prüfen, inwieweit sich sowohl der Fehler als auch der Bias durch die Verwendung eines dreieckigen Arrays reduzieren lassen, soll die bisher durchgeführte Analyse für eine dreieckige Mikrofonanordnung mit 0,05 m Kantenlänge wiederholt werden. Der entscheidende Einfluss der linearen Anordnung auf den Bias zeigt sich auch daran, dass der bei einem Mikrofonpaar auftretende Bias (vgl. Abb. 4.6) durch die Nutzung eines dreieckigen Arrays fast vollständig verschwindet (siehe Abb. 4.8).

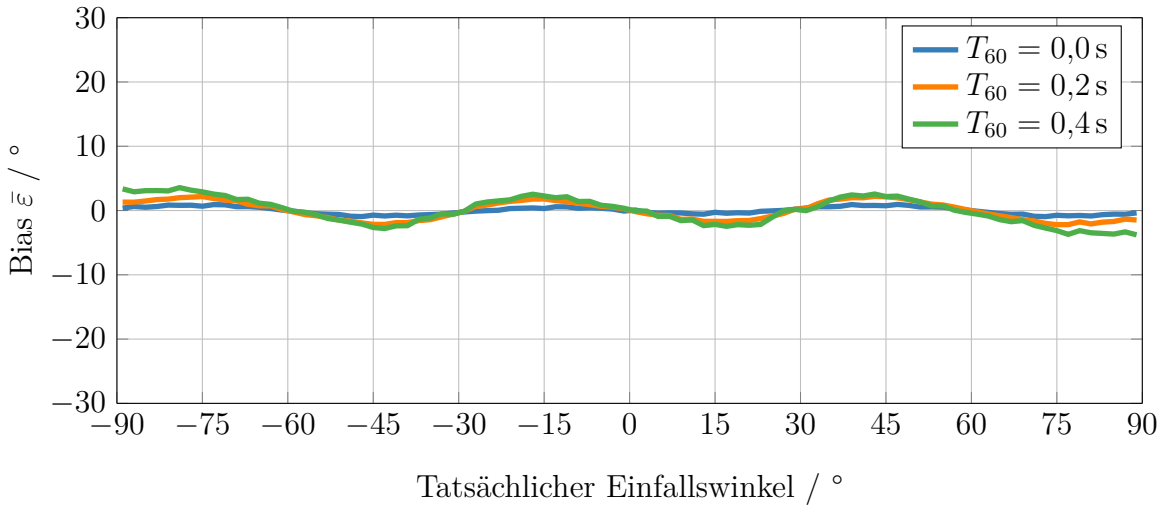


Abbildung 4.8: Richtungsabhängigkeit des Bias der Einfallswinkelschätzung bei der Anwendung von SRPPhat auf die Filterimpulsantworten eines FSB bei der Nutzung einer dreieckigen Mikrofonanordnung mit 0,05 m Kantenlänge.

Die Welligkeit des verbleibenden Bias lässt sich auf die Interpolation bei der Schätzung des Einfallswinkels zurückführen. Aufgrund der Abtastrate von 16 kHz und dem Mikrofonabstand von 0,05 m liegen die Verzögerungen der Mikrofonsignale zwischen −2,33 und 2,33 Abtastwerten. Um dennoch Winkelschätzungen mit ausreichender Auslösung zu erhalten, verwenden alle betrachteten Winkelschätzer eine Interpolation. Die Stützstellen bei ganzzahligen Abtastwerten (±2, ±1 bzw. 0) korrespondieren dabei gerade mit den Nullstellen des Bias bei ±59°, ±29° sowie 0° (vgl. Abb. 4.8).

Die Reduktion des Bias spiegelt sich auch in der Darstellung der kumulativen Histogramme des absoluten Fehlers in Abb. 4.9 wieder. Beim Vergleich mit Abb. 4.5, die den

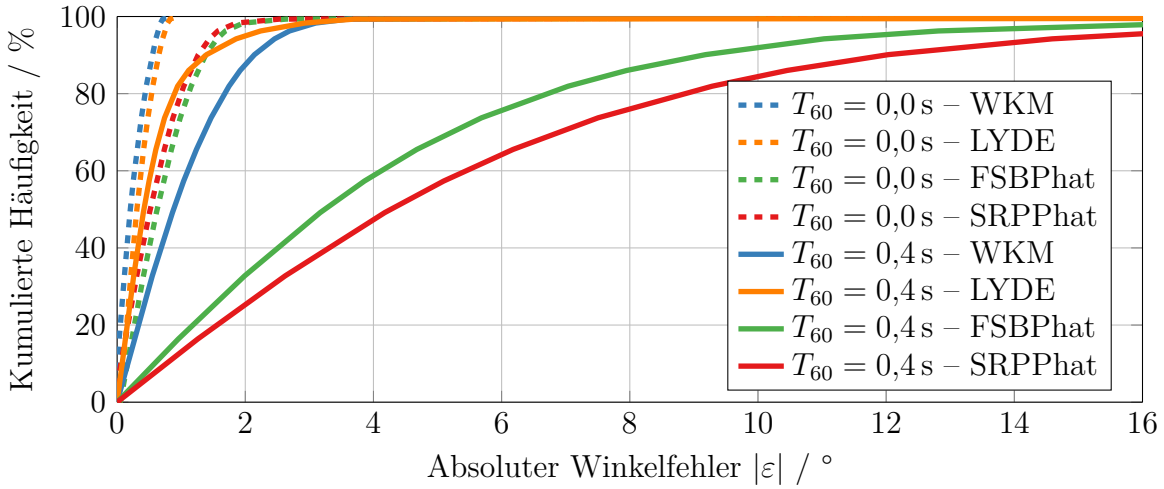


Abbildung 4.9: Kumulative Histogramme des Winkelfehlers ausgewählter Schätzer für ein drei-elementiges Mikrofonarray mit 0,05 m Mikrofonabstand bei verschiedenen Nachhallzeiten.

Fehler der Einfallswinkelschätzung für ein Mikrofonpaar darstellt, gilt es zu beachten, dass in Abb. 4.9, die Skalierung der Abszisse zugunsten einer besseren Darstellung des relevanten Bereiches angepasst wurde. Im jetzt betrachteten Szenario mit drei Mikrofonen pro Array, erzielen alle Winkelschätzer Fehler im akzeptablen Rahmen. Insbesondere der LYDE, aber auch die entwickelte WKM sind sehr robust gegenüber den Störungen durch Nachhall und liefern selbst bei einer Nachhallzeit von 0,4 s in vielen Fällen einen Fehler kleiner als 2°. FSBPhat und SRPPhat schneiden hingegen, wie auch bei der Nutzung von nur zwei Mikrofonen, schlechter ab.

Insgesamt ist die dreieckige Mikrofonanordnung gegenüber einem zwei-elementigen Mikrofonarray klar vorzuziehen, da die Reduktion des Winkelfehlers den zusätzlichen Flächenbedarf und die Erhöhung der Rechenkomplexität eindeutig überwiegt. Grundsätzlich lässt sich der Winkelfehler durch den Einsatz von mehr als drei Mikrofonen noch weiter reduzieren (vgl. [DJH15]). Im Fokus dieser Arbeit stehen jedoch kompakte Arrays mit wenigen Mikrofonen, daher soll diese Option nicht weiter betrachtet werden.

Der bisherige Teil der Analyse ist ausschließlich auf die Qualität der Winkelschätzung bei Störungen durch Nachhall ausgerichtet. Im Bezug auf die Anwendung des Winkelschätzers in realen Umgebungen gilt es jedoch, zusätzlich Störungen durch Rauschen zu berücksichtigen. Zur Modellierung von Hintergrundrauschen, wie es in Büroumgebungen

auftritt, wird wie auch schon in Abschnitt 4.3 isotropes Rauschen verwendet. Ferner beschränken sich die Untersuchungen auf den bisher am besten abschneidenden LYDE und die WKM.

Ausgangspunkt der Simulation ist weiterhin das bereits bekannte Szenario in Kombination mit einem drei-elementigen Mikrofonarray. Die Simulationsergebnisse aus Abb. 4.9 für den rauschfreien Fall (SNR $= \infty$ dB) sollen jetzt mit den Ergebnissen bei einem SNR von 10 dB verglichen werden. Dazu stellt Abb. 4.10 den Fehler der verbleibenden beiden Winkelschätzer bei einem SNR von 10 dB dar. Um die Auswirkungen der Kombination von Nachhall und Rauschen zu verdeutlichen, enthält Abb. 4.10 zusätzlich zu den bislang verwendeten Nachhallzeiten von 0,0 s und 0,4 s auch die bisher nicht dargestellten Ergebnisse für eine Nachhallzeit von 0,2 s.

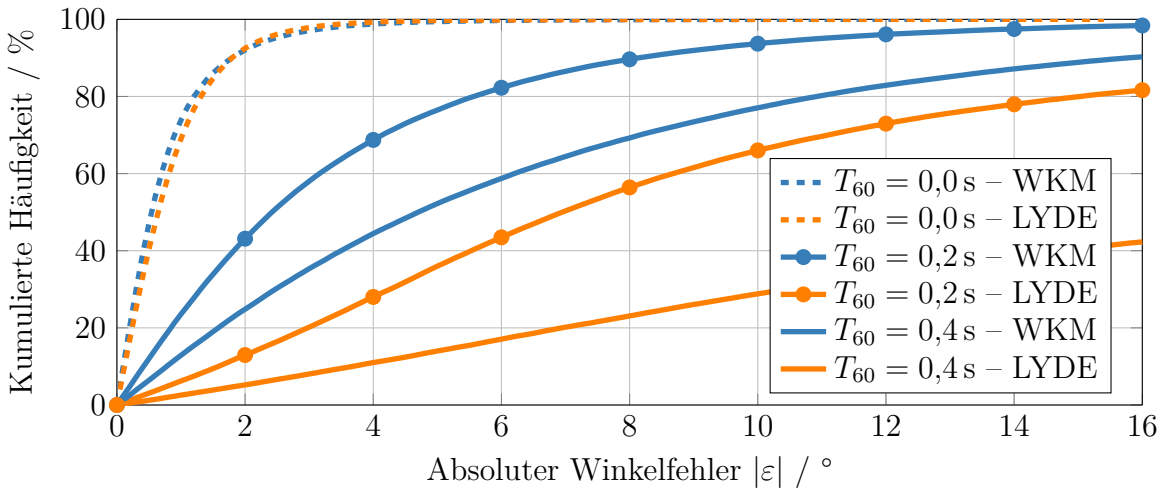


Abbildung 4.10: Kumulative Histogramme des Winkelfehlers ausgewählter Schätzer für ein drei-elementiges Mikrofonarray mit 0,05 m Mikrofonabstand bei verschiedenen Nachhallzeiten und einem SNR von 10 dB.

Der Vergleich von Abb. 4.9 und Abb. 4.10 zeigt unmittelbar die Zunahme des Fehlers durch isotropes Rauschen. Während im rauschfreien Fall der LYDE und die WKM in etwa gleich abschneiden bzw. bei Nachhall sogar eine Präferenz bezüglich des LYDE vorliegt, entsteht durch isotropes Rauschen ein gegensätzliches Bild. Jetzt liefert die WKM präzisere Schätzungen. Besonders deutlich zeigt sich die Robustheit der WKM gegenüber Rauschen dadurch, dass die WKM selbst bei einer Nachhallzeit von 0,4 s noch besser abschneidet als der LYDE bei 0,2 s. Letztendlich ist daher die im Rahmen der Arbeit entwickelte WKM den betrachteten Konkurrenten klar vorzuziehen.

4.5 Bias bei linearen Mikrofonarrays

Die Analyseergebnisse des zurückliegenden Abschnitts dokumentieren die Existenz eines systematischen Fehlers bei der Schätzung des Einfallswinkels (vgl. Abb. 4.6), sofern das betrachtete Mikrofonarray nur über zwei Mikrofone verfügt. Das Ziel dieses Abschnitts ist es deshalb, die Ursachen des auftretenden Bias darzulegen. Ferner werden

die folgenden Ausführungen bestätigen, dass der Bias keine Besonderheit von Arrays mit zwei Mikrofonen ist, sondern durch eine lineare Anordnung der Mikrofone ausgelöst wird.

Die Gemeinsamkeit der bislang untersuchten Algorithmen zur Winkelschätzung besteht darin, dass zunächst eine TDOA-Messung erfolgt und diese anschließend auf einen Einfallswinkel abgebildet wird. Um im weiteren Verlauf eine anschauliche Darstellung zu ermöglichen, beziehen sich die Ausführungen exemplarisch auf die Bestimmung der TDOA durch ein Korrelationsverfahren, wie z. B. GCCPhat bzw. FSBPhat sowie eine Transformation der Signallaufzeitdifferenz gemäß Gl. (4.14). Allerdings sind die Verfahrensschritte, die als Ursache des Bias identifiziert werden, in vergleichbarer Form auch bei den anderen in Abschnitt 4.4 untersuchten Algorithmen vorhanden, sodass die nachfolgenden Darstellungen gleichzeitig auch einen Beleg für das Auftreten des Bias bei diesen Verfahren liefern.

Damit Gl. (4.14) eine Berechnung des Einfallswinkels aus der gemessenen Signallaufzeitdifferenz $\hat{\tau}$ gestattet, muss die Bedingung

$$-\tau_{\max} \leq \hat{\tau} \leq \tau_{\max} \tag{4.27}$$

erfüllt sein, wobei

$$\tau_{\max} = \|\boldsymbol{m}_1 - \boldsymbol{m}_2\|_2 \cdot {c_\mathrm{S}}^{-1} \tag{4.28}$$

gilt. Obwohl es sich bei $\tau_{\max}$ um die maximale TDOA handelt, die aufgrund des vorliegenden Mikrofonabstandes $\|\boldsymbol{m}_1 - \boldsymbol{m}_2\|_2$ und der Schallgeschwindigkeit c_S physikalisch möglich ist, können dennoch größere Messwerte auftreten. Verantwortlich dafür sind die nicht idealen Korrelationseigenschaften der Sprachsignale ebenso wie Störungen durch Nachhall und Rauschen, die für ein Maximum der Phat-Funktion (siehe Gl. (4.12)) außerhalb des physikalisch möglichen Intervalls sorgen können.

Die Bestimmung des Einfallswinkels aus der TDOA besteht daher aus zwei Schritten. Zunächst erfolgt eine Begrenzung (engl. *clipping*) des Messwertes $\hat{\tau}$. Erst die daraus resultierende TDOA

$$\hat{\tau}_\mathrm{c} = \max(-\tau_{\max},\ \min(\hat{\tau},\ \tau_{\max})) \tag{4.29}$$

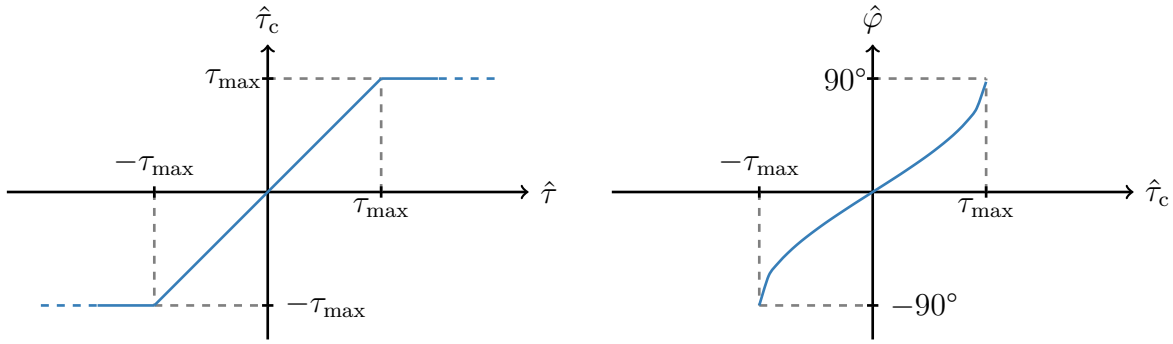
erlaubt eine Berechnung des Einfallswinkels durch Gl. (4.14).

Ausgangspunkt für die weiteren Betrachtungen ist die Annahme, dass die gemessene Signallaufzeitdifferenz $\hat{\tau}$ durch eine Überlagerung der wahren TDOA $\bar{\tau}$ und einem mittelwertfreien, normalverteilten Rauschen mit der Standardabweichung σ_τ modelliert werden kann. Somit lässt sich die Messung $\hat{\tau}$ als Realisierung einer Zufallsvariable auffassen, die die Wahrscheinlichkeitsdichte

$$\mathrm{p}_{\hat{\tau}}(\hat{\tau}; \bar{\tau}, \sigma_\tau) = \frac{1}{\sqrt{2\pi}\sigma_\tau} \cdot \exp\left(\frac{(\hat{\tau} - \bar{\tau})^2}{2\sigma_\tau^2}\right) = \mathcal{N}(\hat{\tau}; \bar{\tau}, \sigma_\tau) \tag{4.30}$$

besitzt.

Der Einfallswinkel φ ist demnach ebenfalls eine Zufallsvariable. Die zugehörige Wahrscheinlichkeitsdichte ergibt sich durch eine Zufallsvariablentransformation, die durch eine Verkettung von Gl. (4.30) und Gl. (4.14) beschrieben wird. Eine graphische Darstellung dieser beiden Gleichungen liefert Abb. 4.11.



(a) Begrenzung der gemessenen TDOA. (b) Berechnung der DOA aus der TDOA.

Abbildung 4.11: Graphische Darstellung der Funktionen zur Bestimmung des Einfallswinkels aus der Signallaufzeitdifferenz.

Die Begrenzung der Messwerte $\hat{\tau}$ auf das Intervall $[-\tau_\mathrm{max}; \tau_\mathrm{max}]$ (vgl. Abb. 4.11a) sorgt für eine Verschiebung der Wahrscheinlichkeitsmasse. Die Masse der Normalverteilung $\mathcal{N}(\hat{\tau}; \bar{\tau}, \sigma_\tau)$, die sich außerhalb des Intervalls befindet, wird auf die obere bzw. untere Intervallgrenze abgebildet. Dementsprechend besitzt $\hat{\tau}_\mathrm{c}$ die Wahrscheinlichkeitsdichte

$$\mathrm{p}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c}; \bar{\tau}, \sigma_\tau) = c_\mathrm{L} \cdot \delta(\hat{\tau}_\mathrm{c}+\tau_\mathrm{max}) + \mathrm{rect}\left(\frac{\hat{\tau}_\mathrm{c}}{2\tau_\mathrm{max}}\right) \cdot \mathrm{p}_{\hat{\tau}}(\hat{\tau}_\mathrm{c}; \bar{\tau}, \sigma_\tau) + c_\mathrm{U} \cdot \delta(\hat{\tau}_\mathrm{c}-\tau_\mathrm{max}), \quad (4.31)$$

$$\text{mit} \quad c_\mathrm{L} = 1 - \mathrm{Q}\left(\frac{-\tau_\mathrm{max} - \bar{\tau}}{\sigma_\tau}\right) \quad \text{und} \quad c_\mathrm{U} = \mathrm{Q}\left(\frac{\tau_\mathrm{max} - \bar{\tau}}{\sigma_\tau}\right). \quad (4.32)$$

Hier bezeichnet $\delta(x)$ den Delta-Impuls, $\mathrm{Q}(y)$ das Integral der Standardnormalverteilung von y bis ∞ und $\mathrm{rect}(x)$ die Rechteck-Funktion mit der Breite 1.

Zur exemplarischen Veranschaulichung des in Gl. (4.31) angegebenen Zusammenhangs, für eine wahre Laufzeitdifferenz $\bar{\tau}$, dient Abb. 4.12. Die Delta-Impulse sind dabei durch Pfeile gekennzeichnet, deren Länge zum Gewicht der Impulse korrespondiert.

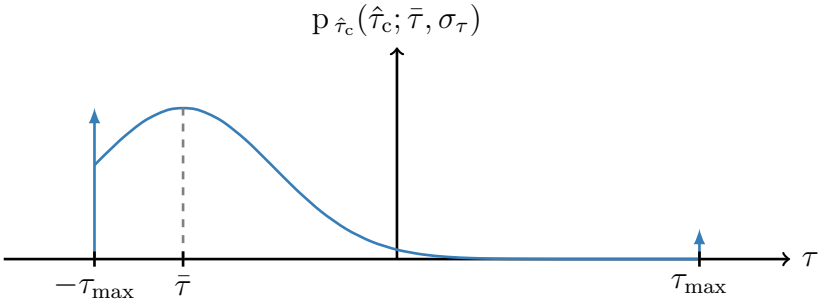


Abbildung 4.12: Verteilungsdichtefunktion der Signallaufzeitdifferenzen $\hat{\tau}_\mathrm{c}$ nach der Begrenzung der Messwerte $\hat{\tau}$ auf das Intervall $[-\tau_\mathrm{max}; \tau_\mathrm{max}]$.

Bereits jetzt liefert die Bestimmung des Erwartungswertes

$$\mathbb{E}[\hat{\tau}_\mathrm{c}] = \int_{-\tau_\mathrm{max}}^{\tau_\mathrm{max}} \hat{\tau}_\mathrm{c} \cdot \mathrm{p}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c}; \bar{\tau}, \sigma_\tau)\, \mathrm{d}\hat{\tau}_\mathrm{c}, \quad (4.33)$$

den Beleg dafür, dass die Begrenzung der TDOA-Messungen $\hat{\tau}$ einen Bias in $\hat{\tau}_\mathrm{c}$ verursacht. Bei der Berechnung des Erwartungswertes bezüglich $\hat{\tau}_\mathrm{c}$ gilt es zudem zu berücksichtigen, dass $\mathrm{p}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c}; \bar{\tau}, \sigma_\tau)$ von der tatsächlichen Laufzeitdifferenz $\bar{\tau}$ abhängt. Dementsprechend ist das Ergebnis von Gl. (4.33) eine Funktion von $\bar{\tau}$. Für die Berechnung des Erwartungswertes sei auf Anhang A.2 verwiesen. An dieser Stelle soll lediglich Abb. 4.13 zur Veranschaulichung des Bias der Signallaufzeitdifferenz $\mathbb{E}[\hat{\tau}_\mathrm{c}] - \bar{\tau}$ dienen.

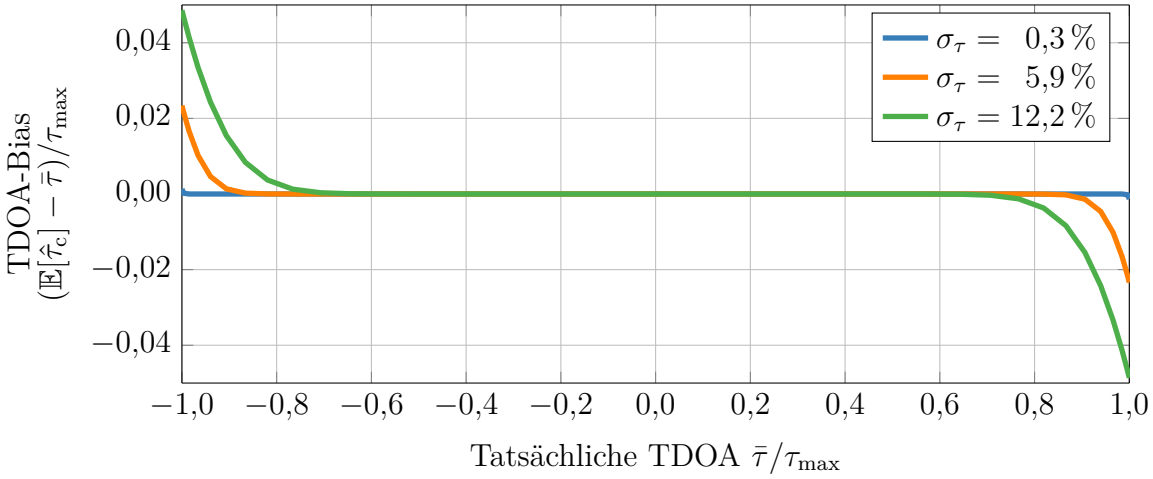


Abbildung 4.13: Durch die Begrenzung der TDOA-Messungen ausgelöster Bias der Signallaufzeitdifferenzen in Abhängigkeit der tatsächlichen TDOA.

Um eine Darstellung der TDOA sowie des Bias unabhängig vom konkreten Abstand der Mikrofone zu gestatten, sind die Werte in Abb. 4.13 ebenso wie auch alle weiteren TDOA stets im Verhältnis zu τ_max angegeben. Die gewählten Standardabweichungen σ_τ sollen dabei ungefähr die Nachhallzeiten von 0,0 s, 0,2 s bzw. 0,4 s aus Abschnitt 4.4 widerspiegeln. Damit auch die Standardabweichungen losgelöst vom Mikrofonabstand betrachtet werden können, sind diese als prozentualer Anteil von τ_max angegeben.

Nach dem die bisherigen Ausführungen eindeutig auf einen Bias bei der Vorverarbeitung der TDOA hinweisen, gilt es nun zu analysieren, wie sich dieser Bias auf den gesuchten Winkels auswirkt. Dafür soll die Wahrscheinlichkeitsdichte des Einfallswinkels in Abhängigkeit der tatsächlichen Verzögerung $\bar{\tau}$ ermittelt werden. Grundlage der dazu verwendeten Transformation ist die Wahrscheinlichkeitsdichte $\mathrm{p}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c}; \bar{\tau}, \sigma_\tau)$ (vgl. Gl. (4.31)) und der in Abb. 4.11b veranschaulichte Zusammenhang zur Berechnung des DOA-Schätzwertes $\hat{\varphi}$ aus der Verzögerung $\hat{\tau}_\mathrm{c}$. Entsprechend der in Anhang A.3 näher erläuterten Schritte, entsteht für den Einfallswinkel die Dichte

$$\mathrm{p}_{\hat{\varphi}}(\hat{\varphi}; \bar{\tau}, \sigma_\tau) = c_\mathrm{L} \cdot \delta\left(\hat{\varphi} + \frac{\pi}{2}\right) + \tau_\mathrm{max} \cdot \mathcal{N}(\sin(\hat{\varphi}) \cdot \tau_\mathrm{max}; \bar{\tau}, \sigma_\tau) \cdot \cos(\hat{\varphi}) + c_\mathrm{U} \cdot \delta\left(\hat{\varphi} - \frac{\pi}{2}\right). \tag{4.34}$$

Exemplarisch für die in Abb. 4.12 veranschaulichte Verteilung der TDOA nach Begrenzung auf das Intervall $[-\tau_\mathrm{max}; \tau_\mathrm{max}]$, visualisiert Abb. 4.14 die entsprechende Verteilung $\mathrm{p}_{\hat{\varphi}}(\hat{\varphi}; \bar{\tau}, \sigma_\tau)$ nach der Transformation zu einem Einfallswinkel.

Die Bestimmung des Bias der Einfallswinkelschätzung erfordert abschließend noch die Berechnung des Erwartungswertes von Gl. (4.34). Auch hier sei für die Details auf Anhang A.3 verwiesen.

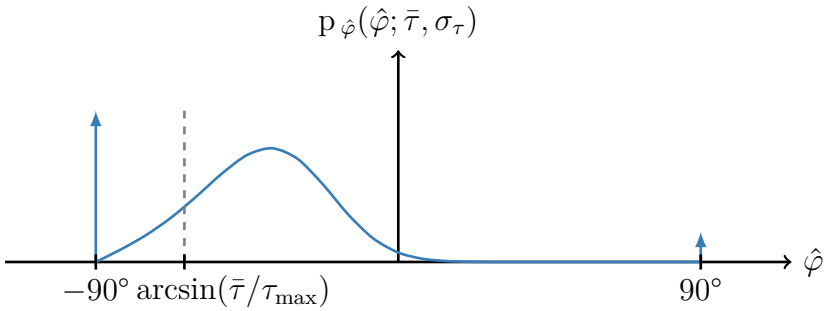


Abbildung 4.14: Verteilungsdichtefunktion des Einfallswinkels bei einer tatsächlichen Verzögerung von $\bar{\tau}$.

Damit ein Vergleich des prognostizierten Bias der Einfallswinkelschätzung mit den Untersuchungsergebnissen aus Abschnitt 4.4 möglich ist, zeigt auch Abb. 4.15 den Bias als Funktion des wahren Einfallswinkels φ. Der dazu erforderliche Einfallswinkel φ, der zur tatsächlichen TDOA $\bar{\tau}$ gehört, lässt sich wie folgt bestimmen:

$$\varphi = \arcsin(\bar{\tau}/\tau_{\max}). \tag{4.35}$$

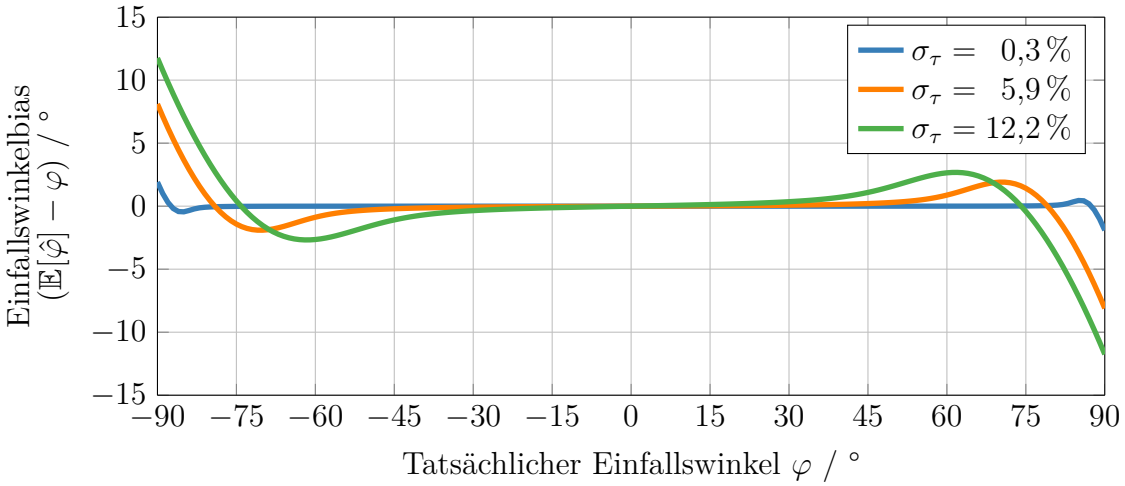


Abbildung 4.15: Durch die Begrenzung der TDOA-Messungen ausgelöster Bias der Einfallswinkel in Abhängigkeit des tatsächlichen Einfallswinkels.

Bei einer Gegenüberstellung des vorhergesagten Bias mit den in Abb. 4.6 dargestellten Untersuchungsergebnissen ergeben sich klare Differenzen. Obwohl die in Abb. 4.15 verwendeten Standardabweichungen in etwa die Nachhallzeiten aus Abb. 4.6 widerspiegeln, fällt der vorhergesagte Bias viel kleiner als bei den Messungen aus. Die größte Übereinstimmung zwischen Vorhersage und Messung liegt bei Einfallswinkeln in der Nähe von ±90° vor. Hier deuten beide auf eine deutliche Abweichung in Richtung der zentralen Position (0°). Für kleinere Winkel sagt das Modell hingegen einen Bias in die entgegengesetzte Richtung vorher. Außerdem prognostiziert das Modell bei Winkeln im Bereich von ±20° gar keinen systematischen Fehler. Zwar sind Abweichungen der tatsächlichen Messwerte vom Modell zu erwarten, da eine Approximation des Fehlers

der Laufzeitmessung durch eine Normalverteilung nur eine grobe Näherung darstellt, gleichwohl sind die auftretenden Unterschiede zu groß um das bislang vorhandene Modell zur Beschreibung des Winkelfehlers zu verwenden.

Die folgenden Ausführungen sollen daher zur Präzisierung des bisherigen Modells dienen. Im Zentrum der Ausführungen stehen erneut die Signallaufzeitdifferenzen $\hat{\tau}_\mathrm{c}$, die die Grundlage für die Bestimmung des Einfallswinkels darstellen. Da diese durch eine Begrenzung der TDOA-Messwerte $\hat{\tau}$ auf das Intervall $[-\tau_\mathrm{max}; \tau_\mathrm{max}]$ entstehen, ergibt sich bereits dort ein Bias (vgl. Gl. (4.33)). Auch wenn die Größe und Ausdehnung dieses Bias von der Standardabweichung σ_τ abhängt, beschränkt er sich hauptsächlich auf die Bereiche in denen die tatsächliche TDOA $\bar{\tau}$ größer als $\pm 0{,}7$ ($\pm 45°$) ist (siehe Abb. 4.13).

Zur Überprüfung, inwieweit die aus den akustischen Signalen extrahierten Signallaufzeitdifferenzen der Vorhersage des Modells entsprechen, dient Abb. 4.16. Sie zeigt exemplarisch für eine TDOA-Schätzung durch FSBPhat den Bias in Abhängigkeit der tatsächlichen Verzögerung $\bar{\tau}$.

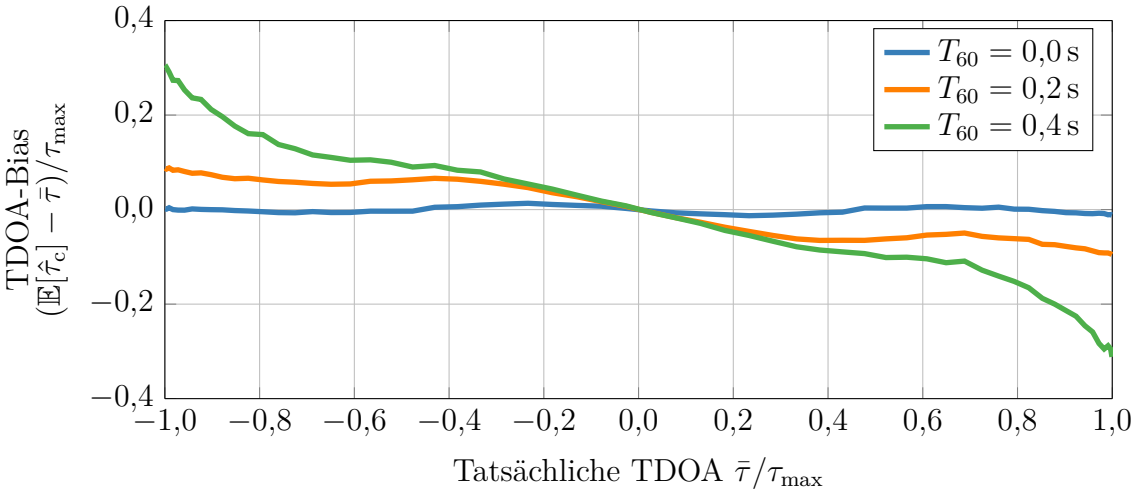


Abbildung 4.16: Bias der TDOA-Schätzung durch FSBPhat für verschiedene Nachhallzeiten bei der Verwendung eines zwei-elementigen Mikrofonarrays (Mikrofonabstand: 0,05 m).

Eine Gegenüberstellung zwischen Modell (siehe Abb. 4.13) und tatsächlich gemessenem Bias (siehe Abb. 4.16) bestätigt unmittelbar, dass bereits beim Bias der TDOA eine deutliche Abweichung vorliegt. Einerseits fällt der tatsächliche Bias viel größer als prognostiziert aus und andererseits erstreckt er sich auf nahezu den gesamten Bereich von -1 bis 1 und nicht nur auf die Verzögerungen, die zu Winkeln größer als $\pm 45°$ gehören. Dementsprechend ist die bei den Einfallswinkeln festgestellte Diskrepanz zwischen der Vorhersage und den Messungen auf die schon bei den Laufzeitdifferenzen vorhandenen Unterschiede zurückzuführen.

Da der Anteil des TDOA-Bias, für den das bisherige Modell noch keine Erklärung liefert, eindeutig von der Nachhallzeit abhängt, sollen nun die Auswirkungen des Nachhalls auf die Schätzung der TDOA detaillierter betrachtet werden. Außerdem besitzt das verwendete Sprachsignal keine idealen Korrelationseigenschaften und kommt deshalb ebenfalls als mögliche Ursache für den Bias infrage. Allerdings kann das Sprachsignal

als Grund für den Bias unmittelbar ausgeschlossen werden, weil dieser selbst dann vorhanden ist, wenn anstatt der akustischen Signale die RIA als Eingangssignale für die TDOA-Schätzung herangezogen werden.

Die Grundlage für alle betrachteten Winkelschätzer bildet die Annahme, dass der TDOA-Schätzwert mit dem direkten Pfad (LOS) von der Quelle zum Mikrofonarray korrespondiert. Ohne Nachhall existiert nur die LOS-Komponente, dementsprechend trifft die zuvor genannte Annahme zu. Durch Reflexionen ergeben sich jedoch neben der LOS-Komponente zahlreiche weitere Signalpfade, die aus unterschiedlichen Richtungen auf das Mikrofonarray einfallen und daher die TDOA-Schätzung ebenfalls beeinflussen. Die Signallaufzeitdifferenzen zwischen den Reflexionen sorgen daher zusätzlich zur bereits modellierten Streuung der TDOA-Schätzung in der Nähe der tatsächlichen Verzögerung für weitere Anteile bei der Verteilung der TDOA. Um diese Anteile abbilden zu können, wird das bisherige Modell $\mathrm{p}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c}; \hat{\tau}, \sigma_\tau)$ (vgl. Gl. (4.31)) mit der Dichte $\mathrm{p}_{\,\mathrm{Hall}}(\hat{\tau}_\mathrm{c}{}')$, die die Verteilung der TDOA der Reflexionen beschreiben soll, kombiniert:

$$\mathrm{p}_{\,\hat{\tau}_\mathrm{c}{}'}(\hat{\tau}_\mathrm{c}{}'; \hat{\tau}, \sigma_\tau) = (1-\alpha) \cdot \mathrm{p}_{\,\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c}{}'; \hat{\tau}, \sigma_\tau) + \alpha \cdot \mathrm{p}_{\,\mathrm{Hall}}(\hat{\tau}_\mathrm{c}{}'). \tag{4.36}$$

Der Faktor α modelliert dabei das Verhältnis zwischen der direkten Signalkomponente und den Reflexionen und hängt daher von der Nachhallzeit ab. Ohne Nachhall ($\alpha = 0$) entsteht somit das bereits in Gl. (4.30) präsentierte Modell. Damit eine eindeutige Unterscheidung zwischen dem bisherigen und dem aktuell betrachteten Modell möglich ist, wird die bislang durch $\hat{\tau}_\mathrm{c}$ gekennzeichnete TDOA nun durch $\hat{\tau}_\mathrm{c}{}'$ dargestellt.

Auf den ersten Blick scheint der Parameter α dem *Direct-to-Reverberant Ratio* (DRR) zu entsprechen. Allerdings charakterisiert das DRR die Energie des direkten Pfades sowie der frühen Reflexionen im Verhältnis zum späten Nachhall [Jeu+11]. Die TDOA-Schätzung wird jedoch insbesondere von den frühen Reflexionen beeinflusst, da diese für sehr viel Energie aus Raumrichtungen sorgen, die nicht mit der gesuchten LOS-Komponente korrespondieren. Der späte Nachhall ist hingegen deutlich unproblematischer. Im Vergleich zu den frühen Reflexionen hat der späte Nachhall eine viel geringere Energie und dementsprechend ist auch der Einfluss auf die TDOA-Schätzung bedeutend schwächer. Darüber hinaus wird in der statistischen Akustik die Phase der Ebenenwellen, die den späten Nachhall repräsentieren, als Zufallsprozess modelliert. Sofern die in [Sch62; LD12] erläuterten Randbedingungen bezüglich Raumgröße und geometrischer Anordnung von Quelle und Mikrofonen erfüllt sind, lässt sich die Phase durch eine Gleichverteilung modellieren. Daraus ergibt sich wiederum, dass auch die Einfallsrichtungen des späten Nachhalls gleichverteilt sind. Demzufolge ist der Erwartungswert der Einfallsrichtungen des späten Nachhalls Null, sodass die Auswirkung der späten Nachhalls auf die TDOA-Schätzung vernachlässigt werden kann. Der Parameter α aus Gl. (4.36) beschreibt deshalb eher das Verhältnis zwischen der Energie der LOS-Komponente und der Energie der frühen Reflexionen.

Zur Untersuchung inwiefern sich der Nachhall auf die Verteilung der TDOA auswirkt, soll das Histogramm der gemessenen Laufzeitdifferenzen für einen gegebenen Einfallswinkel genutzt werden. Exemplarisch für einen Einfallswinkel von $\varphi = -70°$ bzw. die dazu korrespondierende TDOA $\bar{\tau} = -0{,}94$ und eine Nachhallzeit von 0,4 s zeigt Abb. 4.17 das Histogramm der TDOA-Schätzungen der auch in Abschnitt 4.4 untersuchten Szenarien. Die Klassenbreite variiert in dieser Darstellung, da sich die Einteilung der Klassen aus einer gleichmäßigen Unterteilung der zu den TDOA gehörenden DOA ergibt.

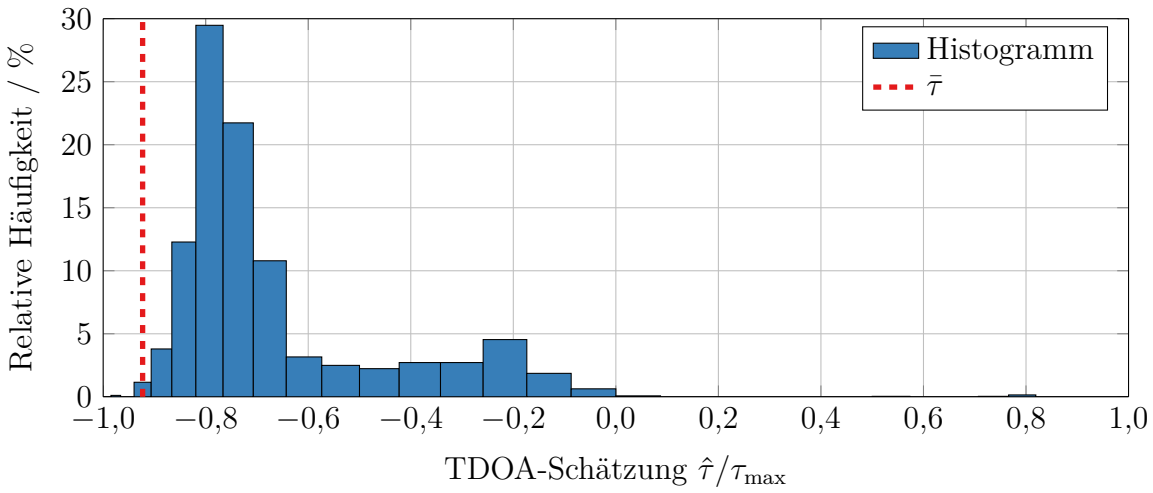


Abbildung 4.17: Histogramm der TDOA-Schätzungen durch FSBPhat für Konfigurationen mit einer tatsächlichen Laufzeitdifferenz $\bar{\tau} = -0{,}94$ ($-70°$) bei einer Nachhallzeit von 0,4 s sowie der Verwendung eines Mikrofonarrays mit zwei Mikrofonen im Abstand von 0,05 m.

Das Histogramm in Abb. 4.17 liefert erneut einen deutlichen Hinweis auf einen Bias, da nahezu alle Schätzwerte kleiner als der wahre Wert $\bar{\tau}$ ausfallen. Die Schätzungen, die um den *Mode* des Histogramms streuen, lassen sich durch eine Normalverteilung annähern und gehören so zu dem bereits durch $\mathrm{p}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c}; \hat{\tau}, \sigma_\tau)$ modellierten Anteil. Allerdings sorgt der Nachhall dafür, dass der Mittelwert nicht mehr $\bar{\tau}$ entspricht, sondern sich ebenfalls in Richtung 0,0 verschiebt. Außerdem ist rechts von der näherungsweise normalverteilten Komponente ein weiterer Anteil zu erkennen. Dieser Anteil verursacht eine rechtsschiefe Verteilung. Je weiter sich jedoch die tatsächliche TDOA 0,0 annähert, desto symmetrischer wird die Verteilung. Wenn die wahre TDOA positiv ist, entsteht hingegen eine linksschiefe Verteilung. Insgesamt deutet das Histogramm aus Abb. 4.17 auf eine systematische, nachhallzeitabhängige Verschiebung des Mittelwertes hin und liefert somit eine Erklärung für den bislang ungeklärten Anteil des TDOA-Bias.

Zur Erläuterung warum der Nachhall zu der in Abb. 4.17 dargestellten Verteilung der TDOA führt, dient Abb. 4.18. Sie skizziert ein beispielhaftes Szenario bei dem die direkte Komponente des Signals (schwarzer Kopf) aus einem Winkel von $-70°$ (vgl. Abb. 4.17) auf das Array (blaue Kreise) einfällt. Das schwarze Rechteck in der Mitte symbolisiert dabei die Wände des Raumes, während die grauen Köpfe in den grauen Rechtecken die Positionen der Spiegel-Quellen beschreiben, die durch eine einfache Spiegelung der ursprünglichen Quelle entstehen. Gemäß den vorausgegangenen Ausführungen haben die Spiegel-Quellen, die die Reflexion an nur einer Wand modellieren, die größte Energie und dementsprechend auch den größten Einfluss auf die TDOA-Schätzung.

Anhand von Abb. 4.18 ist weiterhin zu erkennen, dass die Spiegel-Quellen links unten und rechts eine TDOA verursachen, die betragsmäßig kleiner als die Verzögerung auf dem direkten Pfad ausfällt. Lediglich die obere Spiegel-Quelle hat eine Laufzeitdifferenz die größer als bei der LOS-Komponente ist. Außerdem ist bei der Betrachtung von Abb. 4.18 zu berücksichtigen, dass aufgrund der linearen Anordnung der Mikrofone nur ein eindeutiger Detektionsbereich von 180° vorliegt. Dementsprechend kann die Spiegel-Quelle im linken Raum auch durch eine äquivalente Quelle repräsentiert werden, die

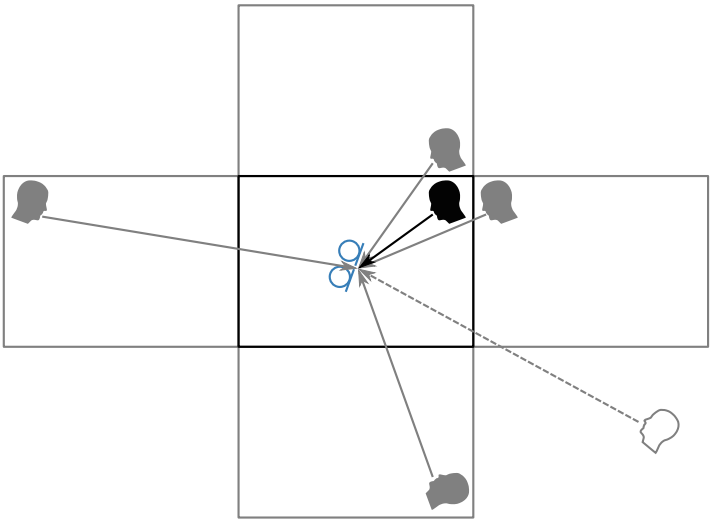


Abbildung 4.18: Einfallsrichtung der LOS-Komponente (schwarzer Kopf) und den Reflexionen erster Ordnung (graue Köpfe). Der graue, nicht-ausgefüllte Kopf symbolisiert die äquivalente Position der Spiegel-Quelle aus dem linken Raum, die aufgrund des begrenzten Sichtfeldes des linearen Arrays entsteht.

durch eine Spiegelung dieser Spiegel-Quelle an der Achse des Mikrofonpaares entsteht. Diese gespiegelte Variante der linken Spiegel-Quelle ist in Abb. 4.18 durch einen grauen, nicht-ausgefüllten Kopf gekennzeichnet. Somit wird die ungleichmäßige Verteilung der Reflexionen auch optisch erkennbar. Darüber hinaus nimmt die ungleichmäßige Verteilung der Reflexionen zu, je näher sich die Verzögerung der LOS-Komponente an ± 1 ($\pm 90°$) befindet und liefert deshalb eine Erklärung warum die TDOA vorzugsweise unterschätzt wird.

Ein exaktes Modell für die Verteilungsdichte $\mathrm{p}_{\mathrm{Hall}}(\hat{\tau}_{\mathrm{c}}')$ lässt sich anhand der graphischen Darstellung jedoch nicht gewinnen. Zur Abschätzung des Einflusses des Nachhalls auf die TDOA-Schätzung soll daher die RIA dienen. Die RIA von einer Signalquelle zu einem Mikrofon kann als gewichtete Summe zeitverzögerter Impulse dargestellt werden [AB79]:

$$h(t) = \sum_{r'=0}^{R} \upsilon_{r'} \cdot \delta(t - t_{r'}). \tag{4.37}$$

Dabei bezeichnet $\upsilon_{r'}$ die Dämpfung auf dem r'-ten Ausbreitungspfad und $t_{r'}$ die Verzögerung dieses Pfades bezüglich der LOS-Komponente ($r' = 0$). Sofern die Verzögerungen $t_{r'}$ und $t_{r''}$, die zu den Raumimpulsantworten von der Quelle zum ersten bzw. zweiten Mikrofon eines Paares gehören, ebenso wie die Dämpfungen $\upsilon_{r'}$ und $\upsilon_{r''}$ vorliegen, ergibt sich daraus die durchschnittliche TDOA des Mikrofonpaares

$$\hat{\tau}_{\mathrm{RIA}} = \frac{\sum\limits_{r'}^{R}\sum\limits_{r''}^{R} \upsilon_{r'} \cdot \upsilon_{r''} \cdot (t_{r'} - t_{r''})}{\sum\limits_{r'}^{R}\sum\limits_{r''}^{R} \upsilon_{r'} \cdot \upsilon_{r''}}. \tag{4.38}$$

Die Auswertung von Gl. (4.38) erfordert demnach die Kenntnis der RIA von der Quelle zu den beiden Mikrofonen.

Zur Untersuchung inwiefern die ungleichmäßige Verteilung der frühen Reflexionen für den festgestellten Bias verantwortlich ist, wird die Spiegel-Quellen-Methode (vgl. Abschnitt 3.2) herangezogen, da diese die Berechnung der notwendigen Verzögerungen ($t_{r'}$ und $t_{r''}$) und Dämpfungen ($\upsilon_{r'}$ und $\upsilon_{r'}$) gestattet. Aus der Differenz zwischen der durchschnittlichen TDOA der RIA ($\hat{\bar{\tau}}_{\mathrm{RIA}}$) und der wahren TDOA $\bar{\tau}$ ergibt sich schließlich eine Abschätzung des nachhallbedingten Bias linearer Mikrofonarrays. Um den Einfluss der frühen Reflexionen zu verdeutlichen, stellt Abb. 4.19 sowohl den Bias dar, wenn in die Berechnung der TDOA alle Reflexionen miteinbezogen werden ($R = R_{\max}$), als auch wenn nur die Reflexionen erster und zweiter Ordnung ($R = 2$) berücksichtigt werden.

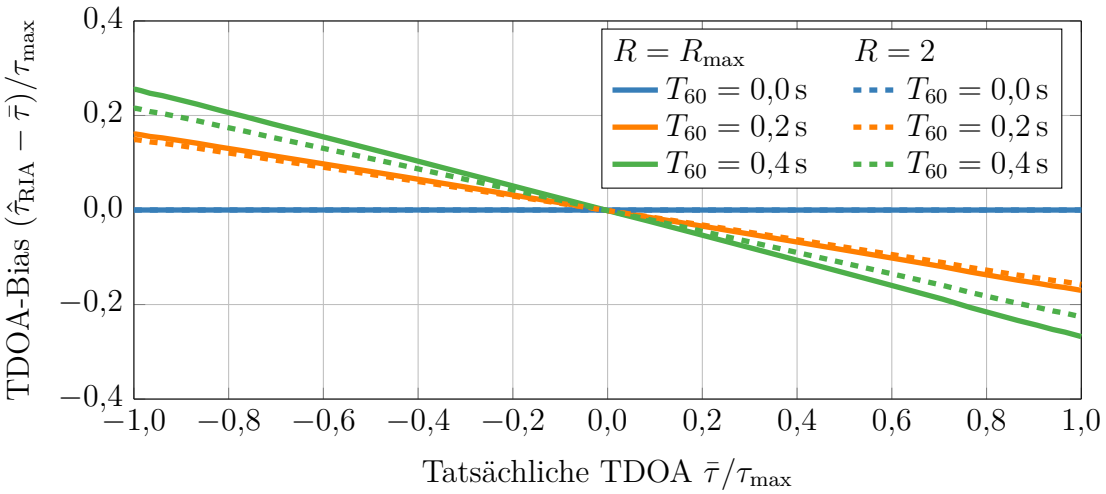


Abbildung 4.19: TDOA-Bias der RIA in Abhängigkeit der tatsächlichen TDOA bei verschiedenen Nachhallzeiten.

Die dominierende Wirkung der frühen Reflexionen zeigt sich speziell anhand der Ergebnisse für eine Nachhallzeit von 0,2 s. In Abb. 4.19 ist nahezu kein Unterschied zwischen dem Bias vorhanden, der sich ergibt, wenn anstatt der gesamten RIA nur die Reflexionen erster und zweiter Ordnung in die Berechnung der TDOA einfließen. Bei einer Nachhallzeit von 0,4 s ist hingegen eine klare Differenz zu erkennen. Durch die Berücksichtigung zusätzlicher Reflexionen ($R \geq 5$) verschwindet jedoch auch die bislang vorhandene Differenz.

Ein Vergleich des in Abb. 4.19 dargestellten Bias der RIA mit den bisherigen Ergebnissen (siehe Abb. 4.16) zeigt außerdem, dass die Vorhersage des Bias durch die durchschnittliche TDOA der RIA, insbesondere bei einer Nachhallzeit von 0,2 s, größer ausfällt als der tatsächliche Bias. Dabei ist jedoch zu beachten, dass das Modell aus Gl. (4.38) nur eine Näherung darstellt, die wichtige Aspekte, wie z. B. den Einfluss des Sprachsignals, gänzlich unberücksichtigt lässt. Da sich aber die Daten aus Abb. 4.16 ungefähr mit den am Beispiel von Abb. 4.18 geschilderten Eigenschaften decken, erscheint die Modellierung des nachhallbedingten Bias durch einen linearen Anteil gerechtfertigt.

Eine exakte Berechnung des Bias würde die Kenntnis der Verteilung $\mathrm{p}_{\mathrm{Hall}}(\hat{\tau}_{\mathrm{c}}')$ erfordern, da diese jedoch unbekannt ist, erfolgt hier eine qualitative Betrachtung. Der

Nachhall besteht aus frühen Reflexionen und spätem Nachhall. Gemäß der vorherigen Ausführungen bzw. anhand von [Sch62; LD12] ergibt sich, dass der späte Nachhall durch eine Gleichverteilung beschrieben werden kann. Daher liefert er keinen Beitrag bei der Berechnung des Erwartungswertes. Der Erwartungswert der frühen Reflexion fällt entsprechend der Simulationen kleiner als er tatsächliche Winkel aus und wird als linear angenommen. Insgesamt wird der Erwartungswert von Gl. (4.36) daher durch

$$\mathbb{E}[\hat{\tau}_\mathrm{c}{}'] \approx (1-\alpha)\cdot\mathbb{E}[\hat{\tau}_\mathrm{c}] \tag{4.39}$$

approximiert. Somit lässt sich der TDOA-Bias als Kombination des bisherigen Verlaufs (siehe Abb. 4.13) und einem linearen, durch den Nachhall ausgelösten, Anteil auffassen. Eine Berechnung des Bias setzt damit die Kenntnis des Parameters α voraus. Da dieser aber das Verhältnis von der LOS-Komponente zu den frühen Reflexionen beschreibt, wird somit die Kenntnis der RIA erforderlich.

Zur Überprüfung, inwieweit das jetzt vorliegende Modell des TDOA-Bias zu dem gemessenen Bias passt, wird der Parameter α so gewählt, dass der Abstand zwischen Messung und Prädiktion minimiert wird. Bei der Gegenüberstellung in Abb. 4.20 zwischen dem gemessenen Bias (vgl. Abb. 4.16) und der Approximation durch das erweiterte Modell (siehe Gl. (4.39)) ergibt sich eine große Übereinstimmung. Daher liefert das Modell eine geeignete Beschreibung des bislang ungeklärten Anteils des Bias.

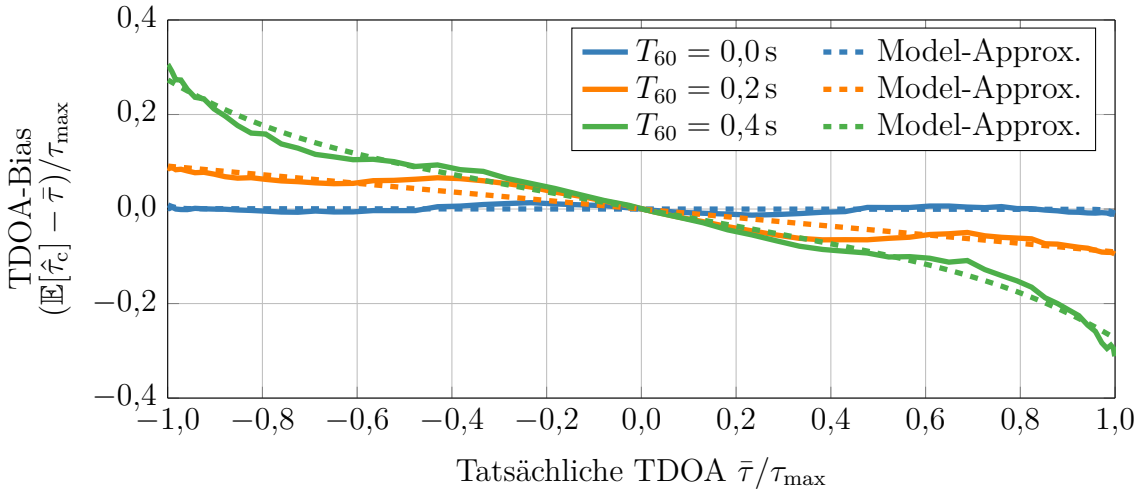


Abbildung 4.20: Approximation des in Abb. 4.13 gezeigten TDOA-Bias durch das entwickelte Modell.

Obwohl sich die bisherigen Betrachtungen auf FSBPhat beziehen, liefern sie gleichzeitig auch einen Beleg für den Bias bei den anderen Winkelschätzern. Die bei FSBPhat explizit durchgeführte Begrenzung der Messwerte findet bei SRPPhat, dem LYDE sowie der WKM implizit statt, weil der *Score* ebenfalls nur für Winkel aus dem Intervall $[-90°; 90°]$ bzw. die dazu korrespondieren TDOA berechnet wird. Der zweite und dominierende Anteil des systematischen Fehlers entsteht durch den Nachhall und ist daher bei allen bisher betrachteten Algorithmen vorhanden. Außerdem sind beide Anteile des Bias eine Folge der linearen Anordnung der Mikrofone und somit auch unabhängig davon, ob das Array zwei oder mehr Mikrofone hat.

Die bislang untersuchten Winkelschätzer, insbesondere GCCPhat und SRPPhat, gelten als hall-robust. Gleichwohl löst die lineare Anordnung der Mikrofone in Kombination mit dem Nachhall bei GCCPhat, SRPPhat, FSBPhat, dem LYDE und der WKM den gezeigten Bias aus. Ferner besteht die Gemeinsamkeit dieser Schätzer darin, dass sie versuchen die LOS-Komponente zu identifizieren ohne dabei explizit die Mehrwegeausbreitung des Signals zu modellieren. Der vorhandene Bias weist jedoch eindeutig darauf hin, dass diese Vorgehensweise bei einer linearen Anordnung der Mikrofone unzureichend ist. Eine Alternative könnte deshalb die Berücksichtigung der Mehrwegeausbreitung direkt bei der Winkelschätzung darstellen. Das kürzlich vorgestellte Verfahren [Jen+16] bezieht die Mehrwegeausbreitung explizit mit ein und bestimmt sowohl die Richtung der LOS-Komponente als auch der frühen Reflexionen. Dadurch könnte der nachhallbedingte und damit der dominierende Anteil des Bias vermieden werden.

Falls das Array aus mehr als zwei Mikrofonen besteht, die nicht auf einer Linie angeordnet sind, verschwindet der Bias, wie bereits durch die Ergebnisse in Abschnitt 4.4 dokumentiert. Für den Teil des Bias, der durch die Begrenzung der TDOA entsteht, leuchtet dies unmittelbar ein, da ein Mikrofonarray, dessen Mikrofone sich nicht auf einer Linie befinden, eine vollständige Rundumsicht gestattet und daher keine Begrenzung der Daten erfordert. Dass jedoch auch der lineare Anteil des Bias entfällt, der aufgrund des Nachhalls auftritt, bedarf einer weiteren Erläuterung.

Im jetzt betrachteten Fall besitzt ein Array mindestens drei Mikrofone, demnach ergeben sich daraus mindestens zwei TDOA-Schätzwerte, die gemeinsam mit der geometrischen Anordnung in die Bestimmung der DOA einfließen. Um dennoch eine Betrachtung des Bias unabhängig vom Aufbau der Mikrofongruppe zu gestatten, wird anstatt der bisher betrachteten TDOA-Schätzwerte nun direkt der Einfallswinkel genutzt. Das bislang verwendete Modell beschreibt die gemessene TDOA als Überlagerung der TDOA der LOS-Komponente und der TDOA der Reflexionen. Der Einfallswinkel lässt sich entsprechend als Überlagerung des Winkels der LOS-Komponente mit den Winkeln der Reflexionen auffassen.

Ferner sorgt die vollständige Rundumsicht dafür, dass es sich bei den Einfallswinkeln nun um eine 2π-periodische Größe handelt. Bei der Berechnung des Mittelwertes ist deshalb die Periodizität zu berücksichtigen und daher der Mittelwert für periodische (zirkuläre) Größen zu verwenden. Die Auswirkung des zirkulären Mittelwertes zeigt sich bei einer erneuten Betrachtung von Abb. 4.18. Während bei einer linearen Anordnung, angesichts des auf 180° beschränkten Detektionsbereiches, eine ungleichmäßige Verteilung der Reflexionen eintritt, kompensieren sich die Richtungen bei der Berechnung des zirkulären Mittelwertes annähernd. Damit ist auch dieser Teil des Bias eine Folge der linearen Anordnung und verschwindet deshalb ebenfalls, wenn die Mikrofone nicht auf einer Linie positioniert sind.

4.6 Fehlermodell der Einfallswinkelschätzung

Die Untersuchungen aus Abschnitt 4.4 haben gezeigt, dass die WKM für die dort betrachteten Szenarien die besten Einfallswinkelschätzungen liefert. Anknüpfend an die Untersuchungsergebnisse der letzten beiden Abschnitte soll jetzt ein Modell für den Fehler der Einfallswinkelschätzung entwickelt werden, um in den folgenden Kapiteln

mithilfe dieses Modells einen Geometriekalibrierungsalgorithmus zu entwickeln. Berücksichtigt werden dabei jedoch ausschließlich die dreieckigen Mikrofonarrays, da der bei linearen Arrays auftretende Bias eine viel zu große Beeinträchtigung dargestellt.

Die bisher zum Vergleich der Winkelschätzer genutzten kumulativen Histogramme ermöglichen bereits eine Einschätzung des Winkelfehlers. Um einen detaillierteren Einblick zu erhalten, wird nun die Verteilung des Fehlers betrachtet. Grundlage für die weiteren Darstellungen bilden die schon in Abschnitt 4.4 verwendeten Daten. Beim Einsatz des dreieckigen Mikrofonarrays ergeben sich im rauschfreien Fall (SNR $= \infty$) die in Abb. 4.21 gezeigten Histogramme. Abb. 4.22 stellt darüber hinaus die Histogramme des Fehlers für ein SNR von 10 dB dar. Bei einem Vergleich von Abb. 4.22 mit Abb. 4.21 gilt es jedoch zu berücksichtigen, dass die Skalierung beider Achsen und der dargestellte Wertebereich angepasst wurden.

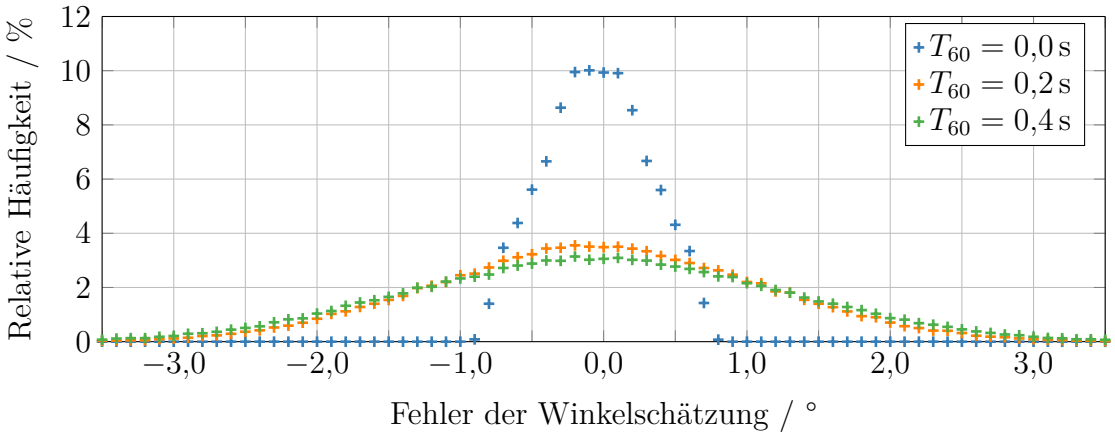


Abbildung 4.21: Histogramme des Fehlers der Winkelschätzung mittels WKM, bei der Nutzung eines dreieckigen Mikrofonarrays mit 0,05 m Kantenlänge für verschiedene Nachhallzeiten und SNR $= \infty$ dB.

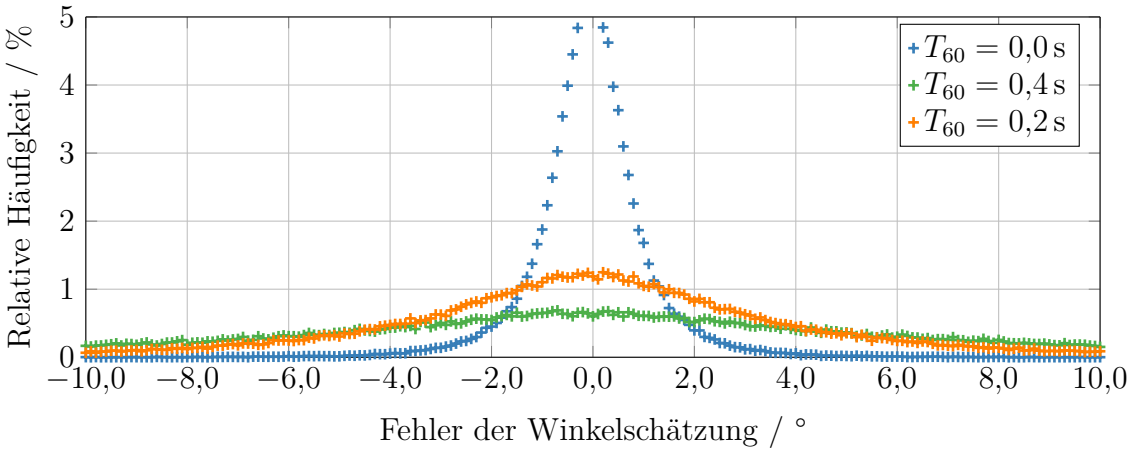


Abbildung 4.22: Histogramme des Fehlers der Winkelschätzung mittels WKM, bei der Nutzung eines dreieckigen Mikrofonarrays mit 0,05 m Kantenlänge für verschiedene Nachhallzeiten und SNR = 10 dB.

Die in Abschnitt 4.5 thematisierten nachhallbedingten Ausreißer bei linearen Arrays sind bei den jetzt betrachteten dreieckigen Arrays nicht mehr zu erkennen. Deshalb erscheint zunächst eine Approximation der Histogramme durch eine Normalverteilung möglich. Allerdings gilt es zu beachten, dass es sich bei Winkeln um periodische bzw. zirkuläre Größen handelt. Bei der Modellierung des Fehlers durch eine nichtperiodische Normalverteilung ist eine Winkelschätzung von 359° bei einem gegeben Winkel von 1° unwahrscheinlicher als eine Schätzung von 3°, da die Nichtberücksichtigung der periodischen Eigenschaften des Winkels zu einer Differenz von 358°, statt der tatsächlich vorhandenen 2° führt. Eine Alternative bildet deshalb die Wrapped-Normalverteilung [MJ09], die eine periodische Aufwicklung der Normalverteilung darstellt. Dadurch verschwindet zwar die erwähnte Problematik mit zirkulären Größen, aber die in der Wahrscheinlichkeitsdichte vorhandene unendliche Summe führt zu einer komplizierteren Handhabung.

Im Bereich der zirkulären Statistik ist daher die Nutzung der von Mises-Verteilung ähnlich verbreitet wie die Normalverteilung in der konventionellen Statistik [MJ09]. Sie besitzt ebenfalls zwei Freiheitsgrade, die durch den Mittelwert und die Konzentration gegeben sind. Die Konzentration stellt dabei das Äquivalent zum Inversen der Varianz dar. Für einen Winkel φ, den Mittelwert μ und eine Konzentration κ entsteht die Wahrscheinlichkeitsdichte

$$\mathrm{p}(\varphi;\mu,\kappa) = \frac{1}{2\pi \cdot I_0(\kappa)} \cdot \exp(\kappa \cdot \cos(\varphi - \mu)) = \mathcal{M}(\varphi;\mu,\kappa). \tag{4.40}$$

Dabei bezeichnet $I_0(\kappa)$ die modifizierte Besselfunktion 0. Ordnung. Die im Exponenten vorhandene Kosinusfunktion sorgt zudem für das erforderliche periodische Abstandsmaß.

Die Verteilungsdichte der von Mises-Verteilung nähert sich darüber hinaus der Normalverteilung an, je größer der Konzentrationsparameter wird. Bei hinreichend großer Konzentration können die Normalverteilung und von Mises-Verteilung daher als gleich betrachtet werden [MJ09].

Allgemein lässt sich die Standardabweichung der von Mises-Verteilung durch

$$\mathrm{std}(\varphi) = \sqrt{1 - \frac{I_0(\kappa)}{I_1(\kappa)}} \tag{4.41}$$

bestimmen [MJ09]. Eine Substitution des Konzentrationsparameters κ in Gl. (4.40) durch $\kappa = 1/\sigma_{\mathrm{M}}^2$ liefert zudem eine alternative parametrische Form der von Mises-Verteilung. Sofern die Konzentration ausreichend groß bzw. σ_{M} hinreichend klein ist, entspricht σ_{M} der Standardabweichung aus Gl. (4.41). Zur intuitiveren Handhabung der von Mises-Verteilung und um eine leichtere Einschätzung der Streuung zu gestatten, soll der Parameter σ_{M} im weiteren Verlauf als Standardabweichung bezeichnet werden. Zwar existiert für kleinere Konzentrationen ein Unterschied zwischen σ_{M} und der tatsächlichen Standardabweichung (vgl. Gl. (4.41)), allerdings sind die im Rahmen der Arbeit auftretenden Konzentrationen so groß, dass die Differenz vernachlässigbar ist.

Abb. 4.23 zeigt eine Darstellung mittelwertfreier von Mises-Verteilungen ($\mu = 0$). Für eine Konzentration von $\kappa = 0$ entsteht eine Gleichverteilung, während die Streuung abnimmt, je größer die Konzentration κ bzw. je kleiner die Standardabweichung σ_{M}.

Zur Überprüfung, inwieweit sich der Schätzfehler durch eine von Mises-Verteilung approximieren lässt, enthält Abb. 4.24 nocheinmal die aus Abb. 4.21 bekannten His-

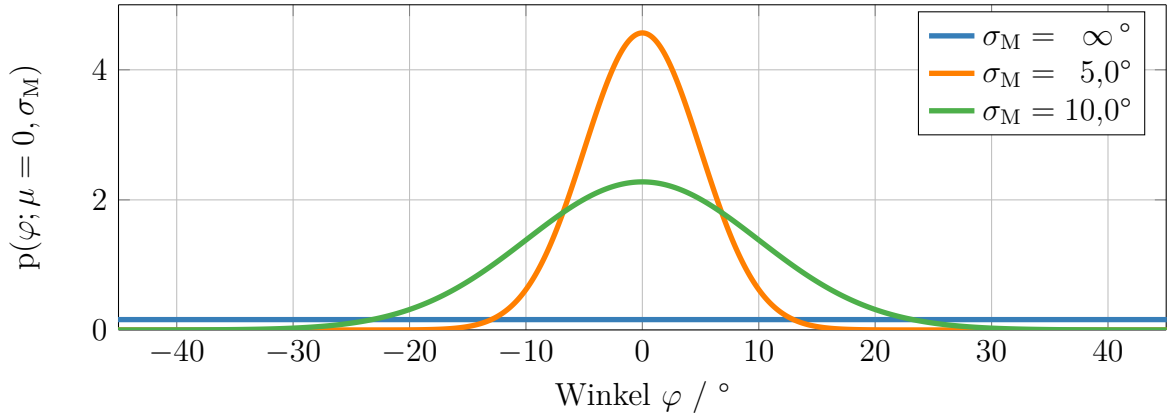


Abbildung 4.23: Wahrscheinlichkeitsdichte der von Mises-Verteilung für $\mu = 0$.

togramme (SNR $= \infty$) ebenso wie die sich aus den Daten ergebenden von Mises-Verteilungen (von Mises-Approximation). Bestimmt wurden die Parameter der von Mises-Verteilunge anhand der Momentenmethode (engl. *method of moments*) [BS06]. Trotz der erkennbaren Abweichungen zwischen den Histogrammen und den an die Daten angepassten, mittelwertfreien von Mises-Verteilungen, erscheint eine von Mises-Verteilung dennoch zur Annäherung des Schätzfehlers geeignet zu sein. Auch die Approximation der Histogramme für ein SNR von 10 dB gelingt mit vergleichbarer Qualität, deshalb wird auf eine Darstellung dieser Fälle verzichtet.

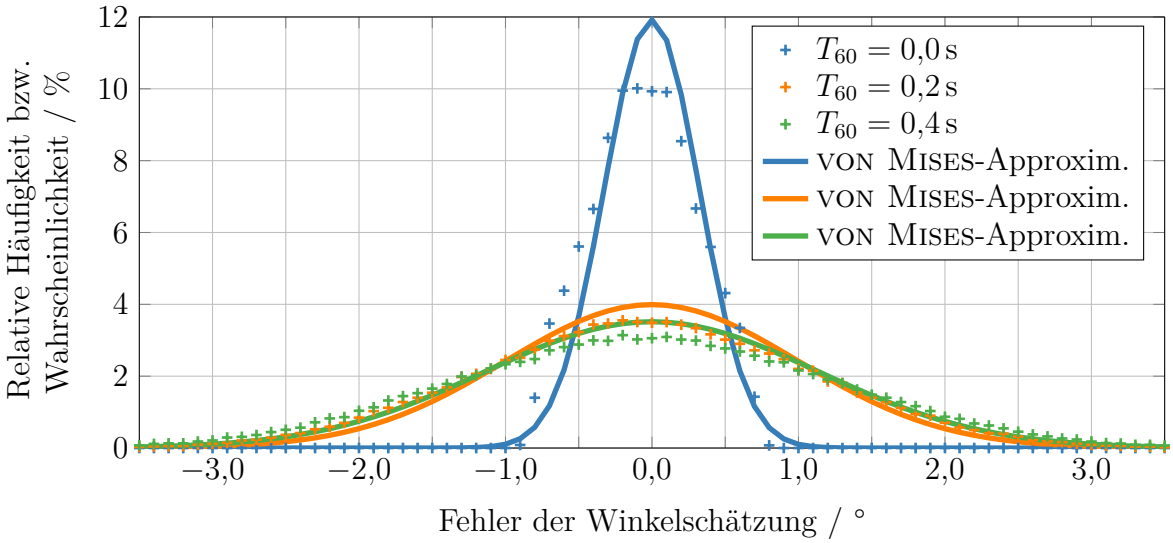


Abbildung 4.24: Histogramme des Fehlers der Winkelschätzung mittels WKM, bei verschiedenen Nachhallzeiten und SNR $= \infty$ dB inklusive der zugehörigen von Mises-Approximationen.

Insgesamt bestätigen die zurückliegenden Betrachtungen, dass die von Mises-Verteilung eine geeignete Approximationen des Fehlers der Einfallswinkelschätzung darstellt, sofern mindestens drei Mikrofone vorhanden sind und die WKM zur Winkelschätzung dient. Im weiteren Verlauf der Arbeit wird der Winkelfehler daher als von Mises-verteilt angenommen.

4.7 Zusammenfassung

Im Zentrum dieses Kapitel stand die Untersuchung der Einfallswinkelschätzung mit dem vorrangigen Ziel, die Ausgangssituation für die Geometriekalibrierung zu klären und einen geeigneten Winkelschätzer aus der Vielzahl der existierenden Ansätze auszuwählen.

Zu Beginn wurden dazu zunächst ausgewählte Winkelschätzer vorgestellt, die sich zur Schätzung des Einfallswinkels aus den Aufnahmen eines kompakten Arrays mit wenigen Mikrofonen eignen. Aufbauend auf den beschriebenen Verfahren erfolgte zudem die Entwicklung der Watson-Kern-Methode (WKM). Diese verwendet zur Bestimmung des Winkels nicht nur Phasendifferenzen, sondern berücksichtigt zusätzlich auch Amplitudendifferenzen. Die durchgeführten Untersuchungen bestätigen außerdem, dass die konventionellen Algorithmen, die implizit eine omnidirektionale Richtcharakteristik voraussetzen, deutlich schlechtere Schätzungen liefern, wenn die tatsächliche Richtcharakteristik von der angenommenen Richtcharakteristik abweicht. Die entwickelte WKM, die die Richtcharakteristik mit in die Schätzung einbezieht, erzielt hingegen auch beim Einsatz gerichteter Mikrofone dieselben oder z. T. sogar bessere Ergebnisse als beim Einsatz omnidirektionaler Mikrofone und ist dementsprechend den konventionellen Ansätzen deutlich überlegen. Darüber hinaus liefert die WKM auch bei der Nutzung omnidirektionaler Mikrofone präzisere Schätzungen als die betrachteten Konkurrenten und dient deshalb im weiteren Verlauf als Winkelschätzer für die Geometriekalibrierung.

Allerdings zeigten die Analysen der Winkelschätzer, dass alle Verfahren einen systematischen Fehler aufweisen, wenn nur zwei Mikrofone zum Einsatz kommen. Im Rahmen weiterer Untersuchungen konnten zwei Ursachen für den Fehler identifiziert werden, die beide eine Folge der linearen Anordnung darstellen. Der Erfassungsbereich eines linearen Arrays ist auf 180° beschränkt, da aber ausgelöst durch Rauschen, Nachhall und die nicht idealen Korrelationseigenschaften der aufgenommenen Signale TDOA-Schätzungen außerhalb des Intervalls $[-\tau_{\max}; \tau_{\max}]$ auftreten, ist eine Begrenzung der Schätzungen erforderlich. Einerseits verursacht diese Begrenzung einen Teil des Bias, andererseits löst auch der Nachhall einen Teil des Bias aus, weil die frühen Reflexionen tendenziell eine geringere TDOA besitzen als die LOS-Komponente und somit bei einem linearen Array eine ungleichmäßige Verteilung vorliegt. Sofern jedoch keine lineare Anordnung vorliegt, entfällt die Notwendigkeit eine Begrenzung der TDOA durchzuführen ebenso wie die ungleichmäßige Verteilung der Reflexionen, dementsprechend verschwindet auch der bislang vorhandene Bias.

Damit die Geometriekalibrierung trotzdem mit möglichst wenigen Mikrofonen auskommt, wurde eine dreieckige Anordnung der Mikrofone innerhalb des Sensorknotens präferiert. Abschließend wurde ein Modell zur Approximation des Fehlers der Einfallswinkelschätzung bei der Nutzung der WKM in Kombination mit einem dreieckigen Array entwickelt. Dieses Modell approximiert den Fehler durch eine von Mises-Verteilung und bildet somit die Grundlage für den im nächsten Kapitel erläuterten Entwurf eines Geometriekalibrierungsalgorithmus.

5 Entwicklung eines Kalibrierungsverfahrens

Nachdem im vorangegangenen Kapitel die Winkelschätzung aus den Aufnahmen eines Mikrofonarrays und damit die Gewinnung der zur Kalibrierung benötigten Informationen betrachtet wurde, steht nun die Entwicklung eines einfallswinkelgestützten Geometriekalibrierungsverfahrens im Fokus. Den Ausgangspunkt dafür bildet das ursprünglich zur Kalibrierung von Infrarotsensornetzen entwickelte Einfallswinkelverfahren [KWL08].

Nach einer kurzen Vorstellung der Verfahrensweise in Abschnitt 5.1 erfolgt in Abschnitt 5.2 unmittelbar eine Untersuchung mit dem Ziel, eingehend zu analysieren, ob der gewählte Algorithmus auch die Kalibrierung eines akustischen Sensornetzes gestattet. Die im Verlauf dieser Analyse festgestellten Probleme, stellen die Verwendung des Verfahrens in seiner bisherigen Form in Frage. Daher wird der Algorithmus in Abschnitt 5.3 unter der Prämisse umformuliert, die entdeckten Schwachstellen zu beseitigen, um danach die Kalibrierung eines ASN zu gewährleisten. Im Rahmen der in Abschnitt 5.4 geschilderten Überprüfung gilt es deshalb zu klären, inwieweit die entwickelten Modifikationen zu der gewünschten Verbesserung des Algorithmus führen.

Bezugnehmend auf das Fehlermodell der Einfallswinkelschätzung aus Abschnitt 4.6 wird darüber hinaus in Abschnitt 5.5 dargelegt, dass es sich bei dem erweiterten Einfallswinkelverfahren um den Maximum-Likelihood-Schätzer (ML-Schätzer) für einen von Mises-verteilten Winkelfehler handelt. Das Ziel der Untersuchungen in Abschnitt 5.6 ist es, die Verlässlichkeit des Kalibrierungsalgorithmus gegenüber Störungen der Einfallswinkel zu analysieren. Obwohl sich diese Arbeit vorrangig mit der Kalibrierung planarer Sensorkonfigurationen befasst, zeigt Abschnitt 5.7, dass das entwickelte Konzept auch die Kalibrierung dreidimensionaler Sensoranordnungen ermöglicht.

5.1 Vorstellung eines einfallswinkelbasierten Kalibrierungsverfahrens

Der in diesem Abschnitt betrachtete Algorithmus [KWL08] benötigt zur Kalibrierung der Sensoranordnung Einfallswinkelschätzungen. In [KWL08] stammen diese von Infrarotsensoren, die die Wärmestrahlung einer sich durch den Raum bewegenden Person erfassen. Die Abstraktion der Sensorinformationen zu Einfallswinkeln erlaubt es daher, diesen Ansatz auch zur Kalibrierung eines akustischen Sensornetzes in Betracht zu ziehen, zumal akustische Sensoren durch Anwendung der in Kapitel 4 präsentierten Methoden ebenfalls Einfallswinkelschätzungen gestatten.

Das Ziel des Einfallswinkelverfahrens besteht in der Bestimmung der Positionen $\boldsymbol{s}_i = [x_i \;\; y_i]^{\mathrm{T}}$ und Orientierungen θ_i für alle $i = 1, \ldots, I$ Sensorknoten eines Sensornetzes. Um dieses Ziel zu erreichen, benötigt der Algorithmus die Einfallswinkel $\varphi_{i,d}$ für alle D Ereignisse, $d = 1, \ldots, D$. Grundlage der Kalibrierung bildet die in Abb. 5.1 veranschaulichte geometrische Beziehung, die zwischen jedem Sensor (blauer Punkt) und jedem Ereignis (grauer Kopf) durch Ausnutzung von Winkelbeziehungen hergestellt werden kann. Die Angabe der dazu benötigten Sensorposition $\boldsymbol{s}_i$ und Ereignisposition $\boldsymbol{e}_d = [a_d \;\; b_d]^{\mathrm{T}}$ erfolgt in einem beliebigen, durch die Basisvektoren $\vec{\boldsymbol{x}}$ und $\vec{\boldsymbol{y}}$ aufgespannten, Bezugskoordinatensystem.

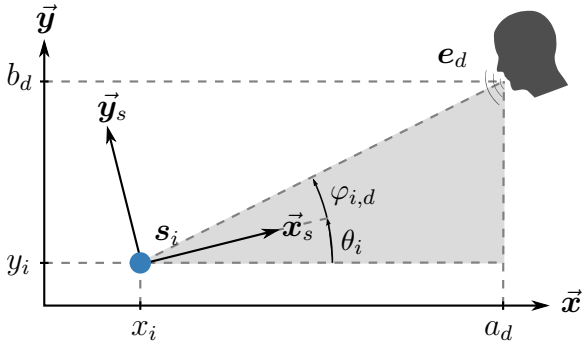


Abbildung 5.1: Geometrische Beziehung zwischen Sensor und erfasstem Ereignis.

Der zur Kalibrierung notwendige geometrische Zusammenhang zwischen der Sensorposition und der Ereignisposition lässt sich mithilfe des Einfallswinkels $\varphi_{i,d}$ herstellen. Die Messung dieses Winkels erfolgt im lokalen Koordinatensystem des i-ten Sensors, welches eine Rotation von θ_i zum Bezugskoordinatensystem aufweist und von den Basisvektoren $\vec{\boldsymbol{x}}_s$ und $\vec{\boldsymbol{y}}_s$ aufgespannt wird. Die geometrische Beziehung entsteht letztendlich durch das von der Verbindungsgeraden zwischen Sensor und Ereignis sowie den Parallelen der Koordinatenachsen aufgespannte Dreieck. Die Form dieses Dreiecks wird durch

$$\tan(\theta_i + \varphi_{i,d}) = \frac{b_d - y_i}{a_d - x_i} \tag{5.1}$$

definiert. Mit den Einheitsvektoren $\boldsymbol{z}_1 = [1 \;\; 0]^{\mathrm{T}}$ bzw. $\boldsymbol{z}_2 = [0 \;\; 1]^{\mathrm{T}}$, der Sensorposition $\boldsymbol{s}_i$ sowie der Ereignisposition $\boldsymbol{e}_d$ entsteht zunächst

$$\tan(\theta_i + \varphi_{i,d}) = \frac{\boldsymbol{z}_2^{\mathrm{T}} (\boldsymbol{e}_d - \boldsymbol{s}_i)}{\boldsymbol{z}_1^{\mathrm{T}} (\boldsymbol{e}_d - \boldsymbol{s}_i)}. \tag{5.2}$$

Daraus ergibt sich unter Verwendung von Additionstheoremen, gemäß der in Anhang B.1 geschilderten Umformungen, schließlich die Zielfunktion

$$f_{\mathrm{Tan}}(\boldsymbol{s}_i, \theta_i, \boldsymbol{e}_d; \varphi_{i,d}) = (\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}} \begin{bmatrix} -\tan(\theta_i) - \tan(\varphi_{i,d}) \\ 1 - \tan(\theta_i)\tan(\varphi_{i,d}) \end{bmatrix} \stackrel{!}{=} 0, \tag{5.3}$$

die den fundamentalen Zusammenhang zur Lösung des Geometriekalibrierungsproblems beschreibt. Dieser in Gl. (5.3) hergestellte Zusammenhang zwischen dem i-ten Sensor und dem d-ten Ereignis gilt für alle Kombinationen aus Sensoren und Ereignissen. Die

Matrizen $\boldsymbol{S} = [\boldsymbol{s}_1 \ \ldots \ \boldsymbol{s}_I]$, $\boldsymbol{\theta} = [\theta_1 \ \ldots \ \theta_I]$ und $\boldsymbol{E} = [\boldsymbol{e}_1 \ \ldots \ \boldsymbol{e}_D]$ sowie die $I \times D$ Matrix $\boldsymbol{\Phi}$, die die Einfallswinkel aller Sensoren zu allen Ereignissen beinhaltet, gestatten zusammen mit der Annahme, dass jeder Sensor jedes Ereignis erfasst, die vektorwertige Darstellung der Zielfunktion

$$\boldsymbol{f}_{\mathrm{Tan}}(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{E}; \boldsymbol{\Phi}) = \left[f_{\mathrm{Tan}}(\boldsymbol{s}_1, \theta_1, \boldsymbol{e}_1; \varphi_{1,1}) \ \ldots \ f_{\mathrm{Tan}}(\boldsymbol{s}_I, \theta_I, \boldsymbol{e}_D; \varphi_{I,D}) \right]^{\mathrm{T}}. \tag{5.4}$$

Sofern $\boldsymbol{S}$ und $\boldsymbol{\theta}$ die tatsächlichen Positionen und Ausrichtungen der Sensoren beschreiben, die Ereignispositionen $\boldsymbol{E}$ vorliegen und die Einfallswinkel aller Sensoren zu allen Ereignissen $\boldsymbol{\Phi}$ keine Störungen aufweisen, liefert die Zielfunktion den Wert Null.

Bei der vorliegenden Problemstellung sind hingegen lediglich die Einfallswinkel bekannt. Daher wird das Geometriekalibrierungsproblem auf die Lösung eines nichtlinearen Gleichungssystems zurückgeführt. Da es sich bei Gl. (5.4) um ein nichtlineares Gleichungssystem handelt, kann die Frage, wann eine Lösung existiert, nicht allgemein beantwortet werden [HMV05]. Deshalb orientieren sich die Autoren von [KWL08] an den Eigenschaften linearer Gleichungssysteme. Diese besitzen unendlich viele Lösungen, wenn weniger Gleichungen als Unbekannte vorhanden sind (unterbestimmtes Gleichungssystem). Eine eindeutige Lösung existiert hingegen, falls die Anzahl der unabhängigen Gleichungen der der Unbekannten entspricht. Sofern es mehr unabhängige Gleichungen als Unbekannte gibt (überbestimmtes Gleichungssystem), besitzt das Gleichungssystem keine Lösung. Allerdings lässt sich z. B. mit der *Methode der kleinsten Quadrate* (engl. *least-squares* (LS)) eine Lösung bestimmen, die die Summe der quadratischen Fehler minimiert. Dadurch wird einerseits die Lösung überbestimmter Gleichungssysteme möglich, andererseits sorgt die LS-Lösung für einen Ausgleich von Störungen der Messwerte. Die Autoren von [KWL08] setzten daher voraus, dass mindestens so viele Gleichungen vorliegen müssen wie das Gleichungssystem Unbekannte aufweist.

Die Anzahl der Unbekannten des Gleichungssystems (5.4) beträgt $3 \cdot (I-1) + 2 \cdot D$. Davon entfallen $3 \cdot (I-1)$ Unbekannte auf die Positionen und Orientierungen der I Sensoren, wobei durch die feste Wahl der Parameter eines Sensors die Definition des Bezugskoordinatensystems erfolgt. Im Rahmen dieser Arbeit wird stets vorausgesetzt, dass das Bezugskoordinatensystem dem Koordinatensystem des ersten Sensors entspricht. Die verbleibenden $2 \cdot D$ Unbekannten korrespondieren zu den Ereignispositionen, die ebenfalls nicht bekannt sind und deshalb auch durch den Kalibrierungsalgorithmus bestimmt werden müssen. Infolge der getroffenen Annahmen existiert eine Lösung, sofern die Sensoren

$$D \geq \frac{3\,(I-1)}{I-2} \tag{5.5}$$

Ereignisse erfassen.

Neben der Anzahl der Ereignisse muss jedoch zusätzlich auch die räumliche Lage dieser berücksichtigt werden. Die bisherigen Überlegungen gehen davon aus, dass jede weitere Beobachtung unabhängig von den bereits vorhandenen ist und dementsprechend neue Informationen zum Gleichungssystem beiträgt. Damit der Schätzwert des Einfallswinkels eines Sensors diese Bedingung erfüllt, muss sich der aktuelle Wert von den bisherigen Winkeln dieses Sensors unterscheiden. Allerdings sorgen die Störungen in einem realen Szenario dafür, dass die Einfallswinkelschätzungen um den tatsächlichen Winkel streuen.

Daher sollte die Differenz zwischen den Winkeln zweier Ereignisse größer als die Streuung der Einfallswinkel ausfallen. Da die Differenz der Winkel von der zugehörigen Distanz zwischen den Ereignispositionen abhängt ergibt sich zusätzlich zu der Bedingung aus Gl. (5.5) die Anforderung, dass die Ereignispositionen eine ausreichende räumliche Diversität aufweisen müssen.

Eine analytische Lösung des nichtlinearen Gleichungssystems scheidet angesichts der Komplexität des Gleichungssystems aus. Stattdessen kommt das Newton-Verfahren zur iterativen Suche von Nullstellen zum Einsatz. Dazu wird ausgehend von einem Startpunkt eine Linearisierung der Zielfunktion (siehe Gl. (5.4)) mithilfe einer Tangente erstellt und die Nullstelle dieser bestimmt. Die Nullstelle der Tangenten dient anschließend als Ausgangspunkt für eine erneute Linearisierung der Zielfunktion. Diese Prozedur wird solange wiederholt, bis die Nullstelle der ursprünglichen Zielfunktion vorliegt. Durch die Zusammenfassung aller unbekannten Parameter ($\boldsymbol{S}$, $\boldsymbol{\theta}$, $\boldsymbol{E}$) zu $\boldsymbol{\Lambda}$ lässt sich die Iterationsgleichung als

$$\boldsymbol{\Lambda}^{(r+1)} = \boldsymbol{\Lambda}^{(r)} - \left(\boldsymbol{J}_{\mathrm{Tan}}\left(\boldsymbol{\Lambda}^{(r)}\right)\right)^{-1} \cdot \boldsymbol{f}_{\mathrm{Tan}}\left(\boldsymbol{\Lambda}^{(r)}; \boldsymbol{\Phi}\right) \tag{5.6}$$

formulieren. Dabei bezeichnet $\cdot^{(r)}$ den r-ten Iterationsschritt und $\boldsymbol{J}_{\mathrm{Tan}}(\boldsymbol{\Lambda}^{(r)})$ die Jacobi-Matrix der Zielfunktion $\boldsymbol{f}_{\mathrm{Tan}}(\boldsymbol{\Lambda}^{(r)}; \boldsymbol{\Phi})$ an der Stelle $\boldsymbol{\Lambda}^{(r)}$. Die im r-ten Iterationsschritt verbleibende Entfernung zur Nullstelle ist durch

$$\epsilon^{(r)} = \left\| \boldsymbol{f}_{\mathrm{Tan}}\left(\boldsymbol{\Lambda}^{(r)}; \boldsymbol{\Phi}\right) \right\|_2 \tag{5.7}$$

gegeben und dient als Kriterium zur Beendigung des Newton-Verfahrens, sobald $\epsilon^{(r)}$ eine bestimmte Schwelle unterschreitet.

Zu Beginn erfordert das Newton-Verfahren einen Startwert für die gesuchten Parameter. Die Schwierigkeit besteht jedoch in der Wahl eines adäquaten Startwertes. Einerseits ist die Konvergenz des Newton-Verfahrens nur dann gewährleistet, wenn der Startwert nah genug am Optimum liegt, andererseits handelt es sich bei Gl. (5.3) um eine nicht konvexe Funktion, sodass auch lokale Optima möglich sind. Obwohl der Startwert eine entscheidende Rolle für den Erfolg des Newton-Verfahrens spielt, verwenden die Autoren von [KWL08] trotz der Vielzahl der Parameter eine zufällige Initialisierung. Sofern das Newton-Verfahren divergiert, starten sie einen weiteren Versuch. Allerdings akzeptieren sie durch die erläuterte Strategie auch, dass am Ende ggf. nur ein lokales Optimum erreicht wird.

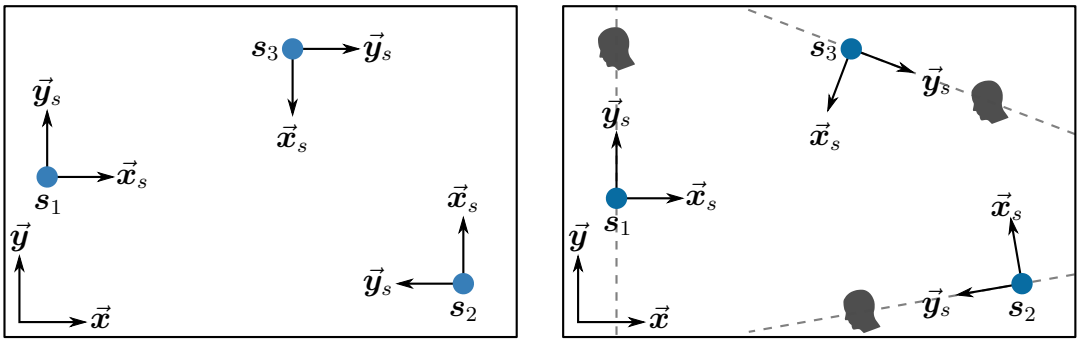
5.2 Analyse eines einfallswinkelbasierten Kalibrierungsverfahrens

Nachdem der zurückliegende Abschnitt die Funktionsweise des Einfallswinkelverfahrens dargestellt hat, soll nun geklärt werden, welche Probleme bei der Anwendung dieses Verfahrens zur Kalibrierung eines akustischen Sensornetzes zu erwarteten sind. Die folgende Analyse bestätigt einerseits die bereits in [KWL08] identifizierten Problematiken und deckt andererseits weitere Nachteile auf. Zusätzlich steht die Erforschung der Ursachen der festgestellten Probleme im Vordergrund, damit im anschließenden Abschnitt geeignete Gegenmaßnahmen entwickelt werden können.

5.2.1 Numerische Probleme der Zielfunktion

Ein Nachteil des zuvor vorgestellten Einfallswinkelverfahrens sind die massiven Konvergenzprobleme des zur Lösung des Gleichungssystems (5.4) eingesetzten Newton-Verfahrens. Diese Konvergenzproblematik bei der Lösung des Gleichungssystems wird zwar bereits in [KWL08] erwähnt, dort aber auf ungünstige Startwerte des Newton-Verfahrens zurückgeführt. Die im Rahmen dieser Arbeit durchgeführten Untersuchungen des Einfallswinkelverfahrens identifizieren hingegen die Tangensfunktionen des Gleichungssystems als Auslöser. Diese besitzen bei allen ungeraden Vielfachen von $\pm 90°$ Polstellen, die wiederum dazu führen, dass die Jacobi-Matrix, die die partiellen Ableitungen der Zielfunktion enthält, in diesem Bereich sehr große Werte aufweist. Als Resultat treten numerische Probleme bei der Inversion der Jacobi-Matrix auf, die sich im Verlauf der Iterationsschritte akkumulieren und zur Divergenz des Newton-Verfahrens führen können.

Bei näherer Betrachtung der Zielfunktion (vgl. Gl. (5.3)) ergeben sich zwei Situationen, die zum Auftreten der Polstellen führen. Diese sind in Abb. 5.2 illustriert. Eine Ursache der Konvergenzproblematik liegt in der Positionierung der Sensoren. In einem typischen Szenario, wie in Abb. 5.2a dargestellt, in dem die Sensoren orthogonal zu den Wänden ausgerichtet sind, verursacht allein die Ausrichtung Konvergenzprobleme. Die Sensoren zwei und drei weisen eine Rotation von $90°$ bzw. $-90°$ zu dem durch den ersten Sensor festgelegten Bezugskoordinatensystem auf. Da die Rotation als Argument der Tangensfunktion auftritt, entsteht somit unmittelbar das Konvergenzproblem.



(a) Beeinträchtigung des Konvergenzverhaltens durch die Ausrichtung der Sensoren.

(b) Beeinträchtigung des Konvergenzverhaltens durch die Lage der erfassten Ereignisse.

Abbildung 5.2: Beispiele für Szenarien, bei denen Polstellen in der Zielfunktion das Konvergenzverhalten des Newton-Verfahren beeinträchtigen.

Darüber hinaus treten weitere Tangensfunktionen auf, deren Argument der Einfallswinkel ist (siehe Gl. (5.3)). Somit hat die Lage der erfassten Ereignisse in Kombination mit der Sensorausrichtung gleichermaßen einen entscheidenden Einfluss auf die numerische Stabilität. Abb. 5.2b zeigt durch gestrichelte Linien die Bereiche an, in denen die Lage der Ereignisse die Konvergenz des Newton-Verfahrens beeinträchtigt.

Neben den Bereichen, in denen ein oder gar mehrere Polstellen die Konvergenz des Newton-Verfahrens verhindern, sind auch die Bereiche in der Nähe der Polstellen problematisch. Ausgelöst durch die großen Steigungen der Tangensfunktionen in diesen Bereichen ergeben sich dort zum Teil erhebliche Kalibrierungsfehler. Weiterhin ist zu berücksichtigen, dass die Wahl des Koordinatensystems in dieser Arbeit willkürlich durch den ersten Sensor erfolgt. Da die numerischen Probleme eine Folge bestimmter geometrischer Anordnungen sind, hängt das Auftreten der geschilderten Probleme auch von der Wahl des Koordinatensystems ab.

5.2.2 Skalierungsinvarianz

Eine weitere Problematik des Einfallswinkelsverfahrens ist bei erneuter Betrachtung von Abb. 5.1 bzw. der zugehörigen Gl. (5.1) zu erkennen. Die zur Kalibrierung genutzte geometrische Beziehung stützt sich auf Winkelbeziehungen und ist daher invariant zur Skalierung der Koordinaten. Deutlich wird diese Problematik durch die Multiplikation sämtlicher Koordinaten in Gl. (5.1) mit einem beliebigen Skalierungsfaktor ν. Dabei entsteht

$$\tan(\theta_i + \varphi_{i,d}) = \frac{\nu \cdot b_d - \nu \cdot y_i}{\nu \cdot a_d - \nu \cdot x_i}. \tag{5.8}$$

Allerdings fällt sofort auf, dass der Faktor ν sowohl im Zähler als auch im Nenner der Gleichung auftritt und deshalb herausfällt. Somit entsteht unmittelbar die ursprüngliche Formulierung des geometrischen Zusammenhangs.

Infolge der Skalierungsinvarianz liefert das Newton-Verfahren in der bisher beschriebenen Form lediglich die triviale Lösung $\mathbf{\Lambda} = \mathbf{0}$. Um dieses unerwünschte Verhalten zu unterbinden, ist eine Berücksichtigung zusätzlicher Informationen notwendig. In [KWL08] wird deshalb angenommen, dass die Positionen von zwei Sensoren bekannt sind. Die Kenntnis dieser Positionen legt die Skalierung eindeutig fest und erlaubt dementsprechend die Rückgewinnung der Positionen und Orientierungen aller anderen Sensoren durch das Newton-Verfahren.

5.2.3 Rotationsinvarianz

Darüber hinaus zeigt die nähere Betrachtung der Lösungen des Newton-Verfahrens, dass verschiedene Ergebnisse existieren. Aufgrund der nicht-konvexen Zielfunktion sind zwar lokale Minima zu erwarten, allerdings besitzen die Lösungen keine signifikanten Unterschiede bei den zugehörigen Funktionswerten. Daher handelt es sich im Sinne der Zielfunktion um äquivalente Lösungen. Zu den auftretenden Lösungen gehören neben der gesuchten Sensoranordnung auch Lösungen die eine Rotation bzw. Spiegelung aufweisen.

Zur näheren Erläuterung dieser Beobachtung dient das in Abb. 5.3a dargestellte exemplarische Szenario. Die Anwendung des Newton-Verfahrens mit unterschiedlichen Startwerten liefert in diesem Fall eine Menge verschiedener Ergebnisse. Diese lassen sich in zwei Gruppen einteilen und sollen anhand der Abbildungen 5.3b und 5.3c genauer erläutert werden.

Bei den Lösungen, die in die erste Gruppe fallen, entsprechen die ermittelten Sensor- und Ereignispositionen, abgesehen von numerischen Abweichungen, den wahren Positionen. Lediglich bei den Orientierungen sind deutliche Abweichungen vorhanden. Abb. 5.3b skizziert eine mögliche Lösung, bei der sowohl der zweite als auch der dritte Sensor eine Rotation von 180° im Vergleich zur tatsächlichen Sensorkonfiguration (vgl. Abb. 5.3a) haben. Abgesehen vom ersten Sensor, dessen Orientierung dazu dient das Referenzkoordinatensystem zu definieren und deshalb nicht durch das Newton-Verfahren ermittelt wird, können alle anderen Sensororientierungen entweder der wahren Orientierung entsprechen oder eine zusätzliche 180° Rotation besitzen. Somit ergeben sich 2^{I-1} verschiedene, äquivalente Lösungen.

Die zweite Gruppe der auftretenden Lösungen beschreibt gespiegelte Anordnungen der Sensoren und Ereignisse. Um diese Lösungen besser charakterisieren zu können, enthält Abb. 5.3c neben dem in schwarz dargestellten Kalibrierungsergebnis auch die wahre Anordnung der Sensoren (grau). Die Position und Rotation des ersten Sensors bleiben dabei unverändert, weil sie das Referenzkoordinatensystem festlegen und demnach nicht kalibriert werden. Die Positionen der anderen Sensoren sowie auch der nicht dargestellten Ereignisse entstehen durch eine Spiegelung am Koordinatenursprung des ersten Sensors. Allerdings können die Sensoren ebenfalls entweder mit der wahren Orientierung oder in der um 180° rotierten Version auftreten. Dadurch ergeben sich noch einmal 2^{I-1} Lösungen, sodass das Gleichungssystem insgesamt 2^I äquivalente Lösungen hat.

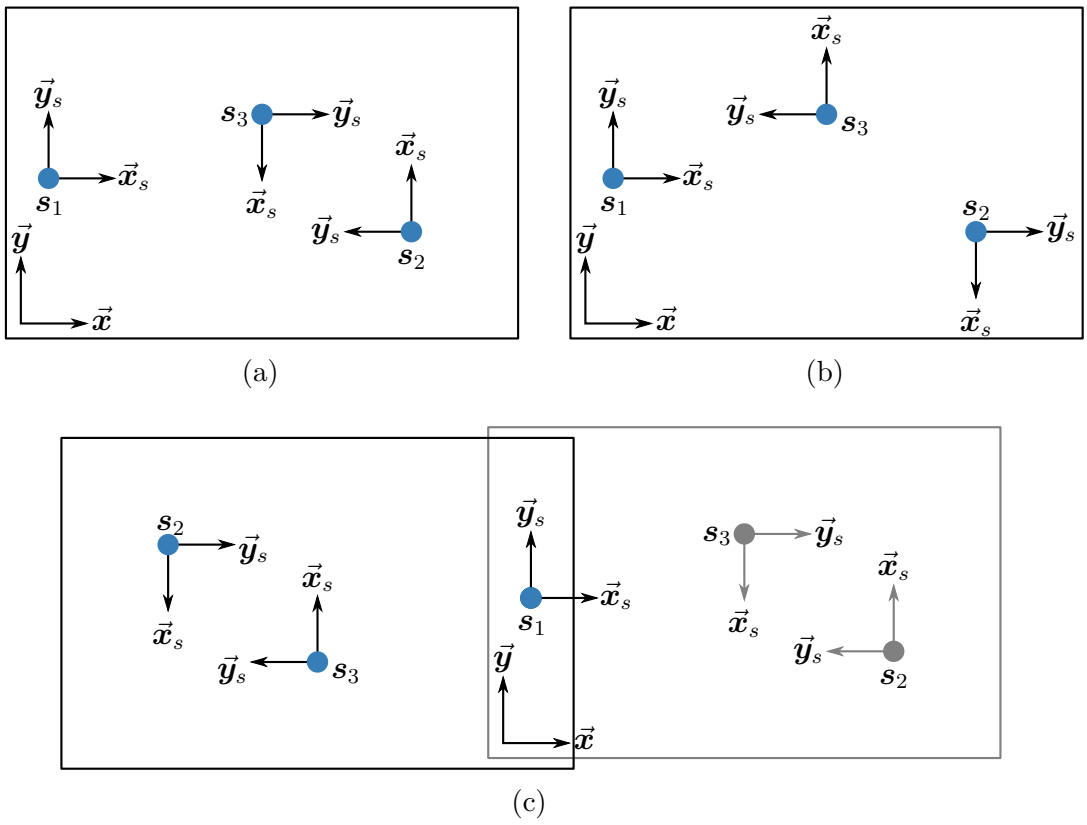


Abbildung 5.3: Beispielhaftes Kalibrierungsszenario (a) sowie mögliche Lösungen durch das Einfallswinkelverfahren (b) und (c).

Verantwortlich für die Rotationsinvarianz der Lösungen ist, wie auch bei den numerischen Problemen (vgl. Abschnitt 5.2.1), die Tangensfunktionen in Gl. (5.3). Die Sensorrotation tritt ausschließlich als Argument der Tangensfunktion auf, die allerdings nur eine Periodizität von 180° aufweist. Somit besitzt das zur Lösung des Geometriekalibrierungsproblems genutzte Gleichungssystem neben dem erwünschten Minimum bei der tatsächlichen Sensororientierung auch weitere Minima bei einer zusätzlichen Rotation von ±180° bzw. Vielfachen davon. Die am Beispiel von Abb. 5.3c erläuterten gespiegelten Konfigurationen entstehen lediglich dadurch, dass die Position und Orientierung des ersten Sensors vorgegeben ist. Statt wie bei den anderen Sensoren zu einer zusätzlichen 180° Rotation zu führen, sorgt die Periodizität der Tangensfunktion hier für die Spiegelung. Da die rotierten bzw. gespiegelten Lösungen eine Folge der Periodizität der Tangensfunktion sind, tritt die Rotationsinvarianz unabhängig von den Eigenschaften der Sensoren auf. Dementsprechend ist sie sowohl vorhanden, wenn eine lineare Mikrofonanordnung genutzt wird, die einen eindeutigen Detektionsbereich von nur 180° besitzt, als auch bei Mikrofonen, die nicht auf einer Linie angeordnet sind und daher eine vollständige Rundumsicht (360°) gestatten.

5.3 Konzeption eines modifizierten Einfallswinkelverfahrens

Die bisher identifizierten Nachteile des Einfallswinkelverfahrens stellen eine deutliche Einschränkung dar. Damit der Algorithmus dennoch eine präzise Kalibrierung eines akustischen Sensornetzes erlaubt, ist es zwingend erforderlich, Strategien zur Kompensation dieser Nachteile zu entwickeln. Die im vorherigen Abschnitt durchgeführten Analysen identifizieren die Formulierung der Zielfunktion mittels Tangensfunktionen als Hauptursache der Nachteile. Deshalb steht in diesem Abschnitt auch die Konzeption einer zuverlässigeren Zielfunktion im Vordergrund. Zuvor wird jedoch kurz die Problematik der Initialisierung des Newton-Verfahrens diskutiert und darüber hinaus werden verschiedene Ansätze zur Vermeidung der Skalierungsinvarianz skizziert.

Die Lösung des Gleichungssystems durch das Newton-Verfahren erfordert zunächst die Initialisierung der zu bestimmenden Parameter. Da die Konvergenz des Newton-Verfahrens nur dann gegeben ist, wenn die Startwerte nah genug am Optimum liegen, spielt die Wahl der Startwerte eine entscheidende Rolle. Trotzdem sieht [KWL08] eine zufällige Initialisierung vor. Das verwendete Schema wird dort allerdings nicht näher spezifiziert. Ferner zeigen die im Rahmen dieser Arbeit durchgeführten Simulationen, dass die in Anlehnung an [KWL08] verwendete Initialisierung der Unbekannten $\boldsymbol{\Lambda}$ mittels Gleich- oder Normalverteilung nur sehr unzureichend funktioniert. Um stattdessen eine Initialisierung zu erhalten, die sich bereits in der Nähe des Optimums befindet, wurden Szenarien generiert, die in etwa den in Abb. 2.1 dargestellten Konfigurationen entsprechen und diese anschließend als Startwert für das Newton-Verfahren verwendet.

Weiterhin erfordert das Einfallswinkelverfahren die Kenntnis der Lage von zwei Sensoren, um dadurch die Skalierung eindeutig zu definieren. Damit auch ohne die Kenntnis der Positionen von zwei Sensoren eine Lösung möglich ist, können bspw. zwei der Sensorpositionen mit einem festen, aber beliebigen Abstand, z. B. $\boldsymbol{s}_1=[0\ \ 0]^{\mathrm{T}}$ und $\boldsymbol{s}_2=[1\ \ 0]^{\mathrm{T}}$, gewählt werden. Aufgrund der beliebigen Wahl des Abstandes, besitzt

das Ergebnis einen Skalierungsfehler, der dem Verhältnis zwischen der tatsächlichen und der gewählten Distanz dieser Sensoren entspricht. Eine Alternative zur Fixierung der Skalierung stellt die Erweiterung des Gleichungssystems (5.4) um eine zusätzliche Gleichung dar, die analog zum bisherigen Ansatz den Abstand zwischen zwei Sensoren festlegt. Allerdings birgt eine Fixierung der Skalierung durch lediglich zwei Sensoren das Risiko, unabhängig davon in welcher Form sie erfolgt, dass sich die Fehler der Winkelschätzung dieser beiden Sensoren besonders stark auf das Kalibrierungsergebnis auswirken. Deshalb soll auch eine Festlegung der Skalierung durch eine zusätzliche Gleichung, die die Summe der Distanzen aller Sensorenpaare definiert, in Erwägung gezogen werden. Dazu wird das Gleichungssystem (5.4) um die Gleichung

$$1 - \sum_{i=1}^{I} \sum_{j=i}^{I} \|\boldsymbol{s}_i - \boldsymbol{s}_j\|_2 = 0 \tag{5.9}$$

erweitert. Eine Analyse, wie sich diese drei Skalierungsvarianten auf die Kalibrierung auswirken, folgt in Abschnitt 5.4.

Ausgangspunkt für die folgenden Überlegungen zur Weiterentwicklung der Zielfunktion ist Gl. (5.3). Der wesentliche Auslöser der resultierenden Probleme sind die Polstellen der Tangensfunktionen. Die Darstellung der Tangensfunktion als Quotient von Sinus und Kosinus sowie die anschließende Multiplikation der Zielfunktion mit den Kosinus-Termen führt zu der modifizierten Zielfunktion

$$f_{\mathrm{SinCos}}(\boldsymbol{s}_i, \theta_i, \boldsymbol{e}_d; \varphi_{i,d}) = (\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}} \begin{bmatrix} -\sin(\theta_i) & -\cos(\theta_i) \\ \cos(\theta_i) & -\sin(\theta_i) \end{bmatrix} \begin{bmatrix} \cos(\varphi_{i,d}) \\ \sin(\varphi_{i,d}) \end{bmatrix}. \tag{5.10}$$

Als Resultat dieser Umformungen (siehe Anhang B.2) verschwinden die Polstellen. Infolgedessen kommt es zu einer deutlichen Verbesserung der Konvergenz des Newton-Verfahrens, weil die partiellen Ableitungen, welche in der Jacobi-Matrix auftreten, durch die Entfernung der Polstellen nun beschränkt sind.

Außerdem besteht die veränderte Zielfunktion f_{SinCos} aus Sinus- und Kosinusfunktionen, die im Gegensatz zur Tangensfunktion jeweils 2π-periodisch sind. Diese Tatsache könnte leicht zu der Fehleinschätzung führen, dass jetzt auch die Rotationsinvarianz verschwindet. Dass die Rotationsinvarianz allerdings weiterhin vorhanden ist, wird durch die folgenden Betrachtungen deutlich.

Dazu wird der im lokalen Koordinatensystem des Sensors gemessene Einfallswinkel $\varphi_{i,d}$ zunächst durch den Einheitsvektor

$$\boldsymbol{g}_{i,d} = \begin{bmatrix} \cos(\varphi_{i,d}) & \sin(\varphi_{i,d}) \end{bmatrix}^{\mathrm{T}} \tag{5.11}$$

ausgedrückt. Zudem hat die in Gl. (5.10) auftretende Matrix große Ähnlichkeit mit einer Rotationsmatrix. Mit der Rotationsmatrix

$$\boldsymbol{R}_{\mathrm{xy}}(\theta_i) = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{bmatrix}, \tag{5.12}$$

die eine Drehung in der xy-Ebene um den Winkel θ_i beschreibt lässt sich Gl. (5.10) durch

$$f_{\mathrm{SinCos}}(\boldsymbol{s}_i, \theta_i, \boldsymbol{e}_d; \varphi_{i,d}) = (\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}} \boldsymbol{R}_{\mathrm{xy}}\left(\theta_i + \frac{\pi}{2}\right) \boldsymbol{g}_{i,d}. \tag{5.13}$$

darstellen.

Letztendlich handelt es sich bei Gl. (5.13) um ein Skalarprodukt. Die auftretende Rotationsmatrix dient lediglich zur Transformation des im lokalen Koordinatensystems des Sensorknotens angegebenen Beobachtungsvektors $\boldsymbol{g}_{i,d}$ in das Bezugskoordinatensystem, sodass das Skalarprodukt zwischen $\boldsymbol{e}_d - \boldsymbol{s}_i$ und $\boldsymbol{R}_{\mathrm{xy}}(\theta_i + \frac{\pi}{2})\, \boldsymbol{g}_{i,d}$ entsteht. Die Gruppierung der Terme ist indes nicht eindeutig. Daher lässt sich Gl. (5.13) auch als Drehung des Verbindungsvektors $\boldsymbol{e}_d - \boldsymbol{s}_i$ in das Sensorkoordinatensystem und die Berechnung des Skalarproduktes zwischen $(\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}}\, \boldsymbol{R}_{\mathrm{xy}}(\theta_i + \frac{\pi}{2})$ und $\boldsymbol{g}_{i,d}$ interpretieren.

Die Formulierung der Zielfunktion als Skalarprodukt gestattet zudem einen anschaulichen Beweis der Rotationsinvarianz. Das Newton-Verfahren, welches die Nullstelle der Zielfunktion bestimmt, ermittelt somit einen Punkt, an dem das Skalarprodukt der Vektoren $\boldsymbol{e}_d - \boldsymbol{s}_i$ und $\boldsymbol{R}_{\mathrm{xy}}(\theta_i + \frac{\pi}{2})\, \boldsymbol{g}_{i,d}$ den Wert Null annimmt. Diese Bedingung ist genau dann erfüllt, wenn beide Vektoren orthogonal zueinander stehen. Allerdings ist die Forderung der Orthogonalität zu einem gegebenen Vektor nicht eindeutig. Da die bislang gezeigten Veränderungen nur die numerische Stabilität verbessern, nicht aber die Mehrdeutigkeiten einschränken, ergeben sich weiterhin 2^I äquivalente Lösungen.

Die Schreibweise des Optimierungskriteriums als Skalarprodukt zeigt neben der Rotationsinvarianz auch eine implizite Annahme des Verfahrens auf. Der Vektor $\boldsymbol{g}_{i,d}$ ist ein Einheitsvektor, dessen Betrag durch die Rotation unangetastet bleibt, da es sich um eine normerhaltende Transformation handelt. Lediglich $\boldsymbol{e}_d - \boldsymbol{s}_i$ besitzt einen von Eins abweichenden Betrag. Somit liefert eine Abweichung zwischen dem beobachteten Einfallswinkel und der aus der Geometrie prädizierten Richtung bei größeren Entfernungen zwischen Sensor und Ereignis auch einen größeren Beitrag zur Zielfunktion. Dementsprechend werden bei der Optimierung bevorzugt die Ereignisse mit einer großen Entfernung zum Sensorknoten berücksichtigt, da dort Abweichungen zwischen Messung und Prädiktion die größte Auswirkung auf Positionsschätzung haben.

Um das Problem der Rotationsinvarianz zu beheben, wird erneut Gl. (5.13) betrachtet. Die darin enthaltene Rotationsmatrix sorgt einerseits für eine Transformation der Vektoren $\boldsymbol{e}_d - \boldsymbol{s}_i$ und $\boldsymbol{g}_{i,d}$ in ein gemeinsames Koordinatensystem. Andererseits führt die zusätzliche Drehung um $\frac{\pi}{2}$ dazu, dass die Vektoren im Optimum nicht in dieselbe Richtung zeigen, sondern orthogonal zueinander stehen und deshalb die ungewollte Mehrdeutigkeit eintritt. Die Eindeutigkeit des Skalarproduktes ist jedoch nur gegeben, sofern die beteiligten Vektoren in dieselbe bzw. genau entgegengesetzte Richtung zeigen.

Wird die zusätzliche Drehung von $\frac{\pi}{2}$ entfernt, entsteht eine Formulierung, bei der die Vektoren im Optimum in dieselbe Richtung zeigen. Allerdings liefert das modifizierte Skalarprodukt im Optimum jetzt den Betrag des Vektors $\boldsymbol{e}_d - \boldsymbol{s}_i$ als Ergebnis:

$$(\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}}\, \boldsymbol{R}_{\mathrm{xy}}(\theta_i)\, \boldsymbol{g}_{i,d} = \|\boldsymbol{e}_d - \boldsymbol{s}_i\|_2 \,. \tag{5.14}$$

Die Umformung dieser Gleichung führt zu

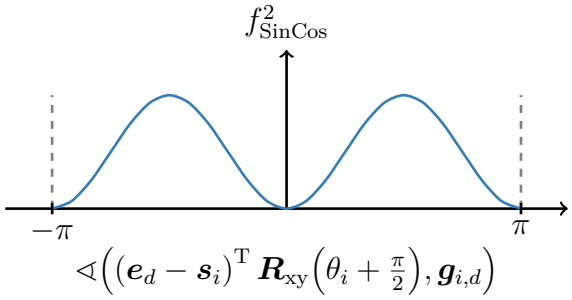
$$1 - \frac{(\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}}}{\|\boldsymbol{e}_d - \boldsymbol{s}_i\|_2} \boldsymbol{R}_{\mathrm{xy}}(\theta_i)\, \boldsymbol{g}_{i,d} = 0 \tag{5.15}$$

und gestattet damit die Formulierung einer weiteren Zielfunktion

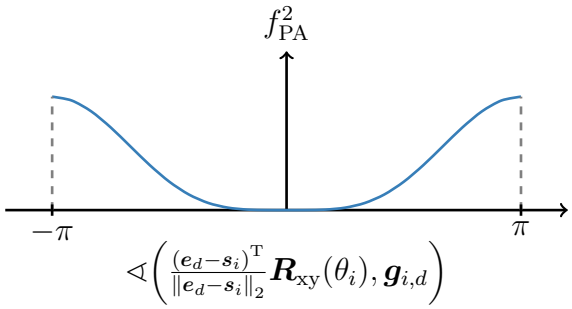
$$f_{\mathrm{PA}}(\boldsymbol{s}_i, \theta_i, \boldsymbol{e}_d; \varphi_{i,d}) = 1 - \frac{(\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}}}{\|\boldsymbol{e}_d - \boldsymbol{s}_i\|_2} \boldsymbol{R}_{\mathrm{xy}}(\theta_i)\, \boldsymbol{g}_{i,d}, \tag{5.16}$$

die ebenfalls die Lösung des Geometriekalibrierungsproblems mittels Newton-Verfahren ermöglicht.

Im Unterschied zu den bisherigen Varianten liefert das Newton-Verfahren jetzt eine eindeutige Lösung, weil die Maximierung des Skalarproduktes zwischen Beobachtung und Prädiktion nur durch eine einzige geometrische Konfiguration erfüllbar ist. Zur Veranschaulichung der geschilderten Weiterentwicklung der Kostenfunktion von der orthogonalen Formulierung f_{SinCos} hin zur parallelen Variante f_{PA} dient Abb. 5.4. Sie zeigt schematisch den Verlauf der jeweiligen Kostenfunktionen in Abhängigkeit des von Messung und Prädiktion eingeschlossenen Winkels. Da das verwendete Newton-Verfahren die LS-Lösung bestimmt, ist in den Abbildungen 5.4a und 5.4b das Quadrat der Funktionen aus Gl. (5.13) bzw. Gl. (5.16) dargestellt. Zudem sind die Verläufe der beiden Funktionen zur einheitlichen Darstellung auf das jeweilige Maximum normiert worden. Der Operator $\sphericalangle(\boldsymbol{x}, \boldsymbol{y})$ kennzeichnet dabei den zwischen den Vektoren $\boldsymbol{x}$ und $\boldsymbol{y}$ eingeschlossenen Winkel.



(a)



(b)

Abbildung 5.4: Schematische Darstellung des Verlaufs der Kostenfunktionen f_{SinCos} (a) bzw. f_{PA} (b).

Anhand der vorliegenden Darstellung wird noch einmal deutlich, dass f_{SinCos} zwei Minima besitzt, die wiederum die Rotationinvarianz verursachen. Die umformulierte Variante f_{PA} hat hingegen nur noch ein Minimum. Durch die Beseitigung der Rotationsinvarianz ergeben sich bei der Lösung des Geometriekalibrierungsproblems keine Lösungen mehr, die eine zusätzliche 180°-Rotation der Sensoren aufweisen oder eine gespiegelte Anordnung beschreiben. Bei fehlerfreien Einfallswinkeln ist dementsprechend nur noch ein globales Optimum vorhanden. Außerdem sorgt die Normierung des Vektors $\boldsymbol{e}_d - \boldsymbol{s}_i$ für eine gleichmäßige Berücksichtigung der Abweichungen von allen Beobachtungen.

5.4 Analyse des erweiterten Einfallswinkelverfahrens

Die zuvor entwickelten Erweiterungen des Einfallswinkelverfahrens haben das Ziel, die Rotationsinvarianz der Lösung und die Beeinträchtigung des Konvergenzverhaltens durch die Tangensfunktionen zu beseitigen. Weiterhin bietet die Ergänzung des Gleichungssystems um eine zusätzliche Gleichung eine alternative Möglichkeit zur Fixierung der Skalierung. Zur Überprüfung des Erfolgs der vorgenommenen Modifikation dient die nachfolgende Analyse.

Die Grundlage der Analyse bilden zufällig generierte Szenarien. Diese ähneln der in Abb. 2.1b gezeigten Anordnung und spiegeln damit typische Sensoranordnungen wider. Sie bestehen jeweils aus einem rechteckigen Raum mit einer Grundfläche zwischen $7{,}50 \times 5{,}50\,\mathrm{m}^2$ und $8{,}50 \times 6{,}50\,\mathrm{m}^2$. In jedem Raum erfolgt eine gleichverteilte Positionierung von 4 Sensoren mit einer maximalen Entfernung von 0,50 m zu der nächstgelegenen Wand. Darüber hinaus sind die Sensoren so ausgerichtet, dass sie ungefähr zur Raummitte zeigen. Ferner besitzen die Ereignispositionen einen minimalen Abstand von 0,50 m zu jeder Wand und stammen ebenfalls aus einer Gleichverteilung. Gemäß Gl. (5.5) sind zur Lösung des Problems mindestens 5 Ereignisse erforderlich. Die zufällige Wahl führt jedoch in einigen Fällen auch zu ungünstigen Positionierungen der Ereignisse, sodass im weiteren Verlauf 6 Ereignisse genutzt werden, um eine zuverlässigere Lösung des Gleichungssystems zu ermöglichen. Eine exemplarische Anordnung der Sensoren und Ereignisse, die sich aufgrund der beschriebenen Rahmenbedingungen ergibt, ist in Abb. 5.5 visualisiert.

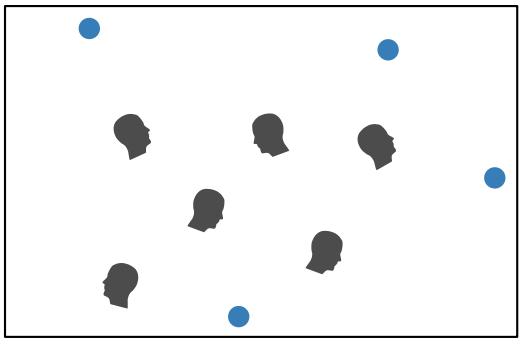


Abbildung 5.5: Exemplarische Anordnung der Sensoren (blaue Punkte) und Ereignisse (graue Köpfe), die zur Evaluierung der verschiedenen Formulierungen der Zielfunktionen dienen.

Als Eingabedaten für das Newton-Verfahren dienen zunächst störungsfreie Einfallswinkel. Die Optimierung wird terminiert, sobald die Kosten, die sich gemäß Gl. (5.7) ergeben, einen geeigneten Schwellwert unterschreiten. Im Rahmen dieser Arbeit beträgt der Schwellwert $1 \cdot 10^{-7}$. Zusätzlich wird das Newton-Verfahren auch dann beendet, wenn die Änderung sämtlicher unbekannter Parameter, in zwei aufeinander folgenden Schritten $1 \cdot 10^{-5}$ unterschreitet oder eine maximale Anzahl von Iterationen erreicht ist. Die durchschnittliche Anzahl der Iterationen bis zum Erreichen der ersten beiden Abbruchkriterien beträgt ca. 30 bis 50. Damit auch bei ungünstigen Szenarien und/oder Startwerten ausreichend viele Schritte möglich sind, beträgt die Obergrenze in diesem Experiment 500 Iterationen.

Sofern eines der ersten beiden Kriterien zur Terminierung führt, wird das Newton-Verfahren als konvergiert, anderenfalls als divergiert bezeichnet. Da eine Konvergenz nur dann garantiert ist, wenn der Startwert nahe genug an der gesuchten Nullstelle liegt, erfordert eine divergente Ausführung einen erneuten Versuch mit einem anderen Startwert. Allerdings führt die mehrfache Anwendung des Newton-Verfahrens zu einer erheblichen Steigung des Rechenaufwandes, da bei einer divergenten Ausführung die maximale Anzahl der Iterationen berechnet wird. Somit ist eine hohe Konvergenzrate nach möglichst wenigen Versuchen anzustreben.

Die Rate divergenter Ausführungen in einem Experiment mit 10 000 zufälligen Szenarien ist in Abb. 5.6 in Abhängigkeit der Anzahl der Initialisierungsversuche dargestellt. Erwartungsgemäß schneidet die aus [KWL08] stammende Formulierung $\boldsymbol{f}_{\mathrm{Tan}}$ aufgrund der Polstellen der Tangensfunktionen am schlechtesten ab. Trotz fehlerfreier Einfallswinkel divergieren im ersten Anlauf fast 16 % der Lösungsversuche. Durch erneute Initialisierungen lässt sich der Anteil der fehlgeschlagenen Lösungen zwar reduzieren, allerdings stagniert dieser bei ca. 2 %. Die Weiterentwicklung $\boldsymbol{f}_{\mathrm{SinCos}}$, die stattdessen Sinus- und Kosinusfunktionen nutzt, ist in der Lage, die Anzahl der divergenten Versuche deutlich zu reduzieren. Die darauf aufbauende Modifikation, die auch die Rotationsinvarianz behebt, führt zu einer weiteren Reduktion der fehlgeschlagenen Anläufe. Der Grund für die weitere Steigerung der Konvergenzrate dürfte jedoch nicht die Beseitigung der Rotationsinvarianz sein, sondern die damit verbundene veränderte Normierung. Während bei $\boldsymbol{f}_{\mathrm{SinCos}}$ eine abstandsabhängige Gewichtung vorliegt, berücksichtigt $\boldsymbol{f}_{\mathrm{PA}}$ alle Ereignisse gleichermaßen und sorgt somit für einen äquivalenten Verlauf der Kostenfunktion in jeder Dimension (vgl. Abb. 5.4). Insgesamt liegen somit durch die Weiterentwicklung der Zielfunktion zu $\boldsymbol{f}_{\mathrm{PA}}$ bereits nach nur einem Versuch genauso viele erfolgreiche Ausführungen vor, wie bei $\boldsymbol{f}_{\mathrm{Tan}}$ nach mehr als fünf Initialisierungen.

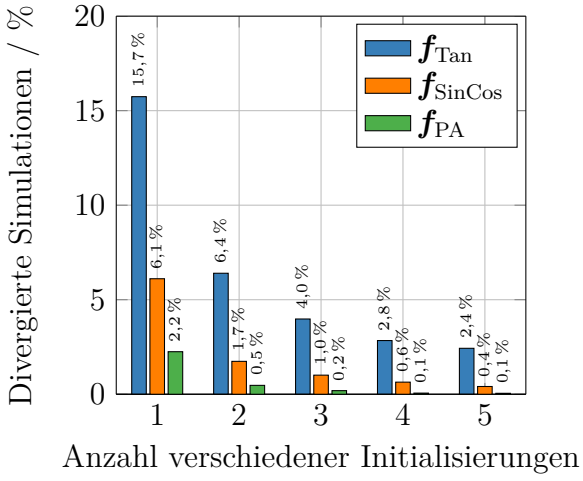


Abbildung 5.6: Prozentualer Anteil divergenter Lösungsversuche bei verschiedenen Zielfunktionen in Abhängigkeit der Anzahl der Initialisierungsversuche.

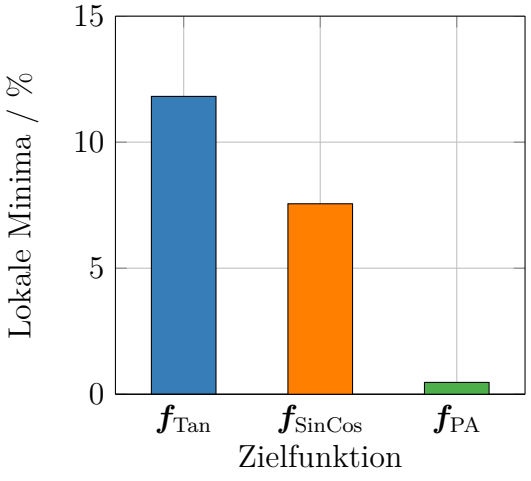


Abbildung 5.7: Prozentualer Anteil lokaler Minima innerhalb der konvergierten Lösungsversuche (siehe Abb. 5.6).

Wie zuvor erwähnt, beruhen die präsentierten Ergebnisse auf idealen Einfallswinkeln. Der weitere Verlauf dieses Abschnittes wird zeigen, dass Störungen der Winkel den Anteil der divergenten Optimierungsversuche signifikant steigern. Damit liefern diese Untersuchungen ein zusätzliches Argument für die Verwendung von $\boldsymbol{f}_{\mathrm{PA}}$.

Die Konvergenz des Newton-Verfahrens bedeutet lediglich, dass ein Minimum der Zielfunktion gefunden wurde. Aufgrund der nichtlinearen Zielfunktion sind jedoch neben der gesuchten geometrischen Anordnung lokale Minima ebenso möglich. Deshalb stellt Abb. 5.7 den prozentualen Anteil lokaler Minima innerhalb der konvergierten Ausführungen des Newton-Verfahrens aus Abb. 5.6 dar. Ausgelöst durch die Rotationsinvarianz der Zielfunktionen $\boldsymbol{f}_{\mathrm{Tan}}$ und $\boldsymbol{f}_{\mathrm{SinCos}}$, ergeben sich bei diesen beiden Formulierungen automatisch Minima, die dieselben Kosten wie die tatsächliche Anordnung aufweisen. Diese Minima bleiben jedoch in der vorliegenden Auswertung unberücksichtigt, da sie anhand der Kosten nicht von dem gesuchten Optimum zu unterscheiden sind.

Eine Möglichkeit zur Erkennung, ob nach erfolgreichem Abschluss des Newton-Verfahrens ein lokales oder ein globales Minimum vorliegt, bietet die Auswertung von Gl. (5.7) und der anschließende Vergleich des Resultates mit einem Schwellwert. Beispielhaft für die Zielfunktion $\boldsymbol{f}_{\mathrm{PA}}$ erfolgt mithilfe der in Tab. 5.1 dargestellten Konfusionsmatrix eine Bewertung der erläuterten Vorgehensweise.

Klasse \ Klassifikation	Lokales Minimum	Globales Minimum
Lokales Minimum	0,4 %	0,0 %
Globales Minimum	0,1 %	99,5 %

Tabelle 5.1: Konfusionsmatrix zur Bewertung der Klassifikation, ob das Newton-Verfahren ein lokales bzw. ein globales Minimum gefunden hat.

Die Ergebnisse dokumentieren, dass durch einen geeigneten Schwellwert eine zuverlässige Erkennung des globalen Minimums gegeben ist. Allerdings hängt das Ergebnis von Gl. (5.7) auch von den Störungen der Einfallswinkel ab. Daher erzielt der Schwellwertvergleich nur dann eine verlässliche Klassifikation, wenn fehlerfreie Einfallswinkelschätzungen vorliegen. Falls die Winkelschätzungen hingegen Störungen aufweisen, eignet sich Gl. (5.7) nur noch zur Erkennung von Lösungen, die sehr weit vom Optimum entfernt sind und nicht mehr zur Erkennung des globalen Minimums.

Insgesamt bestätigen die Ergebnisse, dass $\boldsymbol{f}_{\mathrm{PA}}$ am zuverlässigsten konvergiert und die Wahrscheinlichkeit nur ein lokales Minimum zu erreichen, im Vergleich zu $\boldsymbol{f}_{\mathrm{Tan}}$, klar reduziert wird. Letztendlich ist die entwickelte Formulierung $\boldsymbol{f}_{\mathrm{PA}}$ deshalb deutlich besser zur Bestimmung der Geometrie geeignet als die aus [KWL08] stammende Variante $\boldsymbol{f}_{\mathrm{Tan}}$.

Unabhängig von der eingesetzten Zielfunktion ist die Fixierung der Skalierung essentiell. Daher beschäftigt sich der folgende Teil der Analyse mit der Fragestellung, welche der in Abschnitt 5.3 präsentierten Varianten zur Festlegung der Skalierung am besten geeignet ist. Die Grundlage der Untersuchungen bildet die bereits aus dem vorausgegangenen Experiment bekannte Versuchsanordnung, wobei lediglich die präferierte Formulierung der Zielfunktion $\boldsymbol{f}_{\mathrm{PA}}$ Berücksichtigung findet.

Bei der Nutzung störungsfreier Einfallswinkel liefern alle Skalierungsvarianten eine perfekte Rekonstruktion der Geometrie und gestatten daher keine Bewertung. Deswegen weisen die jetzt verwendeten Einfallswinkel im Gegensatz zum vorangegangenen Experiment eine Störung auf. Diese wird durch eine von Mises-Verteilung mit einer Standardabweichung σ_{M} von ca. 0,9° generiert. Gemäß der Ergebnisse aus Abschnitt 4.6 entspricht dies ungefähr einer Nachhallzeit von 0,2 s, sofern die WKM zur Einfallswinkelschätzung bei einem drei-elementigen Array dient.

Da das Einfallswinkelverfahren die Sensorpositionen relativ zum ersten Sensor liefert, erfordert die Bestimmung des Positionierungsfehlers zunächst eine Abbildung der Sensorposition in das Koordinatensystem in dem die Referenzpositionen vorliegen. Nach der Abbildung in das Referenzkoordinatensystem (siehe Abschnitt 9.1) ergibt sich für die betrachteten Möglichkeiten zur Fixierung der Skalierung der in Abb. 5.8 veranschaulichte mittlere Positionierungsfehler ϵ_{P}. Außerdem führen die Störungen der Einfallswinkel dazu, dass mehr Ausführungen divergieren. Deshalb enthält die Legende in Abb. 5.8 zusätzlich die Raten der konvergierten Ausführungen. Diese kennzeichnen zugleich die maximal erreichbaren Werte der kumulativen Histogramme. Allerdings erreichen die in Abb. 5.8 gezeigten Histogramme die angegeben Werte nicht ganz, da der Fehler einiger Ausführungen außerhalb des dargestellten Wertebereiches liegt.

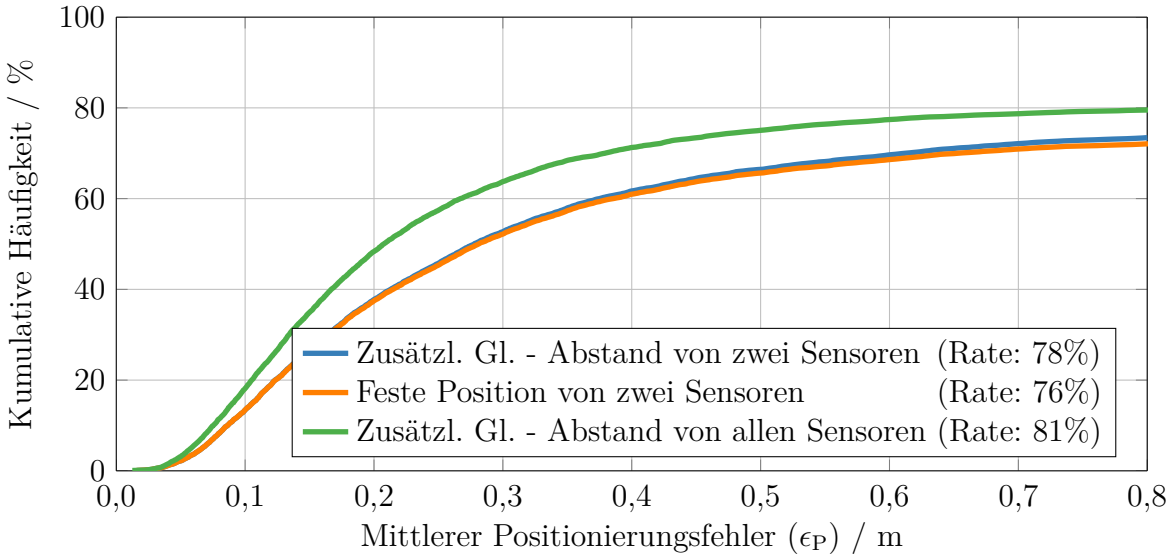


Abbildung 5.8: Kumulative Häufigkeitsverteilung des mittleren Positionierungsfehlers bei der Verwendung verschiedener Varianten zur Fixierung der Skalierung des Geometriekalibrierungsproblems.

Da sich gemäß der Ergebnisse aus Abb. 5.6 bereits mit nur einer Initialisierung hohe Konvergenzraten erzielen lassen, basieren die jetzt begutachteten Experimente ebenfalls auf einer Initialisierung. Der Vergleich zwischen Abb. 5.6 und Abb. 5.8 zeigt allerdings eine deutliche Zunahme der fehlgeschlagenen Lösungsversuche, die somit auf die Störungen der Einfallswinkel zurückzuführen ist.

Bei dem eigentlich im Mittelpunkt stehenden Vergleich des Positionierungsfehlers der verschiedenen Varianten zur Fixierung der Skalierung schneidet die feste Positionierung von zwei Sensoren in etwa genauso gut ab, wie die vorgeschlagene Fixierung

des Abstandes dieser Sensoren durch eine zusätzliche Gleichung. Der Ansatz, alle Sensoren zur Formulierung einer Skalierungsgleichung heranzuziehen, führt hingegen zu einer Reduktion des Fehlers. Ein weiteres Argument für die Verwendung einer Skalierungsgleichung, die alle Sensoren umfasst, liefert die dort ebenfalls größer ausfallende Konvergenzrate. Die Entscheidung, welche Möglichkeit zur Fixierung der Skalierung gewählt werden sollte, hängt jedoch vom Einsatzgebiet des Kalibrierungsalgorithmus ab. Sofern das Ziel lediglich in einer relativen Kalibrierung besteht und die Rückgewinnung der Skalierung durch einen der in Kapitel 7 oder Kapitel 8 präsentierten Ansätze erfolgt, ist die Skalierungsgleichung, die den Abstand aller Sensoren fixiert, zu bevorzugen. Wenn die Skalierung hingegen durch die a priori Kenntnis einer Distanz erfolgen soll, scheidet der bisherige Favorit aus, da anderenfalls die Summe der paarweisen Abstände vorliegen müsste, sodass durch die Anwendung der MDS (vgl. Abschnitt 2.2.1) das Geometriekalibrierungsproblem unmittelbar gelöst wäre.

5.5 Interpretation des erweiterten Einfallswinkelverfahrens als Maximum-Likelihood-Schätzer

Das bislang präsentierte Kalibrierungsverfahren sowie dessen Weiterentwicklungen aus Abschnitt 5.3 stützten sich ausschließlich auf die geometrische Anschauung der Problemstellung (vgl. Abb. 5.1). Die Ausführungen in diesem Abschnitt beleuchten das Geometriekalibrierungsproblem nun aus wahrscheinlichkeitstheoretischer Sichtweise. Sie zeigen den Zusammenhang zwischen der geometrischen Anschauung des Problems und der Formulierung als Maximum-Likelihood-Schätzung (ML-Schätzung).

Ausgangspunkt für die folgenden Schilderungen sind die Analysen aus Kapitel 4. Diese belegen, dass der Fehler der Winkelschätzung im allgemeinen von verschiedenen Faktoren abhängt. Sofern aber ein dreieckiges Array mit 0,05 m Kantenlänge zum Einsatz kommt und die entwickelte WKM zur Winkelschätzung dient, lässt sich der Fehler durch eine mittelwertfreie von Mises-Verteilung approximieren (siehe Abb. 4.24). Dementsprechend entsteht der vom Sensor $\boldsymbol{s}_i$ zum Ereignis $\boldsymbol{e}_d$ gemessene Einfallswinkel $\varphi_{i,d}$ aus einer additiven Überlagerung des tatsächlichen Einfallswinkels $\mu_{i,d}$ und dem mittelwertfreien Fehler, der einer von Mises-Verteilung folgt (vgl. Abb. 5.9) .

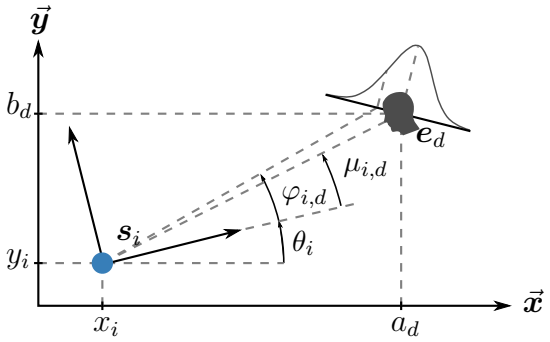


Abbildung 5.9: Interpretation der geometrischen Beziehung der Geometriekalibrierung als ML-Problem.

Angesichts dieser Modellierung ist die Wahrscheinlichkeit, dass der Sensor $\boldsymbol{s}_i$ die Beobachtung $\varphi_{i,d}$ emittiert, durch die Dichte

$$\mathrm{p}(\varphi_{i,d};\mu_{i,d},\kappa_{i,d}) = \frac{1}{2\pi \cdot I_0(\kappa_{i,d})} \cdot \exp(\kappa_{i,d} \cdot \cos(\varphi_{i,d} - \mu_{i,d})) \tag{5.17}$$

gegeben. Bei den im Rahmen dieser Arbeit betrachteten Problemstellungen besteht ein Szenario aus mehreren Sensoren und Ereignissen. Dadurch kann unmittelbar ein gemeinsames ML-Problem aller Sensoren und Ereignisse

$$\langle \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\kappa}} \rangle = \underset{\boldsymbol{\mu},\boldsymbol{\kappa}}{\operatorname{argmax}} \prod_{i=1}^{I} \prod_{d=1}^{D} \mathrm{p}(\varphi_{i,d};\mu_{i,d},\kappa_{i,d}), \tag{5.18}$$

mit $\boldsymbol{\mu} = [\mu_{1,1} \ \cdots \ \mu_{I,D}]^{\mathrm{T}}$ und $\boldsymbol{\kappa} = [\kappa_{1,1} \ \cdots \ \kappa_{I,D}]^{\mathrm{T}}$, formuliert werden. Zunächst erscheint diese Formulierung wenig hilfreich, da doppelt so viele Freiheitsgrade wie Beobachtungen vorliegen und vorerst keine Abhängigkeit zu den gesuchten geometrischen Größen gegeben ist. Dennoch bildet Gl. (5.18) den Grundstein für die Entwicklung eines ML-Schätzers zur Lösung des Geometriekalibrierungsproblems.

Unter Berücksichtigung der angenommenen VON MISES-Verteilung für den Schätzfehler ergibt sich die *Log-Likelihood*-Funktion

$$\mathrm{L}(\boldsymbol{\mu},\boldsymbol{\kappa}) = \sum_{i=1}^{I} \sum_{d=1}^{D} \kappa_{i,d} \cdot \cos(\varphi_{i,d} - \mu_{i,d}) - \sum_{i=1}^{I} \sum_{d=1}^{D} 2\pi \cdot I_0(\kappa_{i,d}) \tag{5.19}$$

zur Maximierung von Gl. (5.18). Die Konzentrationsparamter $\kappa_{i,d}$, die die Unsicherheiten in den Beobachtungen ausdrücken, sind gemäß der Untersuchungen aus Kapitel 4 hauptsächlich von der Nachhallzeit abhängig und können deshalb als konstant für alle Sensoren und Ereignisse betrachtet werden. Damit vereinfacht sich Gl. (5.18) zu

$$\hat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu}}{\operatorname{argmax}} \sum_{i=1}^{I} \sum_{d=1}^{D} \cos(\varphi_{i,d} - \mu_{i,d}). \tag{5.20}$$

Weiterhin kann der Kosinus von der Differenz zweier Winkel als Skalarprodukt interpretiert werden. Sofern der zu $\mu_{i,d}$ gehörende Vektor durch $\widetilde{\boldsymbol{g}}_{i,d}$ gegeben ist und für die Darstellung des gemessenen Einfallswinkels $\varphi_{i,d}$ als Vektor $\boldsymbol{g}_{i,d}$ der aus Gl. (5.11) bekannte Zusammenhang dient, ergibt sich

$$\cos(\varphi_{i,d} - \mu_{i,d}) = \widetilde{\boldsymbol{g}}_{i,d}^{\mathrm{T}}\, \boldsymbol{g}_{i,d}. \tag{5.21}$$

Diese Darstellung als Skalarprodukt erlaubt es unmittelbar eine Abhängigkeit zu den gewünschten Parametern herzustellen, indem der Mittelwertvektor $\widetilde{\boldsymbol{g}}_{i,d}$ mithilfe des geometrischen Zusammenhangs

$$\widetilde{\boldsymbol{g}}_{i,d}^{\mathrm{T}} = \frac{(\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}}}{\|\boldsymbol{e}_d - \boldsymbol{s}_i\|_2} \boldsymbol{R}_{\mathrm{xy}}(\theta_i) \tag{5.22}$$

ausgedrückt wird. Insgesamt gilt es somit das Optimierungsproblem

$$\langle \hat{\boldsymbol{S}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{E}} \rangle = \underset{\boldsymbol{S},\boldsymbol{\theta},\boldsymbol{E}}{\operatorname{argmax}} \sum_{i=1}^{I} \sum_{d=1}^{D} \frac{(\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}}}{\|\boldsymbol{e}_d - \boldsymbol{s}_i\|_2} \boldsymbol{R}_{\mathrm{xy}}(\theta_i)\, \boldsymbol{g}_{i,d} \tag{5.23}$$

zu lösen, das jetzt nicht mehr von den Mittelwerten $\boldsymbol{\mu}$, sondern nur noch von den Parametern der Sensoren bzw. Ereignisse abhängt. Darüber hinaus sorgt die Darstellung der mittleren Richtung $\widetilde{\boldsymbol{g}}_{i,d}$ durch den geometrischen Zusammenhang (vgl. Gl. (5.22)) dafür, dass sich die Anzahl der Parameter von vormals $I \cdot D$ (siehe Gl. (5.21)) auf $3 \cdot (I-1) + 2 \cdot D$ reduziert, während die Anzahl der Messungen weiterhin $I \cdot D$ beträgt.

Der Zusammenhang zwischen dem ML-Schätzer und der aus der geometrischen Anschauung entwickelten Zielfunktion (5.16) wird deutlich, indem das Maximierungsproblem aus Gl. (5.23) zunächst durch das äquivalente Minimierungsproblem

$$\langle \hat{\boldsymbol{S}},\, \hat{\boldsymbol{\theta}},\, \hat{\boldsymbol{E}} \rangle = \underset{\boldsymbol{S},\boldsymbol{\theta},\boldsymbol{E}}{\operatorname{argmin}} \sum_{i=1}^{I} \sum_{d=1}^{D} 1 - \frac{(\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}}}{\|\boldsymbol{e}_d - \boldsymbol{s}_i\|_2} \boldsymbol{R}_{\mathrm{xy}}(\theta_i)\, \boldsymbol{g}_{i,d} \tag{5.24}$$

dargestellt wird. Der Vergleich von Gl. (5.24) und der Zielfunktion (5.16) bestätigt, dass es sich bei dem ausschließlich aufgrund der geometrischen Anordnung entwickelten Gleichungssystem um einen ML-Schätzer für von Mises-verteilte Winkelfehler handelt. Die Lösung des Gleichungssystems durch das Newton-Verfahren bietet zudem eine effektive Möglichkeit zur Berechnung des Schätzwertes. Darüber hinaus ergibt sich durch die Interpretation des erweiterten Einfallswinkelverfahrens als ML-Schätzer ein weiteres Argument dafür, die Winkelschätzung mit drei Mikrofonen durchzuführen, weil nun die dem Schätzer zugrunde liegende Verteilung eine sehr gute Approximation des tatsächlichen Winkelfehlers ermöglicht.

5.6 Analyse des Maximum-Likelihood-Schätzers

Für den praktischen Einsatz des ML-Schätzers ist primär die zu erwartende Kalibrierungsgenauigkeit relevant. Der Vergleich der verschiedenen Varianten zur Fixierung der Skalierung aus Abschnitt 5.4 mithilfe synthetisch erzeugter Störungen liefert bereits erste Erkenntnisse und weist darauf hin, dass schon eine geringe Störung der Einfallswinkel zu ausgeprägten Fehlern bei der Positionsbestimmung der Sensoren führen kann. Eine detaillierte Analyse der Auswirkungen des Fehlers der Winkelschätzung auf das Kalibrierungsergebnis ist daher unbedingt notwendig. Die folgenden Untersuchungen sollen deshalb einerseits die Auswirkungen unterschiedlich stark ausgeprägter Winkelfehler darstellen, andererseits sollen sie aber auch eine Bewertung ermöglichen, inwiefern sich eine Diskrepanz zwischen der angenommenen von Mises-Verteilung und der tatsächlichen Distribution des Winkelfehlers auf die Leistungsfähigkeit der Kalibrierung auswirkt. Dementsprechend werden sowohl künstlich erzeugte Winkelschätzungen, die exakt dem zugrunde liegenden Modell entsprechen, als auch Einfallswinkelschätzungen von drei- ebenso wie von zwei-elementigen Arrays berücksichtigt.

Grundlage der Analyse bilden die aus Abschnitt 5.4 bekannten Szenarien, die aus einem ca. $8{,}00 \times 6{,}00\,\mathrm{m}^2$ großen Raum bestehen, in dem sich vier Mikrofonarrays mit einem maximalen Abstand von 0,50 m zur nächstgelegen Wand befinden. Damit eine Untersuchung mittels Simulationen eine belastbare Aussage gestattet, ist auch hier eine ausreichend große Menge zufälliger Szenarien unerlässlich. Allerdings erfordert die Erzeugung von 12-kanaligen Aufnahmen mit der Spiegel-Quellen-Methode zur Simulation von vier drei-elementigen Arrays einen beträchtlichen Rechenaufwand, der sogar den der eigentlichen Kalibrierung übersteigt.

Durch den Vergleich der Winkelschätzer aus Abschnitt 4.4 liegen für ein einzelnes Mikrofonarray die Schätzfehler eines breiten Spektrums verschiedener Kombinationen aus Mikrofon- und Quellpositionen sowie Nachhallzeiten vor. An dieser Stelle wird deshalb auf diese Daten zurückgegriffen und ein Modell darauf trainiert, welches den Fehler der Einfallswinkelschätzung nachbildet, wenn mithilfe der Spiegel-Quellen-Methode erzeugte Audiosignale verwendet werden. Infolgedessen lässt sich der erforderliche Rechenaufwand einschränken, sodass mehr Sensor-Ereignis-Konstellationen berücksichtigt werden können. Das verwendete Modell beschreibt den Winkelfehler durch ein Histogramm, das von der Nachhallzeit, dem tatsächlichen Einfallswinkel und der Entfernung zwischen Ereignisquelle und Sensorknoten abhängt. Somit bildet es die wichtigsten Einflussfaktoren des Winkelfehlers ab, ohne dass dabei jedes Mal eine Simulation der Audiosignale durch die Spiegel-Quellen-Methode erforderlich ist.

Da das gewählte Szenario aus vier Sensoren besteht, sind gemäß Gl. (5.5) mindestens fünf Ereignisse zur Lösung des Geometriekalibrierungsproblems nötig. Doch schon die in Abb. 5.8 präsentierten Ergebnisse belegen, dass es bei fehlerbehafteten Winkelschätzungen nicht ausreichend ist, eine Lösung mit dieser geringen Anzahl von Beobachtungen zu ermitteln, da die Quote der divergierenden Lösungsversuche anderenfalls beträchtlich ist. Als Konsequenz aus diesen Erkenntnissen verwendet die aktuelle Untersuchung 20 Ereignisse, um somit einen Ausgleich der Störungen mittels LS zu erlauben.

Um darüber hinaus einen Vergleich zwischen Szenarien mit zwei bzw. drei Mikrofonen pro Sensorknoten durchzuführen, ist eine gesonderte Berücksichtigung der Sensororientierung notwendig. Sofern nur eine lineare Anordnung der Mikrofone vorliegt, können abhängig von der Positionierung und Ausrichtung der Sensorknoten Ereignispositionen außerhalb des eindeutigen Detektionsbereichs von $\pm 90°$ auftreten, die den Einsatz zusätzlicher Methoden zur Auflösung der Mehrdeutigkeiten erfordern. Diese Methoden würden jedoch einen direkten Vergleich mit der dreieckigen Anordnung erschweren. Die Orientierung aller Sensorknoten ist in den folgenden Experimenten deshalb stets so gewählt, dass alle Ereignisse innerhalb des eindeutigen Bereichs der Arrays mit zwei Mikrofonen liegen.

Die Basis für den vorgenommenen Vergleich bilden 1000 zufällige Konfigurationen aus Sensoren und Ereignissen, die die zuvor genannten Rahmenbedingungen erfüllen. Zur Bewertung der jeweiligen Lösungen des erweiterten Einfallswinkelverfahrens dient erneut der mittlere Positionierungsfehler (siehe Abschnitt 9.1). Abb. 5.10 stellt diesen für Nachhallzeiten von 0,0 s, 0,2 s und 0,4 s sowohl bei der Verwendung eines Arrays mit zwei als auch mit drei Mikrofonen dar. Zusätzlich enthält Abb. 5.10 eine Darstellung des Kalibrierungsfehlers, wenn die Winkelschätzung nicht aus den Audiosignalen stammt, sondern der Fehler der Winkelschätzung aus der zur jeweiligen Nachhallzeit korrespondierenden von Mises-Verteilung gezogen wird. Um diese Ergebnisse zu kennzeichnen, ist hier statt der T_{60}-Zeit die zugehörige Standardabweichung der von Mises-Verteilung (σ_{M}) angegeben. Ferner enthält die Legende erneut die Rate der konvergierten Lösungsversuche.

Während ohne Nachhall der Positionierungsfehler für alle Möglichkeiten in etwa gleich ausfällt, ergeben sich mit ansteigender Nachhallzeit immense Differenzen. Insbesondere der bei zwei Mikrofonen auftretende systematische Fehler der Einfallswinkelschätzung (vgl. Abb. 4.6) verursacht ein signifikant schlechteres Abschneiden dieser Variante, sowohl bei den erzielten Positionierungsfehlern als auch bei den Konvergenzraten. Während

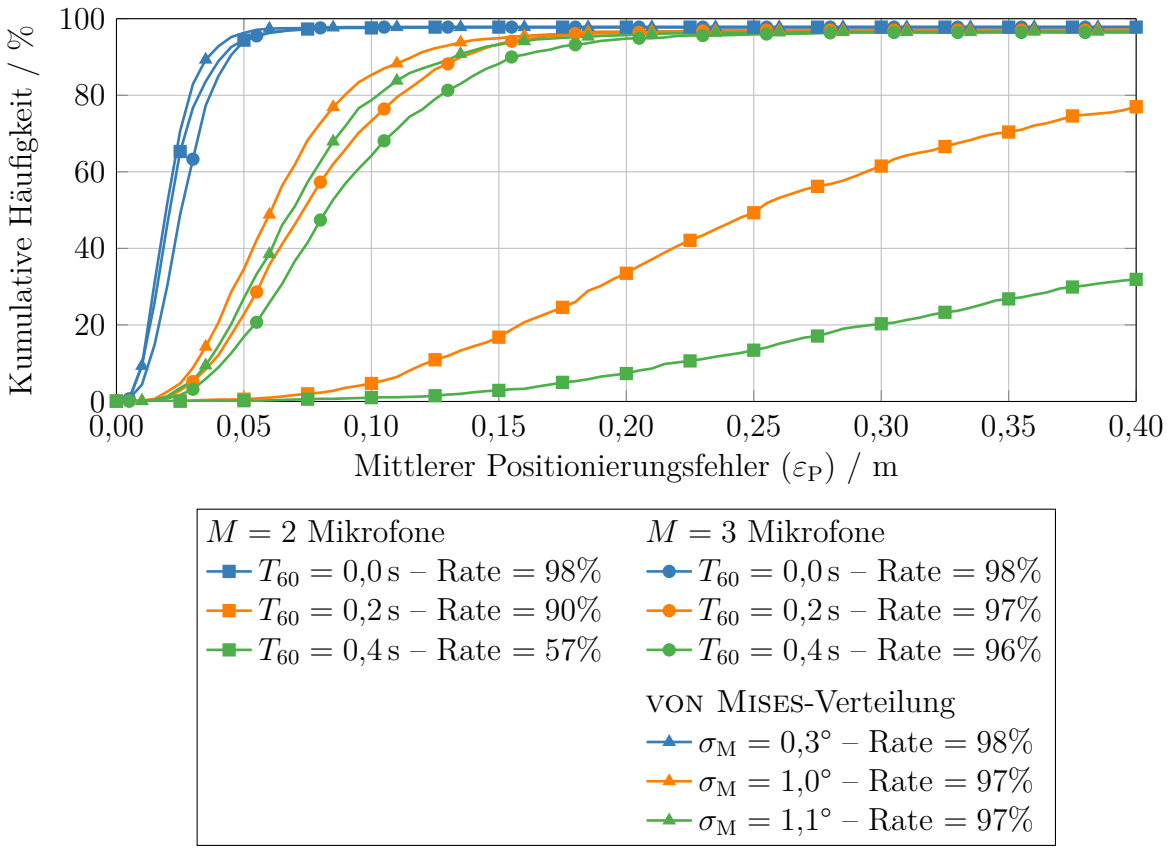


Abbildung 5.10: Kumulative Häufigkeitsverteilung des mittleren Positionierungsfehlers für verschiedene Ausprägungen des Winkelfehlers bei einer Geometriekalibrierung durch das erweiterte Einfallswinkelverfahren.

bei drei Mikrofonen die Konvergenzraten immer über 95 % liegen, tritt bei nur zwei Mikrofonen eine substanzielle Verschlechterung auf, sodass bei einer Nachhallzeit von 0,4 s nur noch in 57 % der Simulationen eine Kalibrierung möglich ist. Die abnehmenden Kovergenzraten lassen sich auf die zunehmenden Winkelfehler zurückführen, da bei den Simulationen für alle Nachhallzeiten dieselben Sensor- und Ereigniskonstellationen genutzt wurden.

Falls jeder Sensorknoten hingegen drei Mikrofone aufweist, sind nicht nur deutlich geringere Kalibrierungsfehler möglich, sondern auch der nachhallbedingte Anstieg des Fehlers fällt ebenfalls kleiner aus. Weiterhin gestattet die Darstellung des Kalibrierungsfehlers bei einer Modellierung der Störung des Einfallswinkels durch eine von Mises-Verteilung eine Bewertung der Approximation des Schätzfehlers durch eine von Mises-Verteilung. Die Standardabweichungen von $\sigma_M = 0{,}3°$, $\sigma_M = 1{,}0°$ und $\sigma_M = 1{,}1°$ korrespondieren dabei ungefähr mit den Nachhallzeiten von 0,0 s, 0,2 s bzw. 0,4 s. Die geringen Unterschiede bestätigen, dass die Annäherung des Winkelfehlers durch eine von Mises-Verteilung gerechtfertigt ist. Außerdem eröffnet sich dadurch die Möglichkeit, anstatt einer aufwändigen Simulation von mehrkanaligen Audiodaten sowie einer anschließenden Winkelschätzung durch die WKM, die Winkelfehler direkt synthetisch durch eine von Mises-Verteilung zu generieren.

5.7 Kalibrierung dreidimensionaler Anordnungen

Die zurückliegenden Untersuchungen bestätigen, dass das Einfallswinkelverfahren durch die vorgenommenen Erweiterungen auch die Kalibrierung eines akustischen Sensornetzes gestattet. Allerdings beschränken sich die bisherigen Betrachtungen auf zweidimensionale Anordnungen, d. h., sowohl die Sensorknoten als auch die zur Kalibrierung verwendeten Ereignisse liegen in einer Ebene. Sofern die Mikrofonarrays z. B. auf einem Tisch platziert sind und die Quellsignale von Sprechern stammen, die an diesem sitzen, ist die bislang getroffene Annahme näherungsweise erfüllt. Je größer aber die Abweichung zwischen dem wahren Szenario und der Approximation ist, desto größer wird der Bedarf für eine dreidimensionale Kalibrierung. Weiterhin entsteht mit zunehmender Entfernung zwischen der Quelle und der Mikrofonebene ein systematischer Fehler bei der Winkelschätzung, sofern diese nur in der Ebene erfolgt. Andererseits besitzt schon die zweidimensionale Problemstellung zahlreiche Unbekannte, deren Bestimmung eine Herausforderung darstellt. Zusätzlich gilt es, bei einer Erweiterung für den dreidimensionalen Fall die Rechenkomplexität zu berücksichtigen, welche durch die steigende Anzahl von unbekannten Parametern signifikant anwächst. Angesichts dieser limitierenden Faktoren dient dieser Abschnitt lediglich dazu, die grundsätzliche Erweiterungsfähigkeit des entwickelten Konzeptes auf ein dreidimensionales Szenario zu zeigen.

Bei einer dreidimensionalen Betrachtung der Kalibrierung steigen zunächst die Anforderungen an eine Mikrofongruppe. Während im Zweidimensionalen die Winkelschätzung in der xy-Ebene ausreicht (Azimuth), erfordert ein dreidimensionales Szenario zusätzlich die Bestimmung des Elevationswinkels. Damit ein Sensorknoten sowohl Azimuth als auch Elevation bereitstellen kann, ist eine dreidimensionale Anordnung der Sensoren innerhalb des Arrays notwendig. Aufgrund der bisherigen Erkenntnisse, sollten für die Schätzung des Winkels in einer Ebene mindestens drei Mikrofone zur Verfügung stehen. Daher ergibt sich nun eine Minimalkonfiguration von vier Mikrofonen pro Sensorknoten.

Die Darstellung der Zielfunktion des zweidimensionalen Szenarios als Skalarprodukt (vgl. Gl. (5.16)) zwischen dem geschätzten und dem prädizierten Einfallswinkel bietet unmittelbar die Möglichkeit zur Verallgemeinerung des Konzeptes auf drei Dimensionen. Zur Veranschaulichung der Vorgehensweise und des geometrischen Zusammenhangs im Dreidimensionalen dient Abb. 5.11.

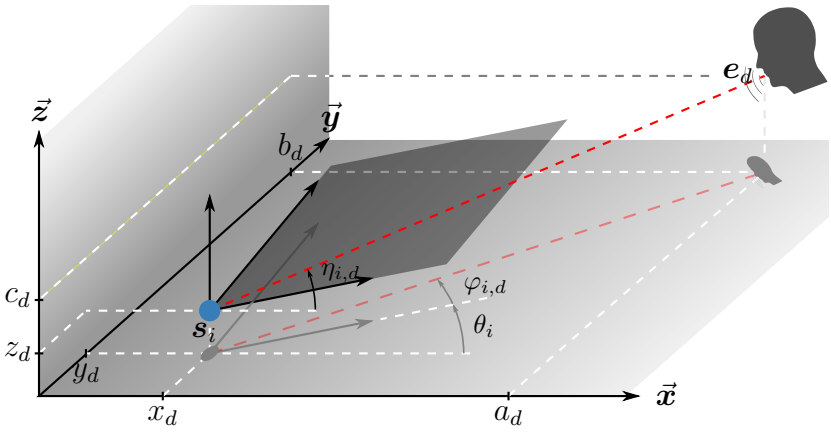


Abbildung 5.11: 3D Szenario: Geometrische Beziehung zwischen Sensor und Ereignis.

Zunächst gilt es hervorzuheben, dass die Sensorposition $\boldsymbol{s}_i = [x_i \;\; y_i \;\; z_i]^{\mathrm{T}}$ und das erfasste Ereignis $\boldsymbol{e}_d = [a_d \;\; b_d \;\; c_d]^{\mathrm{T}}$ jetzt ebenfalls durch dreidimensionale Ortsvektoren gegeben sind. Die Kombination von Azimuth $\varphi_{i,d}$ und Elevation $\eta_{i,d}$ ermöglicht es, die Beobachtung als Richtungsvektor

$$\boldsymbol{g}_{i,d} = \Big[\cos(\varphi_{i,d}) \cdot \cos(\eta_{i,d}) \quad \sin(\varphi_{i,d}) \cdot \cos(\eta_{i,d}) \quad \sin(\eta_{i,d})\Big]^{\mathrm{T}} \tag{5.25}$$

darzustellen. Dieser gibt wie auch bisher die Signaleinfallsrichtung im lokalen Koordinatensystem des Sensors $\boldsymbol{s}_i$ an.

Prinzipiell kann das lokale Sensorkoordinatensystem eine Rotation zur xy-Ebene und eine Neigung bezogen auf die xz-Ebene aufweisen. Die Rotation in der xy-Ebene wird, wie schon im zweidimensionalen Fall, mithilfe des Winkels θ_i ausgedrückt. Die Betrachtung einer Neigung zur xz-Ebene erscheint jedoch für die Geometriekalibrierung nicht erforderlich, weil die Sensoren typischerweise auf ebenen Flächen positioniert sein dürften. Ohne eine Neigung der Sensoren lässt sich, genau wie im zweidimensionalen Szenario, der aus der geometrischen Anordnung prädizierte Richtungsvektor durch eine Rotation in der xy-Ebene in das lokale Koordinatensystem der Sensorgruppe überführen. Dazu wird die bisherige Rotationsmatrix für den zweidimensionalen Fall zu

$$\boldsymbol{R}_{\mathrm{xy}}(\theta_i) = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) & 0 \\ \sin(\theta_i) & \cos(\theta_i) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.26}$$

erweitert, sodass schließlich die Zielfunktion

$$\boldsymbol{f}_{\mathrm{PA,3D}}(\boldsymbol{s}_i, \theta_i, \boldsymbol{e}_d; \varphi_{i,d}, \eta_{i,d}) = \frac{(\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}}}{\|\boldsymbol{e}_d - \boldsymbol{s}_i\|_2} \boldsymbol{R}_{\mathrm{xy}}(\theta_i)\, \boldsymbol{g}_{i,d} \tag{5.27}$$

entsteht. Angesichts der verwendeten Vektorschreibweise gleicht die Zielfunktion der in Gl. (5.16) eingeführten Variante für das zweidimensionale Szenario. Der Unterschied besteht darin, dass es sich bei den Positionen $\boldsymbol{s}_i$ und $\boldsymbol{e}_d$ um dreidimensionale Ortsvektoren handelt, die Beobachtungen aus Azimuth $\varphi_{i,d}$ und Elevation $\eta_{i,d}$ bestehen und die Rotationsmatrix jetzt ebenfalls dreidimensional ist.

Infolge der weitgehenden Übereinstimmung zu den Ausführungen in Abschnitt 5.5 lässt sich auch für den dreidimensionalen Fall zeigen, dass ein ML-Schätzer vorliegt. Ausgangspunkt dafür ist die von Mises-Fisher-Verteilung. Sie stellt die Verallgemeinerung der von Mises-Verteilung für drei und mehr Dimension dar [MJ09]. Die Verteilung eines dreidimensionalen Einheitsvektors $\boldsymbol{g}_{i,d}$ ist damit durch die Dichtefunktion

$$\mathrm{p}(\boldsymbol{g}_{i,d}; \boldsymbol{\mu}_{i,d}, \kappa_{i,d}) = \frac{\kappa_{i,d}}{4\pi \cdot \sinh(\kappa_{i,d})} \cdot \exp\Big(\kappa_{i,d} \cdot \boldsymbol{\mu}_{i,d}^{\mathrm{T}}\, \boldsymbol{g}_{i,d}\Big) \tag{5.28}$$

gegeben. Als Exponent tritt direkt das Skalarprodukt zwischen Mittelwert $\boldsymbol{\mu}_{i,d}$ und Beobachtung $\boldsymbol{g}_{i,d}$ auf. Somit entsteht abgesehen von der Normierungskonstanten dieselbe Darstellung, die auch aus der Kombination von Gl. (5.17) und Gl. (5.21) hervorgeht. Somit kann, wie schon für den zweidimensionalen Fall erläutert, der Mittelwert durch die gesuchten Parameter des Geometriekalibrierungsproblems ausgedrückt werden, sodass am Ende erneut das aus Gl. (5.23) bekannte Optimierungsproblem entsteht.

Aufgrund des identischen Ausdrucks, abgesehen von den Dimensionen der Vektoren und Matrizen, ermöglicht das Newton-Verfahren (vgl. Abschnitt 5.1) auch die Lösung des dreidimensionalen Geometriekalibrierungsproblems. Trotz der ausgeprägten Gemeinsamkeiten führt die Erweiterung auf drei Dimensionen insbesondere zu Veränderungen, die erhebliche Auswirkungen auf die Rechenkomplexität haben. Die Anzahl der minimal erforderlichen Ereignisse steigt auf

$$D \geq \frac{4\,(I-1)}{I-3}. \tag{5.29}$$

Die Konsequenz der vermeintlich geringfügigen Änderungen dieser Ungleichung im Vergleich zu Gl. (5.5) wird bei einer erneuten Betrachtung eines Szenarios aus 4 Sensoren deutlich. Anstatt bisher 5 Ereignissen sind zur Kalibrierung nun 12 Ereignisse notwendig. Dadurch wächst die Anzahl der zu kalibrierenden Unbekannten drastisch von vormals 19 auf jetzt 48 an. Damit einher geht eine deutliche Erhöhung des für die Kalibrierung erforderlichen Zeitaufwandes. Die Laufzeit des Newton-Verfahrens wird durch die in jedem Iterationsschritt erforderliche Inversion der Jacobi-Matrix dominiert. Da die Laufzeit in etwa kubisch mit der Anzahl der Unbekannten anwächst, ist der Aufwand für eine Iteration des Newton-Verfahrens der 3D-Kalibrierung ca. 16-fach größer als bei der 2D-Kalibrierung.

Zur Untersuchung der dreidimensionalen Variante des erweiterten Einfallswinkelverfahrens werden erneut die bereits in Abschnitt 5.6 betrachteten Sensor- und Ereigniskonstellationen verwendet. Im Unterschied zu Abschnitt 5.6 variieren die Höhen der Sensoren jetzt zwischen 0,8 m und 1,8 m, was bspw. einer Platzierung auf einem Tisch bzw. auf einem Schrank entspricht. Die Höhen der Quellpositionen liegen hingegen zwischen 1,2 m und 1,8 m und beschreiben damit bspw. sitzende oder auch stehende Sprecher. Obwohl für die Kalibrierung der 4 Sensoren 12 Ereignisse ausreichen, werden 25 Ereignisse berücksichtigt, um belastbare Ergebnisse zu erzielen. Unter der Annahme, dass die Fehler des Azimuth- und Elevationswinkels statistisch unabhängig sind, ermöglicht das in Abschnitt 5.6 trainierte Modell des Winkelfehlers nun auch die Erzeugung von Winkelschätzungen für dreidimensionale Szenarien.

Bei der Kalibrierung dreidimensionaler Szenarien steigt der Rechenaufwand nicht nur aufgrund der wachsenden Anzahl der zu kalibrierenden Parameter, sondern auch weil das Newton-Verfahren wesentlich mehr Iterationen zur Bestimmung der Parameter erfordert. Während im Zweidimensionalen ca. 100 Iterationen in den meisten Fällen ausreichen, erfordert die Kalibrierung dreidimensionaler Sensorkonfigurationen mehr als 1000 Iterationen. Außerdem zeigen sich Auswirkungen auf die Konvergenz des Newton-Verfahrens. Zur Untersuchung des Konvergenzverhaltens dient Abb. 5.12. Sie zeigt die Rate divergenter Ausführungen in Abhängigkeit von der Anzahl der Initialisierungsversuche und der Nachhallzeit.

Die dargestellten Ergebnisse dokumentieren eindeutig, dass die Zunahme des Winkelfehlers aufgrund einer Erhöhung der Nachhallzeit eine immense Steigerung der Anzahl der divergierenden Simulationen verursacht. Bis zu einer Nachhallzeit von 0,2 s reichen wenige Initialisierungsversuche, damit die Anzahl der erfolgreichen Ausführungen des Newton-Verfahrens deutlich mehr als 90 % beträgt. Bei einer Nachhallzeit von 0,4 s schlagen hingegen etwas mehr als 30 % der Kalibrierungen fehl.

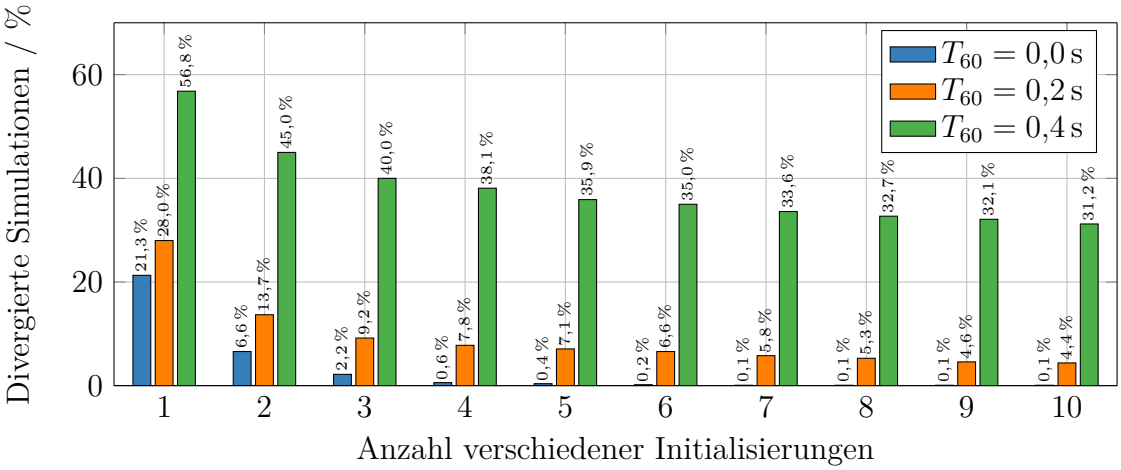


Abbildung 5.12: Prozentualer Anteil divergenter Lösungsversuche bei der Kalibrierung dreidimensionaler Sensorkonfigurationen in Abhängigkeit der Anzahl der Initialisierungsversuche.

Im Vergleich zu den Ergebnissen aus Abb. 5.6 scheint die Konvergenzrate im dreidimensionalen Fall deutlich geringer zu sein. Allerdings gilt es zu beachten, dass bei den Abb. 5.6 zugrunde liegenden Untersuchungen fehlerfreie DOA genutzt wurden, während bei der aktuellen Analyse Störungen der Einfallswinkel vorhanden sind. Zur Einordnung der Ergebnisse aus Abb. 5.12 ist stattdessen Abb. 5.8 besser geeignet. Sie zeigt, dass bei einer Nachhallzeit von 0,2 s ca. 20 % der Ausführungen divergieren. Somit wächst die Rate der divergenten Kalibrierungen im dreidimensionalen Fall nur um 10 % an.

Zur Untersuchung des Positionierungsfehlers bei der Kalibrierung dreidimensionaler Sensorkonfigurationen wurden bis zu 10 Initialisierungen durchgeführt, damit ausreichend viele erfolgreiche Kalibrierungen vorliegen. Die kumulativen Histogramme, die sich dabei ergeben, stellt Abb. 5.13 dar.

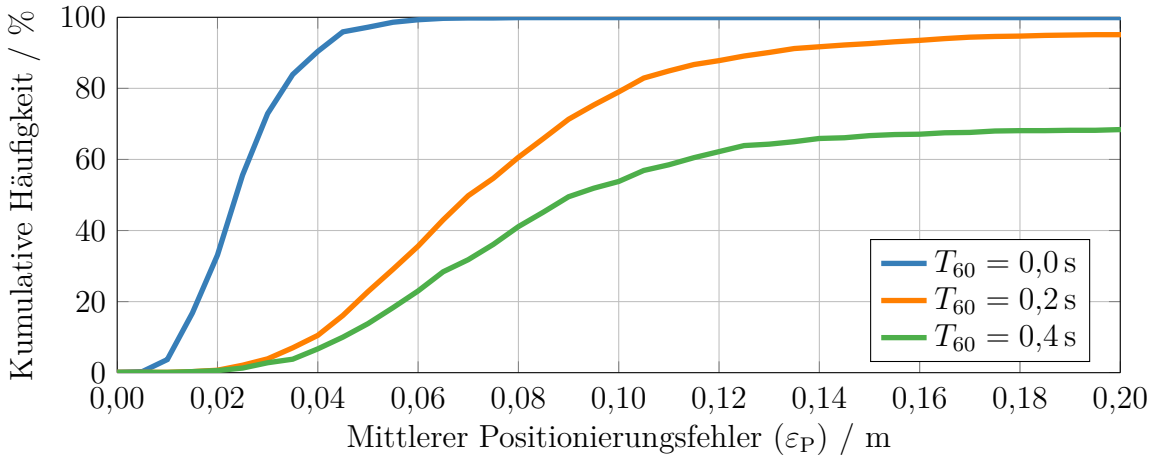


Abbildung 5.13: Kumulative Häufigkeitsverteilung des mittleren Positionierungsfehlers der Geometriekalibrierung dreidimensionaler Sensorkonfigurationen bei verschiedenen Nachhallzeiten.

Die Betrachtung der Ergebnisse zeigt unmittelbar, dass die zu den Nachhallzeiten von 0,2 s und 0,4 s gehörenden Histogramme bei ca. 95 % bzw. 69 % stagnieren. Bei den verbleibenden 5 % bzw. 31 % handelt es sich um divergierte Lösungen. Eine nähere Begutachtung der Lösungen, die die größten Fehler haben, belegt zudem, dass es sich dabei vornehmlich um Konfigurationen handelt, bei denen nur eine geringe Ausdehnung in der dritten Dimension vorhanden ist oder einzelne Winkelschätzungen sehr große Störungen aufweisen. Andererseits fällt der Fehler bei der Kalibrierung im Dreidimensionalen teilweise geringer aus als für das zweidimensionale Szenario (vgl. Abb. 5.10). Für diesen Trend sind zwei Gründe zu nennen. Zum einen ist durch die Unabhängigkeit der Fehler von Azimuth und Elevation eine Kompensation des Fehlers möglich, wenn lediglich einer der beiden Winkel eine größere Störung aufweist. Zum anderen gestattet die erhöhte Anzahl der Beobachtungen einen besseren Ausgleich von Fehlern der Winkelschätzung.

Insgesamt belegen die Simulationen somit, dass das entwickelte Konzept auch zur Kalibrierung von dreidimensionalen Anordnungen geeignet ist. Aufgrund der enormen Steigerung des Rechenaufwandes durch die gestiegene Anzahl der zu kalibrierenden Parameter sowie die Erhöhung der Iterationsanzahl im Newton-Verfahren steigt die Ausführungszeit im Vergleich zum zweidimensionalen Fall in etwa um das 30-fache an.

5.8 Zusammenfassung

In diesem Kapitel wurde ein Algorithmus zur Geometriekalibrierung akustischer Sensornetze entwickelt und anschließend eingehend analysiert. Grundlage für die eigenen Entwicklungen war das ursprünglich zur Kalibrierung von Infrarotsensoren konzipierte Verfahren [KWL08], da dieses keine sensorspezifischen Informationen verwendet, sondern lediglich auf Einfallswinkelschätzungen zurückgreift. Zentrale Komponente dieses Algorithmus bildet eine trigonometrische Beziehung, die mithilfe des gemessenen Einfallswinkels einen Zusammenhang zwischen der Position eines Sensors und der Position eines Ereignisses herstellt. Durch die Formulierung dieser Beziehung für alle Sensoren und Ereignisse entsteht ein nichtlineares Gleichungssystem und die Lösung diese Gleichungssystems durch das Newton-Verfahren liefert schließlich die gesuchten Sensorpositionen und -orientierungen.

Bei der nähren Untersuchung der beschriebenen Vorgehensweise ergaben sich allerdings mehrere Schwachstellen. Einerseits traten bei der Lösung des Gleichungssystems erhebliche numerische Probleme auf, andererseits sorgten rotierte und gespiegelte Lösungen für unerwünschte Mehrdeutigkeiten (Rotationinvarianz). Darüber hinaus zeigte sich, dass die geometrische Beziehung skalierungsinvariant ist und daher neben den Einfallswinkeln zusätzliche Informationen zur Fixierung der Skalierung benötigt werden. Hauptursache der identifizierten Probleme waren die Tangensfunktionen der geometrischen Beziehung. Diese verursachen einerseits durch ihre Polstellen die numerischen Probleme und lösen andererseits aufgrund ihrer Periodizität von nur 180° die Rotationsinvarianz aus.

Angesichts der Beeinträchtigungen des Kalibrierungprozesses durch die Tangensfunktionen, wurde das Einfallswinkelverfahren unter der Prämisse weiterentwickelt die bekannten Schwachstellen zu reparieren. In einem ersten Schritt wurden dazu die

Tangensfunktionen als Quotient von Sinus- und Kosinusfunktionen ausgedrückt. Die dadurch entstandene Formulierung besitzt keine Polstellen mehr und gestattet gleichzeitig eine anschauliche Interpretation des Kalibrierungsvorgangs. Bei der Kalibrierung wird demnach zunächst der Richtungsvektor, der zum gemessenen Einfallswinkel korrespondiert, vom lokalen Koordinatensystem des Sensors in das globale Koordinatensystem transformiert. Neben dieser Rotation, die von der Sensororientierung bestimmt wird, sorgt eine weitere Drehung um 90° dafür, dass der resultierende Vektor orthogonal zu der aus der Geometrie prädizierten Einfallsrichtung steht. Allerdings ist die Forderung der Orthogonalität nicht eindeutig, deshalb existiert für die Ausrichtung jedes Sensors neben der Lösung, die die wahre Orientierung widerspiegelt, eine äquivalente Lösung, die eine um ±180° rotierte Version des Sensors beschreibt. Durch eine erneute Umformulierung des geometrischen Zusammenhangs entstand schließlich eine Formulierung des Optimierungsproblems, die keine zusätzliche Rotation um 90° beinhaltet, sodass im Optimum Messung und Prädiktion nun in dieselbe Richtung zeigen. Da die Forderung, dass beide Vektoren in dieselbe Richtung zeigen, im Gegensatz zur bis dahin geforderten Orthogonalität, eindeutig ist, hat die weiterentwickelte Version des Einfallswinkelverfahrens nur noch ein globales Optimum.

Die Wirksamkeit der geschilderten Modifikation wurde zudem anhand von Simulationen überprüft. Während bei der ursprünglichen Formulierung des Kalibrierungsproblems mittels Tangensfunktion ca. 15 % der Lösungsversuche divergieren, reduzieren die beiden Weiterentwicklungen die Rate der divergierenden Versuche auf 6 % bzw. 2 %. Darüber hinaus sorgen die vorgenommenen Modifikationen auch für eine Steigerung der Wahrscheinlichkeit von 88,2 % auf 99,5 %, dass nach Abschluss des Newton-Verfahrens das globale Optimum erreicht wird.

Weiterhin wurden drei Möglichkeiten zur Fixierung der Skalierung begutachtet. Den geringsten Positionierungsfehler erzielte dabei die Erweiterung des Gleichungssystems um eine zusätzliche Gleichung, die die Summe der aller paarweisen Abstände der Sensoren auf eine feste Distanz normiert. Allerdings eignet sich diese Methode nur zur relativen Kalibrierung und erfordert eine anschließende Bestimmung der Skalierung durch die in Kapitel 7 oder Kapitel 8 präsentierten Ansätze. Falls hingegen der Abstand zwischen zwei Sensoren vorliegt, ergibt sich bei der Erweiterung des Gleichungssystems durch eine Gleichung, die diesen Abstand definiert, eine etwas höhere Kovergenzrate als bei einer festen Positionierung von zwei Sensoren.

Außerdem konnte durch die Ausführungen in diesem Kapitel belegt werden, dass es sich bei dem aufgrund der geometrischen Anschauung entwickelten Kalibrierungsverfahren um den ML-Schätzer handelt, sofern der Winkelfehler einer von Mises-Verteilung folgt. Obwohl die Modellierung des Winkelfehlers durch eine von Mises-Verteilung nur eine Approximation des Fehlers darstellt, erzielt das erweiterte Einfallsfallwinkelverfahren auch bei einer Nachhallzeit von 0,4 s noch einen Positionierungsfehler der in 90 % der Fälle unter 0,15 m liegt. Die Qualität der Approximation wird auch dadurch bestätigt, dass selbst wenn der Winkelfehler exakt der angenommenen von Mises-Verteilung folgt, sich das Kalibrierungsergebnis nur geringfügig verbessert. Sehr deutliche Auswirkungen auf den Positionierungsfehler hat hingegen der systematische Fehler der Einfallswinkelschätzung bei der Verwendung von nur zwei Mikrofonen pro Array. Dort übersteigt der Positionierungsfehler schon bei einer Nachhallzeit von 0,2 s in 50 % der Untersuchungen 0,20 m.

Am Ende dieses Kapitels wurde außerdem die Geometriekalibrierung dreidimensionaler Sensorkonfigurationen betrachtet. Aufgrund der schon bei der Kalibrierung zweidimensionaler Anordnungen verwendeten Vektornotation war unmittelbar eine Erweiterung auf drei Dimensionen möglich. Außerdem handelt es sich auch hier um einen ML-Schätzer, sofern der Winkelfehler einer VON MISES-FISHER-Verteilung folgt, die die Verallgemeinerung der VON MISES-Verteilung für drei und mehr Dimensionen darstellt. Darüber hinaus steigen die Anforderungen an die jeweiligen Sensorknoten, weil zur Messung von Azimuth und Elevation vorzugsweise jeweils drei Mikrofone verwendet werden sollten. Zusätzlich wächst der Rechenaufwand angesichts einer Steigung der zu kalibrierenden Parameter und einer Erhöhung der dazu notwendigen Iterationen des Newton-Verfahrens. Auch wenn die Rate der divergierenden Lösungsversuche durch Störungen der Einfallswinkel signifikant ansteigt, bestätigen die Untersuchungen dennoch, dass das erweiterte Einfallswinkelverfahren auch zur Kalibrierung dreidimensionaler Sensorkonfigurationen geeignet ist.

6 *Random Sample Consensus*

Die Schwachpunkte der ursprünglichen Version des Einfallswinkelverfahrens konnten mithilfe der im vorangegangenen Kapitel vorgestellten Erweiterungen ausgeräumt werden. Allerdings belegen die durchgeführten Untersuchungen (vgl. Abschnitt 5.6) auch, dass eine zu große Abweichung des Fehlers der Einfallswinkelschätzung von der angenommenen von Mises-Verteilung einerseits die Konvergenz des Newton-Verfahrens beeinträchtigt und andererseits einen erheblichen Anstieg des Kalibrierungsfehlers verursacht. Beim Einsatz von drei Mikrofonen pro Sensorknoten liegt näherungsweise eine von Mises-Verteilung vor. Deshalb erzielt das erweiterte Einfallswinkelverfahren, selbst bei einer Nachhallzeit von 0,4 s, in mehr als 90 % der Fälle, einen mittleren Positionierungsfehler von weniger als 0,15 m. Falls die verwendeten Sensorknoten hingegen über nur zwei Mikrofone verfügen, besitzt der Winkelfehler neben einem Anteil, der durch eine von Mises-Verteilung approximiert werden kann, eine zusätzliche Komponente. Diese Komponente sorgt dafür, dass bei einer Nachhallzeit von 0,4 s nur noch in ca. 10 % der Fälle ein mittlerer Positionierungsfehler kleiner als 0,15 m erreicht wird.

Unabhängig vom Aufbau der Sensorknoten bildet das ebenfalls in Kapitel 5 erläuterte Gleichungssystem, das basierend auf den Einfallswinkelschätzungen einen Zusammenhang zwischen den Sensorpositionen und den Positionen der Ereignisse herstellt, die Grundlage zur Kalibrierung der gesuchten Sensorkonfigurationen. Durch eine LS-Lösung des Gleichungssystems können zudem mehr Beobachtungen in die Bestimmung der Geometrie einfließen als erforderlich sind. Dadurch ermöglicht der LS-Ansatz die Kompensation von Fehlern bei der Einfallswinkelschätzung und wirkt so einem Anstieg des Kalibrierungsfehlers aufgrund eines zunehmenden Winkelfehlers entgegen. Voraussetzung für eine verlässliche Kompensation des auftretenden Fehlers sind jedoch Beobachtungen, die sich durch eine Verteilung aus der Exponentialfamilie beschreiben lassen. Wenn diese Voraussetzung erfüllt ist, kann der ML-Schätzwert, der ein optimales Ergebnis darstellt, durch ein verallgemeinertes LS-Verfahren bestimmt werden [CFY76]. Die beim Einsatz von drei Mikrofonen vorliegende von Mises-Verteilung gehört zur Exponentialfamilie, sodass der entwickelte LS-Ansatz dem ML-Schätzer entspricht. Bei der Verwendung von nur zwei Mikrofonen treten dagegen Beobachtungen mit Winkelfehlern auf, die nicht mehr durch eine von Mises-Verteilung modelliert werden können (vgl. Abb. 4.17). Dabei handelt es sich vorwiegend um Beobachtungen mit größeren Abweichungen zum tatsächlichen Einfallswinkel, die bei einer reinen von Mises-Verteilung nur sehr selten auftreten würden. Daher werden diese Beobachtungen als Ausreißer bezeichnet. Da der verwendete LS-Ansatz jedoch alle Beobachtungen gleichermaßen berücksichtigt und eine Sensorkonfiguration bestimmt, die im Sinne von Gl. (5.7) möglichst geringe Kosten besitzt, beeinträchtigen Ausreißer eine LS-Lösung.

Die Nutzung von Sensorknoten mit drei Mikrofonen vermeidet zwar Ausreißer bei der Einfallswinkelschätzung, die durch die Mikrofonanordnung ausgelöst werden, allerdings sind im Hinblick auf reale Szenarien weitere Auslöser von Ausreißern zu berücksichtigen. Die bislang zur Simulation der Sprachsignale verwendete Spiegel-Quellen-Methode modelliert eine omnidirektionale Ausbreitung der Sprachsignale von den Quellpositionen. Der Mund eines echten Sprechers hat hingegen eine Abstrahlcharakteristik, die dafür sorgt, dass das Signal hauptsächlich in seiner Blickrichtung ausgesandt wird [Wei08]. In realen Szenarien treten daher Situationen auf in denen keine LOS-Komponente im Empfangssignal der Mikrofone vorhanden ist, weil sich die Mikrofone bspw. hinter dem Sprecher befinden (vgl. Abb. 2.1b). Die Einfallswinkelschätzung liefert somit anstatt des wahren Winkels nur die Richtung einer Reflexion. Dementsprechend verursacht auch eine fehlende LOS-Komponente Ausreißer. Außerdem können Hintergrundgeräusche, wie z. B. das Zuschlagen einer Tür, ebenfalls Ausreißer auslösen.

Da bei der Einfallswinkelschätzung in realen Szenarien Ausreißer zu erwarten sind und diese zudem eine präzise Kalibrierung der Geometrie beeinträchtigen, wird in diesem Kapitel ein Konzept vorgestellt, um die Robustheit des bislang entworfenen Kalibrierungsverfahrens gegenüber Ausreißern zu steigern. Die Basis dafür ist der *Random Sample Consensus* (RANSAC) [FB81]. Dieser Algorithmus beschreibt ein Paradigma, das eine zuverlässige Schätzung von Modellparametern auch dann gestattet, wenn die vorliegenden Messdaten eine signifikante Menge von Ausreißern beinhalten.

Während ein konventioneller LS-Ansatz alle vorliegenden Beobachtungen gleichermaßen berücksichtigt und daher anfällig gegenüber Ausreißern ist, verfolgt der RANSAC einen komplementären Ansatz. In einem iterativen Verfahren verwendet er zunächst nur eine Auswahl der Messdaten zur Ermittlung der Modellparameter. Anschließend erfolgt anhand des ermittelten Modells eine Bewertung, welche der übrigen Daten ebenfalls mit dem Modell kompatibel sind. Das Modell, das mit den meisten Datensätzen in Einklang steht, ist hoffentlich frei von Ausreißern, sodass die Nutzung eines LS-Ansatzes zur Berechnung der Modellparameter innerhalb des RANSAC kein Problem darstellt.

Die ursprüngliche Veröffentlichung des RANSAC nennt als Einsatzgebiete die Bildverarbeitung und die automatische Kartographie [FB81]. Allerdings ist die Anwendbarkeit keinesfalls auf die genannten Themenbereiche beschränkt. Sowohl bei der Kalibrierung von Systemen mit mehreren Kameras [BD10] als auch im Bereich der akustischen Geometriekalibrierung [Hen+09; Val+10a], ermöglicht der RANSAC eine gegenüber Messausreißern robuste Schätzung von Modellparametern.

In der Literatur ist die Nutzung des RANSAC in Kombination mit akustischen Kalibrierungsalgorithmen stark an die Gegebenheiten in der Bildverarbeitung angelehnt, weil der RANSAC mit positionsbasierten Kalibrierungsverfahren (vgl. Abschnitt 2.2.3) kombiniert wird. Eine Anwendung in Kombination mit einem winkelgestützten Verfahren ist bisher nicht bekannt, daher stellt Abschnitt 6.1 zunächst das Konzept des RANSAC dar. Danach befasst sich Abschnitt 6.2 mit der Realisierung eines RANSAC für das erweiterte Einfallswinkelverfahren. Zudem werden in den Abschnitten 6.3 bis 6.5 Modifikationen des Konzeptes diskutiert. Das Ziel dieser Modifikationen besteht einerseits darin, den Kalibrierungsfehler zu reduzieren und andererseits in der Berücksichtigung von sehr vielen Messdaten, ohne dass dabei der Berechnungsaufwand zu stark ansteigt. Abschließend soll mithilfe der in Abschnitt 6.6 beschriebenen Simulationen geklärt werden, inwiefern sich die jeweiligen Veränderungen auf das Gesamtergebnis auswirken.

6.1 Das Konzept

Das Konzept des RANSAC [FB81] wurde entwickelt um die zuverlässige Schätzung von Modellparametern selbst dann zu gewährleisten, wenn die vorliegenden Messdaten eine signifikante Menge von Ausreißern beinhalteten. Klassische Ansätze, wie z. B. LS, verwenden zur Bestimmung der gesuchten Modellparameter alle vorhandenen Beobachtungen und ermöglichen dadurch eine gute Kompensation der Störungen, sofern diese durch eine Verteilung aus der Exponentialfamilie beschrieben werden können [CFY76]. Wenn jedoch Beobachtungen mit großen Abweichungen zum wahren Wert sehr viel häufiger auftreten als es bei einer Verteilung aus der Exponentialfamilie der Fall ist, werden diese Beobachtungen als Ausreißer bezeichnet. Allerdings sorgen Ausreißer bei einem LS-Ansatz für einen signifikanten Anstieg des Fehlers, da die ermittelten Modellparameter den quadratischen Abstand minimieren. Das Ziel des RANSAC besteht deshalb darin, aus den vorhandenen Messungen diejenigen zu identifizieren, die keine Ausreißer darstellen und nur diese zur Berechnung des Modells zu verwenden. Um dieses Ziel zu erreichen, werden zunächst ausgewählte Beobachtungen zur Berechnung der Modellparameter herangezogen. Die Güte des gewonnenen Modells wird anschließend anhand der Anzahl der übrigen Beobachtungen, die ebenfalls zu diesem Modell passen, ermittelt. Dieser Vorgang wird so lange wiederholt, bis schließlich ein Modell vorliegt, das mit einer ausreichend großen Anzahl von Messdaten in Einklang steht bzw. eine maximale Anzahl von Iteration erreicht ist. Außerdem kann die Berechnung der Modellparameter in jedem Iterationsschritt durch einen LS-Ansatz erfolgen, da der Auswahl der Beobachtungen die Hypothese zugrunde liegt, dass die ausgewählten Datensätze frei von Ausreißern sind.

Zur Veranschaulichung der Funktionsweise des RANSAC wird im weiteren Verlauf die Schätzung der Parameter einer Geradengleichung betrachtet. Die Parametrisierung der Geraden erfolgt dabei durch den Ordinatenabschnitt sowie die Abszisse. Die Beobachtungen, die zur Bestimmung dieser Parameter notwendig sind, bestehen aus einem x-Wert und dem zugehörigen y-Wert. Alle Beobachtungen bilden die Menge Ω. Die Störungen von einem Teil dieser Beobachtungen lassen sich bspw. durch eine Normalverteilung modellieren, während die restlichen Beobachtungen relativ große Abweichungen ausweisen und dementsprechend Ausreißer darstellen.

Für die Schätzung der gesuchten Parameter der Geradengleichung reicht eine Teilmenge der Beobachtungen (Ω_{sel}), die mindestens $c_{\min} = 2$ Beobachtungen beinhaltet. Die Auswahl dieser Teilmenge erfolgt zufällig. Anschließend werden, basierend auf den in Ω_{sel} enthaltenen Beobachtungen, die Modellparameter (Ordinatenabschnitt und Abszisse) gewonnen. Als nächstes wird für sämtliche in Ω enthaltenen Beobachtungen eine binäre Entscheidung getroffen, ob diese mit dem Modell zu vereinbaren sind oder nicht. Diese Charakterisierung erfolgt z. B. durch die Abweichung zwischen dem vom Modell prädizierten und dem gemessenen y-Wert. Alle Beobachtungen, deren Abweichung einen Schwellwert ζ_{Fit} unterschreiten, werden anschließend als zum Modell passend eingeordnet und bilden den Konsens Ω_{fit}.

Die Anzahl der im Konsens enthaltenen Elemente $|\Omega_{\text{fit}}|$ entscheidet über das weitere Vorgehen. Wenn der Konsens weniger Elemente enthält als der Schwellwert ζ vorsieht, wird zunächst geprüft, ob bereits eine der Abbruchbedingungen erfüllt ist. Falls die maximale Anzahl der Iteration ϕ erreicht ist oder ein Modell mit ausreichend vielen

Beobachtungen in Einklang steht, endet der RANSAC. Anderenfalls beginnt der RANSAC von Neuem, mit der zufälligen Wahl von Beobachtungen Ω_{sel}. Wenn der Konsens hingegen aus mehr Beobachtungen besteht als vom Schwellwert ζ gefordert, dienen die Elemente des Konsenses zur Bestimmung verbesserter Modellparameter. Die Anzahl der dazu verwendeten Beobachtungen ($|\Omega_{\text{fit}}|$) wird anschließend zum Schwellwert ζ für die weiteren Anläufe des RANSAC. Eine grafische Darstellung der erläuterten Schritte zeigt Abb. 6.1.

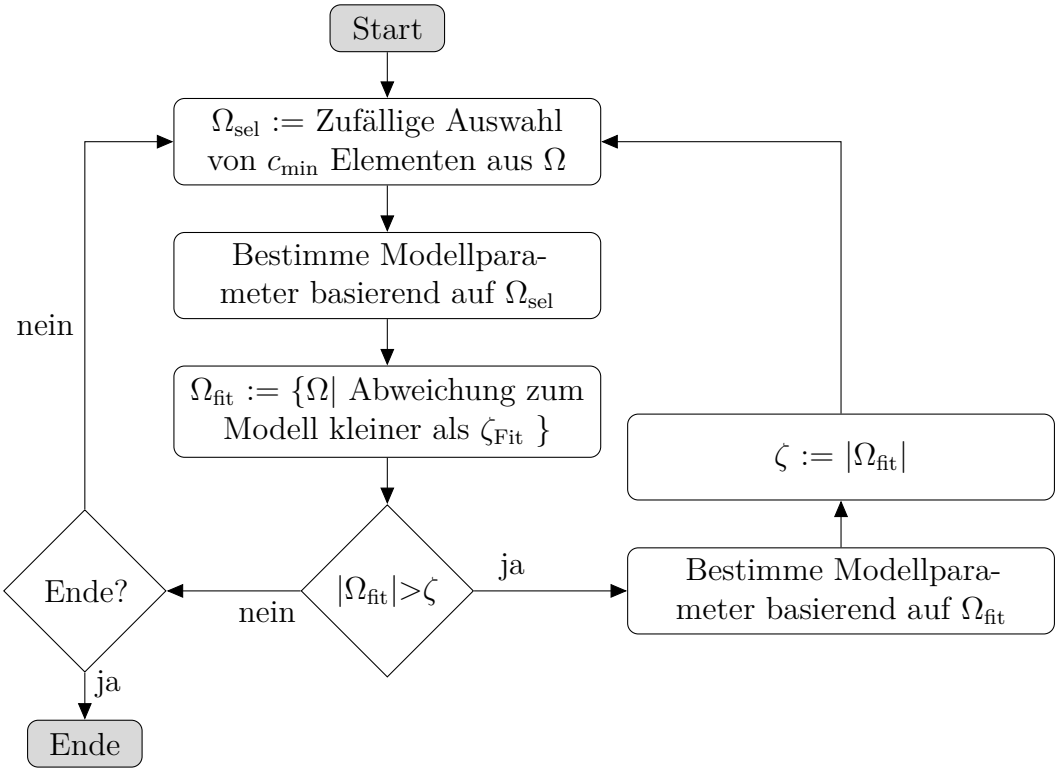


Abbildung 6.1: Schematischer Ablauf des RANSAC-Algorithmus.

Der vorgestellte RANSAC besitzt drei Parameter. Dazu gehören der Schwellwert ζ_{Fit}, zur Klassifikation ob eine Beobachtung mit dem vorliegenden Modell kompatibel ist, der Schwellwert ζ, der die Minimalanzahl von mit dem Modell zu vereinbarenden Messdaten beschreibt sowie die maximale Anzahl der Iterationen ϕ. Die Wahl der beiden Schwellwertparameter ζ_{Fit} und ζ hängt stark von der jeweiligen Problemstellung und dem zur Bewertung der Beobachtungen eingesetzten Kriterium ab. Zur Abschätzung der Anzahl der Iterationen, die erforderlich sind, um ein zuverlässiges Modell zu erhalten, existiert hingegen eine Approximation. Diese erfordert zunächst die Einteilung der Beobachtungen in zwei Klassen. Die erste Klasse repräsentiert Messdaten, die zu einem geeigneten Modell führen, während die zweite Klasse alle anderen Beobachtungen und damit insbesondere die Ausreißer beinhaltet. Die Anzahl der Iterationen hängt gemäß [FB81] von der Wahrscheinlichkeit ν_{OK}, dass eine Beobachtung zu einem tragfähigen Modell führt, ab und sollte

$$\phi = {\nu_{\text{OK}}}^{-(c_{\min})} \tag{6.1}$$

betragen.

6.2 Kombination mit dem Einfallswinkelverfahren

Die Anwendung des RANSAC auf eine Problemstellung besteht aus drei Teilbereichen. Diese umfassen erstens die zufällige Auswahl von Beobachtungen, zweitens die Bestimmung der Modellparameter und drittens für jede Beobachtung eine Entscheidung zu treffen, ob diese mit dem aktuellen Modell zu vereinbaren ist. Die Verwendung des RANSAC zur Steigerung der Robustheit des erweiterten Einfallswinkelverfahrens gegenüber störungsbehafteten Messungen, insbesondere Ausreißern, erfordert indes keine gesonderte Betrachtung des zweiten Teilbereiches, da hier das bereits ausführlich in Kapitel 5 erläuterte erweiterte Einfallswinkelverfahren zum Einsatz kommt. Daher befassen sich die folgenden Abschnitte lediglich mit dem ersten bzw. dritten Teilbereich. Abschnitt 6.2.1 beschreibt dabei zunächst eine Strategie zur Auswahl von Beobachtungen, bevor in Abschnitt 6.2.2 ein Ansatz für die Bewertung der Modellparameter bzw. des Kalibrierungsergebnisses präsentiert wird.

6.2.1 Auswahl der Beobachtungen

Der erste Schritt des RANSAC sieht die Auswahl einer zufälligen Teilmenge von Beobachtungen vor, die anschließend die Schätzung der Modellparameter gestattet. Für eine erfolgreiche Durchführung des Einfallswinkelverfahrens ergeben sich zwei Anforderungen an eine zu entwickelnde Strategie zur Auswahl der Beobachtungen. Einerseits muss in Abhängigkeit der Anzahl der Sensoren eine Mindestanzahl von Beobachtungen vorliegen (vgl. Gl. (5.5)). Andererseits müssen die zu den Einfallswinkeln korrespondierenden Ereignispositionen eine ausreichende räumliche Diversität besitzen (siehe Abschnitt 5.1). Allerdings liegen im Rahmen der Problemstellung ausschließlich Einfallswinkelschätzungen vor, die z. B. aus dem Sprachsignal einer sich durch den Raum bewegenden Person gewonnen wurden. Durch die unbekannte Lage der Quellpositionen innerhalb der Trajektorie, die sowohl Abschnitte des Stillstandes als auch Phasen der Bewegung beinhalten kann, besteht die Herausforderung darin, die Einfallswinkel von möglichst weit über den Raum verteilten Positionen auszuwählen.

Obwohl die Positionen der akustischen Ereignisse unbekannt sind, ermöglichen die Einfallswinkel dennoch eine Einschätzung der Entfernungen zwischen den zugehörigen Positionen. Dazu werden zunächst die Einfallswinkel aller Sensoren zum Zeitpunkt d zu dem Vektor

$$\boldsymbol{\varphi}_d = \begin{bmatrix} \varphi_{1,d} & \dots & \varphi_{I,d} \end{bmatrix}^{\mathrm{T}} \tag{6.2}$$

zusammengefasst. Alle Beobachtungen $\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_D$ beschreiben somit Punkte in einem I-dimensionalen Vektorraum. Je größer der Abstand der Vektoren $\|\boldsymbol{\varphi}_d - \boldsymbol{\varphi}_v\|_2$, desto größer ist auch der Abstand zwischen den zugehörigen Positionen der Ereignisse $\boldsymbol{e}_d$ bzw. $\boldsymbol{e}_v$. Insgesamt sind daher zur Bestimmung einer geometrischen Anordnung $c_{\min}$ Punkte aus einem Vektorraum auszuwählen, die möglichst große paarweise Abstände aufweisen.

Eine vergleichbare Problemstellung tritt beim Clustering von Daten auf. Zur initialen Selektion von Clusterzentren mit großen paarweise Abständen kommt dort der K-means++-Algorithmus zum Einsatz [AV07]. Aufgrund der weitgehenden Übereinstimmung zwischen beiden Problemen, soll K-means++ auch zur Auswahl von Beobachtungen für das Einfallswinkelverfahren dienen.

Bei der Wahl des ersten Clusterzentrums haben zunächst alle Punkte die gleiche Wahrscheinlichkeit ausgewählt zu werden. Für die Auswahl der folgenden Punkte ist die Wahrscheinlichkeit eines Punktes ausgewählt zu werden jedoch proportional zum Quadrat der Entfernung zu den bisher gezogenen Zentren. Infolgedessen besitzen die ausgewählten Punkte eine große räumliche Diversität, weil der Algorithmus Punkte mit geringen Abständen zu den bisherigen Clusterzentren seltener als Punkte mit größeren Abständen selektiert. Die Anwendung dieser Strategie zur Auswahl von Beobachtungen für das erweiterte Einfallswinkelverfahren ist in Algo. 1 dargestellt. Die Distanz zwischen zwei Beobachtungen $\boldsymbol{\varphi}_d$ und $\boldsymbol{\varphi}_v$ errechnet sich dabei aus dem Betragsquadrat der Differenz der beiden Vektoren. Es gilt jedoch zu beachten, dass es sich bei den vorliegenden Größen um Winkel handelt und somit die 2π-periodische Distanz $\operatorname{atan2}(\cos(\varphi_{i,d} - \varphi_{i,v}); \sin(\varphi_{i,d} - \varphi_{i,v}))$ der Elemente zu verwenden ist. Die Funktion $\operatorname{atan2}(\boldsymbol{y}; \boldsymbol{x})$ bezeichnet dabei die verallgemeinerte Arkustangensfunktion, sodass die Inversion des Tangens nicht auf eine Halbebene beschränkt ist, sondern für alle vier Quadranten definiert ist.

Algorithmus 1 K-means++-Algorithmus zur Auswahl von Beobachtungen.

```
1: procedure Wähle Beobachtungen(c_min, [φ_1, ..., φ_D])
2:     Wähle erste Beobachtung gleichverteilt aus [φ_1, ..., φ_D].
3:     for index ← 2, c_min do
4:         for d ← 1, D do
5:             Berechne minimale Distanz von φ_d zu bisher ausgewählten Beobachtungen
6:         end for
7:         Wähle weitere Beobachtung, proportional zur minimalen Distanz
8:     end for
9: end procedure
```

Eine Bewertung der vorgeschlagenen Strategie, die Beobachtungen durch den K-means++-Algorithmus zu bestimmen, erfordert außerdem ein geeignetes Referenzverfahren. Als Referenz dient eine entlang des zeitlichen Verlaufes der Trajektorie gleichmäßige Auswahl von Einfallswinkeln, die sowohl die räumliche Lage als auch die ggf. vorhandenen Phasen des Stillstandes des Sprechers unberücksichtigt lässt. Die Grundlage für den Vergleich zwischen der entwickelten Vorgehensweise und dem Referenzverfahren bildet das bereits aus Abschnitt 5.4 bekannte Szenario mit vier in Wandnähe platzierten Sensoren. Die Ereignispositionen sind jetzt jedoch nicht mehr durch einzelne Punkte gegeben, sondern stammen von einer zusammenhängenden Trajektorie. Die Generierung dieser erfolgt mithilfe eines *Random-Walk*-Modells [Pea05], das für eine Bewegung des Sprechers mit konstanter Geschwindigkeit (1,5 m/s) und zufälliger Richtungsänderung sorgt. Damit die Trajektorie neben den Phasen der Bewegung auch aus Abschnitten des Stillstandes besteht, kommt ein *Hidden* Markov *Model* (HMM) zum Einsatz. Das verwendete *Random-Walk*-Modell hat jedoch keine Informationen darüber, wo sich die Wände des Raumes befinden. Um zu verhindern, dass eine Trajektorie entsteht, die die durch die Wände vorgegebenen Grenzen verlässt, wird beim Erreichen einer Wand ein Teil der bereits erzeugten Trajektorie verworfen und ein alternativer Verlauf generiert.

Die im Weiteren betrachteten Ergebnisse der durchgeführten Untersuchungen basieren auf 100 verschiedenen Sensorkonfigurationen. Für jede Sensorkonfiguration wurden wiederum 100 zufällige Trajektorien, mit einer Länge von jeweils 2 min erzeugt, sodass insgesamt 10 000 Szenarien vorhanden sind. Um die Auswirkungen von Phasen des Stillstandes hervorzuheben, wird einerseits eine Trajektorie mit kontinuierlicher Bewegung des Sprechers und andererseits eine, die auch Standphasen enthält, berücksichtigt. Die Zustandsübergangswahrscheinlichkeiten des HMM sind dabei so gewählt, dass sich ca. drei Standphasen mit einer durchschnittlichen Länge von 20 s ergeben.

Zur Veranschaulichung der Wirkungsweise des K-means++-Algorithmus zur Selektion der Einfallswinkel dient Abb. 6.2. Sie stellt ein beispielhaftes Szenario dar. Bei der als Referenz dienenden gleichverteilten Auswahl konzentrieren sich die ausgesuchten Ereignisposition in der oberen linken Ecke des Raumes, weil sich der Sprecher häufig in diesem Bereich aufhält. Im Gegensatz dazu, wählt der K-means++-Algorithmus auch Positionen in weniger oft besuchten Bereichen des Raumes aus und erreicht deshalb eine deutlich bessere räumliche Diversität.

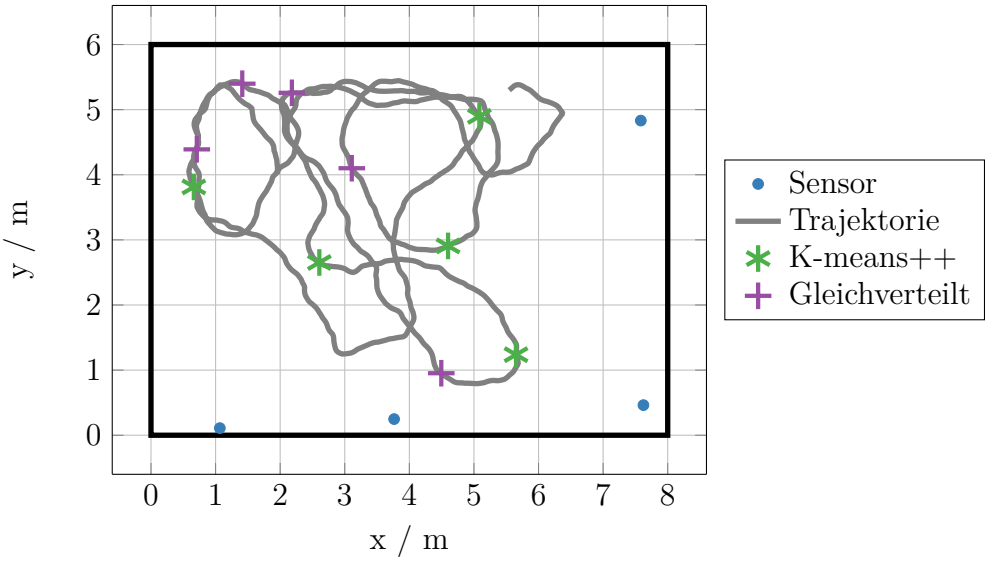
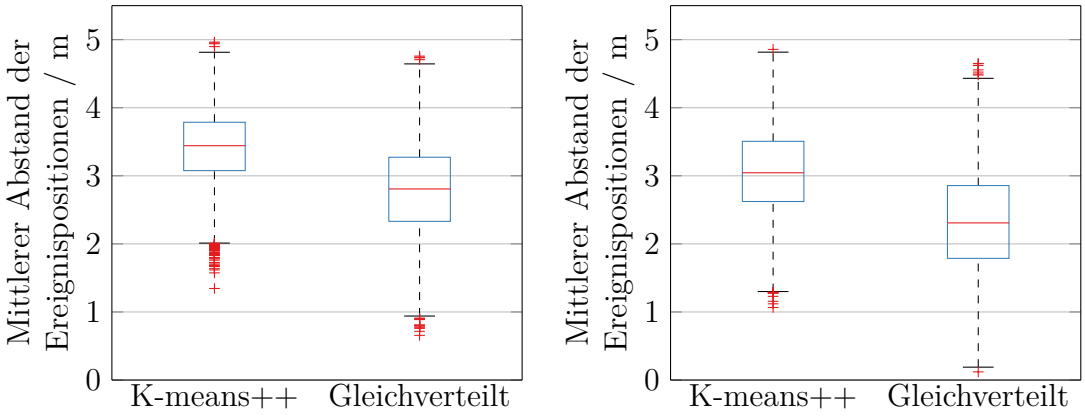


Abbildung 6.2: Beispielhaftes Szenario mit gleichverteilter und K-means++-gestützter Auswahl der Einfallswinkelschätzungen.

Die Quantifizierung der räumlichen Diversität der zu den ausgewählten Einfallswinkeln gehörenden Positionen erfolgt mithilfe des mittleren Abstands der Positionen. Abb. 6.3 stellt die Untersuchungsergebnisse für Trajektorien ohne Standphasen (a) und Trajektorien mit Stand- und Bewegungsphasen (b), sowohl bei der Nutzung der vorgeschlagenen Selektion mit dem K-means++-Algorithmus als auch bei gleichverteilter Auswahl gegenüber. Zur Visualisierung der Ergebnisse dienen sogenannte *Boxplots*. Eine blaue Box kennzeichnet dabei den Bereich, in dem die mittleren 50 % der Daten liegen. Die darin befindliche rote Linie markiert den Median. Die Antennen wiederum entsprechen dem 1,5 fachen der Größe der Box und beschreiben somit das 2,7 fache der Standardabweichung. Die roten Kreuze stellen schließlich die Werte außerhalb des 2,7 fachen der Standardabweichung dar.

Aus diesen Ergebnissen geht eindeutig hervor, dass der K-means++-Algorithmus, unabhängig von der Beschaffenheit der Trajektorie, Einfallswinkel mit größeren mittleren Abständen zwischen den Ereignispositionen liefert als das Referenzverfahren. Falls die Trajektorien keine Standphasen beinhalten, führt auch die gleichverteilte Auswahl zu Einfallswinkeln von Positionen mit ausreichender räumlicher Diversität. Allerdings fällt der Abstand beim Einsatz von K-means++ deutlich größer aus.

Sofern die Trajektorien zusätzlich aus Standphasen bestehen, liefert eine gleichverteilte Auswahl z. T. Einfallswinkel, deren zugehörige Positionen nur eine unzureichende Diversität besitzen, während der K-means++-Algorithmus auch hier einen mittleren Abstand von mehr als 1,00 m erzielt. Insgesamt bestätigen die Simulationen daher, dass der Einsatz von K-means++ die Diversität gegenüber einer gleichverteilten Auswahl steigert und somit auch bei Trajektorien mit längeren Standphasen oder einer ungünstigen Lage innerhalb des Raumes geeignete Einfallswinkel zur Durchführung des erweiterten Einfallswinkelverfahrens selektiert.



(a) Trajektorien ohne Stillstand (b) Trajektorien mit Stillstand

Abbildung 6.3: Vergleich der räumlichen Diversität bei gleichverteilter bzw. K-means++-basierter Auswahl der Beobachtungen.

6.2.2 Bewertung des Modells

Der entscheidende und zugleich herausfordernste Schritt des RANSAC ist die Bewertung der Qualität eines Modells. Der Bewertungsprozess erfordert für jede Beobachtung eine individuelle Entscheidung, ob diese mit dem vorliegenden Modell vereinbart werden kann oder nicht. Gleichzeitig gilt es zu beachten, dass das Bewertungskriterium innerhalb jeder Iteration für alle Beobachtungen zu evaluieren ist und somit, neben der eigentlichen Schätzung der Modellparameter, den Teilschritt darstellt, der die Rechenkomplexität des RANSAC beeinflusst.

Nach Abschluss des erweiterten Einfallswinkelverfahrens mit den ausgewählten Beobachtungen liegt eine Schätzung der Positionen und Orientierungen der Sensoren vor. Die Kenntnis dieser ermöglicht anschließend die Lokalisierung von allen Ereignissen, die z. B. mit dem in [BAS95] präsentierten Verfahren erfolgen kann. Dazu werden die Einfallswinkel als Geraden, ausgehend von den Sensorpositionen, dargestellt und deren

Schnittpunkte ermittelt. Der Einfallswinkel $\varphi_{i,d}$ beschreibt somit in Kombination mit der Sensorposition $\boldsymbol{s}_i$ und der Sensororientierung θ_i die Gerade

$$\boldsymbol{\varrho}_{i,d}(\lambda_{i,d}) = \boldsymbol{s}_i + \lambda_{i,d} \begin{bmatrix} \cos(\theta_i + \varphi_{i,d}) \\ \sin(\theta_i + \varphi_{i,d}) \end{bmatrix}, \tag{6.3}$$

die die möglichen Positionen des Ereignisses $\boldsymbol{e}_d$ charakterisiert. Die Positionsschätzung für das d-te Ereignis basierend auf den Sensoren i und j setzt zunächst die Lösung des Gleichungssystems

$$\begin{bmatrix} \lambda_{i,d} \\ \lambda_{j,d} \end{bmatrix} = \begin{bmatrix} \cos(\theta_i + \varphi_{i,d}) & -\cos(\theta_i + \varphi_{i,d}) \\ \sin(\theta_i + \varphi_{i,d}) & -\sin(\theta_i + \varphi_{i,d}) \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{s}_i - \boldsymbol{s}_j \end{bmatrix} \tag{6.4}$$

voraus. Der Schnittpunkt $\boldsymbol{q}_{(i,j),d} = \boldsymbol{q}_{(j,i),d}$ ergibt sich schließlich durch Einsetzen der Lösungen $\lambda_{i,d}$ bzw. $\lambda_{j,d}$ in Gl. (6.3). Der Schätzwert für die Ereignisposition $\tilde{\boldsymbol{q}}_d$ entsteht anschließend durch die gewichtete Kombination der Schnittpunkte aller Sensorpaare:

$$\tilde{\boldsymbol{q}}_d = \sum_i \sum_{j \setminus i} w_{(i,j),d}\, \boldsymbol{q}_{(i,j),d}. \tag{6.5}$$

Die dazu notwendigen Gewichte werden in Anlehnung an [BAS95] bestimmt. Den Ausgangspunkt bildet die Hypothese, dass die Position des Ereignisses zum Zeitpunkt d dem Schnittpunkt $\boldsymbol{q}_{(i,j),d}$ entspricht. Aufgrund dieser Hypothese, lässt sich der Einfallswinkel des Sensors $\boldsymbol{s}_{i'}$ durch $\arg(\boldsymbol{q}_{(i,j),d} - \boldsymbol{s}_{i'})$ berechnen. Die Qualität des Schnittpunktes $\boldsymbol{q}_{(i,j),d}$ ergibt sich anschließend aus dem Produkt der Wahrscheinlichkeiten, dass die jeweiligen Sensoren den berechneten Einfallswinkel beobachten, sofern der tatsächliche Einfallswinkel dem gemessenen Winkel $\varphi_{i',d}$ entspricht. Eine Normierung des beschriebenen Ausdrucks liefert schließlich

$$w_{(i,j),d} = \frac{\prod\limits_{i'} \mathcal{M}\Big(\arg\Big(\boldsymbol{q}_{(i,j),d} - \boldsymbol{s}_{i'}\Big); \varphi_{i',d}, \sigma_{i'}\Big)}{\sum\limits_{i'} \sum\limits_{j' \setminus i'} \prod\limits_{i''} \mathcal{M}\Big(\arg\Big(\boldsymbol{q}_{(i',j'),d} - \boldsymbol{s}_{i''}\Big); \varphi_{i'',d}, \sigma_{i''}\Big)}, \tag{6.6}$$

wobei $\mathcal{M}(\cdot; \varphi_{i'',d}, \sigma_{i''})$ die Verteilungsdichte der VON MISES-Verteilung mit dem Mittelwert $\varphi_{i'',d}$ und der Standardabweichung $\sigma_{i''}$ bezeichnet. Die Verwendung der VON MISES-Verteilung bietet zudem den Vorteil, dass im Grenzfall für eine gegen Null strebende Konzentration eine gleichmäßige Berücksichtigung aller Schnittpunkte vorliegt.

Sofern die exakten Sensorpositionen und -ausrichtungen vorliegen und die Einfallswinkel ebenfalls keine Störungen aufweisen, fallen alle Schnittpunkte $\boldsymbol{q}_{(i,j),d}$ mit der tatsächlichen Ereignisposition $\boldsymbol{e}_d$ zusammen. Ausgelöst durch die Schätzfehler bei der Einfallswinkelbestimmung, treten jedoch verstreute Schnittpunkte auf (vgl. Abb. 6.4).

Da im Rahmen dieser Arbeit auch die Positionen und Ausrichtungen der Sensoren auf den mit Schätzfehlern behafteten Einfallswinkeln basieren, vergrößert der Kalibrierungsfehler die auftretende Streuung. Letztendlich führen Einfallswinkel mit größerem Fehler sowohl direkt als auch indirekt über die Geometriekalibrierung zu einer stärkeren räumlichen Ausbreitung der Schnittpunkte $\boldsymbol{q}_{(i,j),d}$. Die Zunahme der Streuung führt auch dazu, dass die Distanzen zwischen den Schnittpunkten $\boldsymbol{q}_{(i,j),d}$ und dem Schätzwert für die Ereignisposition $\tilde{\boldsymbol{q}}_d$ ansteigen. Zur quantitativen Beschreibung, wie gut eine Beobachtung durch das vorliegende Modell erklärt werden kann, dient deshalb

$$v_d = \frac{1}{I^2 - I} \sum_i \sum_{j \setminus i} \left\| \boldsymbol{q}_{(i,j),d} - \tilde{\boldsymbol{q}}_d \right\|_2. \tag{6.7}$$

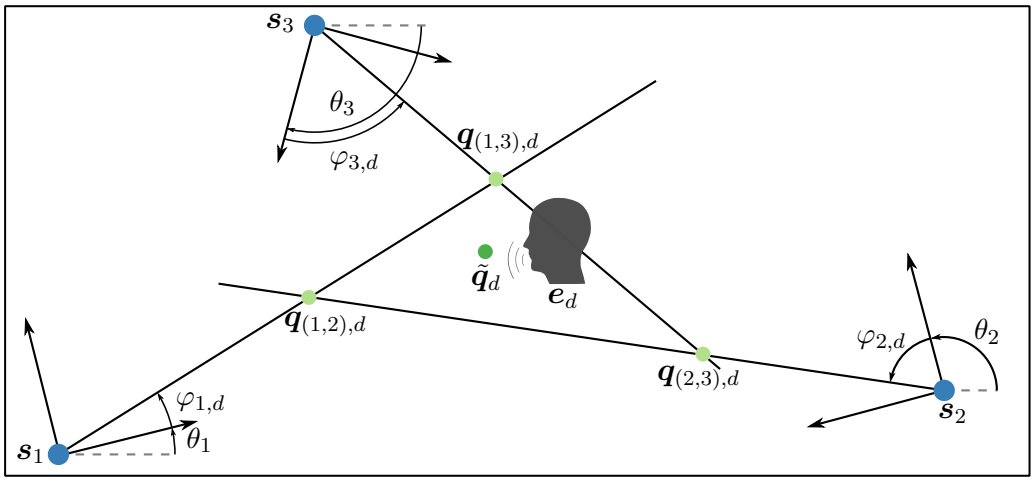


Abbildung 6.4: Lokalisierung eines Ereignisses durch Einfallswinkel.

Überschreitet der durchschnittliche Abstand v_d den Schwellwert ζ_{Fit} (vgl. Abb. 6.1), wird die Beobachtung durch den RANSAC als nicht zum Modell passend klassifiziert. Grundsätzlich sollte der Schwellwert ζ_{Fit} möglichst klein gewählt werden, um ausschließlich Beobachtungen mit geringen Fehlern als passend einzustufen. Andererseits führt eine zu kleine Wahl dazu, dass entweder nur wenige Beobachtungen zum Modell passen und somit eine LS-Lösung keinen ausreichenden Ausgleich der Störungen gestattet oder aber gar keine Lösung erzielt wird, da niemals genügend Beobachtungen den geforderten Schwellwert unterschreiten.

6.3 Modifikationen des Paradigmas

Die zuvor erläuterte Herangehensweise des RANSAC bildet eine einfache und dennoch leistungsstarke Möglichkeit, die Schätzung der geometrischen Anordnung robuster gegenüber Ausreißern zu gestalten. Die Analyseergebnisse aus Kapitel 4 weisen jedoch, insbesondere bei der Nutzung von drei Mikrofonen, hauptsächlich auf moderate Winkelfehler, die sich durch eine VON MISES-Verteilung mit geringer Standardabweichung approximieren lassen, anstatt auf deutliche Ausreißer hin. Grundsätzlich findet diese Problematik bereits im letzten Schritt des RANSAC-Algorithmus Berücksichtigung, weil die abschließende Berechnung der Modellparameter basierend auf der Konsensmenge Ω_{fit} vorgesehen ist und daher einen Ausgleich kleinerer Störungen mittels LS erfolgt. Trotzdem führt die Tatsache, dass alle Beobachtungen eine gewisse Störung aufweisen, zu Problemen für den RANSAC. Dieser findet dadurch oftmals nur ein Modell, das mit einem kleinen Anteil von Messungen zu vereinbaren ist. Diese modellkonformen Messungen ähneln den initial ausgewählten Beobachtungen, umfassen aber nicht alle Beobachtungen die keine Ausreißer darstellen. Als Konsequenz liefert der RANSAC in diesen Situationen keine verlässlichen Modellparameter oder es sind sehr viele Iterationen notwendig, um dennoch geeignete Modellparameter zu erlangen [CMO04].

Eine Möglichkeit diesen Nachteil auszugleichen bietet der *Locally Optimized Random Sample Consensus* (LORANSAC) [CMK03; LMC12]. Dieser entspricht in weiten Teilen

dem konventionellen RANSAC und nutzt genau wie dieser die minimal erforderliche Anzahl an Beobachtungen zur initialen Schätzung des Modells und führt anschließend die Bestimmung der Konsensmenge Ω_{fit} durch. Sofern diese größer als ein vorgegebener Schwellwert ζ ist, erfolgt die Ermittlung der Modellparameter basierend auf der Konsensmenge Ω_{fit}. Im Gegensatz zum konventionellen RANSAC (vgl. Abb. 6.1), der an dieser Stelle erneut beginnt, besitzt der LORANSAC weitere Schritte. Er nutzt das aus der Regression entstehende Modell zur Berechnung einer weiteren Konsensmenge. Die Regression führt zum Ausgleich kleiner Störungen, sodass der zweite Konsens tendenziell mehr Elemente als der Erste besitzt. Übersteigt die Kardinalität des zweiten Konsenses die des Ersten, dient der Zweite zur Berechnung der aktuell bestmöglichen Modellparameter und der LORANSAC beginnt erneut. Letztendlich handelt es sich beim LORANSAC um eine Zwei-Schritt-Strategie, die robuster gegenüber kleineren Störungen der Beobachtungen ist. Diese Strategie reduziert einerseits die Anzahl der Iterationen gegenüber dem konventionellen RANSAC. Andererseits steigt die Rechenkomplexität, weil die rechenintensiven Teilprobleme der Bestimmung der Modellparameter und die Bewertung der selbigen nun zweimal pro Iteration erforderlich sind.

Die Argumentationskette, die zur Entwicklung des zweistufigen LORANSAC führt, lässt sich weiter fortsetzen. Sofern die ersten beiden Schritte des LORANSAC eine sukzessive Vergrößerung der Konsensmenge ermöglichen, besteht die Chance, dass

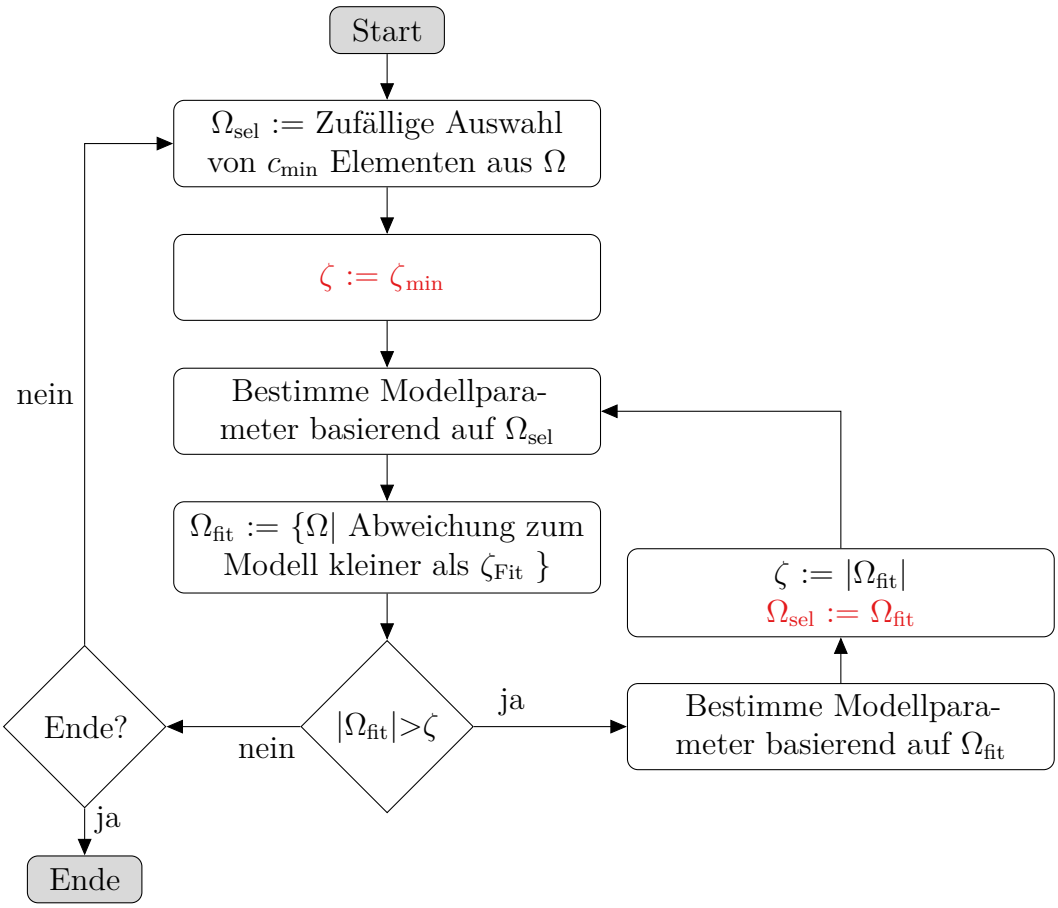


Abbildung 6.5: Schematischer Ablauf des modifizierten LORANSAC-Algorithmus.

ein weiterer Konsens und eine anschließende Schätzung des Modells die gesuchten Parameter weiter verbessert. Im Rahmen dieser Arbeit wird deshalb der LORANSAC so modifiziert, dass ein RANSAC mit beliebig vielen lokalen Iteration entsteht. Dazu wird der Schwellwert ζ im ersten Schritt zu $\zeta_{\min}$ gewählt. In den darauf folgenden Schritten ist er jeweils durch die Kardinalität des Konsenses aus dem vorherigen Schritt gegeben ($|\Omega_{\text{fit}}|$). Insgesamt entsteht der in Abb. 6.5 schematisch dargestellte Algorithmus, der iterativ einen möglichst großen Konsens bestimmt.

Ein Vergleich der vorgeschlagenen Strategie mit dem ursprünglichen RANSAC aus Abb. 6.1 zeigt nur geringfügige Unterschiede des Ablaufes. Die zusätzlichen Schritte, die sich durch die Modifikationen ergeben, sind in Abb. 6.5 zur besseren Veranschaulichung in rot hervorgehoben. Im weiteren Verlauf dieser Arbeit wird stets die angepasste Variante des RANSAC eingesetzt.

6.4 Partitionierung

Der Erfolg des RANSAC hängt neben der geeigneten Wahl der Schwellwerte maßgeblich von der Anzahl der durchgeführten Iterationen ab. Hauptgrund dafür ist, dass es sich beim RANSAC um ein stochastisches Verfahren handelt und somit die zufällige Auswahl der Beobachtungen das spätere Kalibrierungsergebnis beeinflusst. Damit der RANSAC mit ausreichender Sicherheit verlässliche Modellparameter liefert, gilt es, die Abschätzung der Iterationen gemäß Gl. (6.1), zu beachten. Diese Näherung erfordert jedoch, wie bereits erwähnt, eine Einteilung der Beobachtungen in zwei Klassen. Die erste Klasse umfasst dabei alle Messwerte die zu geeigneten Modellparametern führen, wohingegen die zweite Klasse alle anderen Datensätze enthält.

Eine Einteilung der vorliegenden Winkelschätzungen in eine dieser beiden Klassen ist hingegen nicht direkt möglich. Ausreißer lassen sich eindeutig in die Kategorie der Beobachtungen aus denen kein tragfähiges Modell resultiert einordnen. Alle Messungen, die nur eine geringe Störungen aufweisen, gehören jedoch nicht automatisch zur komplementären Kategorie, weil die räumliche Lage der Ereignisposition ebenfalls Einfluss auf das Kalibrierungsergebnis hat. Dadurch können Einfallswinkel, trotz einer sehr geringen Störung, unzureichende Modellparameter zur Folge haben.

Eine ungefähre Abschätzung der Anzahl der notwendigen Iterationen gestattet die idealisierte Betrachtung des auch bisher genutzten Szenarios mit vier Sensoren. Unter der Annahme, dass 50 % der zur Verfügung stehenden Messungen zu einem belastbaren Modell führen, beträgt die empfohlene Anzahl mehr als 1000 Iterationen. Sie fällt u. a. auch deshalb so groß aus, weil alle Winkelschätzungen eine gewisse Störung aufweisen und es somit nicht ausreicht, $c_{\min}$ entsprechend Gl. (5.5) zu wählen, um eine tragfähige Geometrie zu gewinnen. Unter Berücksichtigung des Zeitaufwandes für die Bestimmung eines einzelnen Modells und seiner Bewertung (vgl. Abschnitt 5.3 bzw. Abschnitt 6.2.2), ergibt sich daher ein enormer Berechnungsaufwand.

Die in Abschnitt 6.2.1 vorgeschlagene Methode zur Selektion der Beobachtungen unter Verwendung des K-means++-Algorithmus, steigert die Chance, dass aus den ausgewählten Einfallswinkeln ein geeignetes Modell resultiert. Meist reichen wenige Iterationen des RANSAC, um eine grobe Geometrie zu erhalten. Da diese Geometrie bereits einigermaßen der gesuchten Geometrie entspricht, entsteht auch eine vergleichs-

weise große Konsensmenge. Eine Verfeinerung dieses Ergebnisses erfordert dagegen zahlreiche Anläufe, weil ein Modell gefunden werden muss, welches noch mehr Messungen erklären kann. Bei jedem dieser Versuche können zahlreiche lokale Iteration des RANSAC auftreten, sodass die Suche nach einer noch besseren Sensorkonfiguration mit beträchtlichem Aufwand verbunden ist.

Zur Begrenzung bzw. Reduktion dieses Aufwandes soll deshalb ein mehrstufiges Konzept erprobt werden. Es sieht zunächst die Beschränkung der Iterationen des RANSAC auf eine geringe Anzahl vor. Diese soll ausreichen, um ein grobes Modell zu finden, jedoch vermeiden, dass sehr viele Versuche durchgeführt werden, die nicht zu einer Verbesserung dieses ungefähren Modells führen. Eine alleinige Reduktion der Iterationen ist hingegen kontraproduktiv, da die Wahrscheinlichkeit zuverlässige Modellparameter zu erlangen sinkt. Zur Kompensation sollen mehrere dieser deutlich schnelleren, aber dadurch auch ungenauere Modellparameter liefernden Ausführungen des RANSAC, bspw. durch die Berechnung des Mittelwertes, zu einer gemeinsamen Lösung zusammengefasst werden.

Darüber hinaus eröffnet das vorgeschlagene Konzept, dass im Folgenden als *Partitioned Random Sample Consensus* (PRANSAC) bezeichnet wird, auch eine Möglichkeit zur Handhabung von sehr vielen Beobachtungen. Ein Konzept zur Verarbeitung von vielen Beobachtungen ist bspw. dann notwendig, wenn während der gesamten Kalibrierung fortlaufend eine Einfallswinkelschätzung durchgeführt wird und dementsprechend die Anzahl der vorliegenden Daten kontinuierliche anwächst. Hauptsächlich die iterative Vergrößerung des Konsenses durch die in Abschnitt 6.3 entwickelte Strategie der lokalen Iterationen führt bei steigender Anzahl von verfügbaren Einfallswinkeln zu einem sehr deutlichen Anstieg der Komplexität. Dieser entsteht vornehmlich, weil die iterative Vergrößerung in vielen Schritten erfolgt, die jeweils die Durchführung des erweiterten Einfallswinkelverfahrens mit einer wachsenden Anzahl von Gleichungen voraussetzen. Andererseits lässt sich diesem Trend mit mehreren kurzen RANSAC und der anschließenden Kombination der Ergebnisse entgegenwirken. Sofern die Ausführungen dieser RANSAC auf unterschiedlichen Teilmengen der Beobachtungen erfolgen, kann einerseits die Komplexität reduziert werden und andererseits können durch die Kombination der Teilergebnisse trotzdem alle Beobachtungen in das Gesamtergebnis einfließen.

Zur Veranschaulichung des erläuterten Konzeptes dient das in Abb. 6.6 dargestellte System. Die erste Stufe hat die Aufgabe, die Beobachtungen an die nachfolgenden RANSAC zu verteilen. Wenn sehr viele Beobachtungen vorliegen wird, erneut der K-means++-Algorithmus zur Verteilung verwendet, um sicherzustellen, dass alle Teilemengen eine ausreichende räumliche Diversität besitzen. Falls lediglich eine begrenzte Datenmenge vorhanden ist, arbeiten alle RANSAC auf denselben Daten. Abschließend erfolgt eine Kombination der Teilergebnisse zu einer gemeinsamen Geometrie.

Abbildung 6.6: Ablauf des partitionierten RANSAC (PRANSAC) zur Verarbeitung einer großen Anzahl von Beobachtungen bzw. zur Reduktion der Ausführungszeit.

6.5 Geometriefusion

Voraussetzung für die Realisierung des zuvor erläuterten partitionierten RANSAC (PRANSAC), ist eine Methode zur Kombination der Teilergebnisse. Aufgrund der Definition des Koordinatensystems durch einen beliebigen Sensor, müssen die Ergebnisse der verschiedenen RANSAC nicht notwendigerweise im selben Koordinatensystem liegen. Um dennoch eine gemeinsame Geometrie ermitteln zu können, soll auf Methoden aus der statistischen *Shape*-Analyse zurückgegriffen werden [DM98]. Die statistische *Shape*-Analyse findet z. B. in der Biologie und Geodäsie Anwendung und gestattet die Untersuchung von zufälligen geometrischen Formen/Objekten, unabhängig von der Translation, Rotation und Skalierung. Grundlage dafür sind Transformationen der geometrischen Informationen vom Definitionsbereich (engl. *configurationspace*) in einen Bildbereich (engl. *shape domain*), der eine Beschreibung der Formen (engl. *shapes*) unabhängig von den genannten Größen erlaubt.

Um das eigentliche Ziel zu erreichen, eine mittlere geometrische Anordnung aus mehreren Lösungen zu bestimmen, werden Techniken der Prokrustes-Analyse[1] (engl. Procrustes *analysis*) verwendet. Die Prokrustes-Analyse ist das Teilgebiet der statistischen *Shape*-Analyse, das sich mit der Quantifizierung des Abstandes geometrischer Formen und der Berechnung von mittleren Formen sowie deren Variabilität beschäftigt [DM98]. Da der Fokus dieser Arbeit auf zweidimensionalen Anordnungen liegt, werden nur Methoden der planaren Prokrustes-Analyse betrachtet. Auf die Berücksichtigung mehrdimensionaler Anordnungen wird verzichtet, weil diese kompliziertere Herangehensweisen erfordern.

Zur Beschreibung einer geometrischen Form kommen bei der statistischen *Shape*-Analyse sogenannte Landmarken (engl. *landmarks*), die z. B. durch die Ecken des betrachteten Polygons gegeben sind, zum Einsatz. Durch die Beschränkung auf planare Formen lassen sich die Landmarken durch komplexe Zahlen $\ell_k = x_k^{\mathrm{L}} + \mathrm{j}y_k^{\mathrm{L}}$ darstellen. Eine geometrische Form ist damit durch die Zusammenfassung der Landmarken für $k = 1, \ldots, \mathcal{K}$ zum Vektor

$$\boldsymbol{\ell} = \begin{bmatrix} \ell_1 & \ldots & \ell_{\mathcal{K}} \end{bmatrix} \tag{6.8}$$

gegeben.

Die Berechnung einer mittleren geometrischen Anordnung erfordert zunächst ein Maß zur Quantifizierung des Abstandes von zwei Formen $\boldsymbol{\ell}_1$ und $\boldsymbol{\ell}_2$. Zur Berechnung der Ähnlichkeit unabhängig von Translation, Rotation und Skalierung der Formen, müssen die Formen zunächst möglichst deckungsgleich aufeinander abgebildet werden. Die dazu notwendige Koordinatentransformation muss eine Translation, Rotation und Skalierung gestatten. Eine solche, auch als Starrkörpertransformation (engl. *rigid body transformation* (RBT)) bezeichnete Koordinatentransformation, lässt sich aufgrund der Darstellung der Formen durch komplexwertige Landmarken mithilfe von zwei ebenfalls komplexwertigen Parametern realisieren. Einerseits dient die Multiplikation der Landmarken mit α zur Realisierung einer Rotation sowie Skalierung und andererseits

[1] Prokrustes war ein Riese in der griechischen Mythologie, der Reisenden ein Bett anbot und sie an die Größe des Bettes anpasste. Waren sie zu klein, streckte er sie, waren sie zu groß, trennte er ihnen die überschüssigen Gliedmaßen ab. [Bro72c]

ermöglicht die Addition von β eine Verschiebung der Landmarken. Die Distanz, die nach der möglichst deckungsgleichen Abbildung verbleibt, wird als volle PROKRUSTES-Distanz (engl. *full* PROCRUSTES *distance*) bezeichnet und ist gemäß [DM98] als

$$\gamma(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) = \min_{\beta,\alpha} \|\boldsymbol{\ell}_2 - (\alpha \cdot \boldsymbol{\ell}_1 + \beta)\|_2 \tag{6.9}$$

definiert. Auf eine Beschreibung, wie die Parameter α und β zu ermitteln sind, wird an dieser Stelle verzichtet und auf Abschnitt 8.3 bzw. [DM98] verwiesen. Unter Verwendung der optimalen Parameter für α und β ergibt sich die volle PROKRUSTES-Distanz zu

$$\gamma(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) = \sqrt{1 - \frac{\|\boldsymbol{\ell}_1^{\mathrm{H}}\, \boldsymbol{\ell}_2\|_2^2}{\|\boldsymbol{\ell}_1\|_2^2\, \|\boldsymbol{\ell}_2\|_2^2}}. \tag{6.10}$$

Die mittlere geometrische Anordnung $\hat{\overline{\boldsymbol{\ell}}}$ ist nun diejenige, die die geringste Distanz zu allen anderen Sensorkonfiguration $\boldsymbol{\ell}_\jmath$, $\jmath = 1, \ldots, \mathcal{J}$, besitzt:

$$\hat{\overline{\boldsymbol{\ell}}} = \operatorname*{argmin}_{\overline{\boldsymbol{\ell}}} \sum_{\jmath=1}^{\mathcal{J}} \left(\gamma\left(\boldsymbol{\ell}_\jmath, \overline{\boldsymbol{\ell}}\right)\right)^2. \tag{6.11}$$

Die Lösung von Gl. (6.11) ist nach [Ken94] durch den Eigenvektor zum größten Eigenwert der Summe

$$S = \sum_{\jmath=1}^{\mathcal{J}} \frac{\boldsymbol{\ell}_\jmath\, \boldsymbol{\ell}_\jmath^{\mathrm{H}}}{\boldsymbol{\ell}_\jmath^{\mathrm{H}}\, \boldsymbol{\ell}_\jmath} \tag{6.12}$$

gegeben. Letztendlich können so die von mehreren RANSAC ermittelten geometrischen Anordnungen zusammengefasst werden, selbst wenn die Beschreibungen in unterschiedlichen Koordinatensystemen vorliegen.

6.6 Analyse

Das in den zurückliegenden Abschnitten entwickelte Konzept zur Anwendung des RANSAC hat das Ziel, die Robustheit des erweiterten Einfallswinkelverfahrens gegenüber Störungen bei der Einfallswinkelschätzung zu steigern. Inwiefern der RANSAC diese Erwartungen erfüllen kann, sollen die nachfolgenden Untersuchungen klären. Die Analysen beginnen mit einem Vergleich der Kalibrierungsgenauigkeit zwischen dem erweiterten Einfallswinkelverfahren und der in den RANSAC eingebetteten Variante. Die Kernfunktionalität des RANSAC besteht jedoch in der Auswahl von Beobachtungen, die keine Ausreißer darstellen. Allerdings weisen die vorliegenden Einfallswinkel, insbesondere bei der Simulation von drei Mikrofonen, nur geringe Störungen auf. Daher soll außerdem die Auswirkung von Ausreißern auf das Ergebnis betrachtet werden. Als letztes findet eine Gegenüberstellung der Ergebnisse des konventionellen RANSAC mit denen, die die partitionierte Variante liefert (vgl. Abschnitt 6.4) statt.

Grundlage für alle Experimente sind die schon aus Abschnitt 5.4 bekannten Szenarien mit vier Sensoren. Für die Kalibrierung stehen weiterhin 20 zufällig im Raum verteilte Ereignisse zur Verfügung. Die Modellierung des Fehlers der Winkelschätzung erfolgt

erneut mit dem auch in Abschnitt 5.6 verwendeten Histogrammmodell, das aus den Simulationsergebnissen aus Kapitel 4 entsteht. Im Fall von drei Mikrofonen beschreiben diese Histogramme näherungsweise eine VON MISES-Verteilung. Bei nur zwei Mikrofonen entstehen hingegen Verteilungen vergleichbar zu Abb. 4.17.

Sofern zur Kalibrierung das in den RANSAC eingebettete erweiterte Einfallswinkelverfahren dient, das zusätzlich lokale Iterationen erlaubt (vgl. Abb. 6.5), resultieren daraus die in Abb. 6.7 dargestellten Ergebnisse. Eine Einordnung dieser Resultate gestatten die in Abb. 5.10 präsentierten Kalibrierungsfehler des erweiterten Einfallswinkelverfahrens ohne zusätzlichen RANSAC.

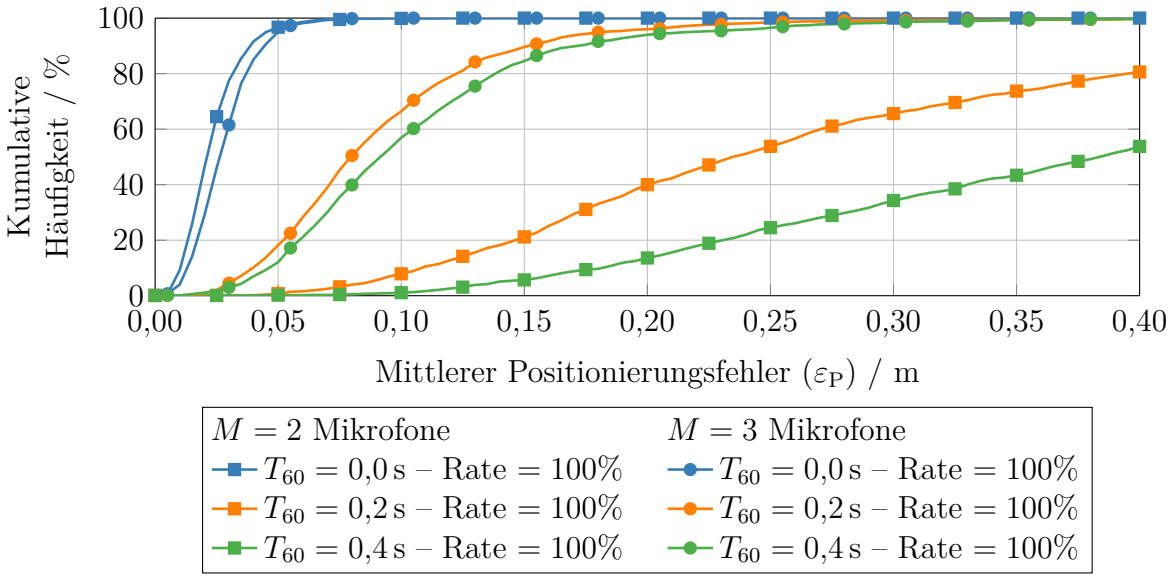


Abbildung 6.7: Kumulative Häufigkeitsverteilung des mittleren Positionierungsfehlers bei der Geometriekalibrierung durch das in den RANSAC eingebettete Einfallswinkelverfahren bei verschiedenen Nachhallzeiten.

Bei der Verwendung von drei Mikrofonen entstehen mit und ohne RANSAC nahezu identische Kalibrierungsfehler. Eine Reduktion des Fehlers war jedoch auch nicht zu erwarten, da beim Einsatz von drei Mikrofonen näherungsweise eine VON MISES-Verteilung des Winkelfehlers vorliegt und das erweiterte Einfallswinkelverfahren dementsprechend den ML-Schätzer darstellt. Damit bestätigen die Untersuchungen, dass ein RANSAC keinen Vorteil bringt, wenn keine Ausreißer vorhanden sind.

Falls dagegen nur zwei Mikrofone zur Verfügung stehen, treten bei der Einfallswinkelschätzung Ausreißer auf (vgl. Abb. 4.6). Diese führen einerseits zu dem signifikanten Anstieg des Kalibrierungsfehlers und andererseits dazu, dass das Einfallswinkelverfahren häufiger divergiert (siehe Abb. 5.10). Anhand eines Vergleichs der Ergebnisse aus Abb. 5.10 mit den Ergebnissen des RANSAC (Abb. 6.7), zeigt sich eindeutig, dass dieser in der Lage, ist die vorhandenen Ausreißer zu erkennen und so den Kalibrierungsfehler zu reduzieren.

Abb. 6.7 enthält darüber erneut die Konvergenzraten für die jeweiligen Untersuchungen. Allerdings entsprechen diese nicht der in Abb. 5.10 verwendeten Definition. Die dort genannten Raten beziehen sich unmittelbar auf ein einzelnes Einfallswin-

kelverfahren. Die hier angegebenen Raten kennzeichnen hingegen den erfolgreichen Abschluss des gesamten RANSAC und belegen damit, dass mindestens ein konvergentes Newton-Verfahren innerhalb des RANSAC vorliegt.

Aufgrund der bisherigen Gegenüberstellung der Ergebnisse mit und ohne RANSAC erscheint der Einsatz des wesentlich aufwändigeren RANSAC-Konzeptes nur bei zwei Mikrofonen notwendig zu sein, da nur dort Ausreißer der Einfallswinkelschätzung vorhanden sind. Allerdings sind Ausreißer bei der Einfallswinkelschätzung nicht nur eine Folge der Mikrofonanordnung bzw. -anzahl. In realen Umgebungen treten Ausreißer z. B. durch eine fehlende LOS-Komponente im Mikrofonsignal oder aufgrund von impulshaften Geräuschen, wie z. B. dem Zuschlagen einer Tür, auf. Um diese Situation nachzubilden, werden 10 % der Einfallswinkelschätzungen nicht mit dem bislang verwendeten Modell generiert, sondern aus einer Gleichverteilung gezogen. Die durch diesen gleichverteilten Anteil modellierten Ausreißer sorgen dafür, dass das erweiterte Einfallswinkelverfahren ohne RANSAC kollabiert. Es liefert nur noch in knapp 20 % der Fälle eine Lösung und wird somit unbrauchbar. Durch den Einsatz des RANSAC ist eine zuverlässige Erkennung dieser Ausreißer möglich und die erzielten Ergebnisse (siehe Abb. 6.8) unterscheiden sich nur geringfügig von den Ergebnissen ohne die Präsenz von Ausreißern (vgl. Abb. 6.7). Angesichts der Anfälligkeit des erweiterten Einfallswinkelverfahrens gegenüber einzelnen Ausreißern ist die Nutzung des RANSAC für reale Anwendungen damit unvermeidbar.

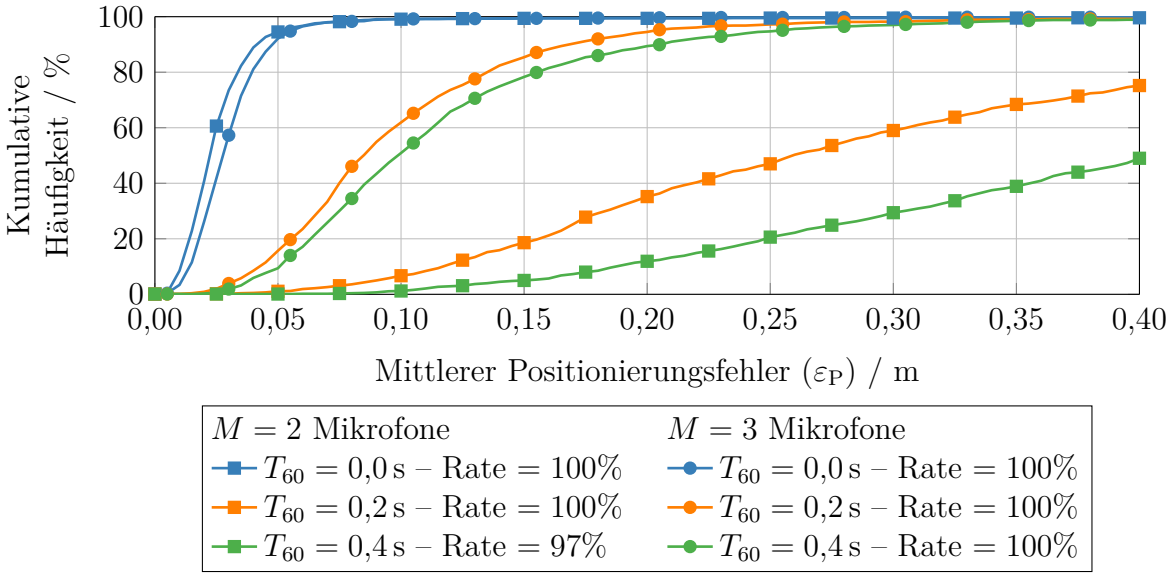


Abbildung 6.8: Kumulative Häufigkeitsverteilung des mittleren Positionierungsfehlers des in den RANSAC eingebetteten Einfallswinkelverfahrens bei zwei- bzw. drei-elementigen Sensorknoten und verschiedenen Nachhallzeiten, wenn zusätzlich 10 % der Beobachtungen gleichverteilte Ausreißer darstellen.

Die Entwicklung des PRANSAC (siehe Abschnitt 6.4) soll sowohl die Rechenkomplexität reduzieren als auch die Handhabung größerer Mengen von Beobachtungen gestatten. Allerdings lässt sich ein zeitlicher Gewinn nur schwer quantifizieren und im aktuell betrachteten Setup stehen lediglich 20 Beobachtungen zur Verfügung. In dieser

Untersuchung sollen daher ausschließlich die Auswirkungen auf den Kalibrierungsfehler berücksichtigt werden. Für eine weitere Analyse des PRANSAC sei zunächst auf Abschnitt 8.1 oder Abschnitt 9.4 verwiesen. Die Grundlage für den jetzt durchgeführten Vergleich bilden dieselben Daten, die auch schon für den Vergleich der Kalibrierung mit und ohne RANSAC zum Einsatz kamen. Sofern die Winkelschätzung durch drei Mikrofone erfolgt, liegen deutlich präzisere Winkelschätzungen als beim Einsatz von nur zwei Mikrofonen vor. Daher ist bereits die bisherige Variante des RANSAC in der Lage, eine zuverlässige Kalibrierung zu ermöglichen, die sich durch eine Partitionierung des RANSAC nicht weiter steigern lässt. Bei den Winkelschätzungen, die nur auf zwei Mikrofonen basieren, führt der PRANSAC zu einer Reduktion des Fehlers, wie die Ergebnisse in Tab. 6.1 dokumentieren. Dabei wird noch einmal bestätigt, dass bei geringen Fehlern der Einfallswinkelschätzung ($T_{60} = 0{,}0\,\mathrm{s}$) keine Verbesserung zu erzielen ist und diese erst bei starken Störungen ($T_{60} = 0{,}4\,\mathrm{s}$) eintritt. Überzeugende Argumente für den Einsatz des PRANSAC werden erst die in Abschnitt 8.1 beschriebenen Experimente zur Kalibrierung audio-visueller Sensornetze liefern.

Nachhallzeit (T_{60})		0,0 s	0,2 s	0,4 s
Durchschnittlicher mittlerer Positionierungsfehler	mit RANSAC	0,02 m	0,32 m	0,78 m
	mit PRANSAC	0,02 m	0,26 m	0,62 m

Tabelle 6.1: Vergleich des durchschnittlichen mittleren Positionierungsfehlers bei der Kalibrierung mit RANSAC bzw. PRANSAC, wenn nur zwei Mikrofone zur Verfügung stehen.

6.7 Zusammenfassung

Das erweiterte Einfallswinkelverfahren verwendet alle vorhandenen Einfallswinkelschätzungen zur Bestimmung der LS-Lösung der Sensoranordnung und gestattet dadurch die Kompensation von Störungen, die durch eine Verteilung aus der Exponentialfamilie beschrieben werden können. Bei der Nutzung von drei Mikrofonen zur Einfallswinkelschätzung liegt mit der von Mises-Verteilung eine Verteilung vor, die zur Exponentialfamilie gehört. Dementsprechend ermöglicht das Einfallswinkelverfahren eine präzise Kalibrierung. Wenn jedoch nur zwei Mikrofone zur Verfügung stehen oder aufgrund der Ausbreitungsbedingungen keine LOS-Komponente im Signal vorhanden ist, treten zusätzlich Winkelschätzungen mit großen Fehlern (Ausreißer) auf. Diese Ausreißer verursachen bei einem LS-Ansatz einen signifikanten Anstieg des Kalibrierungsfehlers. Damit dennoch eine gegenüber Ausreißern robuste Kalibrierung erzielt werden konnte, wurde das Einfallswinkelverfahren in den RANSAC eingebettet.

Der RANSAC nutzt weiterhin das Einfallswinkelverfahren zur Bestimmung der Sensorkonfiguration, allerdings wird nur eine zufällig ausgewählte Teilmenge der Winkelschätzungen verwendet. Alle Winkel, die mit der daraus ermittelten Sensorkonfiguration in Einklang stehen, bilden den Konsens. Anhand der Größe des Konsenses wird wiederum die Qualität der Lösung bewertet und die zufällige Auswahl von Einfallswinkeln

so lange wiederholt, bis eine Sensoranordnung vorliegt, die mit ausreichend vielen Winkelschätzungen zu vereinbaren ist.

Die zentralen Komponenten des RANSAC sind neben der eigentlichen Schätzung der Sensorkonfiguration, die zufällige Selektion der Beobachtungen sowie ein Kriterium zur Beurteilung, ob die Einfallswinkel eines Ereignisses zu einer gegebenen Geometrie passen. Damit das Einfallswinkelverfahren basierend auf den ausgewählten Winkeln eine Geometrie ermitteln kann, ist zudem eine ausreichende räumliche Diversität der zu den Winkeln korrespondierenden Ereignispositionen notwendig. Da die Positionen jedoch unbekannt sind, wurde stattdessen die Änderung der Einfallswinkel zwischen zwei Ereignissen als Indikator für den Abstand verwendet und der K-means++-Algorithmus zur Auswahl von möglichst weit voneinander entfernt liegenden Ereignissen eingesetzt. Die durchgeführten Untersuchungen dokumentieren, dass sich durch den K-means++-Algorithmus ein größerer mittlerer Abstand der Ereignispositionen erzielen lässt als bei einer gleichverteilten Selektion. Besonders deutlich wird die Steigerung der räumlichen Diversität bei Trajektorien, die Abschnitte beinhalten, in denen sich die Quelle des Ereignisses nicht bewegt. Dort liefert eine gleichmäßige Auswahl z. T. unzureichende Abstände, während die K-means++-basierte Alternative weiterhin einen mittleren Abstand von mehr als 1,00 m gewährleistet.

Zur Bewertung, ob die Einfallswinkel eines Ereignisses zu einer ermittelten Geometrie passen, wurde ein Triangulationsverfahren verwendet, das aufgrund der Positionen und Orientierungen der Sensoren sowie den Einfallswinkeln, Schnittpunkte bestimmt. Fehler bei der Einfallswinkelschätzung, ebenso wie bei der Geometriekalibrierung, sorgen allerdings dafür, dass die Schnittpunkte eines Ereignisses nicht in einem Punkt zusammen fallen, sondern in der Nähe dieses Punktes streuen. Anhand der Streuung der Schnittpunkte wurde wiederum bewertet, ob die Einfallswinkel eines Ereignisses mit der Geometrie in Einklang stehen.

Obwohl bei der Nutzung des RANSAC durch die Erkennung von Ausreißern meist nur Winkelschätzungen, deren Fehler durch eine von Mises-Verteilung modelliert werden kann, in die LS-Lösung einfließen, beeinträchtigen auch diese Fehler das Kalibrierungsergebnis. Das Konzept des LORANSAC sieht deshalb vor, dass innerhalb jeder RANSAC-Iteration eine zusätzliche lokale Iteration erfolgt. Dabei wird, wie auch bisher, basierend auf einer zufällig ausgewählten Teilmenge, ein Konsens bestimmt. Allerdings dient dieser unmittelbar zur Berechnung neuer Modellparameter und erst der daraus resultierende Konsens bildet das Ergebnis einer RANSAC-Iteration. Dieses zweistufige Vorgehen wurde dahingehend erweitert, dass beliebig viel lokale Iterationen möglich sind, die eine schrittweise Präzisierung der Lösung gestatten. Andererseits wächst durch die schrittweise Präzisierung jedoch auch der Rechenaufwand. Um dem entgegenzuwirken, wurde außerdem eine partitionierte Variante des RANSAC konzipiert. Diese verringert nicht nur den Anstieg der Rechenkomplexität, sondern ermöglicht darüber hinaus auch die Handhabung von sehr vielen Daten in einem RANSAC. Damit am Ende des partitionierten RANSAC (PRANSAC) die individuellen Schätzungen der geometrischen Anordnungen zu einer gemeinsamen Lösung vereinigt werden konnten, wurde zudem ein Konzept zur Berechnung des Mittelwertes der Sensorkonfiguration entwickelt. Dazu kamen Methoden der Prokrustes Analyse zum Einsatz, die es ermöglichen, den Mittelwert verschiedener Geometrien selbst dann zu ermitteln, wenn die Teilergebnisse in verschiedenen Koordinatensystem vorliegen.

Abschließend wurden in einer Analyse die Auswirkungen des RANSAC und der vorgeschlagenen Weiterentwicklungen auf das Kalibrierungsergebnis untersucht. Bei der Verwendung von nur zwei Mikrofonen pro Sensorknoten weisen die Einfallswinkelschätzungen, insbesondere bei größerem Nachall, Ausreißer auf (vgl. Kapitel 4). Dementsprechend erzielt der RANSAC im Vergleich zum LS-Ansatz eine Reduktion des Kalibrierungsfehlers. Werden dagegen drei Mikrofone verwendet, entspricht die Verteilung des Winkelfehlers ungefähr der dem Einfallswinkelverfahren zugrunde liegenden von Mises-Verteilung. Daher liegt der Kalibrierungsfehler weiterhin auch bei einer Nachhallzeit von 0,4 s noch in mehr als 90 % der Fälle unter 0,20 m. Da in realen Umgebungen jedoch bspw. aufgrund einer fehlenden LOS-Komponente auch beim Einsatz von drei Mikrofonen Ausreißer bei der Einfallswinkelschätzung zu erwarten sind, wurden in einem weiteren Experiment 10 % der Einfallswinkel durch gleichverteilte Ausreißer ersetzt. Diese Ausreißer sorgen dafür, dass das Einfallswinkelverfahren ohne RANSAC zusammenbricht und nur in ca. 20 % der Fälle überhaupt eine Lösung liefert. Bei der Nutzung des RANSAC haben diese zusätzlichen Ausreißer nahezu keinen Einfluss auf den Kalibrierungsfehler. Daher ist der RANSAC ein wichtiger Schritt, um die Kalibrierung akustischer Sensornetze in realen Umgebungen zu ermöglichen. Der PRANSAC trägt unterdessen nur bei der Verwendung von zwei Mikrofonen zu einer Reduktion des Kalibrierungsfehlers bei, da bei drei Mikrofonen keine Ausreißer vorhanden sind.

7 Skalierung

Durch die Kombination des RANSAC mit der weiterentwickelten Variante des Einfallswinkelverfahrens ist insgesamt ein Algorithmus entstanden, der die Kalibrierung eines akustischen Sensornetzes ermöglicht, auch wenn die DOA-Schätzungen aufgrund von Nachhall gestört sind und zusätzlich Ausreißer beinhalten. Lediglich die Problematik eines unbekannten Skalierungsfaktors, der infolge einer ausschließlich auf Winkeln basierenden Kalibrierung entsteht, ist weiterhin vorhanden. Die Bestimmung der Skalierung durch die a priori Kenntnis einer Distanz [KWL08] bietet zwar eine Lösungsmöglichkeit, widerspricht damit aber der angestrebten vollständigen Automatisierung des Kalibrierungsprozesses. In Anbetracht dessen beschäftigt sich dieses Kapitel ebenso wie das Nächste mit verschiedenen Ansätzen, die Skalierung automatisch festzulegen.

Die grundsätzliche Problematik eines Skalierungsfaktors stellt indes keine Besonderheit des genutzten Algorithmus dar. Sie tritt in ähnlicher Form auch bei anderen winkelgestützten Ansätzen, wie z. B. der Kalibrierung von Kameranetzen [BD10], auf. Im visuellen Umfeld trägt die Kenntnis der intrinsischen Kameraparameter[1] zur Lösung des Problems bei. Die Berücksichtigung dieser Parameter erlaubt es jeder Kamera, die Größe bzw. Entfernung zu den erfassten Objekten zu ermitteln. Angesichts der daraus gewonnen Distanzinformationen ist letztendlich eine Kalibrierung möglich, ohne dass ein Skalierungsfaktor verbleibt [Nis04].

Damit die im visuellen Bereich genutzten Techniken zur Fixierung der Skalierung von akustischen Sensornetzen wiederverwendet werden könnten, müssten aus den Audiosignalen der jeweiligen Sensorknoten neben den Einfallswinkeln auch Distanzschätzungen zur Signalquelle extrahiert werden. Allerdings bietet ein Sensorknoten, sofern dieser nur aus zwei Mikrofonen besteht, im Gegensatz zu einer Kamera, zunächst keine direkte Möglichkeit, die Distanz zur Quelle zu schätzen. Erst der Einsatz von weiteren Mikrofonen gestattet eine Entfernungsschätzung durch TDOA-Messungen. Existierende Verfahren, wie z. B. [Val+10b], verwenden dazu Mikrofone mit einem Abstand von 0,20 m und stehen daher im Widerspruch zu der Anforderung dieser Arbeit, die Kalibrierung mithilfe von kompakten Arrays zu realisieren. Für die Alternative [Esa+12] reicht zwar ein Dodekaeder-Array mit nur 0,08 m Durchmesser aus, aber trotz des Einsatzes von 60 Mikrofonen beträgt der RMSE der Distanzschätzung noch ca. 0,33 m.

Erfolgversprechender erscheint hingegen die Nutzung von Signallaufzeitdifferenzen zwischen Mikrofonen aus verschiedenen Arrays (Inter-Array-TDOA), da diese eine wesentlich größere Distanz aufweisen und somit präzisere Schätzungen liefern. Voraussetzung für die Verwendung von Mikrofonen aus verschiedenen Arrays ist jedoch

[1] Die intrinsischen Kameraparameter modellieren den Zusammenhang zwischen einem Punkt in der Bildebene und dessen Position im Weltkoordinatensystem [FP03].

eine Abtastsynchronisation der beteiligten Sensorknoten, auf die bislang verzichtet werden konnte. Trotzdem beschreibt Abschnitt 7.1 einen Ansatz, der Inter-Array-TDOA-Schätzungen verwendet. Die Berücksichtigung dieses Ansatzes soll einerseits die deutlich bessere Ausgangssituation, die sich bei der Nutzung eines Sensornetzes mit Abtastsynchronisation ergibt, aufzeigen und andererseits als Referenz für den in Abschnitt 7.2 erläuterten Algorithmus dienen. Dieser nutzt erneut ausschließlich Einfallswinkel und verlangt deshalb keine Abtastsynchronisation. Um dennoch zusätzliche Informationen gewinnen zu können, werden Arrays berücksichtigt, die über mehr als zwei Mikrofone verfügen. Die zusätzlichen Mikrofone sollen jedoch nicht zur Steigerung der Präzision einer gemeinsamen Winkelschätzung dienen, sondern die individuelle DOA-Schätzung von mehreren Teilarrays gestatten. Die Kenntnis der Anordnung der Teilarrays innerhalb der jeweiligen Sensorknoten soll anschließend, ähnlich der bereits in Abschnitt 5.4 verwendeten Abstandsgleichungen, die notwendigen Bedingungen zur Festlegung der Skalierung liefern.

7.1 Skalierung durch Signallaufzeitdifferenzen

Den Ausgangspunkt für die im weiteren Verlauf erläuterte Vorgehensweise, den Skalierungsfaktor aus Inter-Array-TDOA-Messungen zu ermitteln, bilden die Kalibrierungsergebnisse des in den RANSAC eingebetteten erweiterten Einfallswinkelverfahrens. Zur Fixierung der Skalierung innerhalb des Newton-Verfahrens dient dabei die zusätzliche Gleichung, die die paarweisen Abstände aller Sensoren beschreibt (vgl. Gl. (5.9)). Nach Abschluss des RANSAC liegen somit zwar die Sensorpositionen $\widetilde{\boldsymbol{s}}_i$, $i = 1, \ldots, I$, vor, aber die Skalierung dieser hängt von der beliebig gewählten Distanz ab.

Zur Gewinnung des verbleibenden Skalierungsfaktors mithilfe von Signallaufzeitdifferenzen dienen Konzepte, die sowohl bei der akustischen Lokalisation [GS08] als auch im Bereich der Funk-Ortung [El +13] Anwendung finden. Die Gemeinsamkeit der Verfahren besteht darin, dass die Messung der TDOA einen Rückschluss auf die Streckendifferenz gestattet. Übertragen auf zwei Mikrofone m' und n' an den Positionen $\boldsymbol{m}_{m'}$ und $\boldsymbol{m}_{n'}$ ergibt sich die Zeitdifferenz zwischen dem Eintreffen des von der Position $\boldsymbol{e}$ ausgesandten Signals zu

$$\tau_{(n',m')} = \frac{\|\boldsymbol{e} - \boldsymbol{m}_{n'}\|_2 - \|\boldsymbol{e} - \boldsymbol{m}_{m'}\|_2}{c_\mathrm{S}}. \tag{7.1}$$

Sofern sowohl die Schallgeschwindigkeit c_S als auch die Positionen der Mikrofone innerhalb des Sensornetzes vorliegen, ermöglichen TDOA-Messungen die Lokalisation der Signalquelle:

$$\hat{\boldsymbol{e}} = \underset{\boldsymbol{e}}{\operatorname{argmin}} \sum_{m'=1}^{M'} \sum_{n'=m+1}^{M'} \left\| \|\boldsymbol{e} - \boldsymbol{m}_{n'}\|_2 - \|\boldsymbol{e} - \boldsymbol{m}_{m'}\|_2 - c_\mathrm{S} \cdot \tau_{(n',m')} \right\|_2. \tag{7.2}$$

Um eine Verwechselung mit der Anzahl der Mikrofone innerhalb eines Arrays (M) zu vermeiden, wird die Gesamtanzahl aller Mikrofone des Sensorsystems durch M' gekennzeichnet. Analog dazu werden m' und n' zur Indizierung verwendet.

Wie schon zuvor erwähnt, resultieren aus der bisherigen Kalibrierung beliebig skalierte Positionen der Sensorknoten $\widetilde{\boldsymbol{s}}_i$ sowie davon unbeeinflusste Orientierungen θ_i. Mit diesen

Informationen gestattet das auch schon im RANSAC eingesetzte Lokalisationsverfahren (vgl. Abschnitt 6.2.2) eine Bestimmung der Ereignispositionen $\widetilde{\boldsymbol{e}}_d$. Allerdings beinhalten auch die Ereignispositionen den Skalierungsfaktor ν, weil zur Berechnung dieser die beliebig skalierten Sensorpositionen herangezogen werden.

Außerdem liefert der RANSAC zunächst nur die Positionen der Sensorknoten. Die Bestimmung des Skalierungsfaktors erfordert hingegen die Positionen der Mikrofone. Aufgrund des bekannten Aufbaus der Sensorknoten ergeben sich diese jedoch unmittelbar aus den Positionen und Orientierungen der Sensoren. Zusammen mit einer Funktion zur Abbildung des globalen Mikrofonindexes m' auf die dazu korrespondierenden Indices i und m lässt sich die Position eines Mikrofons durch

$$\boldsymbol{m}_{m'} = \boldsymbol{m}_{i,m} = \nu \cdot \widetilde{\boldsymbol{s}}_i + \boldsymbol{o}_m(\theta_i) \tag{7.3}$$

beschreiben. Dabei bezeichnet $\boldsymbol{o}_m(\theta_i)$ den Vektor vom Zentrum des i-ten Sensorknotens zum m-ten Mikrofon.

Um eine kompakte Notation zu gestatten, werden im weiteren Verlauf die Indices m' und n' synonym für Indices i und m bzw. j und n genutzt. Mit dieser Notation ergibt sich aus der Kombination von Gl. (7.1) und Gl. (7.3) schließlich der Zusammenhang:

$$\tilde{\tau}_{(n',m')}(d,\nu) = \frac{\left\|\nu \cdot \widetilde{\boldsymbol{e}}_d - \nu \cdot \widetilde{\boldsymbol{s}}_j + \boldsymbol{o}_n(\theta_j)\right\|_2 - \left\|\nu \cdot \widetilde{\boldsymbol{e}}_d - \nu \cdot \widetilde{\boldsymbol{s}}_i + \boldsymbol{o}_m(\theta_i)\right\|_2}{c_\mathrm{S}}. \tag{7.4}$$

In Kombination mit den gemessenen TDOA $\tau_{(n',m')}(d)$, $m' = 1, \ldots, M'$ und $n' = 1, \ldots, M'$, entsteht daraus das Optimierungsproblem

$$\hat{\nu} = \underset{\nu}{\operatorname{argmin}} \sum_{d=1}^{D} \sum_{m'=1}^{M'} \sum_{n'=m'+1}^{M'} \left\| \tilde{\tau}_{(n',m')}(d,\nu) - \tau_{(n',m')}(d) \right\|_2. \tag{7.5}$$

Der Skalierungsfaktor ν ist demnach so zu wählen, dass die Abweichungen zwischen der gemessenen und der aus dem Geometriekalibrierungsergebnis prädizierten Zeitdifferenz minimal wird. Dabei ist es allerdings nicht notwendig, alle paarweisen Mikrofonkombinationen zu berücksichtigen, da die Signallaufzeitdifferenz $\tau_{(n',m')}(d)$ der TDOA $\tau_{(m',n')}(d)$ mit invertiertem Vorzeichen entspricht und deshalb keine zusätzlichen Informationen liefert. Darüber hinaus schließt die in Gl. (7.5) verwendete Notation sowohl die TDOA zwischen Mikrofonen innerhalb eines Arrays als auch die arrayübergreifenden Messungen mit ein. Die durchgeführten Untersuchungen zeigen jedoch, dass sich eine Berücksichtigung der TDOA zwischen Mikrofonen desselben Arrays z. T. negativ auf das Gesamtergebnis auswirkt. Daher werden diese, auch ohne eine explizite Kennzeichnung in der verwendeten Notation, nicht zur Berechnung des Skalierungsfaktors herangezogen.

Verantwortlich für die Beeinträchtigung der Skalierungsfaktorschätzung durch die Nutzung von Intra-Array-TDOA-Messungen sind die geringen Abstände der Mikrofone innerhalb der jeweiligen Arrays. Bereits bei der Lokalisierung von Personen und Ereignissen mithilfe von TDOA-Messungen werden bevorzugt größere Mikrofonabstände verwendet [Val+10b; Hen+09], um eine präzisere Positionsschätzung zu gestatten. Da der zur Bestimmung des Skalierungsfaktors genutzte Ansatz (siehe Gl. (7.5)) auf dem schon bei der Lokalisierung verwendeten Optimierungsproblem basiert (vgl. Gl. (7.2)), treten dementsprechend erneut dieselben Probleme auf.

Die Basis für die Analyse des Konzeptes zur Gewinnung der Skalierung durch arrayübergreifende TDOA-Messungen bilden die Simulationsergebnisse des in den RANSAC eingebetteten Einfallswinkelverfahrens aus Abschnitt 6.6. Die TDOA, die sich bspw. durch Einsatz von GCCPhat aus den Audiosignalen der Mikrofone extrahieren lassen, werden bei der folgenden Untersuchungen synthetisch generiert. Dabei wird die verbreitete Annahme verwendet, dass die Störungen normalverteilt sind, auch wenn in realen Umgebungen aufgrund der Mehrwegeausbreitung z. T. multimodale Verteilungen auftreten [Oua+12]. Obwohl die Nachhallzeit den Fehler der TDOA-Schätzung beeinflusst, wurde die Standardabweichung des TDOA-Fehlers unabhängig von der Nachhallzeit zu 0,08 m gewählt. Dies entspricht dem in [PF14b] angegebenen Fehlern beim Einsatz von GCCPhat bzw. SRPPhat in einem Raum mit einer Nachhallzeit von ca. 0,7 s.

Abb. 7.1 stellt den Positionierungsfehler für die erläuterte Skalierung durch TDOA-Messungen dar. Dabei kennzeichnen die Balken den Mittelwert des Positionierungsfehlers aus 1000 Experimenten und die zugehörigen Antennen zeigen die Standardabweichung. Um eine Einschätzung zu gestatten, welcher Anteil des Fehlers bereits vom Einfallswinkelverfahren stammt, ist neben dem Gesamtfehler (TDOA-Skalierung) auch der Anteil dargestellt, wenn das Ergebnis des Einfallswinkelverfahrens mithilfe eines Orakels bzw. a priori Wissen optimal skaliert (Orakel) wird. Das Ziel besteht somit darin, eine möglichst geringe Zunahme des Fehlers durch die TDOA-Skalierung zu erreichen.

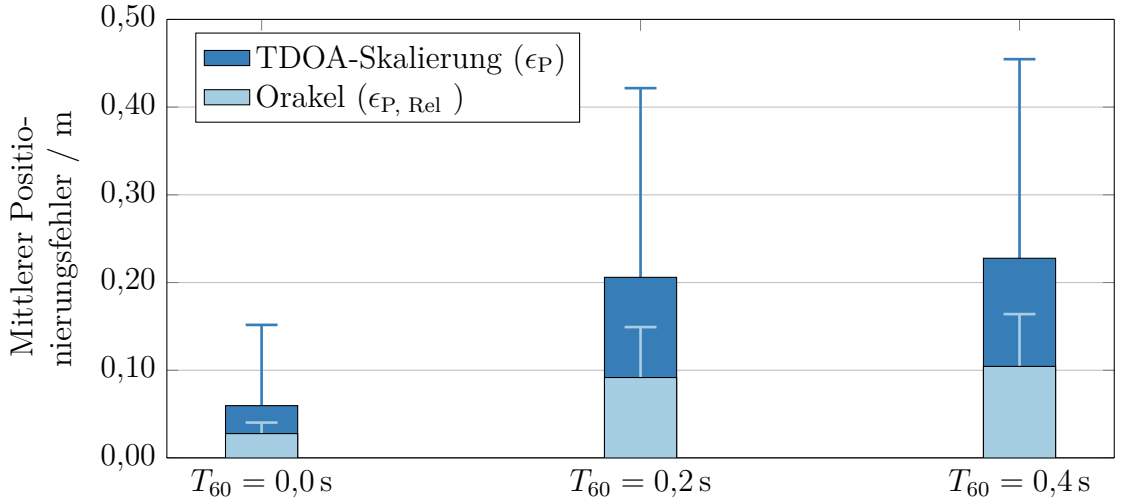


Abbildung 7.1: Mittlerer Positionierungsfehler des in den RANSAC eingebetteten Einfallswinkelverfahrens bei einer Skalierung durch TDOA-Messungen.

Aus den dargestellten Simulationsergebnissen geht hervor, dass der von der Skalierung verursachte Fehler ungefähr eine Verdoppelung des Gesamtfehlers bewirkt. Ferner belegen die außerdem im Rahmen dieser Arbeit durchgeführten Untersuchungen, dass eine Änderung der Standardabweichung des normalverteilten TDOA-Fehlers den Gesamtfehler (TDOA-Skalierung) nur um einige Zentimeter ändert. Verantwortlich dafür sind die ebenfalls vorhandenen Fehler in der geometrischen Anordnung der Mikrofone und Ereignisse, die schon aus dem Einfallswinkelverfahren und der darauf aufbauenden Lokalisierung resultieren. Diese besitzen einen größeren Einfluss als die TDOA-Fehler, sodass deren Einfluss nur eine untergeordnete Rolle spielt.

Begünstigt wird die sehr deutliche Auswirkung der Skalierung auf den Gesamtfehler durch das gewählte Szenario. Die Lage der Sensoren am Rand des Raumes (siehe Abb. 2.1b bzw. Abb. 5.5) führt dazu, dass ein Skalierungsfehler dort einen größeren mittleren Positionierungsfehler hervorruft als ein vergleichbar großer Fehler bei Sensoren, die sich in der Mitte des Raumes konzentrieren (vgl. Abb. 2.1a). Der Grund für dieses Verhalten ist die räumliche Ausdehnung der Sensorkonfiguration. Mit zunehmender Ausdehnung wächst die Entfernung einzelner Sensoren vom Koordinatenursprung und damit vom Zentrum der Skalierung. Je weiter ein Sensor vom Zentrum der Skalierung entfernt ist, desto größer ist auch die von der Skalierung verursachte Positionsänderung und dementsprechend steigt bei einer unpräzisen Skalierung auch der daraus resultierende Positionierungsfehler.

7.2 Skalierung mithilfe von Teilarrays

Als Alternative zur Bestimmung des Skalierungsfaktors durch Signallaufzeitdifferenzen steht in diesem Abschnitt eine Methode im Fokus, die ausschließlich mit Einfallswinkeln arbeitet und deshalb keine Abtastsynchronisation zwischen den Sensorknoten erfordert. Sowohl die Simulationen in Abschnitt 5.4 als auch die Betrachtungen in [KWL08] belegen, dass die a priori Kenntnis des Abstandes zwischen zwei Sensorknoten geeignet ist, um die Skalierung der geometrischen Anordnung festzulegen.

Eine manuelle Messung der erforderlichen Distanz steht indes im Widerspruch zu einer automatischen Geometriekalibrierung. Sofern jeder Sensorknoten jedoch über mehr als zwei Mikrofone verfügt, eröffnet sich eine Möglichkeit, die ohne eine manuelle Messung auskommt. Einerseits können die zusätzlichen Mikrofone bei einer gemeinsamen Winkelschätzung zu einer Reduktion des Schätzfehlers beitragen. Andererseits gestatten sie aber auch die Durchführung von mehreren unabhängigen Winkelschätzungen für verschiedene Teilarrays des Sensorknotens. Unter der Voraussetzung, dass die Anordnung der Mikrofone innerhalb eines Sensorknotens gegeben ist, lässt sich dieser als Ansammlung von mehreren Mikrofonarrays, mit jeweils individuellen Winkelschätzungen und bekannten Abständen, auffassen. Somit liefert jeder Sensorknoten Distanzinformationen, die in Form zusätzlicher Gleichungen (vgl. Abschnitt 5.3) in die Kalibrierung einfließen und damit die Skalierung festlegen.

Die Interpretation der Sensorknoten als Zusammenschluss von Teilarrays, bei denen die Abstände durch zusätzliche Gleichungen definiert sind, gestattet unmittelbar die Lösung des Geometriekalibrierungsproblems mithilfe des bisherigen Algorithmus (vgl. Kapitel 5). Gleichzeitig führt die Unterteilung der Sensorknoten in Teilarrays zu einer Steigerung der Anzahl der zu kalibrierenden Parameter, da die Beschreibung jedes Teilarrays durch eine individuelle Position und Orientierung erfolgt. Demgegenüber stehen jedoch die Skalierungsgleichungen, die die Abstände der Teilarrays fixieren und somit die Freiheitsgerade einschränken. Eine LS-Lösung des Gleichungssystems basierend auf imperfekten Beobachtungen sorgt allerdings dafür, dass alle Gleichungen näherungsweise eingehalten werden. Obwohl also die Skalierungsgleichungen die exakte Anordnung der Teilarrays innerhalb der Sensorknoten beschreiben, werden die Positionen und Orientierungen dieser bei der Lösung des Gleichungssystems verändert, um die Störungen der Einfallswinkel auszugleichen.

Angesichts der genannten Nachteile soll stattdessen eine modifizierte Variante des Einfallswinkelverfahrens konzipiert werden. Dazu wird die bisher genutzte geometrische Beziehung abgeändert, sodass die Lage eines Teilarrays relativ zur Position und Orientierung des Sensorknotens ausgedrückt werden kann. Dadurch lässt sich sowohl eine Steigerung der Anzahl der zu kalibrierenden Parameter verhindern als auch die exakte Berücksichtigung des Aufbaus eines Sensorknotens gewährleisten.

Zur Erläuterung des entwickelten Konzeptes dient die in Abb. 7.2 dargestellte zirkuläre Mikrofongruppe, die sich aus $M = 5$ Mikrofonen zusammensetzt. Diese Mikrofone wiederum bilden $H = M(M-1)/2$ zwei-elementige Teilarrays. Die weiteren Ausführungen beschränken sich zwar auf zirkuläre Arrays und deren Teilung in Arrays mit nur zwei Mikrofonen, gleichwohl lassen sich die präsentierten Schritte auch für andere Konstellationen durchführen.

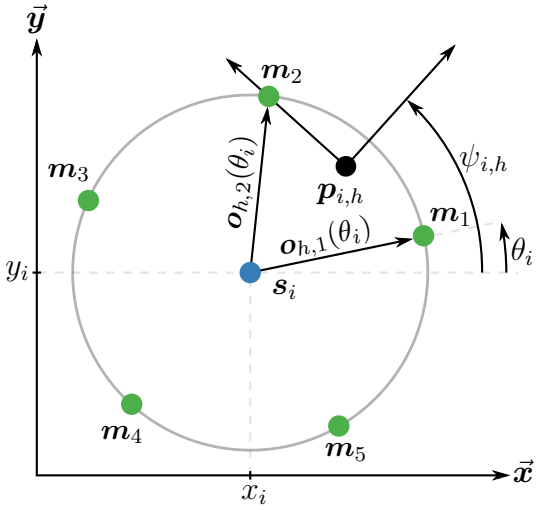


Abbildung 7.2: Aufbau eines zirkulären Mikrofonarrays.

Die in Gl. (5.1) formulierte geometrische Beziehung zwischen Sensorknoten und Ereignisposition bzw. die darauf basierende Weiterentwicklung aus Gl. (5.16) gilt gleichermaßen auch für jedes Teilarray. Die Anwendung des Zusammenhangs für das Teilarray $\boldsymbol{p}_{i,h}$, welches das Ereignis $\boldsymbol{e}_d$ unter dem Winkel $\varphi_{i,h,d}$ erfasst, liefert

$$f_{\mathrm{PA}}(\boldsymbol{p}_{i,h}, \psi_{i,h}, \boldsymbol{e}_d; \varphi_{i,h,d}) = \frac{(\boldsymbol{e}_d - \boldsymbol{p}_{i,h})^{\mathrm{T}}}{\|\boldsymbol{e}_d - \boldsymbol{p}_{i,h}\|_2} \boldsymbol{R}_{\mathrm{xy}}(\psi_{i,h}) \begin{bmatrix} \cos(\varphi_{i,h,d}) \\ \sin(\varphi_{i,h,d}) \end{bmatrix}. \tag{7.6}$$

Sofern die Berücksichtigung des Aufbaus eines Sensorknoten nur durch zusätzliche Distanzgleichungen erfolgt, liegt, wie zuvor erwähnt, die bereits in Kapitel 5 gelöste Problemstellung vor.

Aus der näheren Betrachtung von Abb. 7.2 ergeben sich unmittelbar die notwendigen Zusammenhänge, die es erlauben, die Kenntnis des exakten geometrischen Aufbaus der Sensorknoten mit in die Kalibrierung einzubeziehen. Für die hier beispielhaft betrachtete Unterteilung eines zirkulären Arrays in Mikrofonpaare, ergibt sich die Position eines Paares zu

$$\boldsymbol{p}_{i,h} = \boldsymbol{s}_i + \frac{\boldsymbol{o}_{h,1}(\theta_i) + \boldsymbol{o}_{h,2}(\theta_i)}{2}. \tag{7.7}$$

Dabei bezeichnet $\boldsymbol{o}_{h,1}(\theta_i)$ bzw. $\boldsymbol{o}_{h,2}(\theta_i)$ den Vektor vom Zentrum des i-ten Sensorknotens zum ersten bzw. zweiten Mikrofon des h-ten Paares. Analog dazu lässt sich die Orientierung des Paares als

$$\psi_{i,h} = \theta_i + \frac{\sphericalangle(\boldsymbol{o}_{h,1}(\theta_i), \boldsymbol{o}_{h,2}(\theta_i))}{2} \tag{7.8}$$

ausdrücken. Der Ausdruck $\sphericalangle(\boldsymbol{o}_{h,1}(\theta_i), \boldsymbol{o}_{h,2}(\theta_i))$ bezeichnet dabei den zwischen $\boldsymbol{o}_{h,1}(\theta_i)$ und $\boldsymbol{o}_{h,2}(\theta_i)$ eingeschlossenen Winkel.

Die Verwendung dieser beiden Zusammenhänge sorgt dafür, dass die Anzahl der Unbekannten pro Sensorknoten weiterhin nur 3 beträgt und nicht auf $3 \cdot H$ ansteigt, wie es bei einer Berücksichtigung des Aufbaus des Sensorknotens nur durch zusätzliche Distanzgleichungen der Fall gewesen wäre. Das aus der Kombination von Gl. (7.6), Gl. (7.7) und Gl. (7.8) resultierende Gleichungssystem

$$\boldsymbol{f}_{\mathrm{PA}}(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{E}; \boldsymbol{\Phi}) = \begin{bmatrix} f_{\mathrm{PA}}(\boldsymbol{p}_{1,1}, \psi_{1,1}, \boldsymbol{e}_1; \varphi_{1,1,1}) & \dots & f_{\mathrm{PA}}(\boldsymbol{p}_{I,H}, \psi_{I,H}, \boldsymbol{e}_D; \varphi_{I,H,D}) \end{bmatrix} \tag{7.9}$$

besitzt somit genauso viele Unbekannte, wie das in Kapitel 5 verwendete. Gleichzeitig steigt durch die Verwendung von Teilarrays die Anzahl der Gleichungen um den Faktor H auf $I \cdot D \cdot H$. Die Ausnutzung der Abstände zwischen den Teilarrays (siehe Gl. (7.7) und Gl. (7.8)) sorgt außerdem dafür, dass die Skalierungsinvarianz aufgehoben wird. Somit liefert die Lösung des Gleichungssystems (7.9) die Positionen und Orientierungen der Sensoren, ohne das die Positionsangaben einen unbekannten Skalierungsfaktor beinhalten.

Die Analyse der dargelegten Fixierung der Skalierung mithilfe von aus dem Aufbau der Sensorknoten abgeleiteten Distanzen erfolgt auf Basis derselben Szenarien, wie auch schon in Kapiteln 5 und 6 sowie Abschnitt 7.1. Ein Sensorknoten besteht jetzt allerdings aus einem zirkulären Array mit $M = 5$ Mikrofonen und hat einen Radius von 0,05 m. Zur Beschränkung des Simulationsaufwandes wird der Fehler der DOA-Schätzung der einzelnen Paare weiterhin mit dem aus Abschnitt 5.6 bekannten Modell generiert.

Die erzielten Ergebnisse der zuvor beschriebenen Untersuchung sind in Abb. 7.3 zusammengefasst. Auch hier wird einerseits der Gesamtfehler (ϵ_{P}) und andererseits der Kalibrierungsfehler, nachdem die ermittelte Geometrie zunächst durch ein Orakel skaliert wurde ($\epsilon_{\mathrm{P,\,Rel}}$), dargestellt. Dabei beschreibt der Fehler nach der Skalierung durch ein Orakel ($\epsilon_{\mathrm{P,\,Rel}}$) den Anteil des Gesamtfehlers, der ausschließlich durch eine fehlerhafte Positionierung der Sensoren und nicht von einer ungenauen Skalierung verursacht wird. Die Differenz zwischen $\epsilon_{\mathrm{P,\,Rel}}$ und dem Gesamtfehler ϵ_{P} kennzeichnet somit die von der imperfekten Skalierung hervorgerufene Komponente des Positionierungsfehlers.

Der relative Fehler ($\epsilon_{\mathrm{P,\,Rel}}$) fällt bei einer Nachhallzeit von 0,0 s bzw. 0,2 s geringer aus als bei den bisherigen Untersuchungen mit nur zwei Mikrofonen pro Sensorknoten (siehe Abb. 6.7). Diese Reduktion des Kalibrierungsfehlers lässt sich darauf zurückführen, dass durch die deutlich erhöhte Anzahl von Mikrofonpaaren die LS-Lösung des Gleichungssystems zu einer besseren Kompensation von Winkelfehlern führt. Wenn die Nachhallzeit jedoch auf 0,4 s anwächst, entsteht ein gegensätzliches Bild. Jetzt liefern die zirkulären Arrays wesentlich schlechtere Ergebnisse, als die bisher eingesetzten Paare. Diese Beobachtung erscheint zunächst unerwartet, da mehr Gleichungen zu einem besseren Ausgleich der Störungen und damit auch zu einem präziseren Kalibrierungsergebnis führen sollten. Auslöser für den Anstieg des Kalibrierungsfehlers ist

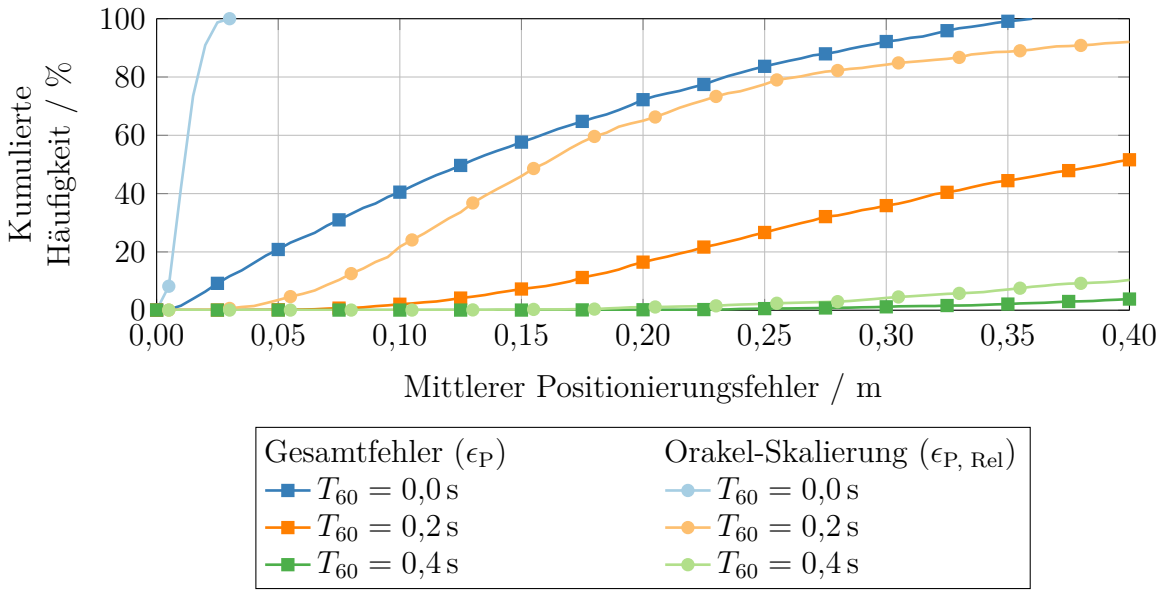


Abbildung 7.3: Kumulative Häufigkeitsverteilung des mittleren Positionierungsfehlers des erweiterten Einfallswinkelverfahrens mit RANSAC, beim Einsatz zirkulärer Arrays zur Fixierung der Skalierung sowie der Unterteilung der zirkulären Arrays in Mikrofonpaare.

die deutliche Zunahme des systematischen Anteils des Winkelfehlers (vgl. Kapitel 4) aufgrund der gestiegenen Nachhallzeit. Daher erzeugen die Winkelschätzungen innerhalb der Sensorknoten sich widersprechende Angaben, die sowohl durch den RANSAC als auch durch die innerhalb des RANSAC verwendete LS-Lösung nicht mehr ausgeglichen werden können.

Für die in diesem Abschnitt im Vordergrund stehende Bewertung des Kalibrierungsergebnisses, inklusive Skalierung, ist allerdings der Gesamtfehler (ϵ_P) ausschlaggebend. Somit belegen die Untersuchungen, dass selbst bei sehr geringen Winkelfehlern ($T_{60} = 0{,}0\,s$) enorme Kalibrierungsfehler durch eine ungenaue Skalierung der geometrischen Anordnung auftreten. Im Gegensatz zu den Ergebnissen aus Abb. 7.3 zeigen die Ergebnisse aus Abschnitt 5.4, dass Sensoren mit ausreichend großer Entfernung untereinander eine Fixierung der Skalierung ermöglichen. Die hier untersuchte Betrachtung der Sensorknoten als Zusammenschluss von Teilarrays gestattet dagegen keine verlässliche Rückgewinnung der Skalierung. Die Wiederholung des zuvor erläuterten Experimentes mit dem für drei-elementige Arrays trainierten Modell des Winkelfehlers bestätigt zudem, dass die Distanz der Mikrofone, im Vergleich zum Fehler der Winkelschätzung, einen größeren Einfluss besitzt (siehe Abb. 7.4).

Der Gesamtfehler ϵ_P fällt bei der in Abb. 7.4 dargestellten Untersuchung erwartungsgemäß geringer als zuvor aus. Allerdings wird dieses Ergebnis im Wesentlichen durch die ebenfalls gesunkenen relativen Fehler $\epsilon_{P,\,Rel}$ verursacht, die wiederum auf die Reduktion der Winkelfehler zurückzuführen sind. Die vom Skalierungsfehler ausgelöste Differenz zwischen $\epsilon_{P,\,Rel}$ und ϵ_P ist näherungsweise unverändert, weil die geringen Abstände der Mikrofonpaare innerhalb eines Sensorknotens keine zuverlässige Triangulation zur Fixierung der Skalierung gestatten.

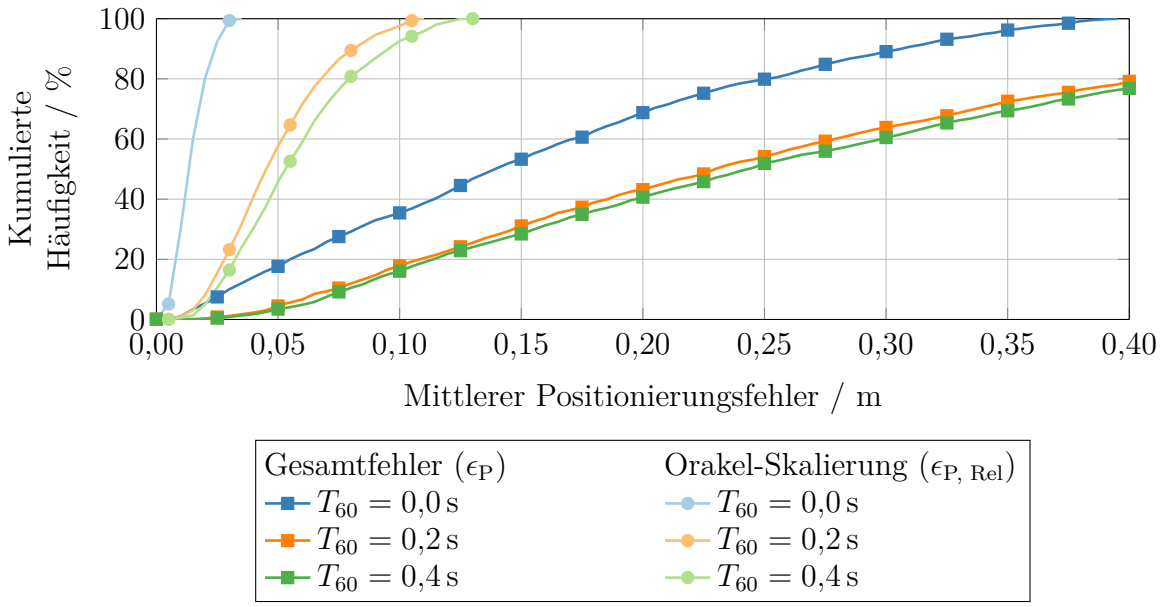


Abbildung 7.4: Kumulative Häufigkeitsverteilung des mittleren Positionierungsfehlers des erweiterten Einfallswinkelverfahrens mit RANSAC, beim Einsatz zirkulärer Arrays zur Fixierung der Skalierung sowie der Unterteilung der zirkulären Arrays in drei-elementige Teilarrays.

7.3 Zusammenfassung

In diesem Kapitel wurden zwei Methoden entwickelt, um den Skalierungsfaktor, der nach Abschluss des erweiterten Einfallswinkelverfahrens verbleibt, anhand der vorliegenden Sprachsignale zu bestimmen. Die Grundlage für die erste Variante bildeten TDOA-Messungen. Durch die Multiplikation mit der Schallgeschwindigkeit ermöglichen diese zwar unmittelbar einen Rückschluss auf die Distanz, allerdings erfordert eine präzise TDOA-Schätzung Mikrofone mit ausreichendem Abstand. Da die Abstände der Mikrofone innerhalb der im Rahmen dieser Arbeit betrachteten kompakten Arrays zu klein für eine verlässliche Schätzung sind, wurden stattdessen Inter-Array-TDOA-Messungen verwendet. Andererseits wird dadurch eine Abtastsynchronisation zwischen den Sensorknoten notwendig, auf die bislang verzichtet werden konnte.

Das zur Schätzung der Skalierung entwickelte Verfahren nutzt Techniken die bereits zur Lokalisierung von Personen und Ereignissen zum Einsatz kommen. Dort lassen sich mithilfe der TDOA und den Positionen der Mikrofone die Positionen der Personen bzw. Ereignisse schätzen. Nach Abschluss der Geometriekalibrierung durch das erweiterte Einfallswinkelverfahren liegen dagegen sowohl Schätzwerte für die Positionen der Mikrofone als auch für die Positionen der Ereignisse vor, die allerdings noch eine unbekannte Skalierung beinhalten. Anstatt die TDOA zur Lokalisierung der Signalquellen zu verwenden, kann stattdessen auch der Skalierungsfaktor ermittelt werden. Dazu wurde zunächst die Differenz zwischen den aus der geometrischen Anordnung prädizierten Laufzeitdifferenzen und den gemessenen TDOA bestimmt und anschließend der Skalierungsfaktor so gewählt, dass die Differenz minimiert wird.

Die Leistungsfähigkeit der entwickelten Vorgehensweise hängt dabei einerseits von der Qualität der TDOA-Schätzungen und andererseits von den Ergebnissen der Geometriekalibrierung ab. Zudem dokumentieren die durchgeführten Untersuchungen, dass sich Fehler der Geometriekalibrierung viel stärker auswirken als eine ungenaue TDOA-Schätzung. Da die bereits von der Geometriekalibrierung stammenden Fehler durch die Schätzung des Skalierungsfaktors nicht mehr kompensiert werden können, wurde zur Bewertung der Ergebnisse die Zunahme des Positionierungsfehlers durch die TDOA-Skalierung gegenüber einer Skalierung durch ein Orakel betrachtet. Ohne Nachhall verursacht die TDOA-Skalierung nur einen geringen Anstieg des mittleren Positionierungsfehlers von 0,03 m auf 0,04 m. Bei Nachhallzeiten von 0,2 s bzw. 0,4 s liegt der Fehler der Geometriekalibrierung bereits im Bereich von ca. 0,10 m und die Skalierung vergrößert den Fehler um weitere 0,10 m.

Als Alternative zur Fixierung der Skalierung anhand von Inter-Array-TDOA-Messungen wurde außerdem eine Strategie entwickelt, die nur Einfallswinkel verwendet und deshalb keine Abtastsynchronisation zwischen den Sensorknoten erfordert. Um trotzdem Informationen über die Skalierung zu erlangen, wurden Arrays berücksichtigt, die über mehr als zwei Mikrofone verfügen. Die zusätzlichen Mikrofone der Sensorknoten wurden jedoch nicht zur Steigerung der Präzision einer gemeinsamen Winkelschätzung genutzt, sondern zur Gewinnung von mehreren Winkelschätzungen pro Sensorknoten eingesetzt. Aufgrund der bekannten Abstände der Teilarrays eines Sensorknotens, gestatten die individuellen Winkelschätzungen eine Triangulation der Ereignisposition. Darüber hinaus lässt sich die Position jedes Teilarrays relativ zur Position und Orientierung des Sensorknotens ausdrücken und gestattet dadurch die Formulierung eines Geometriekalibrierungsalgorithmus, der keine Skalierungsinvarianz mehr besitzt.

Bei der Untersuchung dieses Algorithmus traten jedoch erhebliche Positionierungsfehler auf. Anhand des relativen Fehlers, der die Skalierung der Sensorkonfiguration unberücksichtigt lässt, zeigte sich zudem, dass nur eine geringe Abweichung bei der Sensoranordnung vorlag und dementsprechend hauptsächlich ein Skalierungsfehler vorhanden war. Auslöser für den großen Skalierungsfehler sind die Abstände zwischen den Teilarrays, die im Vergleich zum vorhandenen Winkelfehler viel zu klein ausfallen, um eine zuverlässige Triangulation und somit eine präzise Skalierung zu gewährleisten.

8 Audio-visuelle Geometriekalibrierung

Das in den zurückliegenden Kapiteln dieser Arbeit weiterentwickelte Einfallswinkelverfahren und dessen Einbettung in den RANSAC, haben insgesamt ein robustes Verfahren zur Geometriekalibrierung akustischer Sensornetze entstehen lassen. Allerdings stellen die im letzten Kapitel entwickelten Konzepte noch keine zufriedenstellende Lösung dar, um auch die Skalierung automatisch zu bestimmen. Als Alternative werden in diesem Kapitel deshalb audio-visuelle Kalibrierungsansätze konzipiert und bewertet.

Den Ausgangspunkt für die folgenden Ausführungen zum Entwurf audio-visueller Lösungen bildet ein Telekonferenzsystem, das neben akustischen Sensoren auch über mehrere Kameras verfügt. Damit nach Abschluss der Geometriekalibrierung bspw. die akustische Lokalisierung einer Person zur Auswahl einer geeigneten Kamera genutzt werden kann, müssen die Sensorpositionen beider Modalitäten in einem gemeinsamen Koordinatensystem vorliegen. Das bisher entwickelte Kalibrierungsverfahren für akustische Sensornetze liefert jedoch, ebenso wie die Ansätze für visuelle Sensornetze [BD10], eine Beschreibung der Sensoren in einem modalitätsspezifischen Koordinatensystem. Mit den im Folgenden erläuterten audio-visuellen Konzepten werden daher zwei Ziele verfolgt: Einerseits sollen die zu entwickelnden Verfahren die Gewinnung der zur Skalierung notwendigen Informationen erlauben und andererseits eine Beschreibung aller Sensoren in einem modalitätsübergreifenden Koordinatensystem gestatten.

Das bisher vorgestellte erweiterte Einfallswinkelverfahren zur Kalibrierung eines ASN besitzt eine hohe Anpassungsfähigkeit, da es keine direkten Eigenschaften des akustischen Signals nutzt, sondern lediglich DOA-Messungen verwendet. Im visuellen Bereich sind solche Winkelschätzungen bspw. durch die Detektion der Kopf-Schulter-Partie eines Menschen [DT05] möglich. Somit führt die Verwendung von Winkeln zu einer Abstraktion der Sensordaten, die den Unterschied zwischen den akustischen und visuellen Sensoren verschwinden lässt. Infolgedessen entsteht durch die in Abschnitt 8.1 präsentierte Verallgemeinerung des bisher zur Kalibrierung eines ASN genutzten Einfallswinkelverfahrens ein Kalibrierungsalgorithmus für audio-visuelle Sensornetze.

Das Ziel der Ausweitung der Kalibrierung auf audio-visuelle Sensornetze besteht, wie zuvor erwähnt, ausschließlich in der Bestimmung eines modalitätsübergreifenden Koordinatensystems und der Fixierung der Skalierung der akustischen Modalität. Daher soll der Algorithmus keinesfalls eine Konkurrenz zu etablierten visuellen Kalibrierungsalgorithmen, wie z. B. [BD10] oder [KLB08], darstellen. Im Gegenteil, er setzt sogar voraus, dass die Positionen und Orientierungen der visuellen Sensoren bereits im Vorfeld kalibriert wurden, damit die zuvor genannten Ziele erreicht werden können.

Alternativ zu der in Abschnitt 8.1 präsentierten gemeinsamen Kalibrierung beider Modalitäten, stellt Abschnitt 8.2 eine Möglichkeit vor, die eine unabhängige Kalibrierung beider Modalitäten mit beliebigen Algorithmen ermöglicht. Dementsprechend lassen sich zur Bestimmung der Sensorpositionen auch an das jeweilige Sensornetz angepasste Algorithmen einsetzen, um so ggf. eine präzisere Kalibrierung zu erreichen. Sofern eine individuelle Kalibrierung der Modalitäten vorliegt, gestatten beide die Lokalisation von Ereignissen in modalitätsspezifischen Koordinatensystemen. Die Konstruktion eines modalitätsübergreifenden Koordinatensystems und die Schätzung eines ggf. vorhandenen Skalierungsfaktors erfolgt anschließend mithilfe einer Koordinatentransformation, die die Positionsschätzungen der Ereignisse aus beiden Modalitäten aufeinander abbildet.

Die Schätzung einer Koordinatentransformation ist wiederum nicht nur für den in Abschnitt 8.2 entwickelten Kalibrierungsansatz, sondern auch bei anderen Anwendungsfällen von zentraler Bedeutung. Zu diesen Anwendungsfällen gehören u. a. die Geometriekalibrierungsverfahren [Hen+09] und [Val+10b], aber auch das maschinellen Sehen [AHB87] oder die in Abschnitt 6.4 verwendeten Konzepte der statistischen *Shape*-Analyse. Angesichts der vielfältigen Einsatzmöglichkeiten von Algorithmen zur Schätzung der Parameter einer Koordinatentransformation zeigt Abschnitt 8.3 zunächst eine konventionelle Methode, bevor im Anschluss die Konzepte der statistischen *Shape*-Analyse zur Entwicklung einer intuitiven und zugleich recheneffizient zu realisierenden Alternative führen.

8.1 Gemeinsame Geometriekalibrierung

Die Verwendung von Einfallswinkeln als Informationsquelle zur Durchführung der Geometriekalibrierung erlaubt eine Abstraktion von den konkreten Eigenschaften der genutzten Sensoren. Daher ermöglicht das Einfallswinkelverfahren die Kombination von unterschiedlichen Sensortypen und verschiedenen Modalitäten in einem gemeinsamen Kalibrierungsprozess. Lediglich die Extraktion der Einfallswinkel erfordert spezielle, an das Signal angepasste, Algorithmen. Somit stellt die Kalibrierung der visuellen Sensoren, in einem gemeinsamen Einfallswinkelverfahren mit den akustischen Sensoren, die konsequente Verallgemeinerung des bislang präsentierten Konzeptes dar.

Voraussetzung für eine gemeinsame Kalibrierung ist, dass sowohl die akustischen als auch die visuellen Sensoren die gleichen Ereignisse erfassen. Bei dem hier betrachteten Telekonferenzsystem ist diese Situation dann gegeben, wenn sich eine Person sprechend durch den Raum bewegt. Für die Schätzung der Einfallswinkel in der akustischen Modalität wird auf die aus Kapitel 4 bekannten Techniken zurückgegriffen. Im visuellen Bereich lässt sich die Winkelschätzung durch eine Personendetektion realisieren [DT05].

Angesichts der verwendeten Abstraktion der Sensorinformationen zu Einfallswinkeln, kann die geometrische Beziehung aus Gl. (5.16) nicht nur für die akustischen, sondern auch für die visuellen Sensoren formuliert werden. Mithilfe des Einfallswinkels $\vartheta_{c,d}$ der c-ten Kamera, deren Position und Orientierung durch $\boldsymbol{v}_c$ bzw. δ_c gegeben sind, entsteht somit der Zusammenhang

$$f_{\mathrm{PA}}^{\mathrm{V}}(\boldsymbol{e}_d; \boldsymbol{v}_i, \delta_i, \vartheta_{i,d}) = \frac{(\boldsymbol{e}_d - \boldsymbol{v}_c)^{\mathrm{T}}}{\|\boldsymbol{e}_d - \boldsymbol{v}_c\|_2} \boldsymbol{R}_{\mathrm{xy}}(\delta_c) \begin{bmatrix} \cos(\vartheta_{c,d}) \\ \sin(\vartheta_{c,d}) \end{bmatrix}. \tag{8.1}$$

Aufgrund der als bekannt vorausgesetzten Positionen und Orientierungen der Kameras, besteht der Unterschied zu der Formulierung für die akustischen Sensoren lediglich darin, dass jetzt ausschließlich die Ereignispositionen unbekannt sind, während es in Gl. (5.16) zusätzlich die Mikrofonpositionen und -orientierungen zu ermitteln galt. Analog zu der im akustischen Umfeld verwendeten Notation werden die Positionen und Orientierungen aller Kameras zu den Matrizen $\boldsymbol{V} = [\boldsymbol{v}_1 \ \ldots \ \boldsymbol{v}_C]$ bzw. $\boldsymbol{\delta} = [\delta_1 \ \ldots \ \delta_C]$ zusammengefasst. Gemeinsam mit der Matrix $\boldsymbol{\Theta}$, die alle visuellen Winkelschätzungen beinhaltet, lässt sich letztendlich ein modalitätsübergreifendes Gleichungssystem

$$\boldsymbol{f}_{\mathrm{PA}}^{\mathrm{AV}}(\boldsymbol{\Lambda};\boldsymbol{\Phi},\boldsymbol{\Theta},\boldsymbol{V},\boldsymbol{\delta}) = \left[\boldsymbol{f}_{\mathrm{PA}}(\boldsymbol{\Lambda};\boldsymbol{\Phi}) \quad f_{\mathrm{PA}}^{\mathrm{V}}(\boldsymbol{e}_1;\boldsymbol{v}_1,\delta_1,\vartheta_{1,1}) \quad \ldots \quad f_{\mathrm{PA}}^{\mathrm{V}}(\boldsymbol{e}_D;\boldsymbol{v}_C,\delta_C,\vartheta_{C,D})\right]^{\mathrm{T}} \tag{8.2}$$

zusammenstellen, das sich ebenfalls mit dem Newton-Verfahren lösen lässt.

Die von beiden Modalitäten erfassten Ereignisse gestatten es daraufhin, einen Zusammenhang zwischen dem akustischen und visuellen Teil des Gleichungssystems herzustellen. Weiterhin legen die bekannten Positionen der visuellen Sensoren das Referenzkoordinatensystem und die Skalierung der geometrischen Anordnung fest. Nach Abschluss der Kalibrierung liegen deshalb einerseits die Positionen und Orientierungen der Sensoren beider Modalitäten im Koordinatensystem der visuellen Sensoren vor und andererseits entfällt die bisher auftretende Skalierungsinvarianz. Daher ermöglicht das zur Kalibrierung akustischer Sensornetze entwickelte Framework unmittelbar die Kalibrierung audio-visueller Sensornetze, ohne dabei zusätzliche Informationen zur Bestimmung der Skalierung zu erfordern.

Die zurückliegenden Betrachtungen beschränken sich bislang auf die Zusammenführung der Einfallswinkel der akustischen und visuellen Modalität in einem gemeinsamen Einfallswinkelverfahren. Dieses wiederum kann, analog zu Kapitel 6, ebenfalls mit dem RANSAC kombiniert werden, um eine störungsrobuste Kalibrierung zu gewährleisten.

Trotz der Abstraktion der Sensorinformationen zu Einfallswinkeln ergeben sich bei einer modalitätsübergreifenden Kalibrierung weitere Punkte, die es zu berücksichtigen gilt. Dies betrifft insbesondere die Verknüpfung der Ereignisse aus beiden Modalitäten. Dazu muss zunächst eine Ereignisquelle, z. B. eine Person, vorhanden sein, die sowohl durch die akustischen als auch die visuellen Sensoren erfasst werden kann. Allerdings darf sich nur eine Quelle im Raum befinden, damit eine Zuordnung der Beobachtungen zwischen den Modalitäten gegeben ist. Außerdem erfordert die Zuordnung der Winkelschätzungen aller Sensoren zu einem Ereignis eine Synchronisation der Sensorknoten. Im Gegensatz zu der in Abschnitt 7.1 betrachteten Skalierung anhand von Inter-Array-TDOA-Messungen, die eine arrayübergreifende Abtastsynchronisation erfordert, fallen die Anforderungen an die Synchronisation jetzt deutlich geringer aus. Für die aktuell betrachtete gemeinsame Geometriekalibrierung ist es ausreichend, wenn eine zeitliche Zuordnung der Winkelschätzungen aus den beiden Modalitäten möglich ist. Da die zeitliche Auflösung der Winkelschätzungen im Vergleich zur Abtastrate der akustischen Signale bzw. zur Bildwiederholfrequenz der Kameras viel kleiner ausfällt, reicht eine ungefähre Synchronisation zwischen den akustischen und den visuellen Sensoren aus.

Darüber hinaus spielt der Erfassungsbereich der Sensoren eine entscheidende Rolle, da die zur Kalibrierung genutzte geometrische Beziehung durch die von den Sensoren gemeinsam erfassten Ereignisse hergestellt wird. Die akustischen Sensorknoten bieten

schon bei Einsatz von nur zwei Mikrofonen einen Erfassungsbereich von 180°. Durch die Nutzung des präferierten dreieckigen Mikrofonarrays ist sogar eine vollständige Rundumsicht (360°) gegeben. Eine Kamera verfügt hingegen über ein deutlich eingeschränkteres Sichtfeld. Im akademischen Umfeld eingesetzte Kameralösungen, wie z. B. in [PF14a], besitzen ein horizontales bzw. vertikales Sichtfeld von lediglich $45° \times 35°$. Diese sehr deutliche Einschränkung des Sichtfeldes führt dazu, dass die Erfassung eines Ereignisses durch alle Sensoren nur noch für Teilbereiche des Raumes gegeben ist.

Grundsätzlich reicht es aus, wenn jedes Ereignis von mindestens drei Sensoren detektiert wird, damit es mehr Informationen zum Gleichungssystem beiträgt, als es Unbekannte verursacht. Zur Fixierung der Skalierung durch die visuellen Sensoren ist zudem eine Erfassung jedes Ereignisses durch mindestens zwei Kameras notwendig. Um eine möglichst präzise Lokalisierung zu erzielen, sollten jedoch mehr als zwei Kameras zur Positionsschätzung beitragen. Insgesamt verursachen diese Randbedingungen eine massive Beschränkung des Raumes, in dem sich zur Kalibrierung geeignete Ereignisse befinden dürfen. Gleichzeitig verlangt das Einfallswinkelverfahren aber eine ausreichende räumliche Diversität der Ereignispositionen. Daher spielt die Positionierung der Kameras eine entscheidende Rolle für die Erfolgsaussichten des zuvor dargelegten Ansatzes.

Die Fläche eines Raumes, die zur Kalibrierung zur Verfügung steht, hängt letztendlich von drei Faktoren ab: Dem Sichtfeld, der Anzahl und der Ausrichtung der Kameras. Eine Einschätzung wie groß diese Fläche ist, soll deshalb mithilfe der nachfolgenden Simulation ermöglicht werden. Dazu wird auf die bereits aus Abschnitt 5.4 bekannten Szenarien zurückgegriffen. In jedem dieser Räume werden zufällig vier Kameras hinzugefügt. Die Kameras sind jeweils an den Wänden des Raumes platziert und stets so ausgerichtet, dass das Sichtfeld durch die Wände nicht zusätzlich eingeschränkt wird. Zur Veranschaulichung, welche Flächenanteile durchschnittlich durch wie viele Kameras erfasst werden, dient Abb. 8.1.

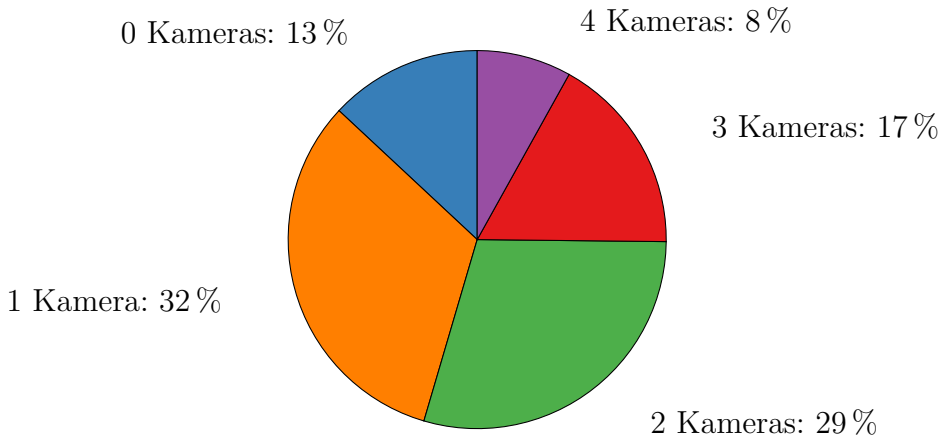


Abbildung 8.1: Durchschnittlich erfasste Fläche des Raumes.

Die Zahlen belegen unmittelbar, dass die angestrebte Erfassung der Ereignisse durch drei oder mehr Kameras durchschnittlich nur für ca. 25 % der Fläche gegeben ist. Außerdem stehen für die Kalibrierung im Mittel nur 54 % der Gesamtfläche zur Verfügung, da hier eine Erfassung durch mindestens zwei Kameras vorliegt.

Für eine Evaluierung, ob unter diesen Voraussetzungen eine zuverlässige Geometriekalibrierung möglich ist, werden zunächst DOA-Schätzungen von beiden Modalitäten benötigt. Zur Modellierung des Fehlers im akustischen Bereich, wird erneut auf das aus Abschnitt 5.6 bekannte Modell zurückgegriffen. Im Folgenden soll daher nur die Entwicklung eines vergleichbaren Modells für die Kameras betrachtet werden.

Wie bereits erwähnt, lässt sich eine visuelle Einfallswinkelschätzung durch eine Personendetektion realisieren. Dazu wird das Histogramm der Gradientenorientierungen (engl. *histogram of oriented gradient* (HOG)) berechnet und mit einer *Support Vector Machine* (SVM) als Klassifikator eine Erkennung der Kopf-Schulter-Partie durchgeführt [DT05]. Zusammen mit den intrinsischen Kameraparametern ergibt sich aus der Lage der Kopf-Schulter-Partie innerhalb des Bildes der Einfallswinkel [FP03]. Die zum Training der SVM genutzten Bilder stammen, wie auch in [DT05], aus der *INRIA Person Database* [Dal06], sodass das eigene Training vergleichbare Ereignisse liefert.

Die Entwicklung eines Fehlermodells der visuellen Einfallswinkelschätzung setzt außerdem Bildmaterial voraus, zu dem auch eine Annotation vorliegt, die angibt, ob und wo sich eine Person im Bild befindet. Im Rahmen dieser Arbeit wird dazu auf die Videosequenzen *seq01-1p-0000* und *seq15-1p-0100* der Datenbank AV16.3 [LOG05] zurückgegriffen. Diese zeigen jeweils eine Person aus drei Kameraperspektiven und besitzen eine Gesamtlänge von ca. 12,5 min bei einer Auflösung von 360×288 Pixeln. Die Anwendung der zuvor erläuterten Personendetektion auf diesen Videos zeigt einen maximalen Fehler von $\pm 16°$. Die Standardabweichung beträgt ca. 1,1°. Zur Modellierung des visuellen Detektionsfehlers wird daher, analog zu Abschnitt 5.6, ein Histogramm aus den Trainingsdaten ermittelt.

Die Durchführung einer gemeinsamen audio-visuellen Geometriekalibrierung basierend auf den Daten des zuvor erläuterten visuellen und dem bereits bekannten akustischen Fehlermodell liefert die in Abb. 8.2 dargestellten Resultate. Dabei sind sowohl die Ergebnisse des konventionellen RANSAC als auch der partitionierten Variante (PRANSAC), jeweils für verschiedene Nachhallzeiten, visualisiert. Als Fehlermaß dient weiterhin der mittlere Positionierungsfehler. Allerdings definieren jetzt die Kamerapositionen das Koordinatensystem, sodass nun keine Koordinatentransformation mehr notwendig ist. Für die Details der Berechnung des Fehlers (ε_{P}) sei erneut auf Abschnitt 9.1 verwiesen.

Die Simulationsergebnisse belegen, dass es trotz des eingeschränkten Erfassungsbereiches der Kameras möglich ist, eine gemeinsame Kalibrierung der akustischen und visuellen Sensoren durch das Einfallswinkelverfahren zu realisieren. Weiterhin liegen aufgrund der bekannten Positionen der Kameras die Positionen aller Sensoren in einem gemeinsamen Koordinatensystem vor und die Skalierungsinvarianz ist ebenfalls nicht mehr vorhanden. Die deutlichere Reduktion des Kalibrierungsfehlers durch den Einsatz des PRANSAC schon bei einer Nachhallzeit von 0,0 s lässt sich darauf zurückführen, dass in der visuellen Modalität bereits eine signifikante Störung vorliegt, wohingegen die akustische Modalität kaum Störungen aufweist. Daher tritt der schon bei den Analysen in Abschnitt 6.6 festgestellte Gewinn auch ohne Nachhall ein und ist sogar deutlich stärker ausgeprägt als im akustischen Umfeld (vgl. Tab. 6.1). Insgesamt erfüllt die gemeinsame Kalibrierung die gesetzten Ziele. Der Fehler fällt jedoch etwas größer als bei der in Abschnitt 7.1 betrachteten Fixierung der Skalierung durch Inter-Array-TDOA-Messungen aus. Dafür erfordert die gemeinsame Kalibrierung im Gegensatz zur Inter-Array-TDOA-Variante aber auch keine arrayübergreifende Abtastsynchronisation.

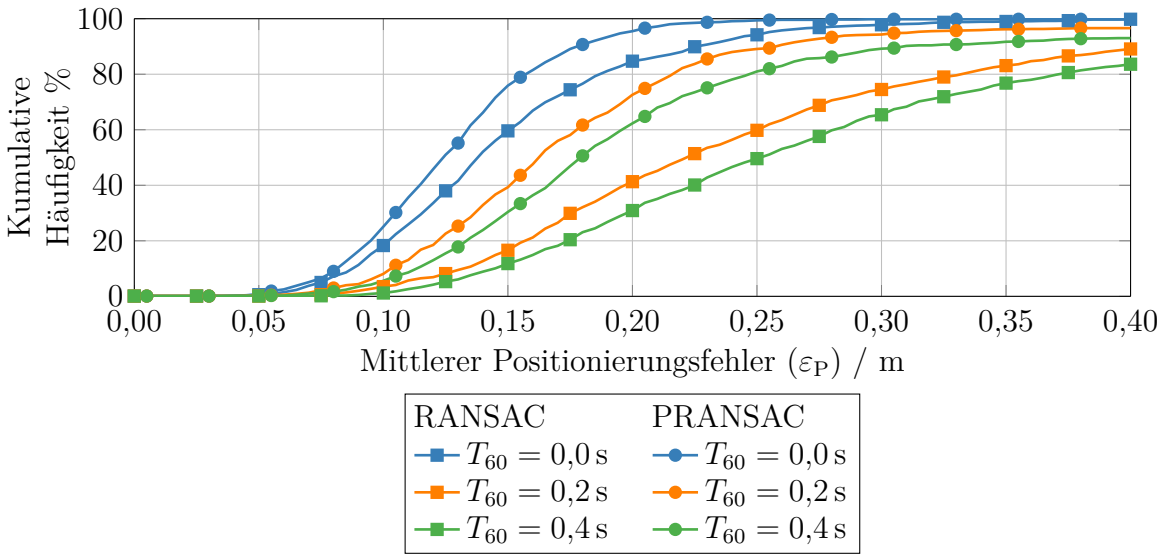


Abbildung 8.2: Kumulative Häufigkeitsverteilung des mittleren Positionierungsfehlers der gemeinsamen audio-visuellen Geometriekalibrierung bei verschiedenen Nachhallzeiten.

8.2 Audio-Video-Abbildung

Die Abstraktion der Sensorinformationen zu Einfallswinkeln ist die Grundlage für die im vorherigen Abschnitt erläuterte modalitätsübergreifende Kalibrierung durch ein gemeinsames Gleichungssystem. Allerdings sorgt die Abstraktion der Sensorinformationen ggf. für einen Informationsverlust, weil die Ausnutzung modalitätsspezifischer Eigenschaften entfällt. Deshalb wird zusätzlich eine alternative Verfahrensweise entworfen. Diese soll ebenfalls die Positionen der akustischen Sensoren im Koordinatensystem des bereits kalibrierten Kameranetzwerks liefern, die Skalierung festlegen und gleichzeitig über die Flexibilität verfügen, bei der Kalibrierung sensorspezifische Informationen ausnutzen zu können.

Damit bei der Kalibrierung des akustischen bzw. des visuellen Sensornetzes sensorspezifische Eigenschaften miteinfließen können, ist die Verwendung von Kalibrierungsverfahren, die an die jeweilige Modalität angepasst sind, notwendig. Andererseits liegen durch die getrennte Kalibrierung der Modalitäten die Positionsbeschreibungen in zwei unabhängigen Koordinatensystemen vor, da die Angabe der Sensorpositionen zumeist relativ zu einem Referenzsensor der jeweiligen Modalität erfolgt. Die Herausforderung besteht somit in der Entwicklung eines Systems, das die in separaten Koordinatensystemen vorliegenden akustischen bzw. visuellen Sensorkonfigurationen in einem gemeinsamen Koordinatensystem vereint. Dazu muss mithilfe einer Transformation eine Beziehung zwischen den Koordinatensystemen hergestellt werden, um daraus anschließend ein gemeinsames Koordinatensystem zu berechnen.

Die individuelle Kalibrierung der jeweiligen Sensorsysteme gestattet es, diese getrennt voneinander für weiterführende Aufgaben einzusetzen. Eine mögliche Anwendung stellt die Lokalisation und Verfolgung eines Ereignisses bzw. Sprechers dar. Sofern, wie auch

bei der gemeinsamen Geometriekalibrierung, eine Person vorhanden ist, die sich sprechend durch einen Raum bewegt, sind beide Modalitäten unabhängig voneinander in der Lage, die Trajektorie, entlang derer sich die Person bewegt, zu ermitteln. Dadurch, dass die individuell ermittelten Trajektorien von derselben Quelle stammen, gleicht die Ausgangssituation der der positionsbasierten Geometriekalibrierung (vgl. Abschnitt 2.2.3). Eine Schätzung der Koordinatentransformation (RBT) zur deckungsgleichen Abbildung der Trajektorien, liefert somit eine Transformation zwischen dem akustischen und dem visuellen Koordinatensystem. Diese ermöglicht anschließend die Abbildung der Sensorpositionen und Orientierungen in ein gemeinsames Koordinatensystem. Zur Veranschaulichung des erläuterten Vorgehens dient das in Abb. 8.3 dargestellte Blockdiagramm.

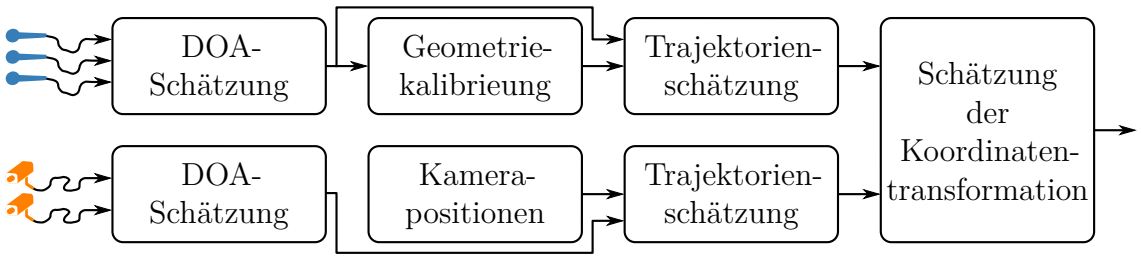


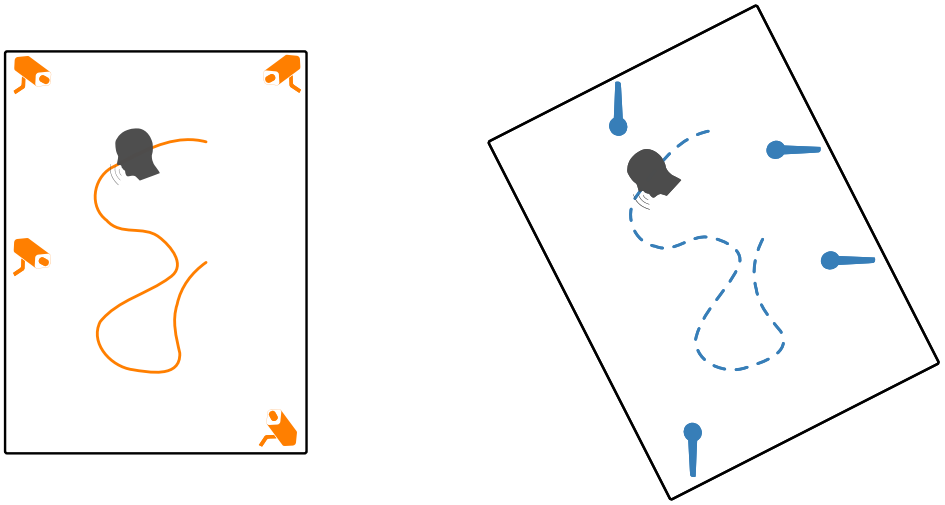
Abbildung 8.3: Schematischer Ablauf der Geometriekalibrierung durch Abbildung der akustischen auf die visuelle Trajektorie.

Ausgangspunkt für die Schätzung der Koordinatentransformation zur Bestimmung eines modalitätsübergreifenden Koordinatensystems bilden die Trajektorien der beteiligten Modalitäten. Zur Kalibrierung der akustischen Sensoren dient das modifizierte, in den RANSAC eingebettete Einfallswinkelverfahren (vgl. Kapitel 6). Bei den visuellen Sensoren wird, wie bereits im vorangegangenen Abschnitt, die Kenntnis der Positionen und Orientierungen vorausgesetzt. Für die Schätzung der benötigten Trajektorien existieren zahlreiche Ansätze. Dazu gehören u. a. Tracking-Algorithmen, wie z. B. [Eve+15] oder [LSM06], die speziell an die akustischen bzw. visuellen Gegebenheiten angepasst sind. Um einen direkten Vergleich zu den anderen Ansätzen dieser Arbeit zu ermöglichen, wird lediglich das bereits im Zuge des RANSAC eingesetzte schnittpunktgestützte Triangulationsverfahren zur Bestimmung der Positionen verwendet.

Eine Illustration der beschriebenen Ausgangssituation liefert Abb. 8.4. Sie zeigt die geometrische Anordnung der visuellen bzw. akustischen Sensoren in zwei getrennten Darstellungen, da die Koordinatensysteme der beiden Modalitäten aufgrund der individuellen Kalibrierung eine unbekannte Rotation und Translation aufweisen. Sofern die Kalibrierung des akustischen Teilsystems ausschließlich unter Verwendung von Einfallswinkeln erfolgt, ist zusätzlich auch die Skalierung zu ermitteln.

Aufgabe der gesuchten Koordinatentransformation ist es, die akustische Trajektorie, die sich aus den Positionen $\boldsymbol{e}_d^{\mathrm{A}}$, $d = 1 \ldots D$, zusammensetzt, bestmöglich auf die zugehörigen Positionen $\boldsymbol{e}_d^{\mathrm{V}}$, $d = 1 \ldots D$, der visuellen Trajektorie abzubilden. Die dazu notwendige RBT besteht aus der Rotationsmatrix $\boldsymbol{R}$, dem Translationsvektor $\boldsymbol{t}$, sowie dem optionalen Skalierungsfaktor ν:

$$\boldsymbol{e}_d^{\mathrm{V}} = \nu \cdot \boldsymbol{R}\, \boldsymbol{e}_d^{\mathrm{A}} + \boldsymbol{t}. \tag{8.3}$$



(a) Geometrische Anordnung der visuellen Sensoren

(b) Geometrische Anordnung der akustischen Sensoren

Abbildung 8.4: Ausgangssituation vor der Koordinatentransformation: Die akustischen und visuellen Sensorkonstellationen sowie deren zugehörigen Trajektorien befinden sich in zwei getrennten Koordinatensystemen.

Die im Sinne des euklidischen Abstandes optimalen RBT-Parameter liefert die Minimierung von

$$\langle \hat{\nu}, \hat{\boldsymbol{R}}, \hat{\boldsymbol{t}} \rangle = \underset{\nu, \boldsymbol{R}, \boldsymbol{t}}{\operatorname{argmin}} \sum_{d=1}^{D} \left\| \boldsymbol{e}_d^{\mathrm{V}} - \nu \cdot \boldsymbol{R}\, \boldsymbol{e}_d^{\mathrm{A}} - \boldsymbol{t} \right\|_2^2 . \tag{8.4}$$

Voraussetzung für die Bestimmung der RBT-Parameter ist jedoch wie bei der gemeinsamen Geometriekalibrierung (vgl. Abschnitt 8.1), dass eine zeitliche Zuordnung der Messungen aus beiden Modalitäten gegeben ist. Dementsprechend erfordert auch die Abbildung der Trajektorien eine Synchronisation. Allerdings wird bei der hier betrachteten Abbildung der Trajektorien ebenfalls keine Abtastsynchronisation benötigt, sondern eine ungefähre Synchronisation, die eine Zuordnung der Positionsschätzungen aus den beiden Modalitäten zu einem Ereignis gestattet, reicht aus.

Sofern eine Zuordnung zwischen den Positionsschätzungen aus beiden Modalitäten gegeben ist, existieren unterschiedliche Verfahrensweisen zur Minimierung von Gl. (8.4). Diese liefern jedoch abgesehen von numerischen Unterschieden dieselben RBT-Parameter [ELF97]. Daher wird an dieser Stelle zunächst nicht näher auf die Verfahren zur Bestimmung der RBT-Parameter eingegangen, sondern auf Abschnitt 8.3 verwiesen, da sich dieser gesondert mit der Schätzung der RBT-Parameter beschäftigt.

Sobald die RBT-Parameter vorliegen, ermöglichen sie die Abbildung der Trajektorie vom akustischen in das visuelle Koordinatensystem. Gleichzeitig gestatten sie auch die Transformation der Positionen und Ausrichtungen der akustischen Sensoren in das visuelle Koordinatensystem. Dazu werden in Gl. (8.3) anstatt der Ereignispositionen die Sensorpositionen bzw. -orientierungen verwendet, sodass letztendlich eine Beschreibung aller Sensoren in einem gemeinsamen Koordinatensystem entsteht (vgl. Abb. 8.5).

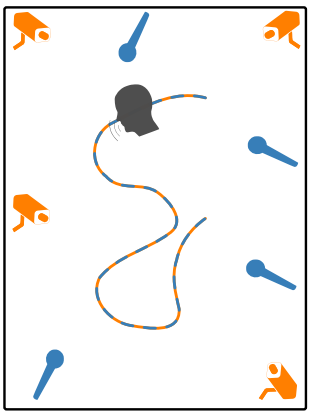


Abbildung 8.5: Nach Anwendung der Koordinatentransformation liegen die Positionen der akustischen Sensoren sowie die zugehörige Trajektorie im visuellen Koordinatensystem vor.

Die bisher skizzierte Idee, die Trajektorie eines Sprechers zunächst durch zwei separate Sensorsysteme zu erfassen, um damit anschließend mithilfe einer Koordinatentransformation ein gemeinsames Koordinatensystem zu erhalten, kommt in [Hen+09] ebenfalls zum Einsatz. Weiterhin gewährleistet dort die Nutzung eines RANSAC eine robuste Bestimmung der RBT-Parameter. Da die bereits mehrfach thematisierten Ausreißer bei der Einfallswinkelschätzung Ausreißer bei den daraus ermittelten Positionsschätzungen zur Folge haben, wird auch die Schätzung der Transformation zwischen dem akustischen und visuellen Koordinatensystem in einen RANSAC eingebettet. Zumal der PRANSAC bei der gemeinsamen Kalibrierung eine deutliche Reduktion des Kalibrierungsfehlers erzielt, wird hier unmittelbar die partitionierte Variante genutzt.

Der grundsätzliche Ablauf des PRANSAC entspricht weitgehend dem in Kapitel 6 geschilderten Vorgehen. Ein Unterschied besteht darin, dass es sich bei den Beobachtungen nun um Punkte der Trajektorie handelt, aus denen eine Schätzung der RBT-Parameter erfolgt. Außerdem ist ein anderes Kriterium zur Bewertung der Modellparameter erforderlich. Angelehnt an [Hen+09], wird hier der Abstand zwischen der akustischen und visuellen Trajektorie nach Anwendung der ermittelten RBT verwendet.

Um die Leistungsfähigkeit des entwickelten Trajektorien-Abbildung bewerten zu können und gleichzeitig einen Vergleich mit der Kalibrierung durch ein modalitätsübergreifendes Gleichungssystem zu gestatten, kommen die schon in Abschnitt 8.1 verwendeten Szenarien zum Einsatz. Zuerst erfolgt eine Kalibrierung der akustischen Sensoren mit dem in den RANSAC eingebetteten Einfallswinkelverfahren. Danach dienen sowohl das akustische Sensornetz als auch die Kameras zur Lokalisation der Ereignisse. Die dabei ermittelten Ereignispositionen ermöglichen anschließend die Bestimmung der RBT-Parameter. Die Ergebnisse der beschriebenen Analyse sind in Abb. 8.6 dargestellt. Diese Abbildung beinhaltet neben den Ergebnissen bei einer Schätzung von Rotation, Translation und Skalierung (RBT-Skalierung) auch die Ergebnisse, wenn ein Orakel den bestmöglichen Skalierungsfaktor liefert (Orakel-Skalierung).

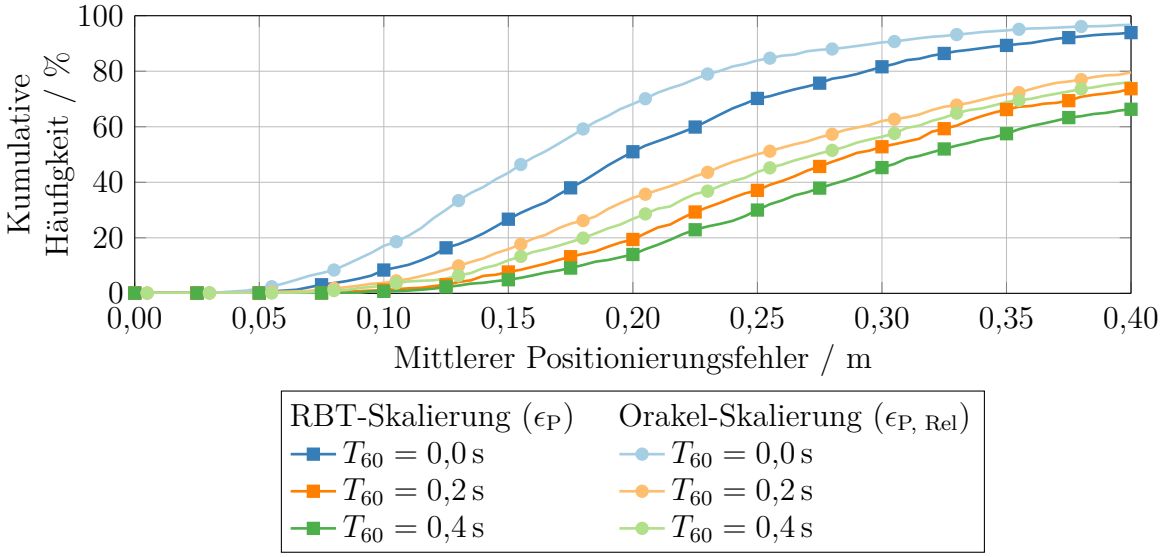


Abbildung 8.6: Kumulative Häufigkeitsverteilung des mittleren Positionierungsfehlers bei der Geometriekalibrierung durch eine Abbildung der akustischen auf die visuellen Positionsschätzungen bei verschiedenen Nachhallzeiten.

Der Vergleich der Ergebnisse mit denen der gemeinsamen Geometriekalibrierung (siehe Abb. 8.2) zeigt, dass der Koordinatentransformationsansatz der gemeinsamen Geometriekalibrierung unterlegen ist. Für das schlechtere Abschneiden der Koordinatentransformation sind zwei Gründe zu nennen. Einerseits hat ein fehlerhafter Skalierungsfaktor deutlichen Einfluss auf den Kalibrierungsfehler, wie der Vergleich der RBT-Skalierung mit dem Orakelexperiment belegt. Andererseits besitzen die zur Schätzung der Koordinatentransformation genutzten akustischen Positionsschätzungen einen beträchtlichen Fehler, da sich diese aus den Positionen und Orientierungen der akustischen Geometriekalibrierung ergeben, die selbst einen Fehler aufweisen. Die mit dem RANSAC kombinierte RBT-Parameterschätzung kann zwar die Abweichung zwischen den akustischen und visuellen Positionsschätzungen minimieren, nicht aber den durch die akustische Geometriekalibrierung ausgelösten Fehler der akustischen Positionen kompensieren. Daher ist die Audio-Video-Abbildung nicht in der Lage, eine präzisere Kalibrierung als die gemeinsame Kalibrierung der Modalitäten zu erzielen.

8.3 Koordinatentransformationsparameterschätzung

Die Schätzung einer Koordinatentransformation oder Starrkörpertransformation (RBT) zur Erstellung einer Abbildung zwischen zwei Mengen von Datenpunkten, die in verschiedenen Koordinatensystemen vorliegen, ist eine Problemstellung, die in sehr vielen Forschungsbereichen auftritt. Neben den im Rahmen dieser Arbeit betrachteten Anwendungsbereichen der Audio-Video-Abbildung (vgl. Abschnitt 8.2) und Geometriefusion (siehe Abschnitt 6.5), gehört insbesondere das maschinelle Sehen zu den wichtigsten Anwendungsgebieten [AHB87]. Dort ermöglicht eine RBT bspw. den Abgleich von Objekten aus verschiedenen Kameraperspektiven. Darüber hinaus dient sie zur Steue-

rung von Greifarmen in der Robotik, erlaubt die Positionsbestimmung von Objekten in Bilddaten und findet auch in der Biologie zur Auswertung von Knochenformen Anwendung [DM98].

Aufgrund der Vielzahl von verschiedenen Bereichen, in denen die RBT zum Einsatz kommt, existieren ebenfalls zahlreiche Ansätze zur Schätzung ihrer Parameter. Aus der in [CMK03] präsentierten Analyse von vier wichtigen Algorithmen geht hervor, dass sich diese in der Qualität der gelieferten Parameter in realistischen Szenarien kaum unterscheiden. Dennoch ergibt sich aufgrund der numerischen Stabilität und der erforderlichen Rechenkomplexität eine leichte Präferenz bezüglich des SVD gestützten Ansatzes aus [AHB87]. Dieser Ansatz liefert jedoch ausschließlich Rotation und Translation und lässt eine ggf. vorhandene Skalierung unberücksichtigt. Im weiteren Verlauf wird deshalb stets die Weiterentwicklung aus [Cha95] berücksichtigt, die außerdem in der Lage ist, die Skalierung zu bestimmen.

Nach einer kurzen Erläuterung der Verfahrensweise aus [Cha95], die zur Bestimmung der RBT-Parameter eine SVD einsetzt, erfolgt die Entwicklung eines eigenen Verfahrens zur Schätzung der RBT-Parameter. Grundlage dafür bilden die Konzepte der statistischen *Shape*-Analyse. Die Transformation in einen *Shape*-Bereich verwendet normalerweise spezielle Transformationsmatrizen, wie z. B. die Helmert-Matrix [DM98]. Die weiteren Ausführungen zeigen eine alternative Möglichkeit, eine solche Transformation zu realisieren. Dabei kommt anstatt der Helmert-Matrix die DFT zum Einsatz. Der Vorteil besteht darin, dass die *Fast* Fourier *Transform* (FFT) eine recheneffiziente Implementierung der DFT gestattet und somit in der Lage ist, den Zeitbedarf für eine RBT-Parameterschätzung zu reduzieren. Insbesondere wenn die RBT-Parameterschätzung, wie bei der Anwendung innerhalb des RANSAC, sehr häufig erfolgt, stellt eine Verkürzung der Rechenzeit einen deutlichen Vorteil dar.

Zur Veranschaulichung der Algorithmen zur Bestimmung der RBT-Parameter dient die im vorherigen Abschnitt betrachtete Abbildung der Positionsschätzungen. Ausgangspunkt ist daher das Optimierungsproblem aus Gl. (8.4). Um die gesuchten Parameter zu ermitteln, nutzt der SVD-Ansatz [Cha95] die Dispersions- bzw. Kovarianzmatrix

$$\boldsymbol{\Gamma} = \frac{1}{D} \sum_{d=1}^{D} \left(\boldsymbol{e}_d^{\mathrm{V}} - \bar{\boldsymbol{e}}^{\mathrm{V}} \right) \left(\boldsymbol{e}_d^{\mathrm{A}} - \bar{\boldsymbol{e}}^{\mathrm{A}} \right)^{\mathrm{T}} . \tag{8.5}$$

Durch die Subtraktion der Schwerpunkte der akustischen und visuellen Koordinaten $\bar{\boldsymbol{e}}^{\mathrm{A}}$ bzw. $\bar{\boldsymbol{e}}^{\mathrm{V}}$, unterscheiden sich die beiden Datensätze nur noch durch eine Rotation und eine ggf. vorhandene Skalierung. Aus der SVD der Kovarianzmatrix

$$\boldsymbol{\Gamma} = \boldsymbol{\Upsilon}\, \boldsymbol{\Sigma}\, \boldsymbol{\Psi}^{\mathrm{T}} \tag{8.6}$$

lässt sich zunächst die Rotationsmatrix

$$\boldsymbol{R} = \boldsymbol{\Upsilon}\, \boldsymbol{\Psi}^{\mathrm{T}} \tag{8.7}$$

zur Abbildung der Positionen aus dem akustischen Koordinatensystem in das visuelle Koordinatensystem ermitteln. Zusammen mit der Varianz der akustischen Koordinaten $(\sigma^{\mathrm{A}})^2$ ergibt sich der Skalierungsfaktor zu

$$\nu = \frac{1}{(\sigma^{\mathrm{A}})^2} \operatorname{tr} \left(\boldsymbol{R}^{\mathrm{T}}\, \boldsymbol{\Gamma} \right) . \tag{8.8}$$

Falls hingegen eine Problemstellung vorliegt, bei der nur Rotation und Translation bestimmt werden sollen, entfällt die Berechnung des Skalierungsfaktors und dieser kann für die weiteren Schritte zu Eins gewählt werden.

Die Bestimmung der Translation kann nun unter Berücksichtigung der bereits ermittelten Rotationsmatrix und des Skalierungsfaktors erfolgen. Dazu wird zunächst der akustische Schwerpunkt $\bar{\boldsymbol{e}}^{\mathrm{A}}$ mithilfe dieser Parameter in das visuelle Koordinatensystem übertragen. Aus der verbleibenden Differenz ergibt sich die Translation

$$\boldsymbol{t} = \bar{\boldsymbol{e}}^{\mathrm{V}} - \nu \cdot \boldsymbol{R}\,\bar{\boldsymbol{e}}^{\mathrm{A}}. \tag{8.9}$$

Das Optimierungsproblem aus Gl. (8.4) bildet nicht nur den Ausgangspunkt für die zuvor erläuterte SVD-basierte Schätzung der RBT-Parameter, sondern auch für die aus dem Bereich der statistischen *Shape*-Analyse stammenden Konzepte. Die Gemeinsamkeit von allen Verfahren aus dem Bereich der statistischen *Shape*-Analyse besteht darin, dass zur Bestimmung der RBT-Parameter, eine Transformation der Formen in einen *Shape*-Bereich erfolgt. Anhand der dabei durchgeführten schrittweisen Entfernung von Rotation, Translation und Skalierung ergeben sich schließlich die gesuchten RBT-Parameter. Sofern es sich bei den Formen, wie bei den hier betrachteten Trajektorien, um planare Objekte handelt, lassen sich diese durch komplexwertige Landmarken (vgl. Abschnitt 6.5) beschreiben.

Eine Möglichkeit diese Formen bzw. die dazugehörigen komplexwertigen Landmarken in einen *Shape*-Bereich zu überführen, stellen die Kendal-Koordinaten dar [DM98]. Sie ergeben sich aus der Multiplikation der Formen mit der Sub-Helmert-Matrix. Diese wiederum geht aus der Helmert-Matrix durch die Entfernung der ersten Zeile hervor. Gemäß des in [DM98] näher erläuterten Schemas zur Erzeugung einer $\mathcal{K}$-dimensionalen Helmert-Matrix haben alle Elemente der ersten Zeile den Wert $1/\sqrt{\mathcal{K}}$. Die weiteren Zeilen bestehen aus

$$\begin{bmatrix} h_\ell & \dots & h_\ell & -\ell h_\ell & 0 & \dots & 0 \end{bmatrix}, \text{ mit } h_\ell = -\frac{1}{\sqrt{\ell(\ell-1)}}, \tag{8.10}$$

wobei ℓ den Index der Zeile bezeichnet. Der Koeffizient h_ℓ wird dabei jeweils $(\ell-1)$-mal wiederholt, gefolgt von $-\ell h_\ell$ sowie ggf. notwendigen Nullen. Dadurch ergibt sich eine Matrix mit folgenden Eigenschaften: Die Elemente der ersten Zeile besitzen den Wert $1/\sqrt{\mathcal{K}}$. Die weiteren Zeilen sind jeweils orthogonal zu allen anderen, summieren sich zu Null und haben einen Betrag von Eins. Für $\mathcal{K} = 4$ entsteht somit die Helmert-Matrix

$$\boldsymbol{H} = \begin{bmatrix} 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} & 0 \\ -1/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} & 3/\sqrt{12} \end{bmatrix}. \tag{8.11}$$

Die Multiplikation eines Vektors mit der Helmert-Matrix erlaubt es daher eine geometrische Dekomposition durchzuführen, sodass die erste Komponente des Multiplikationsergebnisses den Mittelpunkt bzw. geometrischen Schwerpunkt repräsentiert. Daher resultiert aus der Multiplikation mit der Sub-Helmert-Matrix eine mittelwertfreie Darstellung im sogenannten *Preshape*-Bereich. Dementsprechend verbleibt nach der

Transformation in den *Preshape*-Bereich zwischen den Formen nur noch eine Rotation und Skalierung.

Da Rotation und Skalierung nun unabhängig von der Translation gewonnen werden können, ermöglicht die Multiplikation mit der Sub-HELMERT-Matrix eine Entkoppelung des Optimierungsproblems aus Gl. (8.15) in zwei getrennte Teilprobleme. Dieses Prinzip bildet die Basis für die Entwicklung einer weiteren Strategie zur Schätzung der RBT-Parameter. Wie bereits zuvor erwähnt, werden die folgenden Ausführungen zeigen, dass anstelle der HELMERT-Matrix bzw. Sub-HELMERT-Matrix auch die DFT eingesetzt werden kann, weil diese ebenfalls die erforderlichen Eigenschaften besitzt und darüber hinaus recheneffizient zu implementieren ist.

Ausgangspunkt für die Entwicklung des eigenen Ansatzes zur Schätzung der RBT-Parameter sind ebenfalls die akustische und die visuelle Trajektorie $\boldsymbol{e}_d^{\mathrm{A}} = [a_d^{\mathrm{A}} \;\; b_d^{\mathrm{A}}]^{\mathrm{T}}$ bzw. $\boldsymbol{e}_d^{\mathrm{V}} = [a_d^{\mathrm{V}} \;\; b_d^{\mathrm{V}}]^{\mathrm{T}}$, $d = 1, \ldots, D$. Die Trajektorien werden allerdings, wie bei der statistischen *Shape*-Analyse für planare Formen üblich, zunächst durch komplexwertige Landmarken

$$u_d = a_d^{\mathrm{A}} + \mathrm{j} \cdot b_d^{\mathrm{A}} \;\; \text{bzw.} \;\; w_d = a_d^{\mathrm{V}} + \mathrm{j} \cdot b_d^{\mathrm{V}} \tag{8.12}$$

ausgedrückt. Aufgrund der komplexwertigen Notation lässt sich die Koordinatentransformation aus Gl. (8.3) in

$$w_d = \alpha \cdot u_d + \beta \tag{8.13}$$

überführen. Dementsprechend kann die Koordinatentransformation durch die Multiplikation mit dem komplexen Skalar α und einer Addition der ebenfalls komplexen Größe β realisiert werden. Weiterhin gestattet die Zusammenfassung aller Landmarken zu den Formen

$$\boldsymbol{u} = \begin{bmatrix} u_1 & \ldots & u_D \end{bmatrix}^{\mathrm{T}} \;\; \text{bzw.} \;\; \boldsymbol{w} = \begin{bmatrix} w_1 & \ldots & w_D \end{bmatrix}^{\mathrm{T}} \tag{8.14}$$

die Interpretation des Optimierungsproblems aus Gl. (8.3) als die Bestimmung derjenigen Transformationsparameter, die die volle PROKRUSTES-Distanz $\gamma(\boldsymbol{w}, \boldsymbol{u})$ (vgl. Gl. (6.9)) minimieren. Letztendlich gilt es daher, folgendes Optimierungsproblem zu lösen:

$$\langle \hat{\alpha}, \hat{\beta} \rangle = \underset{\alpha,\beta}{\operatorname{argmin}} \, \|\gamma(\boldsymbol{w}, \boldsymbol{u})\|_2^2 = \underset{\alpha,\beta}{\operatorname{argmin}} \, (\boldsymbol{w} - \alpha \cdot \boldsymbol{u} - \beta \cdot \boldsymbol{1})^{\mathrm{H}} \, (\boldsymbol{w} - \alpha \cdot \boldsymbol{u} - \beta \cdot \boldsymbol{1}) \, . \tag{8.15}$$

Ferner lässt sich die Berechnung einer DFT auch durch eine Matrixmultiplikation formulieren. Dabei enthält die Matrix $\boldsymbol{F}$ zur Transformation vom Definitionsbereich in den FOURIER-Bereich die komplexen Exponentialschwingungen. Für eine DFT mit $\mathscr{K} = 4$ Punkten entsteht somit die Matrix

$$\boldsymbol{F} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \mathrm{e}^{-\mathrm{j}2\pi 1/4} & \mathrm{e}^{-\mathrm{j}2\pi 2/4} & \mathrm{e}^{-\mathrm{j}2\pi 3/4} \\ 1 & \mathrm{e}^{-\mathrm{j}2\pi 2/4} & \mathrm{e}^{-\mathrm{j}2\pi 4/4} & \mathrm{e}^{-\mathrm{j}2\pi 6/4} \\ 1 & \mathrm{e}^{-\mathrm{j}2\pi 3/4} & \mathrm{e}^{-\mathrm{j}2\pi 6/4} & \mathrm{e}^{-\mathrm{j}2\pi 9/4} \end{bmatrix} . \tag{8.16}$$

Analog dazu lässt sich die IDFT durch die Multiplikation mit $\frac{1}{\mathscr{K}} \boldsymbol{F}^{\mathrm{H}}$ darstellen und das Produkt aus DFT- und IDFT-Matrix $\frac{1}{\mathscr{K}} \boldsymbol{F}^{\mathrm{H}} \boldsymbol{F}$ liefert eine $\mathscr{K}$-dimensionale Einheitsmatrix.

Unter Berücksichtigung dieser Zusammenhänge kann das zu lösende Optimierungsproblem aus Gl. (8.15) auch als

$$\langle \hat{\alpha}, \hat{\beta} \rangle = \underset{\alpha,\beta}{\operatorname{argmin}} \frac{1}{\mathscr{K}} \left(\boldsymbol{w} - \alpha \cdot \boldsymbol{u} - \beta \cdot \mathbf{1} \right)^{\mathrm{H}} \boldsymbol{F}^{\mathrm{H}} \boldsymbol{F} \left(\boldsymbol{w} - \alpha \cdot \boldsymbol{u} - \beta \cdot \mathbf{1} \right) \tag{8.17}$$

dargestellt werden. Weiterhin ist die Repräsentation der akustischen bzw. visuellen Trajektorie im Fourier-Bereich durch

$$\boldsymbol{u}_{\mathrm{F}} = \boldsymbol{F}\,\boldsymbol{u} \text{ bzw. } \boldsymbol{w}_{\mathrm{F}} = \boldsymbol{F}\,\boldsymbol{w} \tag{8.18}$$

gegeben. Die Einsen in der ersten Zeile von $\boldsymbol{F}$ sorgen zusammen mit der Orthogonalität der Zeilen außerdem dafür, dass das erste Element von $\boldsymbol{u}_{\mathrm{F}}$ bzw. $\boldsymbol{w}_{\mathrm{F}}$ den mit $\mathscr{K}$ multiplizierten geometrischen Schwerpunkt der jeweiligen Trajektorie beschreibt. Daher gestattet auch die DFT eine Entkoppelung des Optimierungsproblems aus Gl. (8.15) bzw. Gl. (8.17) in ein Optimierungsproblem für die Rotation und Skalierung:

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \frac{1}{\mathscr{K}} \left(\{\boldsymbol{w}_{\mathrm{F}}\}_{2:\mathscr{K}} - \alpha \cdot \{\boldsymbol{u}_{\mathrm{F}}\}_{2:\mathscr{K}} \right)^{\mathrm{H}} \left(\{\boldsymbol{w}_{\mathrm{F}}\}_{2:\mathscr{K}} - \alpha \cdot \{\boldsymbol{u}_{\mathrm{F}}\}_{2:\mathscr{K}} \right) \tag{8.19}$$

sowie ein weiteres für die Translation:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{\mathscr{K}} \left(\{\boldsymbol{w}_{\mathrm{F}}\}_{1} - \hat{\alpha} \cdot \{\boldsymbol{u}_{\mathrm{F}}\}_{1} - \beta \right)^{\mathrm{H}} \left(\{\boldsymbol{w}_{\mathrm{F}}\}_{1} - \hat{\alpha} \cdot \{\boldsymbol{u}_{\mathrm{F}}\}_{1} - \beta \right). \tag{8.20}$$

Dabei bezeichnet $\{\cdot\}_1$ die erste Komponente der Fourier-Transformation und $\{\cdot\}_{2:\mathscr{K}}$ analog dazu alle weiteren Komponenten. Insgesamt ergeben sich die gesuchten RBT-Parameter daher, gemäß der in Anhang C.1 näher ausgeführten Schritte, zu:

$$\hat{\alpha} = \frac{\left\{\boldsymbol{u}_{\mathrm{F}}^{\mathrm{H}}\right\}_{2:\mathscr{K}} \{\boldsymbol{w}_{\mathrm{F}}\}_{2:\mathscr{K}}}{\left\{\boldsymbol{u}_{\mathrm{F}}^{\mathrm{H}}\right\}_{2:\mathscr{K}} \{\boldsymbol{u}_{\mathrm{F}}\}_{2:\mathscr{K}}} \quad \text{und} \quad \hat{\beta} = \frac{1}{\mathscr{K}} \left(\{\boldsymbol{w}_{\mathrm{F}}\}_{1} - \hat{\alpha} \cdot \{\boldsymbol{u}_{\mathrm{F}}\}_{1} \right). \tag{8.21}$$

Die erzielte Lösung entspricht damit dem aus [DM98] bekannten Ergebnis, das sich bei der Verwendung von Kendal-Koordinaten ergibt. Der Unterschied zu dem hier entwickelten Verfahren liegt lediglich in der Realisierung der Koordinatentransformation. Während die Gewinnung der Kendal-Koordinaten eine Multiplikation mit der Sub-Helmert-Matrix erfordert, kann die Transformation jetzt mithilfe der FFT realisiert werden. Diese besitzt eine geringere Rechenkomplexität als die Multiplikation mit der Sub-Helmert-Matrix und ermöglicht daher eine effizientere Bestimmung der RBT-Parameter.

Im Umfeld der Geometriekalibrierung kommt jedoch vorrangig die zuvor erläuterte SVD-Variante zum Einsatz [Hen+09; SSP05; AHB87; Val+10b]. Da sowohl diese SVD-Variante als auch die *Shape*-Bereichsverfahren mittels DFT oder Kendal-Koordinaten, abgesehen von numerischen Abweichungen dieselben Ergebnisse liefern, befasst sich der verbleibende Teil dieses Abschnitts mit einem Vergleich der Rechenkomplexität.

Die in Anhang C.2 näher ausgeführten Betrachtungen der asymptotischen Laufzeit des SVD-Ansatzes zeigen, dass sowohl die Zerlegung der Kovarianzmatrix in die Singulärvektoren als auch die anschließende Berechnung der RBT-Parameter nicht von

der Anzahl der Elemente der vorliegenden Datenmenge abhängt. Die Menge der Datenpunkte beeinflusst lediglich die Berechnung der Kovarianzmatrix. Insgesamt beträgt die asymptotische Laufzeit daher $\mathcal{O}(D)$ und hängt damit linear von der Anzahl der Beobachtungen ab.

Bereits die FFT der vorgeschlagenen Strategie zur Bestimmung der RBT-Parameter im *Shape*-Bereich besitzt eine asymptotische Ausführungszeit von $\mathcal{O}(\mathcal{K}\log_2(\mathcal{K}))$ [FJ05], wobei $\mathcal{K}$ die DFT-Länge kennzeichnet, die im vorliegenden Fall mit der Anzahl der Beobachtungen D korrespondiert. Gemäß der ebenfalls in Anhang C.2 durchgeführten asymptotischen Laufzeitabschätzung dominiert die FFT die Gesamtlaufzeit. Dementsprechend hat die entwickelte Methode eine wesentlich größere asymptotische Laufzeit. Somit scheint die SVD-Variante zunächst die geeignetere Wahl zu sein. Es gilt jedoch zu berücksichtigen, dass es sich hier lediglich um das asymptotische Verhalten handelt, bei dem etwaige konstante Anteile ebenso wie sämtliche multiplikativen Faktoren, vernachlässigt wurden. Aufgrund dessen erfolgt ein Vergleich der beiden Alternativen durch eine Messung der tatsächlichen Ausführungszeiten.

Basis für die Laufzeitmessung der beiden Algorithmen ist eine dafür angefertigte C++-Implementierung beider Verfahren. Für die jeweiligen Kernkomponenten der Algorithmen wird allerdings auf existierende Bibliotheken zurückgegriffen. Zur Realisierung der FFT dient die *Fastest* Fourier *Transform in the West* (FFTW), die als eine der schnellsten *Open-Source*-Implementierungen gilt [FJ05]. Die zum Vergleich notwendige SVD stammt dabei aus der zu MATLAB® gehörenden Implementierung der *Basic Linear Algebra Subprograms* (BLAS).

Angelehnt an das bisher betrachtete Szenario der Geometriekalibrierung, werden für die Laufzeitmessung D Trajektorienpunkte in einem Raum der Größe $8{,}00 \times 6{,}00\,\mathrm{m}^2$ generiert. Die anschließende Schätzung der RBT-Parameter mit beiden Verfahren wird auf einem Laptop mit Ubuntu 14.04.1 LTS, Intel(R) Core(TM) i7-2620M CPU @ 2,70 GHz und 4 GB RAM durchgeführt. Abb. 8.7 stellt die Ausführungszeiten beider Algorithmen, in Abhängigkeit der Anzahl der vorliegenden Beobachtungen, gemittelt über 20 000 Simulationen dar.

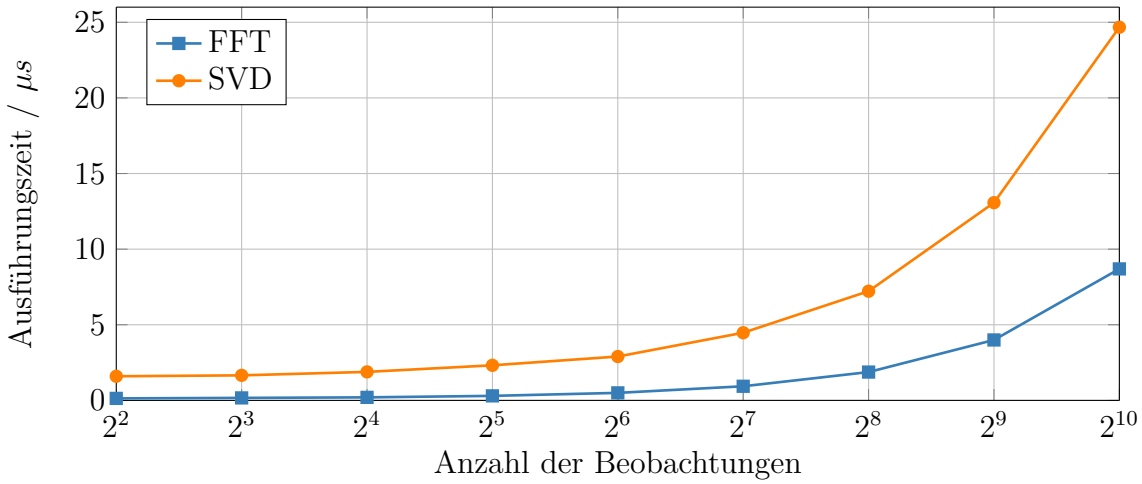


Abbildung 8.7: Vergleich der Ausführungszeiten von SVD- und FFT-basierter Schätzung der RBT-Parameter.

Die Simulationsergebnisse belegen unmittelbar, dass die Schätzung der RBT-Parameter unter Verwendung der FFT deutlich schneller als die konkurrierende Variante mittels SVD arbeitet. Für einen Vergleich zwischen den asymptotischen Laufzeiten und den visualisierten Ergebnissen gilt es zunächst, die logarithmische Darstellung der Anzahl der Beobachtungen in Abb. 8.7 zu berücksichtigen. Der exponentielle Anstieg der Ausführungszeit des SVD-Ansatzes in Abb. 8.7 stimmt somit mit der linearen Laufzeitabschätzung überein. Die Abschätzung des FFT-Ansatzes durch $\mathcal{O}(\mathcal{K}\log_2(\mathcal{K}))$ scheint unterdessen Abb. 8.7 zu widersprechen, da sich ein steilerer Anstieg als bei der SVD ergeben sollte. Dabei ist jedoch zu beachten, dass bei der asymptotischen Laufzeitabschätzung alle konstanten Anteile vernachlässigt wurden. Eine detaillierte Analyse der gemessenen Ausführungszeiten zeigt, dass diese auch für die FFT-Variante dem vorhergesagten Verlauf entsprechen, wenn der vernachlässigte Vorfaktor vor dem Ausdruck $\mathcal{K}\log_2(\mathcal{K})$ passend gewählt wird. Da dieser Faktor viel kleiner als bei der RBT-Parameterschätzung durch die SVD ausfällt, kompensiert er den vermeintlich steileren Verlauf von $\mathcal{K}\log_2(\mathcal{K})$. Insgesamt fällt daher die Ausführungszeit des entwickelten Verfahrens grundsätzlich kleiner als die der konventionellen Alternative aus. Die klare Reduktion der Ausführungszeiten bedeutet damit insbesondere einen Vorteil für die Nutzung der RBT-Parameterschätzung innerhalb eines RANSAC, da die Schätzung dort sehr häufig erfolgt.

8.4 Zusammenfassung

Während sich die bisherigen Kapitel auf die Geometriekalibrierung akustischer Sensornetze konzentriert haben, wurden in diesem Kapitel audio-visuelle Ansätze entwickelt, da bspw. beim Einsatz akustischer Sensornetze in einem Telekonferenzsystem, neben Mikrofonen auch Kameras zur Verfügung stehen. Ziel der modalitätsübergreifenden Kalibrierung ist es, nicht nur weitere Alternativen zur Fixierung der Skalierung zu entwickeln, sondern darüber hinaus eine Beschreibung aller Sensoren in einem gemeinsamen Koordinatensystem zu erhalten, um darauf aufbauend eine Verwendung der Sensoren bspw. zur audio-visuellen Lokalisation zu ermöglichen.

Als erstes wurde die gemeinsame Geometriekalibrierung akustischer und visueller Sensoren durch das Einfallswinkelverfahren betrachtet. Da dieses zur Kalibrierung lediglich Einfallswinkel verwendet und dementsprechend keine sensorspezifischen Information benötigt, gestattet es die Kalibrierung beliebiger Sensoren, sofern diese in der Lage sind Einfallswinkelschätzungen zu liefern. Voraussetzung ist jedoch, dass die Winkelschätzungen der verschiedenen Sensoren zu ein und demselben Ereignis gehören. Diese Bedingung ist allerdings unmittelbar erfüllt, weil im Rahmen dieser Arbeit ein Sprecher als Quelle dient und dieser sowohl akustisch als auch visuell erfasst werden kann. Zur visuellen Bestimmung des Winkels wurde die Position der Kopf-Schulter-Partie innerhalb des Kamerabildes mittels SVM und HOG detektiert. Zudem wurde bei der Kalibrierung die Kenntnis der Positionen und Orientierungen der visuellen Sensoren berücksichtigt, um daraus die notwendigen Informationen zur Fixierung der Skalierung zu gewinnen.

Ein wichtiger Aspekt bei der audio-visuellen Detektion der Ereignisse war außerdem der Erfassungsbereich der Sensoren. Während Mikrofonarrays eine vollständige Rundumsicht bieten, haben Kameras nur ein Sichtfeld von ca. 45°. Die Skalierung der

Geometrie anhand der visuellen Sensoren erfordert aber eine Erfassung der Ereignisse durch mindestens zwei Kameras. Daher sorgt das begrenzte Sichtfeld dafür, dass in den untersuchten Szenarien die zuvor genannte Bedingung nur für ca. die Hälfte der Fläche des Raumes erfüllt ist. Aufgrund der Einschränkung der Fläche, die zur Kalibrierung zur Verfügung steht, sinkt gleichzeitig auch die räumliche Diversität der Ereignispositionen und erschwert somit die Kalibrierung. Die durchgeführten Untersuchungen dokumentieren jedoch, dass das Einfallswinkelverfahren trotzdem eine verlässliche Kalibrierung audio-visueller Sensornetze erzielt.

Die Gegenüberstellung des Kalibrierungsfehlers bei der Verwendung von RANSAC und PRANSAC bestätigt zudem, dass die Berechnung des Mittelwertes verschiedener Lösungen eine deutliche Reduktion des Kalibrierungsfehlers gestattet. Ohne Nachhall liegt der Fehler durch die Nutzung des PRANSAC in mehr als 90 % der Untersuchungen unter 0,20 m. Durch Nachhall sinkt diese Quote zwar auf ca. 60 %, ohne PRANSAC beträgt die Quote jedoch nur noch 40 %. Im Vergleich zur rein akustischen Kalibrierung und der anschließenden Skalierung der Geometrie anhand von TDOA-Informationen (siehe Abschnitt 7.1) fällt der Kalibrierungsfehler zwar etwas größer aus, dafür ist aber auch keine arrayübergreifende Abtastsynchronisation mehr erforderlich.

Als Alternative zur Kalibrierung der akustischen und visuellen Sensoren in einem gemeinsamen Prozess wurde außerdem ein Ansatz vorgestellt, der eine Vereinigung der individuellen Kalibrierungen der jeweiligen Modalitäten ermöglicht. Dadurch wird zwar die Ausnutzung modalitätsspezifischer Eigenschaften möglich, andererseits liegen die Beschreibungen der beiden Modalitäten dann in unterschiedlichen Koordinatensystemen vor und müssen daher in ein gemeinsames Koordinatensystem abgebildet werden. Zur Schätzung der Koordinatentransformation wurde deshalb mit beiden Sensorsystemen zunächst eine Lokalisierung von Ereignissen durchgeführt und die daraus gewonnenen Ereignispositionen anschließend zur Bestimmung der RBT-Parameter verwendet. Obwohl die Schätzung der RBT-Parameter in einen RANSAC eingebettet wurde, fällt der Fehler im Vergleich zur Geometriekalibrierung durch ein modalitätsübergreifendes Gleichungssystem größer aus. Verantwortlich dafür ist einerseits die unpräzise Bestimmung des Skalierungsfaktors bei der Schätzung der Koordinatentransformationsparameter und andererseits der beträchtliche Fehler der Ereignispositionen, der auch durch die Nutzung des RANSAC nicht mehr ausgeglichen werden kann.

Im letzten Abschnitt dieses Kapitels wurde schließlich die Schätzung der Koordinatentransformationsparameter (RBT-Parameter) betrachtet. Dabei wurde zunächst ein konventionelles Verfahren vorgestellt, dass die gesuchten Parameter mithilfe einer SVD ermittelt. Im weiteren Verlauf dienten die Konzepte der statistischen *Shape*-Analyse zur Entwicklung eines eigenen Verfahrens. Die Gemeinsamkeit der im *Shape*-Bereich arbeitenden Verfahren besteht darin, dass diese die gesuchten Transformationsparameter durch die schrittweise Entfernung von Rotation, Translation und Skalierung ermitteln. Eine Möglichkeit dazu bildet die Darstellung in Kendal-Koordinaten, die sich durch die Multiplikation der Landmarken mit der Sub-Helmert-Matrix ergeben. Darüber hinaus besitzt auch die DFT, die bei der Darstellung in Kendal-Koordinaten ausgenutzten Eigenschaften. Dadurch lässt sich die Koordinatentransformation anstatt durch eine Matrixmultiplikation auch durch die FFT realisieren. Ein Vergleich der Rechenkomplexität des konventionellen SVD-Ansatzes und des entwickelten *Shape*-Bereichsansatzes mittels DFT/FFT zeigt eine durchschnittliche Reduktion der Ausführungszeit von 20 %.

9 Experimentelle Untersuchungen

In den zurückliegenden Kapiteln dieser Arbeit wurden verschiedene Algorithmen zur Geometriekalibrierung mithilfe von Einfallswinkeln entwickelt und analysiert. Allerdings beruht die Bewertung der Verfahren ausschließlich auf Simulationen. Das Ziel des aktuellen Kapitels besteht deshalb darin, die Leistungsfähigkeit der bislang konzipierten Strategien anhand von Experimenten in realen Umgebungen zu evaluieren. Erst die Ergebnisse dieser Experimente erlauben es, eine Entscheidung zu treffen, ob sich die entworfenen Algorithmen zur automatischen Kalibrierung eines akustischen Sensornetzes eignen und damit die Anforderungen dieser Arbeit (vgl. Abschnitt 2.3 bzw. 2.4) erfüllen. Dennoch gestatten die Simulationen einen Vergleich der vorliegenden Möglichkeiten und dienen somit zur Vorauswahl der Ansätze für die aktuelle Evaluierung.

Bei der Geometriekalibrierung eines ASN allein durch akustische Informationen erzielt die Skalierung durch TDOA-Messungen (vgl. Abschnitt 7.1) einen geringeren Fehler als die Skalierung mithilfe von Teilarrays (siehe Abschnitt 7.2). Im Bereich der audio-visuellen Kalibrierung liefert die gemeinsame Kalibrierung (siehe Abschnitt 8.1) bessere Ergebnisse als die Abbildung der Trajektorien (siehe Abschnitt 8.2). Für die Untersuchung in realen Szenarien werden nur die präferierten Möglichkeiten berücksichtigt.

Grundvoraussetzung für eine Bewertung der Experimente sind, wie auch schon bei den Simulationen, geeignete Maße zur Quantifizierung des Fehlers. Dazu dienen erneut die in den vergangenen Kapiteln eingesetzten, bislang aber noch nicht näher betrachteten, Fehlermaße. Abschnitt 9.1 erläutert daher zunächst die genutzten Bewertungskriterien. Im Zentrum dieses Kapitels stehen dagegen die Experimente in realen Umgebungen. Die Basis dafür bilden die beiden in Abschnitt 9.2 erläuterten Szenarien. Die Darstellung dieser Szenarien umfasst u. a. die Beschreibung der eingesetzten audio-visuellen Sensornetze, inklusive ihrer räumlichen Anordnung sowie eine Erläuterung der Aufnahmebedingungen. Anschließend befasst sich Abschnitt 9.3 mit der Gewinnung der zur Kalibrierung erforderlichen Daten aus den audio-visuellen Aufnahmen. Erst danach präsentiert Abschnitt 9.4 die daraus resultierenden Kalibrierungsergebnisse.

9.1 Bewertung der Kalibrierungsergebnisse

Die Definition eines adäquaten Maßes zur Beurteilung der ermittelten Sensorkonfigurationen verlangt zuerst die Betrachtung der Informationen, die nach dem Abschluss der Geometriekalibrierung vorliegen. Zu diesen Informationen gehören einerseits die Positionen der Sensorknoten, andererseits aber auch die Orientierungen, da jeder Sensorknoten nicht nur aus einem einzelnen Mikrofon, sondern einem Mikrofonarray besteht.

Die Angabe der Sensorpositionen und -orientierungen unterscheidet sich abhängig davon, ob eine akustische (vgl. Kapitel 5 bis 7) oder eine modalitätsübergreifende Kalibrierung (siehe Kapitel 8) durchgeführt wurde. Eine ausschließlich akustische Kalibrierung liefert eine Beschreibung der Sensorpositionen und -orientierungen relativ zueinander, ohne dass ein Zusammenhang zum Koordinatensystem des umgebenden Raumes oder eines anderen Bezugssystems vorliegt. Aus der audio-visuellen Kalibrierung resultiert hingegen eine Darstellung der akustischen Sensoren im Koordinatensystem der visuellen Sensoren. Das verwendete Fehlermaß muss somit in der Lage sein, in beiden Situationen eine Bewertung des Ergebnisses zu gestatten.

Verantwortlich für die relative Beschreibung der geometrischen Anordnung nach Abschluss der rein akustischen Geometriekalibrierung ist die willkürliche Festlegung des Koordinatenursprungs durch einen bzw. zwei Sensoren (siehe Abschnitt 5.3). Demzufolge erfordert die Bewertung eines Kalibrierungsergebnisses die Transformation des Ergebnisses in das Referenzkoordinatensystem, in dem die tatsächlichen Sensorpositionen bzw. -ausrichtungen vorliegen. Dazu wird, wie im Umfeld der Geometriekalibrierung üblich, eine RBT eingesetzt, die eine Rotation und Translation der Sensoranordnung ermöglicht [GKH13; Hen+09]. Sofern für die in Kapitel 5 bzw. Kapitel 6 betrachteten Algorithmen a priori kein Wissen zur Fixierung der Skalierung vorhanden ist, muss zusätzlich auch die Skalierung angepasst werden.

Die RBT-Parameter zur Abbildung der durch die Geometriekalibrierung gewonnenen Sensorpositionen $\widehat{\boldsymbol{s}}_i$ und -orientierungen $\widehat{\theta}_i$ in das Referenzkoordinatensystem, in dem die tatsächliche Sensorkonfiguration vorliegt, lassen sich mithilfe der aus Abschnitt 8.3 bekannten Verfahren bestimmen. Unter Verwendung der so ermittelten Parameter ergeben sich die Sensorpositionen im Referenzkoordinatensystem zu:

$$\widehat{\boldsymbol{s}}'_i = \nu \cdot \boldsymbol{R}\, \widehat{\boldsymbol{s}}_i + \boldsymbol{t}. \tag{9.1}$$

Die Berücksichtigung des Skalierungsfaktors ν ist jedoch nur dann notwendig, wenn die Skalierung nicht durch die Kalibrierung festgelegt wurde oder der von einer ungenauen Skalierung verursachte Anteil des Fehlers ermittelt werden soll. Mit der Rotationsmatrix $\boldsymbol{R}$ lässt sich anschließend auch die aus dem Kalibrierungsprozess stammende Orientierung $\widehat{\theta}_i$ in die Orientierung im Referenzkoordinatensystem ($\widehat{\theta}'_i$) transformieren.

Bei der modalitätsübergreifenden Kalibrierung hingegen, befinden sich die Positionen bzw. Orientierungen der akustischen Sensoren unmittelbar im visuellen Koordinatensystem. Somit gilt dort:

$$\widehat{\boldsymbol{s}}'_i = \widehat{\boldsymbol{s}}_i \quad \text{bzw.} \quad \widehat{\theta}'_i = \widehat{\theta}_i. \tag{9.2}$$

Nach Abschluss der zuvor erläuterten Schritte liegen sowohl die Ergebnisse der akustischen als auch der modalitätsübergreifenden Kalibrierung in einem bekannten Koordinatensystem vor. Infolgedessen lässt sich nun die Differenz zwischen Kalibrierungsergebnis und Referenz mit einem gemeinsamen Maß bewerten. Bei der Wahl eines Bewertungsmaßes für die Ergebnisse der Geometriekalibrierung existieren jedoch zwei Lager. Einerseits verwenden z. B. [GKH14], [BS05] oder [Con+12] den mittleren quadratischen Fehler (RMSE), wohingegen [Cro+12], [Hen+09] und [KWL08] das arithmetische Mittel der Abstände zwischen dem Kalibrierungsergebnis und den tatsächlichen Positionen nutzen. Im Rahmen dieser Arbeit kommt ebenfalls das arithmetische Mittel zum Einsatz, damit die Abweichungen aller Sensoren den gleichen Einfluss haben.

Unabhängig von der Art der Mittelung, lässt sich der Fehler entweder getrennt für Sensorpositionen und -rotationen oder aber übergreifend über die einzelnen Mikrofone aller Sensorknoten ermitteln. Die für den zweiten Fall erforderlichen Positionen aller Mikrofone ergeben sich unmittelbar aus der Position und Orientierung der jeweiligen Sensorknoten sowie der Anordnung der Mikrofone innerhalb der Arrays. Der anschließend daraus gewonnene Fehler hängt allerdings von der Anzahl der Mikrofone pro Sensorknoten und den Abständen zwischen den Mikrofonen eines Arrays ab. Aufgrund dessen ist die Verwendung der Positionen aller Mikrofone ungeeignet, um einen Vergleich von Inter-Array-Kalibrierungsverfahren zwischen verschiedenen Szenarien herzustellen, sofern sich die Sensorknoten in der Anzahl der Mikrofone oder in ihren räumlichen Abmessungen unterscheiden. Die Bewertung der durchgeführten Simulationen und Experimente erfolgt deshalb durch den mittleren Positionierungsfehler

$$\epsilon_{\mathrm{P}} = \frac{1}{I} \sum_{i=1}^{I} \left\| \widehat{\boldsymbol{s}}_i' - \boldsymbol{s}_i \right\|_2 \tag{9.3}$$

sowie den mittleren Orientierungsfehler

$$\epsilon_{\mathrm{W}} = \operatorname{atan2}\left(\frac{1}{I} \sum_{i=1}^{I} \sin\left(\widehat{\theta}_i' - \theta_i\right); \frac{1}{I} \sum_{i=1}^{I} \cos\left(\widehat{\theta}_i' - \theta_i\right) \right). \tag{9.4}$$

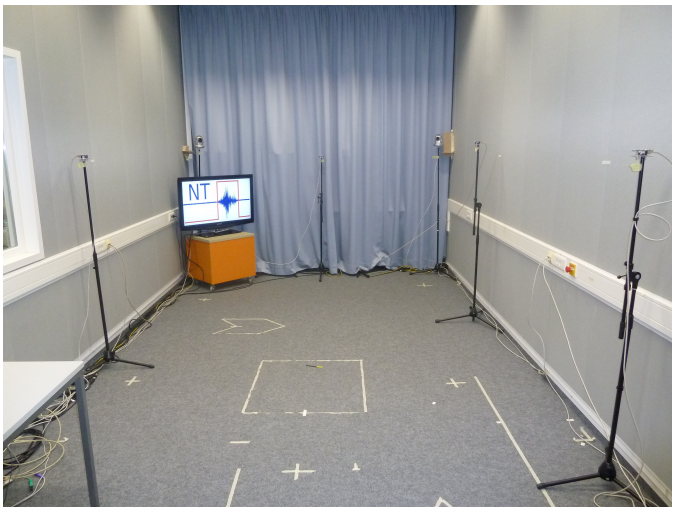
Die Funktion $\operatorname{atan2}(y; x)$ bezeichnet dabei die Verallgemeinerung der Arkustangensfunktion, die die Orientierung des Vektors $[x \;\; y]^{\mathrm{T}}$ liefert.

Obwohl zur Bewertung grundsätzlich der mittlere Positionierungs- bzw. Orientierungsfehler dient, entstehen insgesamt drei Kriterien. Diese unterscheiden sich dadurch, welche Schritte der Auswertung von Gl. (9.3) bzw. Gl. (9.4) vorangehen. Um eine eindeutige Kennzeichnung der drei Fälle zu gewährleisten, werden separate Formelzeichen verwendet. Die zuvor definierten Größen ϵ_{P} und ϵ_{W} beschreiben den Fehler, der nach Anwendung einer Koordinatentransformation ohne Skalierung verbleibt. Wenn allerdings ein Kalibrierungsverfahren zum Einsatz kommt, das nur eine relative Sensorkonfiguration liefert oder der Einfluss der Skalierung nicht berücksichtigt werden soll, wird zusätzlich zur Rotation und Translation auch die Skalierung so angepasst, dass der resultierende Fehler minimiert wird. In diesem Fall bezeichnet $\epsilon_{\mathrm{P,\,Rel}}$ den Positionierungsfehler. Da die Skalierung den Orientierungsfehler nicht beeinflusst, wird dieser weiterhin mit ϵ_{W} bezeichnet. Bei der modalitätsübergreifenden Kalibrierung, welche hingegen keine Transformation des Kalibrierungsergebnisses verlangt, werden die Fehler mit ε_{P} bzw. ε_{W} gekennzeichnet.

9.2 Szenarien

Für die Einschätzung der Leistungsfähigkeit der entwickelten Geometriekalibrierungsverfahren sind, wie in der Einleitung zu diesem Kapitel erwähnt, Experimente in realen Umgebungen unerlässlich. Die Grundlage der in Abschnitt 9.4 durchgeführten Evaluierung bilden zwei Datensätze, die in Räumen, die jeweils sowohl mit Mikrofonarrays als auch Kameras ausgestattet sind und über komplementäre Eigenschaften bzw. Rahmenbedingungen verfügen, aufgenommen wurden.

Bei dem ersten Raum handelt es sich um einen speziell gedämmten Laborraum (*Audiolabor*), der lediglich eine sehr geringe Nachhallzeit aufweist. Dadurch bietet das *Audiolabor* beste Voraussetzungen, um die Kalibrierung bei idealen akustischen Bedingungen zu studieren (siehe Abb. 9.1a). Die Aufnahme des zweiten Datensatzes erfolgte in einem *Smartroom* (vgl. Abb. 9.1b) [PF14a]. Dieser verfügt über eine realistische Nachhallzeit und die Mikrofon- und Kamerapositionierung entspricht einem möglichen Telekonferenzszenario.



(a) *Audiolabor*

(b) *Smartroom*

Abbildung 9.1: Ansicht der realen Szenarien zur Evaluierung der Geometriekalibrierung.

Das *Audiolabor* hat eine Größe von $3{,}50 \times 6{,}50 \times 3{,}10\,\mathrm{m}^3$ und ist mit 4 Mikrofonarrays ausgestattet. Jedes Array wiederum besitzt 4, in einem Quadrat mit 0,05 m Kantenlänge angeordnete Mikrofone. Zusätzlich gehören zur Ausstattung des *Audiolabors* 4 Kameras. Die absorbierenden Wände und der vorhandene Teppichboden sorgen dafür, dass die Nachhallzeit des Raumes lediglich bei ca. 0,16 s liegt [Sch+09].

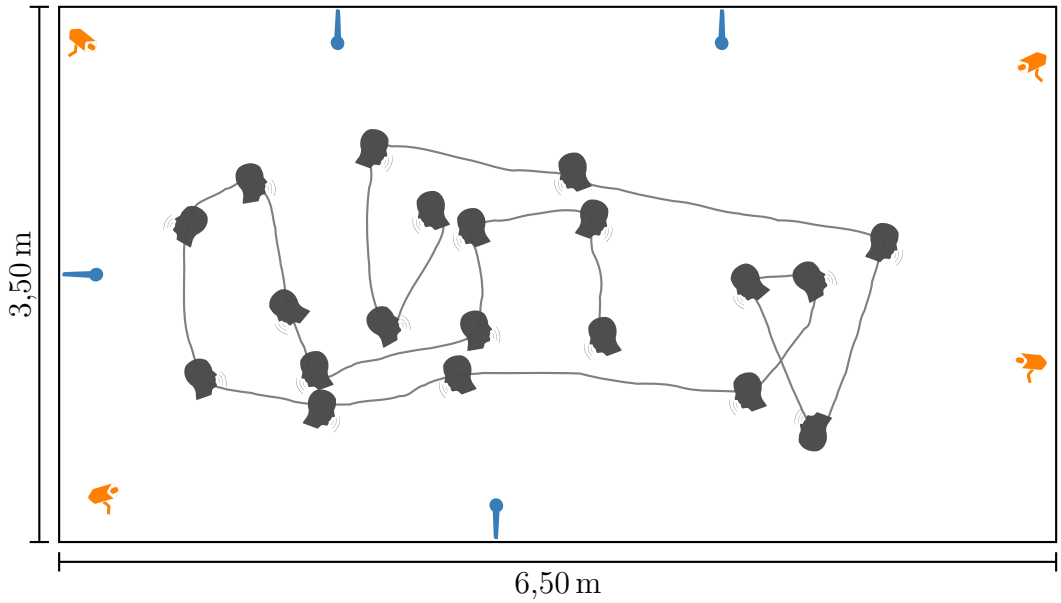


Abbildung 9.2: Schematische Übersicht des Szenarios im *Audiolabor*.

Als Signalquelle dient, wie schon mehrfach erwähnt, eine Person, da diese sowohl durch die akustischen als auch visuellen Sensoren erfasst werden kann. Die Person bewegt sich entlang einer zufälligen Trajektorie durch den Raum und hält an 20 Positionen für eine ca. 4 s bis 5 s andauernde Äußerung inne. Damit steht eine ca. 150 s umfassende audio-visulle Aufnahme für die anschließende Kalibrierung zur Verfügung. Zur Veranschaulichung des schematischen Aufbaus des Sensornetzes und des Verlaufes der Trajektorie dient Abb. 9.2.

Angesichts der Beschränkung der Geometriekalibrierung auf zwei Dimensionen erfolgt auch die DOA- bzw. TDOA Schätzung nur in einer Ebene (siehe Abschnitt 9.3). Um bestmögliche Voraussetzungen für die jeweiligen Schätzungen zu erzielen, sind die Mikrofonarrays auf Schulterhöhe des Sprechers angebracht. Dadurch lässt sich die Höhendifferenz zwischen Signalquelle und Mikrofonen, die ebenfalls die Schätzung beeinflusst, näherungsweise vernachlässigen.

Zur Aufzeichnung der Audiosignale dienen handelsübliche Kondensatormikrofone, die jedoch mit einer speziellen Soundkarte verbunden sind, die eine abtastsynchrone Aufnahme aller Kanäle mit 16 kHz gestattet [SJH14]. Die visuelle Aufnahme der Szene erfolgt durch 4 Netzwerkkameras, die mit einer Auflösung von 320×240 Pixeln und einer Rate von 12 FPS arbeiten. Das sehr geringe Sichtfeld der Kameras von nur ca. 30° führt dazu, dass trotz der Anordnung der Kameras in den Raumecken, nur ein kleiner Bereich in der Raummitte für eine Detektion durch mehr als zwei Kameras zur Verfügung steht. Allerdings sind alle verwendeten Positionen so gewählt, dass sie im Blickfeld von mindestens zwei Kameras liegen.

Als zweites Szenario fungiert der $3{,}90 \times 6{,}80 \times 2{,}60\,\mathrm{m}^3$ große *Smartroom*. Er verfügt über 3 zirkuläre Mikrofonarrays mit jeweils 5 Mikrofonen und einem Radius von 0,05 m. Darüber hinaus ist der *Smartroom* mit 5 Kameras ausgestattet, die allesamt in der Nähe der Wände platziert sind und zur Raummitte blicken. Währenddessen sind die Mikrofonarrays in den mittig im Raum positionierten Konferenztisch integriert (siehe

Abb. 9.1b). Als Signalquelle dient erneut eine Person, die sich nacheinander an 10 um den Konferenztisch angeordneten Positionen aufhält. Die Länge jeder einzelnen Äußerung beträgt ca. 5 s. Abb. 9.3 zeigt die schematische Darstellung des beschriebenen Szenarios.

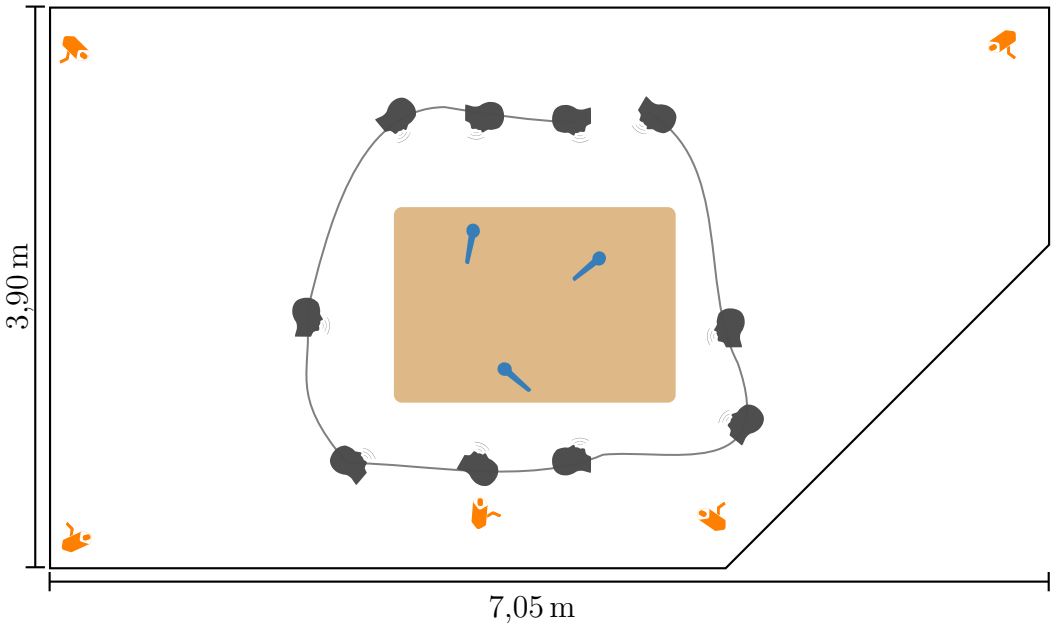


Abbildung 9.3: Schematische Übersicht des Szenarios im *Smartroom* (Nach [PF14a]).

Aufgrund der schallharten Wände und Böden beträgt die Nachhallzeit des *Smartroom* ca. 0,67 s [PF14a]. Die Aufnahme der Mikrofonsignale erfolgt mit zwei synchronisierten 8-Kanal Soundkarten, die mit einer Abtastrate von 48 kHz arbeiten. Die eingesetzten Kameras liefern 10 FPS. Trotz des horizontalen Sichtfeldes, das mit ca. 36° etwas größer als im *Audiolabor* ausfällt, ist auch im *Smartroom* eine Detektion der Person zumeist nur durch zwei Kameras gleichzeitig möglich.

9.3 Extraktion der erforderlichen Informationen

Dieser Abschnitt befasst sich nun mit der Extraktion der für die Evaluierung der verschiedenen Kalibrierungsvarianten notwendigen Informationen.

Um einen Vergleich zwischen *Audiolabor* und *Smartroom* zu gewährleisten, findet als erstes eine Umsetzung der Abtastrate der akustischen Aufnahmen auf 16 kHz statt. Zur Schätzung der Einfallswinkel kommt die in Abschnitt 4.4 favorisierte WKM zum Einsatz. Diese erfordert zuerst eine Transformation der Signale in den Frequenzbereich. Dazu wird die Kurzzeit-Fourier-Transformation (STFT) mit einer Blockgröße von 1024 Punkten sowie einem zu 75 % überlappenden Blackman-Fenster verwendet. Außerdem erlaubt eine energiebasierte Sprachaktivitätsdetektion (engl. *voice activity detection* (VAD)) die Identifikation der Zeitabschnitte, die sich zur Winkelschätzung eignen. Weiterhin werden zur Winkelschätzung nur die Frequenzen berücksichtigt, die 90 % der Signalenergie repräsentieren. Darüber hinaus gestatten die längeren Sprachpausen zwischen den einzelnen Positionen, an denen sich der Sprecher aufhält, eine

Segmentierung der Trajektorie. Die so ermittelten Segmente ermöglichen wiederum die Vereinigung der einzelnen *Likelihood*-Funktionen der WKM innerhalb eines Segmentes, sodass letztendlich eine Winkelschätzung pro Segment und Sensorknoten vorliegt.

Zur Bestimmung der Signallaufzeitdifferenzen von den Quellpositionen zu den Mikrofonen aus verschiedenen Arrays dient GCCPhat [KC76]. Die STFT wird hier mit einer Fenster-Länge von 4096 Punkten durchgeführt und verwendet ansonsten dieselben Parameter wie auch die akustische Richtungsschätzung. Grundsätzlich zeigen die durchgeführten Analysen, dass bereits eine FFT mit 1024 Punkten ausreicht. Allerdings ist eine möglichst präzise Schätzung der TDOA entscheidend für eine erfolgreiche Bestimmung der Skalierung.

Die Nutzung der Kameras zur Richtungsschätzung erfordert zunächst ein Training des zur Personenerkennung eingesetzten HOG-Detektors. Dazu wird, wie in Abschnitt 8.1 erläutert, erneut die *INRIA Person Database* [Dal06] verwendet. Allerdings verursachen im *Audiolabor* bzw. im *Smartroom* befindliche Gegenstände, wie z. B. das Stativmaterial zur Halterung der Mikrofone bzw. Kameras, deutliche Fehldetektionen. Demzufolge ist eine Erweiterung des Trainingsmaterials mit Aufnahmen aus beiden Räumen zwingend notwendig, um Fehldetektionen zu vermeiden.

Die Besonderheit der Aufnahmen aus dem *Smartroom* besteht darin, dass die Trajektorie sowohl Positionen an denen der Sprecher steht als auch Positionen an denen er auf einem Stuhl sitzt beinhaltet. Problematisch für die visuelle Erkennung der Person sind die sitzenden Passagen. Hier liefert der HOG-Detektor keine Detektionen, weil die aus dem vorhandenen Bildmaterial ermittelten HOG eine zu geringe Ähnlichkeit mit den Trainingsdaten besitzen.

Während die Passagen, in denen die Person sitzt, zu Problemen bei der visuellen Erkennung führen, verbessert sich dort die Situation für die akustische Modalität. Verantwortlich dafür ist die geringere Höhendifferenz zwischen dem Mund des Sprechers und den Mikrofonen. Die Höhendifferenz verringert sich von ca. 0,7 m auf 0,4 m, sodass die Durchführung der Winkelschätzung in nur einer Ebene einen geringeren systematischen Fehler verursacht.

9.4 Experimente

Die letzten beiden Abschnitte haben die verwendeten Szenarien und die Gewinnung der zur Kalibrierung erforderlichen Informationen eingehend erläutert. Der aktuelle Abschnitt präsentiert nun die erzielten Kalibrierungsergebnisse. Dabei befasst sich der erste Teil mit der Kalibrierung des Sensornetzes im *Audiolabor*, während der zweite die Kalibrierung im *Smartroom* beschreibt. In beiden Fällen kommt sowohl die akustische Kalibrierung mit anschließender Skalierung anhand von TDOA-Informationen als auch die Kalibrierung mithilfe eines modalitätsübergreifenden Gleichungssystems zum Einsatz. Angesichts der geringen Reduktion des Fehlers bei der rein akustischen Kalibrierung durch den PRANSAC, werden dort lediglich die Ergebnisse des PRANSAC betrachtet und es wird auf einen Vergleich zu dem z. T. ähnlich gut abschneidenden RANSAC verzichtetet. Bei der audio-visuellen Kalibrierung erfolgt hingegen eine Gegenüberstellung beider Varianten, da durch den PRANSAC dort ein deutlicher Gewinn entsteht.

Den Auftakt der experimentellen Untersuchungen bildet die akustische Geometriekalibrierung im *Audiolabor*. Aufgrund der zufälligen Auswahl der Winkelschätzungen innerhalb des RANSAC bzw. PRANSAC ist das Ergebnis sämtlicher Kalibrierungsvarianten nicht mehr deterministisch. Um eine Einschätzung der Bandbreite der zu erwartenden Ergebnisse zu ermöglichen, stellt Abb. 9.4 den Mittelwert und die Standardabweichung von 100 Ausführungen dar. Der linke Teil dieser Abbildung zeigt den mittleren Positionierungsfehler, einerseits bei optimaler Skalierung durch ein Orakel ($\epsilon_{\mathrm{P,\,Rel}}$) und andererseits bei der Verwendung des aus den TDOA-Messungen gewonnen Skalierungsfaktors (ϵ_{P}). Der rechte Teil von Abb. 9.4 visualisiert den mittleren Orientierungsfehler ϵ_{W}. Die Orientierung bzw. der daraus resultierende Fehler liegt bereits nach Abschluss des erweiterten Einfallswinkelverfahrens vor und wird nicht durch die nachträgliche Skalierung beeinflusst. Somit entfällt hier die Betrachtung von zwei separaten Anteilen des Fehlers.

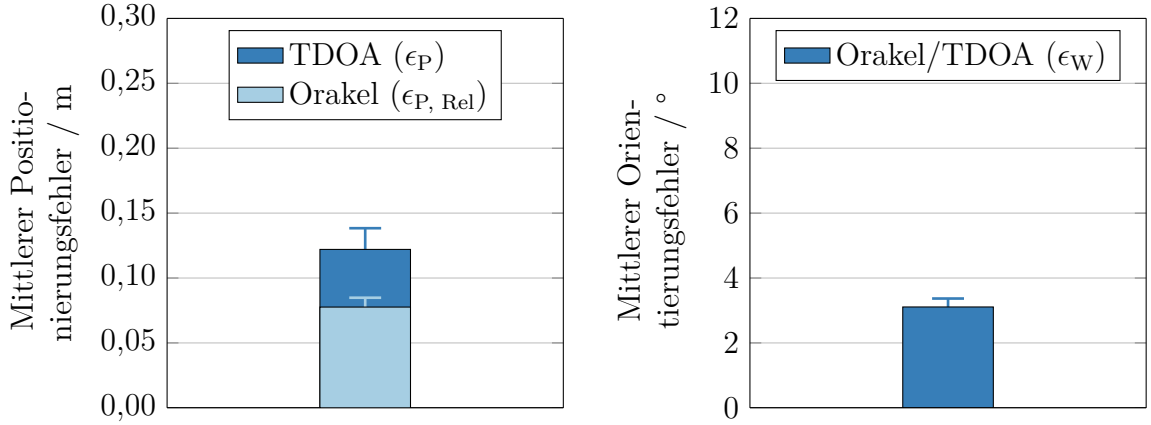


Abbildung 9.4: Fehler der akustischen Geometriekalibrierung im *Audiolabor*.

Die dargestellten Ergebnisse dokumentieren unmittelbar, dass der entwickelte Algorithmus auch in einem realen Szenario ausschließlich durch die aus einem Sprachsignal extrahierten Informationen eine präzise Kalibrierung der Sensorpositionen und -orientierungen gestattet. Darüber hinaus ist die niedrige Streuung der Ergebnisse ein Indikator für die geringe Abhängigkeit der Lösungen von der zufälligen Auswahl der Beobachtungen.

Weiterhin zeigt ein Vergleich mit den Simulationsergebnissen aus Abb. 7.1 eine ungefähre Übereinstimmung zwischen den Simulationen und der Realität. Bei der simulierten Nachhallzeit von 0,2 s fällt der durchschnittliche Positionierungsfehler mit 0,09 m vor bzw. 0,20 m nach der Skalierung zwar etwas größer aus als in dem jetzt betrachteten Experiment, allerdings ist die Nachhallzeit des realen Szenarios mit 0,16 s ebenfalls etwas geringer. Außerdem gilt es zu berücksichtigen, dass bei der Simulation nur 3 Mikrofone verwendet wurden, während im *Audiolabor* jedes Mikrofonarray über 4 Mikrofone verfügt. Der durch die Skalierung verursachte Positionierungsfehler besitzt eine größere Abweichung als der bislang betrachtete Anteil, da der Fehler der TDOA-Schätzung für die Simulationen lediglich mit einer von der tatsächlichen Nachhallzeit unabhängigen Normalverteilung modelliert wurde.

Als Alternative zur Kalibrierung der Sensoren allein durch akustische Informationen erfolgt weiterhin die Betrachtung der gemeinsamen audio-visuellen Kalibrierung. Die Ergebnisse für die Anwendung dieser modalitätsübergreifenden Variante im *Audiolabor* sind in Abb. 9.5 dargestellt. Auch hier wird erneut der Mittelwert bzw. die Standardabweichung des mittleren Positionierungsfehler (links) bzw. Orientierungsfehlers (rechts) von 100 Ausführungen angegeben. Im Gegensatz zum rein akustischen Ansatz beinhaltet die Lösung bei der gemeinsamen Geometriekalibrierung automatisch die Skalierung. Dementsprechend erübrigt sich auch die Aufteilung des Positionierungsfehlers in zwei Anteile (vgl. Abb. 9.4). Ferner erfordert die Berechnung des Fehlers keine Koordinatentransformation mehr, sodass der mittlere Positionierungs- und Orientierungsfehler jetzt durch ε_P bzw. ε_W gegeben ist (siehe Abschnitt 9.2).

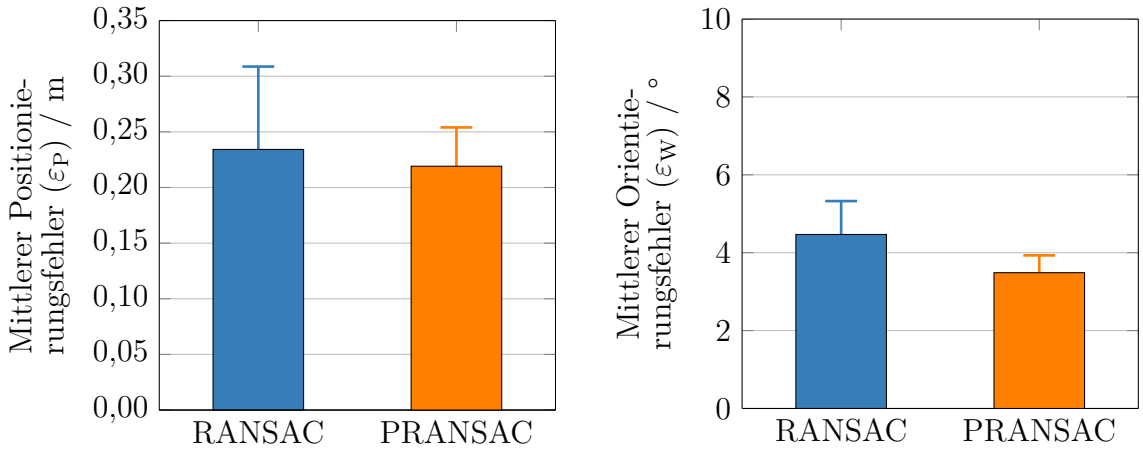


Abbildung 9.5: Fehler der audio-visuellen Geometriekalibrierung im *Audiolabor*.

Die Gegenüberstellung der Ergebnisse des PRANSAC mit denen des RANSAC belegt zum einen die Reduktion des Positionierungs- bzw. Orientierungsfehlers und zum anderen die Verringerung der Varianz der Lösungen. Damit erzielt die Fusion der verschiedenen RANSAC den erwarteten Effekt.

Ein Vergleich der Ergebnisse der modalitätsübergreifenden Kalibrierung mit der akustischen Kalibrierung zeigt jedoch auch einen starken Anstieg der Fehler. Auslöser dafür sind in erster Linie die visuellen Einfallswinkel. Zwar enthalten diese aufgrund des an den Kamerablickwinkel angepassten Modelltrainings nur wenige Fehldetektionen, aber eine nahezu permanent vorhandene Abweichung bei der Bestimmung der Position einer Person innerhalb des Bildes von ca. 10–15 Pixeln verursacht bei den genutzten Kameras einen Winkelfehler von etwas mehr als 2°. Als Konsequenz daraus besitzen die visuellen Winkelschätzungen einen größeren Fehler als die bislang verwendeten Akustischen und verursachen deshalb ein schlechteres Kalibrierungsergebnis.

Außerdem definieren bereits die Kamerapositionen eindeutig das Koordinatensystem, sodass die Berechnung der Fehler keine Koordinatentransformation mehr beinhaltet. Daher entfällt die Anpassung zwischen den ermittelten Positionen der Sensorknoten und den Referenzpositionen, die für eine Minimierung des Fehlers sorgen würde. Sofern die Berechnung des Fehlers dennoch eine Transformation beinhalten würde, ergäbe sich im hier betrachteten Szenario ein um ca. 0,05 m geringerer Fehler.

Die zurückliegenden Experimente im *Audiolabor* bestätigen, dass sowohl der ausschließlich akustische als auch der audio-visuelle Ansatz, die automatische Kalibrierung eines akustischen Sensornetzes in einer realen Umgebung ermöglichen. Allerdings verursachen die unpräziseren Winkelschätzungen der Kameras einen signifikanten Anstieg des Kalibrierungsfehlers. Zumal das *Audiolabor* hinsichtlich der Nachhallzeit ideale Rahmenbedingungen für die akustische Signalverarbeitung bietet, soll mit dem *Smartroom* ein realistisches Anwendungsbeispiel, das gleichzeitig einem möglichen Telekonferenzszenario entspricht, analysiert werden. Der Unterschied zum *Audiolabor* besteht einerseits in der deutlich höheren Nachhallzeit, die damit die Ausgangssituation für die Schätzung von DOA und TDOA aus den Aufnahmen der Mikrofone erschwert und andererseits in einer veränderten Anordnung der Sensoren (vgl. Abb. 9.2 bzw. Abb. 9.3). Bei der Anwendung des akustischen Geometriekalibrierungsverfahrens auf die Aufnahmen aus dem *Smartroom* entstehen die in Abb. 9.6 veranschaulichten Ergebnisse. Dabei wird erneut die aus Abb. 9.4 bekannte Darstellungsweise wiederverwendet.

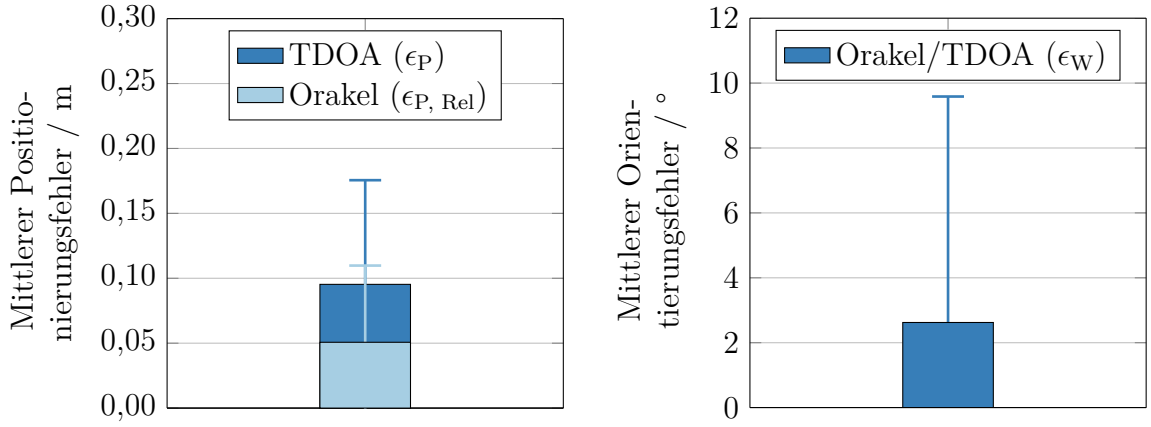


Abbildung 9.6: Fehler der akustischen Geometriekalibrierung im *Smartroom*.

Angesichts des erzielten mittleren Kalibrierungsfehlers von weniger als 0,10 m bzw. 3° liefert das Experiment im *Smartroom* die Bestätigung, dass das entwickelte Verfahren auch bei einer deutlich angestiegenen Nachhallzeit die Kalibrierung eines akustischen Sensornetzes erlaubt. Allerdings fällt bei einer gemeinsamen Betrachtung der Ergebnisse von *Audiolabor* und *Smartroom* unmittelbar die Reduktion des Positionierungsfehlers trotz gestiegener Nachhallzeit auf. Dies widerspricht zunächst den Erwartungen, weil eine höhere Nachhallzeit zu größeren Fehlern bei der akustischen Richtungsschätzung führen sollte, die sich wiederum negativ auf das Kalibrierungsergebnis auswirken sollten. Andererseits besteht ein Mikrofonarray im *Smartroom* aus 5 Mikrofonen und verfügt damit über ein Mikrofon mehr als die im *Audiolabor* genutzten Arrays. Außerdem steht aufgrund der durchschnittlich ca. 0,75 s länger ausfallenden Standphasen des Sprechers ein größerer Zeitraum für die Berechnung der Winkelschätzung zur Verfügung.

Zusätzlich zu den genannten Aspekten, spielt die Positionierung der Sensoren sowie die Position des Sprechers eine entscheidende Rolle. Im *Smartroom* sind die Mikrofonarrays in der Raummitte angeordnet und der Sprecher blickt von allen Positionen stets in Richtung der Mikrofone, damit ein direkter Ausbreitungspfad (LOS) existiert. Die Positionierung der Sensoren im *Audiolabor* sorgt allerdings dafür, dass sich die Mikrofone

z. T. hinter dem Sprecher befinden. Infolgedessen entfällt die direkte Signalkomponente (NLOS) und es entstehen zusätzliche Fehler bei der Winkelschätzung. Sofern es sich dabei um klar erkennbare Ausreißer handelt, ist der RANSAC bzw. PRANSAC zuverlässig in der Lage, diese Beobachtungen auszusortieren. Geringere Abweichungen werden jedoch teilweise nicht erkannt und fließen daher in das Kalibrierungsergebnis ein. Alle ausgeführten Aspekte tragen gemeinsam dazu bei, dass der Positionierungsfehler im *Smartroom* trotz einer größeren Nachhallzeit geringer ausfällt als im *Audiolabor*.

Zum Abschluss der experimentellen Untersuchungen zeigt Abb. 9.7 schließlich noch die Ergebnisse für die audio-visuelle Kalibrierung im *Smartroom*. Auch hier werden erneut die Ergebnisse für den RANSAC und die der partitionierte Variante (PRANSAC) gegenübergestellt.

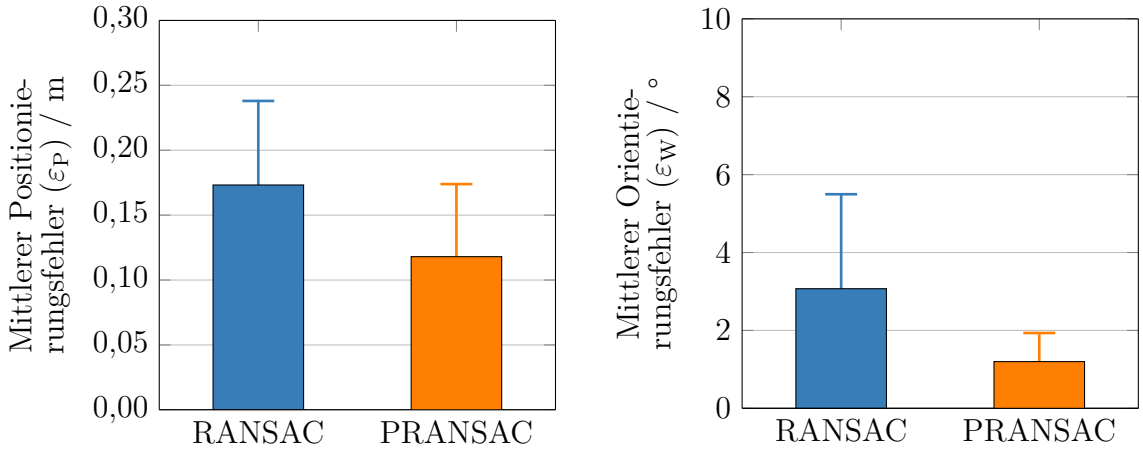


Abbildung 9.7: Vergleich der audio-visuellen Kalibrierung im *Smartroom*, beim Einsatz von RANSAC und PRANSAC.

Bei der Betrachtung des Positionierungsfehlers ergeben sich zunächst keine neuen Erkenntnisse. Auch im *Smartroom* steigt der Positionierungsfehler von der akustischen zur audio-visuellen Kalibrierung an. Hervorzuheben sind lediglich die Auswirkungen des PRANSAC auf den Orientierungsfehler. Selbst wenn nur der RANSAC verwendet wird liegt dieser bereits bei ca. 3°, dennoch trägt der PRANSAC noch einmal zu einer Reduktion um fast 2° bei.

Die präsentierten Ergebnisse für das *Audiolabor* und den *Smartroom* stellen zwar nur eine Momentaufnahme für zwei ausgewählte Szenarien dar, aber weitere, ebenfalls im *Audiolabor* durchgeführte Untersuchungen erzielen eine vergleichbare Präzision auch bei anderen Sensorkonfigurationen. Auf der anderen Seite existieren aber auch geometrische Konstellationen, die eine enorme Herausforderung für die Geometriekalibrierung mit dem entwickelten Verfahren darstellen. Bei der Kalibrierung von Szenarien vergleichbar zum *Smartroom* ist ein Mindestabstand zwischen den Sensorknoten sowie eine nicht zu große Entfernung zwischen Sprecher und den Sensoren notwendig, um sehr spitze Winkel bei der Lokalisierung der Ereignisse zu vermeiden. Die Anforderung, wie spitz die Winkel sein dürfen, hängt vom Fehler der Einfallswinkelschätzungen ab. Je präziser die Winkelschätzung, desto spitzer können die Winkel sein. Im *Smartroom* sollte ein Winkel von ca. 8°–10° nicht unterschritten werden. Problematisch sind die spitzen Winkel,

weil dort bereits ein geringer Fehler große Auswirkungen auf die daraus resultierende Positionsschätzung hat. In Szenarien, die ähnlich zum *Audiolabor* sind, können zwar sehr spitze Winkel aufgrund der Sensoranordnung vermieden werden, allerdings ergeben sich dort Probleme, wenn sich Ereignisse auf den Verbindungsgeraden zwischen den Sensoren befinden. In diesen Fällen entstehen gestreckte Winkel, die wiederum die selben Probleme wie die spitzen Winkel auslösen.

Insgesamt bestätigen die gezeigten Ergebnisse dennoch, dass die entwickelten Algorithmen die Geometriekalibrierung akustischer Sensornetze auch in realen Umgebungen gestatten. Die rein akustische Kalibrierung erzielt einen Kalibrierungsfehler, der den angestrebten Positionierungsfehler von 0,20 m (siehe Abschnitt 2.4) deutlich unterschreitet. Auch die audio-visuellen Verfahren erfüllen die gesetzten Ziele, wenn allerdings auch nur knapp. Aufgrund der deutlich geringeren Kalibrierungsfehler des akustischen Ansatzes im Vergleich zur audio-visuellen Variante, erscheint dieser zunächst als die bessere Wahl. Allerdings erfordert der rein akustische Ansatz zur Bestimmung der Inter-Array-TDOA eine Abtastsynchronisation, die, sofern sie nicht vorhanden ist, durch zusätzliche Verfahren hergestellt werden muss. Ohne Synchronisation verbleibt nur die gemeinsame Kalibrierung von akustischen und visuellen Sensoren. Hauptverantwortlich für das schlechtere Abschneiden der gemeinsamen Geometriekalibrierung akustischer und visueller Sensoren ist die Qualität der visuellen Winkelschätzungen. Da sich diese Arbeit vorrangig mit der akustischen Kalibrierung beschäftigt, muss an dieser Stelle jedoch auch berücksichtigt werden, dass im visuellen Bereich leistungsfähigere Alternativen zur Personenerkennung existieren [Tom+15], die zu einer Reduktion des Fehlers der modalitätsübergreifenden Kalibrierung führen sollten.

9.5 Zusammenfassung

Im Zentrum dieses Kapitels stand die Evaluierung der Geometriekalibrierungsalgorithmen in realen Umgebungen. Dazu wurden zunächst die verwendeten Bewertungsmaße definiert. Dabei galt es insbesondere zu berücksichtigen, dass sich die Angaben der Positionen und Orientierungen der Sensoren abhängig vom eingesetzten Kalibrierungsverfahren unterscheiden. Nach Abschluss einer akustischen Kalibrierung liegt eine relative Beschreibung der Sensorpositionen und -orientierungen vor, die keinen Zusammenhang zum Koordinatensystem des umgebenden Raumes oder dem Bezugssystem, in dem sich die tatsächlichen Sensorpositionen befinden, besitzt. Bei einer audio-visuellen Kalibrierung wird hingegen das Referenzkoordinatensystem durch die bekannten Positionen der visuellen Sensoren festgelegt. Damit trotz dieser Unterschiede in beiden Fällen dieselben Abstandsmaße genutzt werden konnten, wurde bei der akustischen Kalibrierung vor der Berechnung des Fehlers eine Transformation des Kalibrierungsergebnisses in das Referenzkoordinatensystem durchgeführt. Die dazu verwendete Koordinatentransformation minimiert durch Rotation und Translation den Abstand zwischen dem Kalibrierungsergebnis und der tatsächlichen Sensorkonfiguration. Die Abweichung zwischen dem Kalibrierungsergebnis und der wahren Geometrie wurde schließlich anhand des mittleren Positionierungs- und Orientierungsfehlers bewertet.

Die Grundlage, für die im Mittelpunkt dieses Kapitels stehende Untersuchung der Geometriekalibrierung in realen Umgebungen, bildeten zwei Szenarien (*Audiolabor* und

Smartroom). Diese beiden Szenarien verfügen neben den Mikrofonarrays auch über Kameras und gestatteten daher sowohl die Evaluierung der akustischen als auch der modalitätsübergreifenden Kalibrierungsverfahren. Das *Audiolabor* ist mit vier Kameras sowie vier Mikrofonarrays mit jeweils vier Mikrofonen ausgestattet und bietet aufgrund der Nachhallzeit von 0,16 s sehr gute Voraussetzungen für die Winkelschätzung. Andererseits sorgt die Anordnung aller Sensoren an den Wänden des Raumes dafür, dass sich die Mikrofonarrays z. T. hinter dem Sprecher befinden und dementsprechend NLOS-Situationen auftreten, die wiederum für Ausreißer bei der Winkelschätzung verantwortlich sind. Mit dem *Smartroom* wurde deshalb ein alternatives Szenario berücksichtigt. Die Nachhallzeit fällt dort mit 0,67 s deutlich größer aus, dafür verfügt jedes der dort verwendeten drei Mikrofonarrays über fünf Mikrofone. Zudem ist die Anordnung der Mikrofone in der Raummitte und die Positionierung der Kameras an den Wänden, einem Telekonferenzszenario nachempfunden.

Als Informationsquelle für die Gewinnung der zur Kalibrierung notwendigen Einfallswinkelschätzungen, diente in beiden Szenarien ein Sprecher, der sich für ca. 4–5 s an einer Position aufhielt. Im *Smartroom* wurden Aufnahmen von 10 Positionen durchgeführt, während im *Audiolabor* aufgrund der NLOS-Problematik 20 Positionen genutzt wurden. Die Schätzung der Einfallswinkel für die akustische Modalität wurde, wie auch in den bisherigen Kapiteln dieser Arbeit, mit der WKM realisiert. Die Extraktion der Signallaufzeitdifferenzen zur Fixierung der Skalierung erfolgte durch GCCPhat und die visuelle Winkelschätzung wurde mithilfe eines HOG-Detektors sowie einer SVM als Klassifikator realisiert.

Zur rein akustischen Kalibrierung des Sensornetzes wurde das in den RANSAC eingebettete erweiterte Einfallswinkelverfahren verwendet und anschließend die Skalierung der Geometrie anhand von TDOA-Messungen bestimmt. Die audio-visuelle Kalibrierung wurde ebenfalls mit der Kombination von erweitertem Einfallswinkelverfahren und RANSAC realisiert. Allerdings kommt die modalitätsübergreifende Variante des Einfallswinkelverfahrens aufgrund der bekannten Positionen und Orientierungen der visuellen Sensoren, ohne weitere Informationen zur Skalierung aus.

Bei der Betrachtung der erzielten Kalibrierungsfehler zeigte sich, dass der Fehler im *Audiolabor* mit durchschnittlich ca. 0,12 m bzw. 3,1° geringfügig größer als im *Smartroom* (0,09 m bzw. 2,6°) ausfällt, obwohl die Nachhallzeit im *Audiolabor* sehr viel geringer ist. Die Nachhallzeit ist jedoch nicht der einzige Faktor, der einen Einfluss auf die Qualität der Winkelschätzungen und damit auf den Kalibrierungsfehler hat. Zu den weiteren Faktoren gehört einerseits die Anzahl der Mikrofone pro Sensorknoten, die im *Smartroom* größer als im *Audiolabor* ist und andererseits die bereits mehrfach thematisierte NLOS-Problematik, die durch die Sensorkonfiguration im *Audiolabor* begünstigt wird. Der RANSAC ist zwar in der Lage viele der Ausreißer zu erkennen und kann somit vermeiden, dass diese in das Kalibrierungsergebnis einfließen, gleichwohl ist die Erkennung nicht fehlerfrei. Dementsprechend werden teilweise Ausreißer zur Bestimmung der Geometrie berücksichtigt.

Während bei der akustischen Kalibrierung die Fehler für beide Szenarien in etwa gleich groß sind, ergab sich bei der modalitätsübergreifenden Kalibrierung ein deutlicher Unterschied. Zudem zeigte sich eine Zunahme des Kalibrierungsfehlers im Vergleich zur akustischen Kalibrierung. Der größte Anstieg des Fehlers ergab sich, sofern der auch bei der akustische Kalibrierung verwendete RANSAC mit lokalen Iterationen zur

Kalibrierung genutzt wurde. Wesentlich präzisere Kalibrierungen konnten hingegen mithilfe der partitionierten Variante des RANSAC (PRANSAC) erreicht werden. Trotz der Reduktion des Kalibrierungsfehlers durch den PRANSAC fällt dieser weiterhin sowohl im *Smartroom* (0,11 m bzw. 1,2°) als auch im *Audiolabor* (0,21 m bzw. 3,5°) größer als bei der akustischen Kalibrierung aus. Eine Ausnahme sind lediglich die Orientierungsfehler, die durch die Mittelung mehrerer Ergebnisse unter den Fehler bei der akustischen Kalibrierung absinken.

Das Ziel dieser Arbeit (Abschnitt 2.4), Geometriekalibrierungsalgorithmen zu entwickeln deren Kalibrierungsfehler bis zu 0,20 m und 4° beträgt, wurde durch die akustischen Ansätze mehr als erreicht. Auch der Kalibrierungsfehler bei der Untersuchung der audio-visuellen Variante im *Smartroom* unterschreitet die gesetzte Obergrenze deutlich. Lediglich im *Audiolabor* erreicht der Fehler die Grenze aufgrund der ungünstigen Sensorkonstellation.

10 Zusammenfassung

Der Zusammenschluss von Mikrofonen bzw. Mikrofonarrays zu einem akustischen Sensornetz bildet die Basis für zahlreiche Anwendungen im Bereich der akustischen Signalverarbeitung. Dazu gehören u. a. *Beamforming*, die Störgeräuschunterdrückung sowie die Lokalisierung von Ereignissen oder Sprechern. Insbesondere für die Lokalisierung durch Trilateration oder Triangulation, ist die Kenntnis der Sensorpositionen und -orientierungen essentiell. Im Rahmen dieser Arbeit wurden deshalb verschiedene Ansätze zur automatischen Geometriekalibrierung akustischer Sensornetze entworfen und eingehend analysiert. Der Schwerpunkt lag dabei auf der Entwicklung von Verfahren zur Kalibrierung der Positionen und Orientierungen von Sensorknoten, die jeweils aus einem kompakten Mikrofonarray bestehen (Inter-Array-Kalibrierung) und bspw. zur Realisierung der zuvor genannten Anwendungen in einem Telekonferenzsystem dienen.

Um einen universell einsetzbaren Algorithmus zu entwickeln, galt es, sowohl die Anforderungen an das Sensornetz als auch für die Person, die den Kalibrierungsprozess durchführt, auf ein Minimum zu beschränken. Daher sollten keine Hilfsmittel, wie etwa dedizierte Kalibrierungssignale oder spezielle Lautsprecherkonstruktionen zum Einsatz kommen. Darüber hinaus war die Nutzung von aktiven Sensorknoten zu vermeiden und die Anzahl der Mikrofone pro Sensorknoten möglichst gering zu halten. Weiterhin sorgt die räumliche Trennung der Sensoren dafür, dass in vielen Fällen keine Abtastsynchronisation zwischen den verschiedenen Knoten zur Verfügung steht.

Angesichts der vorliegenden Rahmenbedingungen schied eine Kalibrierung mithilfe von TOA-, TOF- oder Inter-Array-TDOA-Messungen aus, da anderenfalls zusätzliche Schritte zur Synchronisation der Sensorknoten erforderlich gewesen wären. Verfahren, die zunächst jeden Sensorknoten getrennt zur Lokalisierung von Ereignissen nutzen und anschließend die Sensoranordnung durch eine Abbildung der ermittelten Ereignispositionen realisieren, bieten zwar einen Ausweg, da sie keine Abtastsynchronisation voraussetzen, erfordern dafür aber ausreichend große Arrays [Hen+09; Red+09; Val+10b; Val+10a]. Signaleinfallswinkel dienen unterdessen im Bereich der akustischen Geometriekalibrierung meist nur als zusätzliche Informationsquelle [PMH11; PF14a], wohingegen eine Kalibrierung allein durch DOA im nicht-akustischen Umfeld bereits Anwendung findet [SZW14; KL08]. Zumal für die Schätzung von Einfallswinkeln kompakte Arrays mit nur zwei Mikrofonen ausreichen und somit keine Abtastsynchronisation zwischen den Sensorknoten erforderlich ist, wurden im Rahmen dieser Arbeit Kalibrierungsalgorithmen realisiert, die DOA-Schätzungen verwenden.

Den Ausgangspunkt für die eigenen Entwicklungen bildete das aufgrund seiner großen Ähnlichkeiten zur betrachteten Problemstellung ausgewählte und eigentlich zur Kalibrierung von Infrarotsensoren entwickelte Einfallswinkelverfahren [KWL08]. Der Kern

dieses Verfahrens besteht darin, eine geometrische Beziehung mithilfe der zu mehreren Ereignissen gemessenen Einfallswinkel herzustellen und daraus ein Gleichungssystem zusammenzustellen, dessen Lösung sowohl die Sensorpositionen und -orientierungen als auch die Ereignispositionen liefert. Als Quellsignal für die Schätzung der erforderlichen Einfallswinkel wurde das Sprachsignal einer Person, die sich durch den Raum bewegt, verwendet, um auf künstliche Kalibrierungssignale verzichten zu können. Die Nutzung von Sprachsignalen eröffnet zusätzlich die Möglichkeit, das Sensornetz schon während der Kalibrierung für Aufgaben, die keine Kenntnis der Geometrie voraussetzen, einzusetzen oder die Kalibrierung parallel zum eigentlichen Betrieb zu präzisieren.

Zur Sondierung der Ausgangssituation der Geometriekalibrierung erfolgte zunächst eine Analyse von vier ausgewählten Ansätzen zur Schätzung des Einfallswinkels aus den Aufnahmen von kompakten Mikrofonarrays. Dabei wurden u. a. das weit verbreitete SRPPhat, aber auch die im Rahmen dieser Arbeit entwickelte WKM berücksichtigt. Die Grundlage der Untersuchungen bildeten mit der Spiegel-Quellen-Methode generierte mehrkanalige Aufnahmen, die sowohl unterschiedliche Nachhallzeiten als auch zahlreiche Kombination aus Array- und Quellposition widerspiegeln.

Beim Einsatz von nur zwei Mikrofonen wiesen alle untersuchten Winkelschätzer einen systematischen Fehler auf, der maßgeblich vom tatsächlichen Einfallswinkel abhängt und darüber hinaus mit zunehmender Nachhallzeit anwächst. Als Auslöser dieses Bias wurde die lineare Anordnung der Mikrofone identifiziert. Sie erfordert einerseits die Begrenzung (engl. *clipping*) der der Einfallswinkelschätzung zugrunde liegenden TDOA-Messungen und sorgt andererseits für eine Beschränkung des eindeutigen Detektionsbereiches auf 180°. Beide Effekte zusammen lösen schließlich den beobachteten Bias aus. Während die Begrenzung der TDOA vorwiegend einen Bias in der Nähe von ±90° verursacht, führt der bei einer linearen Anordnung der Mikrofone begrenzte Detektionsbereich zu einer ungleichmäßigen Verteilung der frühen Reflexionen, die wiederum die systematische Unterschätzung der TDOA auslöst. Am stärksten ausgeprägt ist die Unterschätzung je größer die tatsächliche TDOA ausfällt. Insgesamt liegt der beobachtete Bias häufig deutlich über 5°. Allerdings dokumentieren die Analysen außerdem, dass schon die Erweiterung des Mikrofonarrays mit einem zusätzlichen Mikrofon, hin zu einem dreieckigen Array ausreicht, damit der systematische Anteil des Fehlers verschwindet und der verbleibende Winkelfehler meist deutlich unter 2° liegt.

Die Begutachtung der Winkelschätzer bei verschiedenen Nachhallzeiten und Störungen durch gerichtetes Rauschen belegt weiterhin, dass die WKM besser als die konkurrierenden Ansätze abschneidet. Angesichts dessen wurde die WKM als Winkelschätzer für die Geometriekalibrierung ausgewählt. Ferner bildeten die im Verlauf dieser Untersuchungen gewonnenen Erkenntnisse auch die Grundlage für die Entwicklung eines statistischen Modells des Winkelfehlers. Gemäß der vorliegenden Daten, lässt sich dieser durch eine mittelwertfreie von Mises-Verteilung approximieren, sofern mindestens drei, nicht auf einer Linie angeordnete, Mikrofone zur Verfügung stehen.

Aufbauend auf den Erkenntnissen aus der Analyse der Einfallswinkelschätzung, erfolgte eine eingehende Untersuchung des zur Geometriekalibrierung von Infrarotsensoren entwickelten Einfallswinkelverfahrens. Ziel dieser Untersuchung war die Überprüfung, ob sich der gewählte Algorithmus auch zur Kalibrierung eines akustischen Sensornetzes eignet. Die dazu durchgeführten Simulationen zeigen jedoch ein enormes Konvergenzproblem. Selbst bei ungestörten Winkelschätzungen scheitern ca. 15 % der Lösungsversuche

und weitere 14 % liefern lediglich ein lokales Minimum. Darüber hinaus treten Lösungen auf, die zwar die korrekten Sensorpositionen beschreiben, aber die ermittelten Orientierungen weichen um $\pm\pi$ von der tatsächlichen Sensorausrichtung ab (Rotationsinvarianz). Verantwortlich für die entdeckten Unzulänglichkeiten sind die Tangensfunktionen des Optimierungsproblems. Diese wirken sich aufgrund ihrer Polstellen einerseits negativ auf das Konvergenzverhalten aus und verursachen andererseits Mehrdeutigkeiten, die wiederum die Rotationsinvarianz auslösen. Zusätzlich sorgt die Beschränkung auf Einfallswinkel als Informationsquelle dafür, dass ohne die a priori-Kenntnis mindestens einer Distanz oder die Nutzung zusätzlicher Informationen zwar die Sensoranordnung, nicht aber die Skalierung dieser, ermittelt werden kann (Skalierungsinvarianz).

Angesichts der gravierenden Probleme, die sich bei der Untersuchung des Einfallswinkelverfahrens ergaben, war eine Weiterentwicklung des Algorithmus unbedingt notwendig, um eine verlässliche Kalibrierung von akustischen Sensornetzen zu gewährleisten. Einen wichtigen Aspekt der Weiterentwicklung stellt die Reformulierung des Optimierungsproblems dar. Dazu wurden in einem ersten Schritt die Tangensfunktionen ersetzt und eine Darstellung der Zielfunktion als Skalarprodukt entwickelt. Dadurch konnten nicht nur die Polstellen entfernt werden, sondern es entstand auch ein anschaulicher Beweis für die Rotationsinvarianz, da sich die Lösung des Optimierungsproblems auf die Suche nach einem Vektor, der orthogonal zum gemessenen Einfallswinkel steht, zurückführen lässt und somit keine eindeutige Lösung besitzt. Andererseits entstand aus der Formulierung als Skalarprodukt auch eine weitere Zielfunktion, die stattdessen einen Vektor parallel zum Einfallswinkel bestimmt und daher ein eindeutiges Ergebnis ohne Rotationsinvarianz erzielt. Insgesamt sorgen die erläuterten Modifikation dafür, dass die Quote der erfolgreichen Lösungen bei fehlerfreien Einfallswinkeln von vormals ca. 70 % auf mehr als 99 % ansteigt.

Weiterhin handelt es sich bei dem erweiterten Einfallswinkelverfahren um den ML-Schätzer für die Sensorpositionen und -orientierungen, sofern der Fehler der Einfallswinkelschätzung durch eine von Mises-Verteilung beschrieben werden kann. Obwohl diese Annahme beim Einsatz von Arrays mit 3 Mikrofonen nur näherungsweise erfüllt ist, erzielt das erweiterte Einfallswinkelverfahren, selbst bei einer Nachhallzeit von 0,4 s, in 90 % der Simulationen einen Positionierungsfehler kleiner als 0,15 m. Außerdem zeigt ein Vergleich mit künstlich erzeugten Winkelfehlern, die exakt einer von Mises-Verteilung folgen, dass der Fehler dort nur geringfügig kleiner ausfällt. Falls hingegen nur zwei Mikrofone zur Verfügung stehen und somit die Winkelschätzungen aufgrund des systematischen Fehlers deutliche Abweichungen beinhalten, entsteht ein drastischer Anstieg des Kalibrierungsfehlers. Dieser übersteigt schon bei einer Nachhallzeit von 0,2 s in mehr als 50 % der Untersuchungen 0,2 m.

Zumal sich bereits bei den Simulationen ein sehr deutlicher Einfluss der Störungen der Einfallswinkel auf das Kalibrierungsergebnis abzeichnete, ergab sich im Hinblick auf die Nutzung in realen Szenarien weiterer Handlungsbedarf, da dort Nachhall oder ein fehlender direkter Signalpfad (NLOS) für Ausreißer der DOA-Schätzung sorgen. Zur Entwicklung eines gegenüber Störungen und Ausreißern robusten Gesamtsystems wurde das erweiterte Einfallswinkelverfahren in einen RANSAC eingebettet. Dieser nutzt im Gegensatz zu einem LS-Verfahren zur Berechnung der gesuchten Geometrie lediglich eine Teilmenge der verfügbaren DOA-Messungen. Um dabei möglichst die Winkel zu berücksichtigen, die nur einen kleinen Fehler besitzen, erfolgt eine zufällige

Auswahl einer Teilmenge der Daten und eine anschließende Bewertung, welche der übrigen Beobachtungen mit dem resultierenden Modell im Einklang stehen. Dieser Vorgang wird mehrfach wiederholt und das Modell, das den größten Konsens erzielt, bildet die Grundlage zur Berechnung der finalen Geometrie.

Die Entscheidung, ob die Einfallswinkel eines Ereignisses mit der vorliegenden Geometrie kompatibel sind, hängt maßgeblich vom Schwellwert, der die Entscheidungsgrenze definiert, ab. Einerseits sollte dieser möglichst klein ausfallen, damit lediglich die Messungen ausgewählt werden, die nah am Modell liegen. Andererseits passen dann häufig nur wenige Beobachtungen, sodass am Ende kein präzises Modell entsteht. Der LORANSAC wirkt dieser Problematik entgegen, indem der Konsens zur Bestimmung eines präzisieren Modells genutzt wird und erst der Konsens des daraus gewonnenen Modells zur Bewertung des Ergebnisses dient. Dieses zweistufige Konzept gestattet eine schrittweise Präzisierung der Kalibrierung und wurde deshalb dahingehend weiterentwickelt, dass beliebig viele Schritte zur Verfeinerung möglich sind.

Allerdings sorgen schon die zahlreichen Anläufe zur Bestimmung des Modells innerhalb des konventionellen RANSAC für einen erheblichen Rechenaufwand, der durch die vorgeschlagene schrittweise Präzisierung jedes einzelnen Versuches, noch einmal deutlich ansteigt. Zur Beschränkung des Rechenaufwandes und zur Verarbeitung von vielen Datensätzen wurde daher eine partitionierte Variante des RANSAC (PRANSAC) entworfen. Anstatt eines RANSAC, der zahlreiche Versuche erfordert, damit am Ende ein verlässliches Ergebnis vorliegt, wurden stattdessen mehrere RANSAC mit wenigen Versuchen verwendet und die Teilergebnisse anschließend zu einer gemeinsamen Geometrie vereinigt. Die Fusion der Teilergebnisse erfolgte mit Techniken der statistischen *Shape*-Analyse, die es erlauben die mittlere Geometrie zu bestimmen, selbst wenn die Sensorpositionen und -orientierungen der verschiedenen Teilergebnisse in unterschiedlichen Koordinatensystemen vorliegen.

Die Notwendigkeit das Einfallswinkelverfahren in Kombination mit dem RANSAC zu verwenden, zeigt sich eindrucksvoll sobald die Fehler der Winkelschätzungen neben einem von Mises-verteilten Anteil zufällige Ausreißer beinhalten. Falls 10 % der Winkelschätzungen Ausreißer sind, bricht das erweiterte Einfallswinkelverfahren zusammen und liefert nur noch in ca. 20 % der Fälle eine Lösung. Die in den RANSAC eingebettete Variante wird dagegen kaum von den Ausreißern beeinflusst und erzielt selbst bei einer Nachhallzeit von 0,4 s in 80 % der Untersuchungen einen Fehler kleiner als 0,15 m. Der Einsatz des RANSAC trägt damit entscheidend dazu bei, dass eine automatische Geometriekalibrierung eines akustischen Sensornetzes möglich wird.

Zur Gewinnung von weiteren Informationen, um die nach Abschluss von Einfallswinkelverfahren und RANSAC verbleibende Skalierung ebenfalls mithilfe von Sprachsignalen festzulegen, wurden zwei Alternativen berücksichtigt. Eine Möglichkeit dazu bieten die im Umfeld der Geometriekalibrierung weit verbreiteten TDOA-Messungen, da diese unmittelbar einen Rückschluss auf Distanzen gestatten. Aufgrund der im Rahmen dieser Arbeit betrachteten kompakten Arrays, kamen jedoch nur Inter-Array-TDOA-Messungen in Betracht. Diese erfordern jedoch eine Abtastsynchronisation der Sensorknoten, die beim restlichen Design konsequent vermieden wurde. Darüber hinaus dokumentieren die Simulationen jedoch auch, dass selbst bei perfekter Synchronisation die aus den TDOA-Messungen gewonnene Skalierung den Positionierungsfehler im Vergleich zu einer Skalierung durch ein Orakel um ca. 0,1 m ansteigen lässt.

Als Alternative wurde außerdem eine Verfahrensweise entwickelt, die sich auf Einfallswinkel beschränkt und deshalb ohne Abtastsynchronisation zwischen den Sensoren auskommt. Damit trotzdem Distanzen vorliegen, werden Sensorknoten, die über zusätzliche Mikrofone verfügen, benötigt. Die Mikrofone der Sensorknoten dienen dabei jedoch nicht zur Schätzung eines gemeinsamen Einfallswinkels, sondern jedes Mikrofonpaar wird für eine separate Winkelbestimmung genutzt. Durch die Kenntnis des Aufbaus der Sensorknoten lässt sich die Position jedes Mikrofonpaares relativ zum Zentrum des Sensorknotens ausdrücken und dadurch ein Einfallswinkelverfahren formulieren, das keine Skalierungsinvarianz mehr besitzt. Diese Strategie liefert allerdings nur unzureichende Ergebnisse, die weitestgehend auf eine fehlerhafte Skalierung zurückzuführen sind. Hauptverantwortlich für den Skalierungsfehler sind die Abstände der Mikrofonpaare, die im Vergleich zu den auftretenden Winkelfehlern viel zu klein sind, damit jeder Sensorknoten durch seine Winkelschätzungen eine Triangulation erlaubt, die die Skalierung fixiert.

Da im Umfeld von akustischen Sensornetzen, bspw. bei der Realisierung von Telekonferenzsystemen, neben Mikrofonarrays auch Kameras zur Verfügung stehen, wurden darüber hinaus modalitätsübergreifende Ansätze untersucht. Das Ziel dabei war es nicht nur die zur Skalierung der Positionen der akustischen Sensoren notwendigen Informationen zu beschaffen, sondern auch eine Beschreibung aller Sensoren in einem gemeinsamen Koordinatensystem zu erhalten, um dadurch die Ausnutzung von Synergien, wie etwa einer audio-visuellen Lokalisierung, zu ermöglichen.

Die für das Einfallswinkelverfahren genutzte Abstraktion der Sensorinformationen hin zu Einfallswinkeln ermöglichte es, Kameras, die durch den Einsatz einer HOG-Detektion ebenfalls die Bestimmung des Einfallswinkel gestatten, nahtlos in das bestehende Einfallswinkelverfahren zu integrieren. Im Unterschied zu den akustischen Sensoren wurden die Positionen und Orientierungen der Kameras jedoch als bekannt vorausgesetzt und liefern daher Informationen zur Skalierung. Aufgrund des Fehlers der visuellen Einfallswinkelschätzungen, steigt der Kalibrierungsfehler im Vergleich zur akustischen Kalibrierung und Skalierung mithilfe eines Orakels leicht an. Allerdings fällt der mittlere Fehler mit ca. 0,16 m geringer aus als bei der Skalierung durch TDOA.

Durch die Abstraktion der Sensorinformationen können keine sensorspezifischen Informationen mehr in die Kalibrierung einfließen, deshalb wurde ein weiterer Algorithmus vorgestellt, der es erlaubt, zunächst beide Modalitäten individuell zu kalibrieren und die in getrennten Koordinatensystemen vorliegenden Lösungen mithilfe einer Koordinatentransformation der Trajektorien zu vereinigen. Sofern jedoch das in den RANSAC eingebettete erweiterte Einfallswinkelverfahren zur akustischen Kalibrierung dient, ist die Abbildung der Trajektorien der gemeinsamen Geometriekalibrierung eindeutig unterlegen. Die Gründe liegen einerseits in der unpräzisen Skalierung durch die geschätzte Koordinatentransformation und andererseits sorgen die Positionen und Orientierungen aus der akustischen Geometriekalibrierung für einen signifikanten Fehler bei der Bestimmung der akustischen Trajektorie. Der RANSAC zur Schätzung der RBT-Parameter minimiert allerdings nur die Abweichungen zwischen den beiden Trajektorien und kann deshalb den bereits durch die akustische Kalibrierung ausgelösten Fehler nicht mehr kompensieren. Der Einsatz einer getrennten Kalibrierung bietet daher nur dann einen Vorteil, wenn durch die modalitätsspezifische Kalibrierung der individuelle Kalibrierungsfehler klar reduziert werden kann.

Die zur Abbildung der Trajektorien ebenso wie zur Bestimmung der mittleren Geometrie verwendete RBT-Parameterschätzung besitzt eine hohe Relevanz in vielen Einsatzgebieten. Dementsprechend existieren unterschiedliche Ansätze. Enorme Bedeutung hat in allen Bereichen die statistische *Shape*-Analyse, die zur Bestimmung der RBT-Parameter eine Transformation in den *Shape*-Bereich vorsieht. Durch den Einsatz der DFT konnte eine alternative Darstellung zu den ansonsten verbreiteten KENDAL-Koordinaten realisiert werden, die anstatt einer Matrixmultiplikation nur die Berechnung der FFT erfordert und somit zu einer Reduktion des Rechenaufwandes führt. Laufzeitmessungen belegen außerdem, dass der entwickelte Ansatz der Vorgehensweise, die zur Schätzung der RBT-Parameter eine SVD nutzt, ebenfalls überlegen ist und bei größeren Datensätzen einen Gewinn von deutlich mehr als 20 % erreicht.

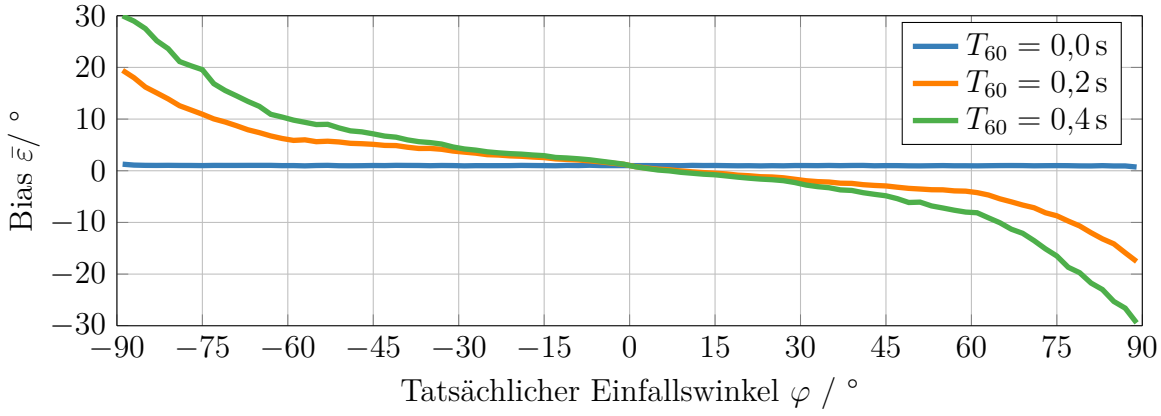
Zur abschließenden Bewertung der Geometriekalibrierungsverfahren, die bei den Simulationen die besten Ergebnisse erzielten, wurden die Algorithmen in zwei realen Umgebungen getestet. Als Informationsquelle diente in beiden Fällen eine Person, die jeweils für ca. 5 s an verschiedenen Positionen des Raumes eine Äußerung tätigte. Das *Audiolabor* bietet mit einer Nachhallzeit von 0,16 s ideale Voraussetzungen für die akustische Signalverarbeitung, wohingegen der *Smartroom* mit einer Nachhallzeit von 0,67 s und einer Anordnung der Mikrofonarrays auf einem Konferenztisch ein realistisches Telekonferenzszenario widerspiegelt. Die Ergebnisse der Experimente dokumentieren jedoch, dass die Nachhallzeit zwar die Qualität der Winkelschätzungen beeinflusst, für das letztendliche Kalibrierungsergebnis aber nur eine untergeordnete Rolle spielt, da sowohl im *Audiolabor* als auch im *Smartroom* der Kalibrierungsfehler des akustischen Ansatzes im Bereich von 0,1 m liegt. Größeren Einfluss haben dagegen die Anordnung der Sensoren und die bei der Winkelschätzung, aufgrund eines fehlenden direkten Ausbreitungspfades (NLOS), auftretenden Ausreißer, die durch den RANSAC nicht von der Kalibrierung ausgeschlossen werden können. Weiterhin steigt der Fehler bei den Experimenten zur modalitätsübergreifenden Kalibrierung an, obwohl die Positionen der Kameras vorliegen und somit durch die visuellen Winkelschätzungen sogar zusätzliche Informationen vorhanden sind. Der Schwerpunkt dieser Arbeit liegt allerdings auf akustischen Verfahren und die Qualität der rudimentären visuellen Winkelschätzung kann nicht mit den akustischen Methoden mithalten und sorgt daher für den Anstieg des Fehlers. Durch den Einsatz von leistungsfähigeren visuellen Ansätzen sind jedoch vergleichbare Ergebnisse wie bei der akustischen Kalibrierung denkbar.

Insgesamt stellen die Weiterentwicklungen des Einfallswinkelverfahrens gemeinsam mit dem RANSAC die entscheidenden Schritte zur Realisierung von automatischen Geometriekalibrierungsverfahren akustischer Sensornetze dar. Als Informationsquelle dienen dabei Sprachsignale, sodass keine Hilfsmittel oder Kalibrierungssignale benötigt werden. Trotz der wesentlich schlechteren Korrelationseigenschaften von Sprachsignalen und der Ausreißerproblematik erreichen die entwickelten Ansätze das Ziel eine Geometriekalibrierung mit einem Fehler von weniger als 0,25 m zu liefern, um somit die automatische Inbetriebnahme von Systemen zur Sprecherlokalisierung zu gestatten. Die angesichts der Kalibrierung durch Einfallswinkel auftretende Skalierungsproblematik konnte indes, bei einer rein akustischen Kalibrierung der Sensornetze, nur mithilfe der ansonsten konsequent vermiedenen Abtastsynchronisation zwischen den Sensoren zufriedenstellend gelöst werden. Einen Ausweg bieten dagegen modalitätsübergreifende Ansätze, die bislang nur wegen der visuellen Winkelschätzung schlechter abschneiden.

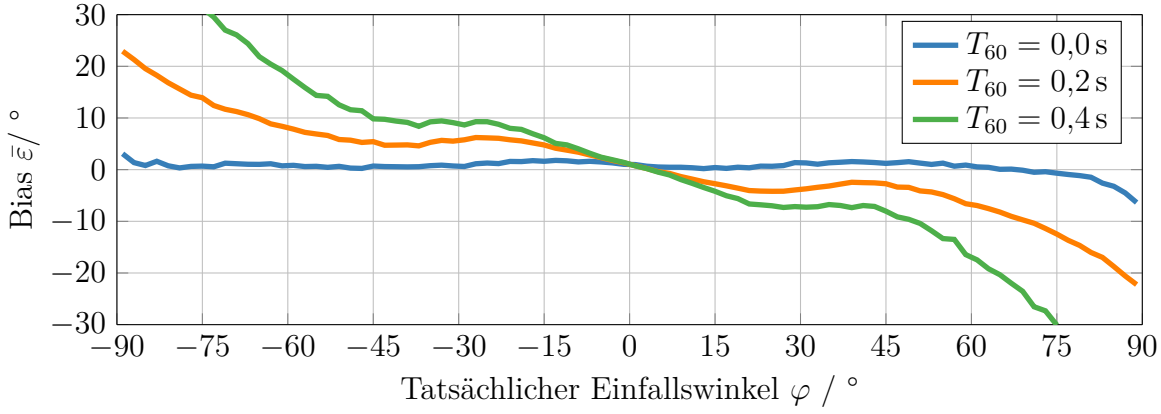
A Einfallswinkelschätzung

A.1 Bias linearer Mikrofonarrays

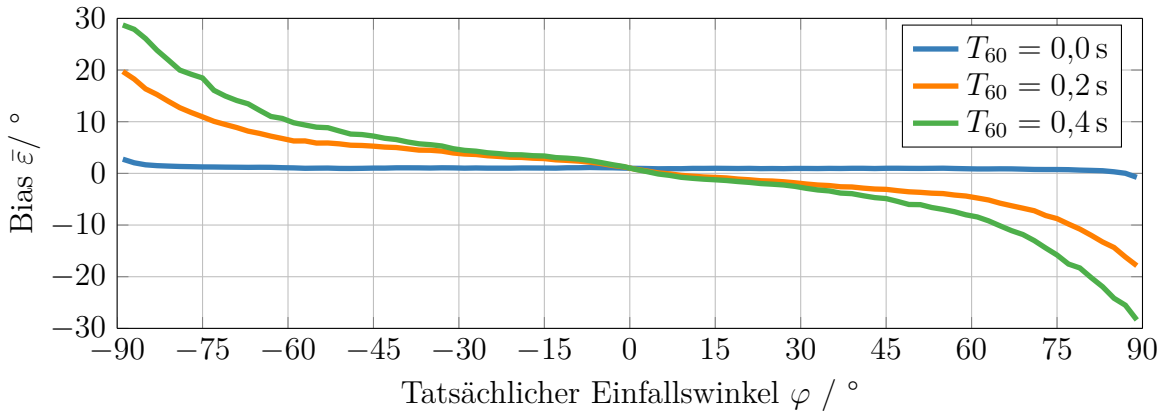
Die Simulationen in Abschnitt 4.4 belegen, dass es beim Einsatz eines linearen Mikrofonarrays zu einem systematischen Fehler bei der Einfallswinkelschätzung kommt, der vom tatsächlichen Einfallswinkel abhängt. Während Abb. 4.6 den Bias exemplarisch für FSBPhat darstellt, zeigt Abb. A.1 diesen für die ebenfalls im Rahmen der Arbeit untersuchten Algorithmen.



(a) LYDE



(b) SRPPhat



(c) WKM

Abbildung A.1: Richtungsabhängigkeit des Bias der Einfallswinkelschätzung eines zweielementigen Mikrofonarrays mit 0,05 m Mikrofonabstand bei Nutzung des LYDE (a), SRPPhat (b) oder der WKM (c).

A.2 Bestimmung des TDOA-Bias

Die Untersuchungen aus Abschnitt 4.4 zeigen die Existenz eines Bias bei der Einfallswinkelschätzung durch lineare Mikrofonarrays. Ein Teil dieses Bias wird dabei durch die Begrenzung der Signallaufzeitdifferenzen auf das Intervall $[-\tau_{\max}; \tau_{\max}]$ verursacht. Die folgenden Ausführungen erläutern nun die analytische Berechnung des bereits in Abb. 4.13 dargestellten Bias. Den Ausgangspunkt dafür bildet die TDOA-Messung $\hat{\tau}$, die gemäß Abschnitt 4.5, als normalverteilte Zufallsvariable mit dem Mittelwert $\bar{\tau}$ und der Standardabweichung σ_τ modelliert wird. Durch die Begrenzung von $\hat{\tau}$ auf das Intervall $[-\tau_{\max}; \tau_{\max}]$ entsteht die Zufallsvariable $\hat{\tau}_\mathrm{c}$, die die Verteilungsdichte

$$\mathrm{p}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c}; \bar{\tau}, \sigma_\tau) = c_\mathrm{L} \cdot \delta(\hat{\tau}_\mathrm{c} + \tau_{\max}) + \mathrm{rect}\left(\frac{\hat{\tau}_\mathrm{c}}{2 \cdot \tau_{\max}}\right) \cdot \mathrm{p}_{\hat{\tau}}(\hat{\tau}_\mathrm{c}; \bar{\tau}, \sigma_\tau) + c_\mathrm{U} \cdot \delta(\hat{\tau}_\mathrm{c} - \tau_{\max}), \tag{A.1}$$

$$\text{mit} \quad c_\mathrm{L} = 1 - \mathrm{Q}\left(\frac{-\tau_{\max} - \bar{\tau}}{\sigma_\tau}\right) \quad \text{und} \quad c_\mathrm{U} = \mathrm{Q}\left(\frac{\tau_{\max} - \bar{\tau}}{\sigma_\tau}\right) \tag{A.2}$$

besitzt (vgl. Gl. (4.31)).

Der Erwartungswert von $\hat{\tau}_\mathrm{c}$ kann, aufgrund der Linearität des Erwartungswertoperators, getrennt für jeden Summanden bestimmt werden:

$$\mathbb{E}[\hat{\tau}_\mathrm{c}] = \mathbb{E}[c_\mathrm{L} \cdot \delta(\hat{\tau}_\mathrm{c} + \tau_{\max})] + \mathbb{E}\left[\mathrm{rect}\left(\frac{\hat{\tau}_\mathrm{c}}{2 \cdot \tau_{\max}}\right) \cdot \mathrm{p}_{\hat{\tau}}(\hat{\tau}_\mathrm{c}; \bar{\tau}, \sigma_\tau)\right] + \mathbb{E}[c_\mathrm{U} \cdot \delta(\hat{\tau}_\mathrm{c} - \tau_{\max})] \tag{A.3}$$

Für den ersten bzw. letzten Term ergeben sich die Erwartungswerte unmittelbar zu

$$\mathbb{E}[c_\mathrm{L} \cdot \delta(\hat{\tau}_\mathrm{c} + \tau_{\max})] = -\tau_{\max} \cdot c_\mathrm{L} \quad \text{bzw.} \quad \mathbb{E}[c_\mathrm{U} \cdot \delta(\hat{\tau}_\mathrm{c} - \tau_{\max})] = \tau_{\max} \cdot c_\mathrm{U}. \tag{A.4}$$

Der Erwartungswert der abgeschnittenen Normalverteilung

$$\mathbb{E}\left[\mathrm{rect}\left(\frac{\hat{\tau}_\mathrm{c}}{2\cdot\tau_\mathrm{max}}\right)\cdot \mathrm{p}_{\hat{\tau}}(\hat{\tau}_\mathrm{c};\bar{\tau},\sigma_\tau)\right] = \int\limits_{-\tau_\mathrm{max}}^{\tau_\mathrm{max}} \hat{\tau}_\mathrm{c}\cdot \mathrm{p}_{\hat{\tau}}(\hat{\tau}_\mathrm{c};\bar{\tau},\sigma_\tau)\,\mathrm{d}\hat{\tau}_\mathrm{c} \tag{A.5}$$

lässt sich bspw. durch die Substitution

$$\hat{\tau}_\mathrm{c}{}' = \frac{\hat{\tau}_\mathrm{c}-\bar{\tau}}{\sigma_\tau} \tag{A.6}$$

lösen. Die Anwendung dieser Substitution liefert schließlich

$$\mathbb{E}\left[\mathrm{rect}\left(\frac{\hat{\tau}_\mathrm{c}}{2\cdot\tau_\mathrm{max}}\right)\cdot \mathrm{p}_{\hat{\tau}}(\hat{\tau}_\mathrm{c};\bar{\tau},\sigma_\tau)\right] = \int\limits_{\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sigma_\tau}}^{\frac{\tau_\mathrm{max}-\bar{\tau}}{\sigma_\tau}} \frac{\sigma_\tau\hat{\tau}_\mathrm{c}{}'+\bar{\tau}}{\sqrt{2\pi}}\cdot\exp\left(-\frac{1}{2}\hat{\tau}_\mathrm{c}{}'^2\right)\mathrm{d}\hat{\tau}_\mathrm{c}{}' \tag{A.7}$$

$$\begin{aligned} &= \frac{\sigma_\tau}{\sqrt{2\pi}}\cdot\int\limits_{\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sigma_\tau}}^{\frac{\tau_\mathrm{max}-\bar{\tau}}{\sigma_\tau}} \hat{\tau}_\mathrm{c}{}'\cdot\exp\left(-\frac{1}{2}\hat{\tau}_\mathrm{c}{}'^2\right)\mathrm{d}\hat{\tau}_\mathrm{c}{}' \\ &\quad + \frac{\bar{\tau}}{\sqrt{2\pi}}\cdot\int\limits_{\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sigma_\tau}}^{\frac{\tau_\mathrm{max}-\bar{\tau}}{\sigma_\tau}} \exp\left(-\frac{1}{2}\hat{\tau}_\mathrm{c}{}'^2\right)\mathrm{d}\hat{\tau}_\mathrm{c}{}' \end{aligned} \tag{A.8}$$

$$\begin{aligned} &= \frac{\sigma_\tau\cdot\left[\exp\left(-\left(\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)^2\right)-\exp\left(-\left(\frac{\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)^2\right)\right]}{\sqrt{2\pi}} \\ &\quad + \frac{\bar{\tau}}{2}\cdot\left[\mathrm{erf}\left(\frac{\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)-\mathrm{erf}\left(\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)\right]. \end{aligned} \tag{A.9}$$

Zwar kann insbesondere der erste Term von Gl. (A.9) auch anders dargestellt werden, allerdings treten dann bei der Auswertung z. T. numerische Probleme auf. Deshalb wird an dieser Stelle auf eine weitere Vereinfachung verzichtet. Insgesamt ergibt sich der gesuchte Erwartungswert durch die Kombination der Teilergebnisse aus Gl. (A.4) und Gl. (A.9) zu

$$\begin{aligned} \mathbb{E}[\hat{\tau}_\mathrm{c}] = &-\tau_\mathrm{max}\cdot c_\mathrm{L} + \tau_\mathrm{max}\cdot c_\mathrm{U} \\ &+ \frac{\bar{\tau}}{2}\cdot\left[\mathrm{erf}\left(\frac{\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)-\mathrm{erf}\left(\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)\right] \\ &+ \frac{\sigma_\tau\cdot\left[\exp\left(-\left(\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)^2\right)-\exp\left(-\left(\frac{\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)^2\right)\right]}{\sqrt{2\pi}}. \end{aligned} \tag{A.10}$$

Die Berechnung des gesuchten Bias der TDOA erfordert schließlich noch die Subtraktion von $\bar{\tau}$ von $\mathbb{E}[\hat{\tau}_\mathrm{c}]$. Eine graphische Darstellung des daraus resultierenden Bias liefert die bereits zuvor erwähnte Abb. 4.13.

A.3 Bestimmung des DOA-Bias

Die Ausführungen im Anhang A.2 haben die Berechnung des Bias für die TDOA dargelegt. In diesem Abschnitt wird nun die Bestimmung des Bias der daraus gewonnenen DOA betrachtet. Dazu wird zunächst anhand der aus Gl. (4.31) bzw. Gl. (A.1) bekannten Verteilungsdichte der TDOA die Verteilungsdichte der DOA ermittelt. Im Anschluss daran dient die Verteilungsdichte der DOA zur Ermittlung des Bias der DOA.

Um aus der Verteilungsdichte der TDOA die Verteilungsdichte der DOA zu berechnen wird zunächst die Verteilungsfunktion der TDOA bestimmt:

$$\mathrm{P}_{\hat{\tau}_\mathrm{c}}(\hat{\tau}_\mathrm{c};\bar{\tau},\sigma_\tau) = \begin{cases} c_\mathrm{L} & \text{für } \hat{\tau}_\mathrm{c} = -\tau_\mathrm{max} \\ c_\mathrm{L} + \frac{1}{2}\cdot\left(\mathrm{erf}\left(\frac{\hat{\tau}_\mathrm{c}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right) - \mathrm{erf}\left(\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)\right) & \text{für } |\hat{\tau}_\mathrm{c}| < \tau_\mathrm{max} \\ c_\mathrm{L} + \frac{1}{2}\cdot\left(\mathrm{erf}\left(\frac{\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right) - \mathrm{erf}\left(\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)\right) + c_\mathrm{U} & \text{für } \hat{\tau}_\mathrm{c} = \tau_\mathrm{max} \end{cases} . \tag{A.11}$$

Mithilfe der Umkehrfunktion von Gl. (4.35)

$$\hat{\tau}_\mathrm{c} = \sin(\hat{\varphi})\cdot\tau_\mathrm{max} \tag{A.12}$$

entsteht die Verteilungsfunktion des Einfallswinkels

$$\mathrm{P}_{\hat{\varphi}}(\hat{\varphi};\bar{\tau},\sigma_\tau) = \begin{cases} c_\mathrm{L} & \text{für } \hat{\varphi} = -\frac{\pi}{2} \\ c_\mathrm{L} + \frac{1}{2}\cdot\left(\mathrm{erf}\left(\frac{\sin(\hat{\varphi})\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right) - \mathrm{erf}\left(\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)\right) & \text{für } |\hat{\varphi}| < \frac{\pi}{2} \\ c_\mathrm{L} + \frac{1}{2}\cdot\left(\mathrm{erf}\left(\frac{\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right) - \mathrm{erf}\left(\frac{-\tau_\mathrm{max}-\bar{\tau}}{\sqrt{2}\cdot\sigma_\tau}\right)\right) + c_\mathrm{U} & \text{für } \hat{\varphi} = \frac{\pi}{2} \end{cases} . \tag{A.13}$$

Durch die Ableitung dieser Funktion ergibt sich die Verteilungsdichte der DOA

$$\mathrm{p}_{\hat{\varphi}}(\hat{\varphi};\bar{\tau},\sigma_\tau) = c_\mathrm{L}\cdot\delta\left(\hat{\varphi}+\frac{\pi}{2}\right) + \tau_\mathrm{max}\cdot\cos(\hat{\varphi})\cdot\mathcal{N}(\sin(\hat{\varphi})\cdot\tau_\mathrm{max};\bar{\tau},\sigma_\tau) + c_\mathrm{U}\cdot\delta\left(\hat{\varphi}-\frac{\pi}{2}\right). \tag{A.14}$$

Die Bestimmung des Erwartungswertes kann analog zu Gl. (A.3) ebenfalls getrennt für alle drei Summanden erfolgen. Für den ersten bzw. letzten Term gilt

$$\mathbb{E}\left[c_\mathrm{L}\cdot\delta\left(\hat{\varphi}+\frac{\pi}{2}\right)\right] = -\frac{\pi}{2}\cdot c_\mathrm{L} \quad \text{und} \quad \mathbb{E}\left[c_\mathrm{U}\cdot\delta\left(\hat{\varphi}-\frac{\pi}{2}\right)\right] = \frac{\pi}{2}\cdot c_\mathrm{U}. \tag{A.15}$$

Der mittlere Term aus Gl. (A.14)

$$\mathbb{E}[\tau_\mathrm{max}\cdot\cos(\hat{\varphi})\cdot\mathcal{N}(\sin(\hat{\varphi})\cdot\tau_\mathrm{max};\bar{\tau},\sigma_\tau)] = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \hat{\varphi}\cdot\tau_\mathrm{max}\cdot\cos(\hat{\varphi})\cdot\mathcal{N}(\sin(\hat{\varphi})\cdot\tau_\mathrm{max};\bar{\tau},\sigma_\tau)\,\mathrm{d}\hat{\varphi} \tag{A.16}$$

wird aufgrund der Komplexität des Integrals nummerisch ausgewertet. Letztendlich entsteht somit durch die Subtraktion von $\hat{\varphi}$ vom Erwartungswert der in Abb. 4.15 veranschaulichte Bias.

B Zielfunktionen

B.1 Formulierung der Zielfunktion

Zur Entwicklung der bereits in Abschnitt 5.1 dargestellten Zielfunktion wird der Geometrischezusammenhang

$$\tan(\theta_i + \varphi_{i,d}) = \frac{\boldsymbol{z}_2^{\mathrm{T}} (\boldsymbol{e}_d - \boldsymbol{s}_i)}{\boldsymbol{z}_1^{\mathrm{T}} (\boldsymbol{e}_d - \boldsymbol{s}_i)} \tag{B.1}$$

zunächst mithilfe von Additionstheoremen in

$$\frac{\tan(\theta_i) + \tan(\varphi_{i,d})}{1 - \tan(\theta_i) \cdot \tan(\varphi_{i,d})} = \frac{\boldsymbol{z}_2^{\mathrm{T}} (\boldsymbol{e}_d - \boldsymbol{s}_i)}{\boldsymbol{z}_1^{\mathrm{T}} (\boldsymbol{e}_d - \boldsymbol{s}_i)} \tag{B.2}$$

überführt. Durch die Auswertung der Skalarprodukte $\boldsymbol{z}_1^{\mathrm{T}} (\boldsymbol{e}_d - \boldsymbol{s}_i)$ und $\boldsymbol{z}_2^{\mathrm{T}} (\boldsymbol{e}_d - \boldsymbol{s}_i)$ sowie eine anschließende Umformung der Gleichung, sodass sich alle Terme auf einer Seite des Gleichheitszeichen befinden, ergibt sich

$$(\tan(\theta_i) + \tan(\varphi_{i,d})) \cdot (a_d - x_i) - (1 - \tan(\theta_i) \cdot \tan(\varphi_{i,d})) \cdot (b_d - y_i) = 0. \tag{B.3}$$

Die gewonnene Formulierung lässt sich erneut als Skalarprodukt darstellen und somit entsteht die in Abschnitt 5.1 genutzte Zielfunktion:

$$f_{\mathrm{Tan}}(\boldsymbol{s}_i, \theta_i, \boldsymbol{e}_d; \varphi_{i,d}) = (\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}} \begin{bmatrix} -\tan(\theta_i) - \tan(\varphi_{i,d}) \\ 1 - \tan(\theta_i) \cdot \tan(\varphi_{i,d}) \end{bmatrix} = 0. \tag{B.4}$$

B.2 Entfernung der Polstellen

Die Basis des in [KWL08] präsentierten Geometriekalibrierungsalgorithmus zur Bestimmung der Positionen und Orientierungen eines Sensornetzes mithilfe von Einfallswinkelschätzungen bildet die Zielfunktion aus Gl. (B.1). Die Optimierung dieser Zielfunktion erlaubt zwar die Bestimmung der gesuchten Parameter, aber die darin enthaltenen Tangensfunktionen haben, aufgrund ihrer Polstellen, negative Auswirkungen auf die Präzision des Ansatzes (vgl. Abschnitt 5.2). Zur Beseitigung des Einflusses der Polstellen erfolgt eine Reformulierung des Optimierungsproblems. Ausgangspunkt dafür ist die bereits aus Gl. (5.3) bzw. Gl. (B.4) bekannte Formulierung:

$$f_{\mathrm{Tan}}(\boldsymbol{s}_i, \theta_i, \boldsymbol{e}_d; \varphi_{i,d}) = (\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}} \begin{bmatrix} -\tan(\theta_i) - \tan(\varphi_{i,d}) \\ 1 - \tan(\theta_i) \tan(\varphi_{i,d}) \end{bmatrix}. \tag{B.5}$$

Sofern die Tangensfunktionen durch Quotienten aus Sinus- und Kosinusfunktionen ersetzt werden entsteht:

$$f_{\mathrm{Tan}}(\boldsymbol{s}_i, \theta_i, \boldsymbol{e}_d; \varphi_{i,d}) = (\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}} \begin{bmatrix} -\frac{\sin(\theta_i)}{\cos(\theta_i)} - \frac{\sin\left(\varphi_{i,d}\right)}{\cos\left(\varphi_{i,d}\right)} \\ 1 - \frac{\sin(\theta_i)}{\cos(\theta_i)} \cdot \frac{\sin\left(\varphi_{i,d}\right)}{\cos\left(\varphi_{i,d}\right)} \end{bmatrix}. \tag{B.6}$$

Die Multiplikation dieser Gleichung mit den im Nenner auftretenden Kosinusfunktionen und eine anschließende Sortierung der Terme führt schließlich zu:

$$f_{\mathrm{SinCos}}(\boldsymbol{s}_i, \theta_i, \boldsymbol{e}_d; \varphi_{i,d}) = (\boldsymbol{e}_d - \boldsymbol{s}_i)^{\mathrm{T}} \begin{bmatrix} -\sin(\theta_i) & -\cos(\theta_i) \\ \cos(\theta_i) & -\sin(\theta_i) \end{bmatrix} \begin{bmatrix} \cos(\varphi_{i,d}) \\ \sin(\varphi_{i,d}) \end{bmatrix}. \tag{B.7}$$

Somit werden die Polstellen aus Gl. (B.5) eliminiert und der Einsatz von Gl. (B.7) zur Geometriekalibrierung reduziert den Anteil divergenter Lösungsversuche (siehe Abb. 5.6). Die Rotationsinvarianz, die ebenfalls Probleme bei der Kalibrierung verursacht, ist jedoch weiterhin vorhanden, wie die Betrachtungen in Abschnitt 5.3 gezeigt haben.

C Koordinatentransformation

C.1 Parameterschätzung in einem *Shape*-Bereich

Den Grundstein für die Schätzung der RBT-Parameter durch eine Transformation der Problemstellung in einen *Shape*-Bereich bildet das bereits aus Gl. (8.15) bekannte Optimierungsproblem. Während konventionelle Verfahren KENDAL-Koordinaten zur Darstellung der Landmarken in einem *Shape*-Bereich verwenden [DM98], erfolgt die Transformation im Rahmen dieser Arbeit mithilfe der DFT.

Unter Verwendung der Matrizen $\boldsymbol{F}$ bzw. $\frac{1}{\mathscr{K}}\boldsymbol{F}^{\mathrm{H}}$ zur Durchführung der DFT bzw. IDFT entsteht aus Gl. (8.15) das auch schon in Gl. (8.17) präsentierte Optimierungsproblem:

$$\langle \hat{\alpha}, \hat{\beta} \rangle = \underset{\alpha,\beta}{\operatorname{argmin}}\, \mathcal{J}\left(\alpha, \beta\right) \tag{C.1}$$

mit

$$\mathcal{J}\left(\alpha, \beta\right) = \frac{1}{\mathscr{K}} \left(\boldsymbol{w} - \alpha \cdot \boldsymbol{u} - \beta \cdot \mathbf{1}\right)^{\mathrm{H}} \boldsymbol{F}^{\mathrm{H}} \boldsymbol{F} \left(\boldsymbol{w} - \alpha \cdot \boldsymbol{u} - \beta \cdot \mathbf{1}\right). \tag{C.2}$$

Die beiden Formen $\boldsymbol{w}$ bzw. $\boldsymbol{u}$, die es durch die komplexwertigen RBT-Parameter α und β aufeinander abzubilden gilt, bestehen dabei jeweils aus $\mathscr{K}$ ebenfalls komplexwertigen Landmarken (vgl. Abschnitt 6.5). Weiterhin lassen sich die Formen in den Mittelwert der Landmarken ($\bar{\boldsymbol{w}}$ bzw. $\bar{\boldsymbol{u}}$), sowie den verbleibenden Rest ($\tilde{\boldsymbol{w}}$ bzw. $\tilde{\boldsymbol{u}}$) aufspalten:

$$\boldsymbol{w} = \bar{\boldsymbol{w}} + \tilde{\boldsymbol{w}} \quad \text{bzw.} \quad \boldsymbol{u} = \bar{\boldsymbol{u}} + \tilde{\boldsymbol{u}}. \tag{C.3}$$

Angesichts der zuvor eingeführten Zerlegung entsteht aus der Kostenfunktion aus Gl. (C.2):

$$\mathcal{J}\left(\alpha, \beta\right) = \frac{1}{\mathscr{K}} \left((\tilde{\boldsymbol{w}}+\bar{\boldsymbol{w}}) - \alpha \cdot (\tilde{\boldsymbol{u}}+\bar{\boldsymbol{u}}) - \beta \cdot \mathbf{1}\right)^{\mathrm{H}} \boldsymbol{F}^{\mathrm{H}} \boldsymbol{F} \left((\tilde{\boldsymbol{w}}+\bar{\boldsymbol{w}}) - \alpha \cdot (\tilde{\boldsymbol{u}}+\bar{\boldsymbol{u}}) - \beta \cdot \mathbf{1}\right). \tag{C.4}$$

Weiterhin liefert die Multiplikation aller Formen mit $\boldsymbol{F}$ die zugehörigen Repräsentationen im FOURIER-Bereich:

$$\tilde{\boldsymbol{u}}_{\mathrm{F}} = \boldsymbol{F}\,\tilde{\boldsymbol{u}}, \quad \tilde{\boldsymbol{w}}_{\mathrm{F}} = \boldsymbol{F}\,\tilde{\boldsymbol{w}}, \quad \bar{\boldsymbol{u}}_{\mathrm{F}} = \boldsymbol{F}\,\bar{\boldsymbol{u}}, \quad \bar{\boldsymbol{w}}_{\mathrm{F}} = \boldsymbol{F}\,\bar{\boldsymbol{w}} \quad \text{und} \quad \beta\boldsymbol{z}_1 = \beta \cdot \boldsymbol{F}\,\mathbf{1} \tag{C.5}$$

Diese Repräsentationen gestatten es wiederum, Gl. (C.4) im FOURIER- bzw. *Shape*-Bereich wie folgt darzustellen:

$$\mathcal{J}\left(\alpha, \beta\right) = \frac{1}{\mathscr{K}} \left((\tilde{\boldsymbol{w}}_{\mathrm{F}}+\bar{\boldsymbol{w}}_{\mathrm{F}}) - \alpha \cdot (\tilde{\boldsymbol{u}}_{\mathrm{F}}+\bar{\boldsymbol{u}}_{\mathrm{F}}) - \beta\cdot\boldsymbol{z}_1\right)^{\mathrm{H}} \left((\tilde{\boldsymbol{w}}_{\mathrm{F}}+\bar{\boldsymbol{w}}_{\mathrm{F}}) - \alpha \cdot (\tilde{\boldsymbol{u}}_{\mathrm{F}}+\bar{\boldsymbol{u}}_{\mathrm{F}}) - \beta\cdot\boldsymbol{z}_1\right). \tag{C.6}$$

Darüber hinaus ist noch hervorzuheben, dass die erste Komponente der Vektoren $\tilde{\boldsymbol{u}}_\mathrm{F}$ und $\tilde{\boldsymbol{w}}_\mathrm{F}$ den Wert Null besitzt, weil die zugehörigen Vektoren $\tilde{\boldsymbol{u}}$ und $\tilde{\boldsymbol{w}}$ aufgrund der durchgeführten Zerlegung mittelwertfrei sind. Bei $\bar{\boldsymbol{w}}_\mathrm{F}$, $\bar{\boldsymbol{u}}_\mathrm{F}$ und $\boldsymbol{z}_1$ tritt hingegen genau der komplementäre Fall auf, da diese die FOURIER-Transformierten von Konstanten darstellen. Dementsprechend entfallen beim Auflösen der Klammern von Gl. (C.6) alle Produkte von mittelwertfreien und mittelwertbehafteten Komponenten. Insgesamt entsteht somit die Zielfunktion:

$$\begin{aligned}\mathcal{J}\left(\alpha,\beta\right) =& \frac{1}{\mathcal{K}}\Big[\tilde{\boldsymbol{w}}_\mathrm{F}^\mathrm{H}\,\tilde{\boldsymbol{w}}_\mathrm{F} - \alpha\cdot\tilde{\boldsymbol{w}}_\mathrm{F}^\mathrm{H}\,\tilde{\boldsymbol{u}}_\mathrm{F} + \bar{\boldsymbol{w}}_\mathrm{F}\,\bar{\boldsymbol{w}}_\mathrm{F} - \alpha\cdot\bar{\boldsymbol{w}}_\mathrm{F}\,\bar{\boldsymbol{u}}_\mathrm{F} - \beta\cdot\bar{\boldsymbol{w}}_\mathrm{F}\,\boldsymbol{z}_1 \\ & - \alpha^\mathrm{H}\cdot\tilde{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\tilde{\boldsymbol{w}}_\mathrm{F} + \alpha^\mathrm{H}\alpha\cdot\tilde{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\tilde{\boldsymbol{u}}_\mathrm{F} - \alpha^\mathrm{H}\cdot\bar{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\bar{\boldsymbol{w}}_\mathrm{F} + \alpha^\mathrm{H}\alpha\cdot\bar{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\bar{\boldsymbol{u}}_\mathrm{F} + \alpha^\mathrm{H}\beta\cdot\bar{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\boldsymbol{z}_1 \\ & - \beta^\mathrm{H}\cdot\boldsymbol{z}_1^\mathrm{H}\,\bar{\boldsymbol{w}}_\mathrm{F} + \beta^\mathrm{H}\alpha\cdot\boldsymbol{z}_1^\mathrm{H}\,\bar{\boldsymbol{u}}_\mathrm{F} + \beta^\mathrm{H}\beta\cdot\boldsymbol{z}_1^\mathrm{H}\,\boldsymbol{z}_1\Big]. \end{aligned} \tag{C.7}$$

Die Bestimmung des Optimums erfordert zunächst die Berechnung der partiellen Ableitungen. Diese ergeben sich zu

$$\frac{\mathrm{d}}{\mathrm{d}\alpha^\mathrm{H}}\mathcal{J}\left(\alpha,\beta\right) = -\tilde{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\tilde{\boldsymbol{w}}_\mathrm{F} + \alpha\cdot\tilde{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\tilde{\boldsymbol{u}}_\mathrm{F} - \bar{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\bar{\boldsymbol{w}}_\mathrm{F} + \alpha\cdot\bar{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\bar{\boldsymbol{u}}_\mathrm{F} + \beta\cdot\bar{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\boldsymbol{z}_1 \tag{C.8}$$

bzw.

$$\frac{\mathrm{d}}{\mathrm{d}\beta^\mathrm{H}}\mathcal{J}\left(\alpha,\beta\right) = -\boldsymbol{z}_1^\mathrm{H}\,\bar{\boldsymbol{w}}_\mathrm{F} + \alpha\cdot\boldsymbol{z}_1^\mathrm{H}\,\bar{\boldsymbol{u}}_\mathrm{F} + \beta\cdot\boldsymbol{z}_1^\mathrm{H}\,\boldsymbol{z}_1. \tag{C.9}$$

Die anschließende Nullstellensuche für die Ableitung aus Gl. (C.9) liefert

$$\beta = \boldsymbol{z}_1^\mathrm{H}\,\bar{\boldsymbol{w}}_\mathrm{F} - \alpha\cdot\boldsymbol{z}_1\,\bar{\boldsymbol{u}}_\mathrm{F}. \tag{C.10}$$

Unter Berücksichtigung dieses Zwischenergebnisses ergibt sich die Nullstelle von Gl. (C.8) zu

$$\alpha = \frac{\tilde{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\tilde{\boldsymbol{w}}_\mathrm{F}}{\tilde{\boldsymbol{u}}_\mathrm{F}^\mathrm{H}\,\tilde{\boldsymbol{u}}_\mathrm{F}}. \tag{C.11}$$

Die Betrachtung der Lösungen zeigt, dass Rotation und Skalierung somit unabhängig von der Translation ermittelt werden können. Erst die Bestimmung der Translation erfordert die Berücksichtigung von Rotation und Skalierung. Außerdem lässt sich die bislang verwendete Aufspaltung der Formen (vgl. Gl. (C.3)) nun wieder rückgängig machen. Die Terme $\tilde{\boldsymbol{w}}_\mathrm{F}$ und $\tilde{\boldsymbol{u}}_\mathrm{F}$ in Gl. (C.10) repräsentieren ausschließlich die Mittelwerte der Formen $\boldsymbol{w}$ und $\boldsymbol{u}$ und ergeben sich daher direkt aus dem ersten Koeffizient der FOURIER-Transformation von $\boldsymbol{w}$ bzw. $\boldsymbol{u}$. Analog dazu können auch $\bar{\boldsymbol{w}}_\mathrm{F}$ und $\bar{\boldsymbol{u}}_\mathrm{F}$ in Gl. (C.11) durch die Frequenzen 2 bis $\mathcal{K}$ der FOURIER-Transformation von $\boldsymbol{w}$ bzw. $\boldsymbol{u}$ dargestellt werden. Demnach ergeben sich die RBT-Parameter zu

$$\alpha = \frac{\left\{\boldsymbol{u}_\mathrm{F}^\mathrm{H}\right\}_{2:\mathcal{K}}\left\{\boldsymbol{w}_\mathrm{F}\right\}_{2:\mathcal{K}}}{\left\{\boldsymbol{u}_\mathrm{F}^\mathrm{H}\right\}_{2:\mathcal{K}}\left\{\boldsymbol{u}_\mathrm{F}\right\}_{2:\mathcal{K}}} \quad\text{und}\quad \beta = \frac{1}{\mathcal{K}}\left(\left\{\boldsymbol{w}_\mathrm{F}\right\}_1 - \alpha\cdot\left\{\boldsymbol{u}_\mathrm{F}\right\}_1\right). \tag{C.12}$$

Bei der Berechnung von β tritt ein zusätzlicher Faktor $\frac{1}{\mathcal{K}}$ auf, da $\bar{\boldsymbol{u}}$ gemäß der Definition den Mittelwert beschreibt und die FOURIER-Transformation mit $\boldsymbol{F}$ die Summe der Elemente liefert.

C.2 Laufzeitanalyse

Dieses Kapitel befasst sich mit der Laufzeitanalyse der RBT-Parameterschätzung. Dazu wird die asymptotische Laufzeit der Schätzung in einem *Shape*-Bereich (vgl. Abschnitt 8.3 bzw. Anhang C.1) mit der konventionellen, SVD-basierten Variante aus [Cha95] verglichen. Die einzelnen Operationen der beiden Algorithmen sind zusammen mit der jeweiligen Laufzeitabschätzung in den Tab. C.1 und C.2 angegeben.

Operation			Laufzeitabschätzung
$\bar{\boldsymbol{e}}^{\mathrm{A}}$	$=$	$\frac{1}{D} \cdot \sum_{d=1}^{D} \boldsymbol{e}_d^{\mathrm{A}}$	$\mathcal{O}(D)$
$\bar{\boldsymbol{e}}^{\mathrm{V}}$	$=$	$\frac{1}{D} \cdot \sum_{d=1}^{D} \boldsymbol{e}_d^{\mathrm{V}}$	$\mathcal{O}(D)$
$\boldsymbol{\Gamma}$	$=$	$\frac{1}{D} \cdot \sum_{d=1}^{D} \left(\boldsymbol{e}_d^{\mathrm{V}} - \bar{\boldsymbol{e}}^{\mathrm{V}}\right)\left(\boldsymbol{e}_d^{\mathrm{A}} - \bar{\boldsymbol{e}}^{\mathrm{A}}\right)^{\mathrm{T}}$	$\mathcal{O}(D)$
$\boldsymbol{\Upsilon}\,\boldsymbol{\Sigma}\,\boldsymbol{\Psi}^{\mathrm{T}}$	$=$	$\mathrm{svd}(\boldsymbol{\Gamma})$	$\mathcal{O}(1)$
$\boldsymbol{R}$	$=$	$\boldsymbol{\Upsilon}\,\boldsymbol{\Psi}^{\mathrm{T}}$	$\mathcal{O}(1)$
$(\sigma^{\mathrm{A}})^2$	$=$	$\frac{1}{D} \cdot \sum_{d=1}^{D} \left(\boldsymbol{e}_d^{\mathrm{A}} - \bar{\boldsymbol{e}}^{\mathrm{A}}\right)^2$	$\mathcal{O}(D)$
ν	$=$	$\frac{1}{(\sigma^{\mathrm{A}})^2} \cdot \mathrm{tr}\left(\boldsymbol{R}^{\mathrm{T}}\,\boldsymbol{\Gamma}\right)$	$\mathcal{O}(1)$
$\boldsymbol{t}$	$=$	$\bar{\boldsymbol{e}}^{\mathrm{V}} - \nu \cdot \boldsymbol{R}\,\bar{\boldsymbol{e}}^{\mathrm{A}}$	$\mathcal{O}(1)$

Tabelle C.1: Laufzeitanalyse: RBT-Parameterschätzung unter Verwendung einer SVD.

Operation			Laufzeitabschätzung
$\boldsymbol{u}_{\mathrm{F}}$	$=$	$\boldsymbol{F}\,\boldsymbol{u}$	$\mathcal{O}\left(\lceil \log_2 D \rceil \cdot 2^{\lceil \log_2(D) \rceil}\right)$
$\boldsymbol{w}_{\mathrm{F}}$	$=$	$\boldsymbol{F}\,\boldsymbol{w}$	$\mathcal{O}\left(\lceil \log_2 D \rceil \cdot 2^{\lceil \log_2(D) \rceil}\right)$
α	$=$	$\frac{\left\{\boldsymbol{u}_{\mathrm{F}}{}^{\mathrm{H}}\right\}_{2:D}\{\boldsymbol{w}_{\mathrm{F}}\}_{2:D}}{\left\{\boldsymbol{u}_{\mathrm{F}}{}^{\mathrm{H}}\right\}_{2:D}\{\boldsymbol{u}_{\mathrm{F}}\}_{2:D}}$	$\mathcal{O}(D)$
β	$=$	$\left\{\boldsymbol{w}_{\mathrm{F}}\right\}_1 - \alpha \cdot \left\{\boldsymbol{u}_{\mathrm{F}}\right\}_1$	$\mathcal{O}(1)$

Tabelle C.2: Laufzeitanalyse: RBT-Parameterschätzung in einem *Shape*-Bereich.

Die Betrachtung sämtlicher Operationen der SVD-Variante zeigt, dass die einzelnen Schritte höchstens linear von der Anzahl der vorliegenden Datenpunkte (D) abhängen. Die SVD besitzt sogar eine konstante Ausführungszeit, weil die Größe der Dispersions- bzw. Kovarianzmatrix unabhängig von der Anzahl der Datenpunkte ist. Insgesamt hängt die asymptotische Laufzeit der RBT-Parameterschätzung mithilfe der SVD daher linear von D ab.

Die Laufzeit des im Rahmen dieser Arbeit entwickelten Ansatzes wird hingegen durch die Berechnung der FOURIER-Transformation der jeweiligen Formen dominiert. Diese wurde bislang stets durch eine Multiplikation mit der Matrix $\boldsymbol{F}$ ausgedrückt. Allerdings bietet der FFT-Algorithmus eine wesentlich effizientere Möglichkeit, die FOURIER-Transformation zu realisieren. Deshalb erfolgt die Transformation durch die

FFTW, die eine Implementierung des FFT-Algorithmus darstellt, deren asymptotische Laufzeit $\mathcal{O}(\lceil \log_2 D \rceil 2^{\lceil \log_2(D) \rceil})$ beträgt [FJ05].

Trotz des Einsatzes der FFTW besitzt das entwickelte Verfahren asymptotisch eine schlechtere Gesamtlaufzeit als das betrachtete Referenzverfahren. Andererseits bleiben bei der asymptotischen Betrachtung der Laufzeit die konstanten Anteile unberücksichtigt, da diese bei großen Datenmengen keine Rolle mehr spielen. Die in Abschnitt 8.3 durchgeführten Untersuchungen belegen jedoch, dass die konstanten Anteile bei dem vorgeschlagenen Algorithmus deutlich keiner als bei der SVD-Variante ausfallen. Infolgedessen gestattet die Schätzung im *Shape*-Bereich eine effizientere Berechnung der RBT-Parameter als die SVD-basierte Alternative.

Formelzeichen

Allgemeine Formelzeichen

a_d	x-Koordinate des d-ten Ereignisses.
a_d^{A}	x-Koordinate des d-ten Ereignisses der akustischen Trajektorie.
a_d^{V}	x-Koordinate des d-ten Ereignisses der visuellen Trajektorie.
b_d	y-Koordinate des d-ten Ereignisses.
b_d^{A}	y-Koordinate des d-ten Ereignisses der akustischen Trajektorie.
b_d^{V}	y-Koordinate des d-ten Ereignisses der visuellen Trajektorie.
c	Index der Kamera $\in [1, C]$.
C	Anzahl der Kameras.
c_d	z-Koordinate des d-ten Ereignisses.
c_{L}, c_{U}	Wahrscheinlichkeit, dass bei der TDOA Werte kleiner als $-\tau_{\max}$ (c_{L}) bzw. größer als $\tau_{\max}$ (c_{U}) auftreten.
$c_{\min}$	Minimale Anzahl der Beobachtungen zur Berechnung eines Modells.
c_{S}	Schallgewschingkeit.
$c_{\mathrm{W}}(\kappa)$	Normalisierungskonstante der WATSON-Verteilung in Abhängigkeit des Konzentration κ des Konzentrationsparameters.
c_α	Form der Richtcharakteristik. Omnidirektional: $c_\alpha = 1$.
d	Index des Ereignisses $\in [1, D]$.
D	Anzahl der Ereignisse.
E_0	Energie der Schallwelle zum Zeitpunkt der Aussendung.
$E(t)$	Energie der Schallwelle zum Zeitpunkt t.
$\boldsymbol{e}_d$	Position des d-ten Ereignisses: $[a_d \;\; b_d]^{\mathrm{T}}$ bzw. $[a_d \;\; b_d \;\; c_d]^{\mathrm{T}}$.
$\widetilde{\boldsymbol{e}}_d$	Position des d-ten Ereignisse vor der Skalierung.
$\boldsymbol{e}_d^{\mathrm{A}}$	Position des d-ten Ereignisses der akustischen Trajektorie.
$\boldsymbol{e}_d^{\mathrm{V}}$	Position des d-ten Ereignisses der visuellen Trajektorie.
$\bar{\boldsymbol{e}}^{\mathrm{A}}$	Mittelwert der akustischen Trajektorie $\boldsymbol{e}_d^{\mathrm{A}}$, $d = 1, \ldots, D$.
$\bar{\boldsymbol{e}}^{\mathrm{V}}$	Mittelwert der visuellen Trajektorie $\boldsymbol{e}_d^{\mathrm{V}}$, $d = 1, \ldots, D$.
$\boldsymbol{E}$	Ereignispositionen: $[\boldsymbol{e}_1 \;\; \ldots \;\; \boldsymbol{e}_D]$.
f_k	Diskrete Frequenz der DFT.
f_{PA}	Zielfunktion des erweiterten Einfallswinkelverfahrens für einen einzelnen Sensor und ein Ereignis.
$\boldsymbol{f}_{\mathrm{PA}}$	Zielfunktion des erweiterten Einfallswinkelverfahrens für alle Sensoren und Ereignisse.
$\boldsymbol{f}_{\mathrm{PA,3D}}$	Zielfunktion des erweiterten Einfallswinkelverfahrens für alle Sensoren und Ereignisse (3D).
$f_{\mathrm{PA}}^{\mathrm{V}}$	Zielfunktion des erweiterten Einfallswinkelverfahrens für eine einzelne Kamera und ein Ereignis.

$\boldsymbol{f}_{\mathrm{PA}}^{\mathrm{AV}}$	Zielfunktion des modalitätsübergreifenden erweiterten Einfallswinkelverfahrens für alle Sensoren und Ereignisse.
f_{s}	Abtastrate.
f_{SinCos}	Zielfunktion des Einfallswinkelverfahrens in Sinus-Kosinus-Notation für einen einzelnen Sensor und ein Ereignis.
$\boldsymbol{f}_{\mathrm{SinCos}}$	Zielfunktion des Einfallswinkelverfahrens in Sinus-Kosinus-Notation für alle Sensoren und ein Ereignis.
f_{Tan}	Zielfunktion des ursprünglichen Einfallswinkelverfahrens für einen einzelnen Sensor und ein Ereignis.
$\boldsymbol{f}_{\mathrm{Tan}}$	Zielfunktion des ursprünglichen Einfallswinkelverfahrens für alle Sensoren und alle Ereignisse.
$\boldsymbol{F}$	Matrix zur Transformation in den FOURIER-Bereich.
$\boldsymbol{g}_{i,d}$	Einfallsvektor vom d-ten Ereignis zum i-ten Sensor.
$\tilde{\boldsymbol{g}}_{i,d}$	Prädiktion des Einfallsvektors vom i-ten Sensor zum d-ten Ereignis.
$\tilde{h}_{(n,m)}$	Dämpfung (Pfadverlust und Mikrofondämpfung) des n-ten Kanals im Verhältnis zum m-ten Kanal.
h_{ℓ}	Element der HELMERT-Matrix $\boldsymbol{H}$.
$h_{m,\mathrm{direkt}}(l)$	LOS-Komponente (direkter Anteil) der RIA zum m-ten Mikrofon.
h	Index des Mikrofonpaares $\in [1,\, H]$.
$\boldsymbol{h}_{\mathrm{direkt}}(\iota, k)$	DFT des direkten Anteils der RIA ($h_{m,\mathrm{Hall}}(l)$).
$h_{m,\mathrm{Hall}}(l)$	Anteil der Reflexionen an der RIA zum m-ten Mikrofon.
$h(t)$	Raumimpulsantwort.
H	Anzahl der Mikrofonpaare.
$\boldsymbol{H}$	HELMERT-Matrix zur Transformation in einen *Shape*-Bereich.
i	Index des Sensors $\in [1,\, I]$.
I	Anzahl der Sensorknoten.
$I_i\,(\kappa)$	Besselfunktion i.ter Ordnung.
j	Index des Sensors $\in [1,\, I]$.
j	Index der Form $\in [1,\, \mathit{\mathit{J}}]$.
$\mathit{\mathit{J}}$	Anzahl der Formen (*Shapes*).
$\mathcal{J}\,(\alpha, \beta)$	Optimierungsproblem zur Bestimmung der RBT-Parameter α und β.
$\boldsymbol{J}_{\mathrm{Tan}}\left(\boldsymbol{\Lambda}^{(r)}\right)$	Jacobi-Matrix der Zielfunktion $\boldsymbol{f}_{\mathrm{Tan}}$ an der Stelle $\boldsymbol{\Lambda}^{(r)}$.
k	DFT-Frequenzindex, $k = 0, \ldots, L/2$.
k	Index der Landmarken $\in [1,\, \mathit{\mathit{J}}]$.
$\mathcal{K}$	Anzahl der Landmarken einer Form/eines *Shape*s.
l	Zeitindex der im Abstand von f_{s} gewonnen Abtastwerte.
ℓ	Index der Zeile der HELMERT-Matrix.
ℓ_{k}	Landmarke k.
L	DFT-Blocklänge.
$\mathrm{L}(\boldsymbol{\mu}, \boldsymbol{\kappa})$	*Log-Likelihood*-Funktion.
$\boldsymbol{\ell}$, $\boldsymbol{\ell}_{\mathit{j}}$	Form/*Shape* bestehend aus den Landmarken ℓ_{k}, $\mathit{k} = 1, \ldots, \mathcal{K}$.
$\overline{\boldsymbol{\ell}}$	Mittlere Form/Mittlerer *Shape*.
m	Index des Mikrofons (innerhalb eines Sensorknotens) $\in [1,\, M]$.
m'	Index des Mikrofons (innerhalb des akustischen Sensornetzes) $\in [1,\, M']$.
M	Anzahl der Mikrofone eines Sensorknotens.

M'	Anzahl aller Mikrofone des akustischen Sensornetzes.
$\boldsymbol{m}_m$	Position des m-tes Mikrofons.
n	Index des Mikrofons (innerhalb eines Sensorknotens) $\in [1,\, M]$.
n'	Index des Mikrofons (innerhalb des akustischen Sensornetzes) $\in [1,\, M']$.
$n_m(l)$	Rauschen des m-ten Mikrofons.
$\tilde{\mathbf{n}}(\iota, k)$	STFT des Rauschens aller Mikrofone.
$\boldsymbol{o}_m(\theta_i)$	Position des m-ten Mikrofons relativ zum Zentrum des i-ten Sensors.
$\boldsymbol{o}_{h,1}(\theta_i)$	Position des 1.-ten Mikrofons des h-ten Paares relativ zum Zentrum des i-ten Sensors.
$\mathrm{p}(\cdot)$	Wahrscheinlichkeitsdichtefunktion.
$\mathrm{P}(\cdot)$	Verteilungsfunktion.
$\mathrm{phat}_{(n,m)}(\cdot)$	Phat-Funktion des Mikrofonpaares bestehend aus den Mikrofonen m und n.
$P_{\mathrm{LYDE}}(\cdot)$	*Score* des LYDE.
$P_{\mathrm{MUSIC}}(\cdot)$	*Score* von MUSIC.
$P_{\mathrm{SRPPhat}}(\cdot)$	*Score* von SRPPhat.
$\boldsymbol{p}_{i,h}$	Position des h-ten Mikrofonpaares des i-ten Sensorknotenes.
$\boldsymbol{q}_{(i,j),d}$	Schnittpunkt der Geraden $\boldsymbol{\varrho}_{i,d}(\lambda_{i,d})$ und $\boldsymbol{\varrho}_{j,d}(\lambda_{j,d})$.
$\tilde{\boldsymbol{q}}_d$	Gewichteter Mittelwert der Schnittpunkte $\boldsymbol{q}_{(i,j),d}$, $i = 1, \ldots, I$, $j = 1, \ldots, I$ und $d = 1, \ldots, I$.
r	Iterationsindex des Newton-Verfahrens.
r'	Reflexionensindex $\in [0; R]$.
r''	Reflexionensindex $\in [0; R]$.
$R_{\max}$	Alle Reflexionen der Raumimpulsantwort.
R	Anzahl der Reflexionen.
$\boldsymbol{R}$	Rotationsmatrix der RBT.
$\boldsymbol{R}_{\mathrm{xy}}(\theta_i)$	Rotationsmatrix in der xy-Ebene um den Winkel θ_i.
$s(l)$	Quellsignal.
$S(\iota, k)$	STFT des Quellsignals $s(l)$.
$\boldsymbol{s}_i$	Position des i-ten Sensors: $[x_i \;\; y_i]^{\mathrm{T}}$ bzw. $[x_i \;\; y_i \;\; z_i]^{\mathrm{T}}$.
$\widehat{\boldsymbol{s}}'_i$	Schätzwert der Position des i-ten Sensors im Referenzkoordinatensystem.
$\widetilde{\boldsymbol{s}}_i$	Position des i-ten Sensors vor der Skalierung.
$\widehat{\boldsymbol{s}}_i$	Schätzwert für die Position des i-ten Sensors.
$\boldsymbol{S}$	Sensorpositionen: $[\boldsymbol{s}_1 \;\; \ldots \;\; \boldsymbol{s}_I]$.
t	Zeit (kontinuierlich).
$\boldsymbol{t}$	Translationsvektor der RBT.
$t_{r'}$	Dämpfung auf den r'-ten Pfade.
T_{60}	Nachhallzeit.
u_d	Darstellung der Position des d-ten Ereignisses der akustischen Trajektorie als *landmark*.
$\boldsymbol{u}$	Form/*Shape* der akustischen Trajektorie: $[u_1 \;\; \ldots \;\; u_D]^{\mathrm{T}}$.
$\bar{\boldsymbol{u}}$	Mittelwert der Form/des *Shapes* der akustischen Trajektorie $\boldsymbol{u}$.
$\tilde{\boldsymbol{u}}$	Mittelwertbefreiter Anteil der Form/des *Shapes* der akustischen Trajektorie $\boldsymbol{u}$.

$\boldsymbol{u}_{\mathrm{F}}$	Repräsentation der akustischen Trajektorie im *Shape*/FOURIER-Bereich.
$\bar{\boldsymbol{u}}_{\mathrm{F}}$	Repräsentation des Mittelwertes der akustischen Trajektorie im *Shape*/FOURIER-Bereich.
$\tilde{\boldsymbol{u}}_{\mathrm{F}}$	Repräsentation des mittelwertbefreiten Anteils der akustischen Trajektorie im *Shape*/FOURIER-Bereich.
v	Index des Ereignisses $\in [1,\, D]$.
v_d	Streuungsmaß der Schnittpunkte $\boldsymbol{q}_{(i,j),d}$, $i = 1, \ldots, I$, $j = 1, \ldots, I$ und $d = 1, \ldots, D$.
$\boldsymbol{v}_c$	Position der c-ten Kamera.
$\boldsymbol{V}$	Kamerapositionen: $[\boldsymbol{v}_1,\ \ldots\ \boldsymbol{v}_C]$.
w_d	Darstellung der Position des d-ten Ereignisses der visuellen Trajektorie als *landmark*.
$w_{(i,j),d}$	Gewicht des Schnittpunktes $\boldsymbol{q}_{(i,j),d}$.
$\boldsymbol{w}$	Form/*Shape* der visuellen Trajektorie: $[w_1\ \ \ldots\ w_D]^{\mathrm{T}}$.
$\bar{\boldsymbol{w}}$	Mittelwert der Form/des *Shapes* der visuellen Trajektorie $\boldsymbol{w}$.
$\tilde{\boldsymbol{w}}$	Mittelwertbefreiter Anteil der Form/des *Shapes* der visuellen Trajektorie $\boldsymbol{w}$.
$\boldsymbol{w}_{\mathrm{F}}$	Repräsentation der visuellen Trajektorie im *Shape*/FOURIER-Bereich.
$\bar{\boldsymbol{w}}_{\mathrm{F}}$	Repräsentation des Mittelwertes der visuellen Trajektorie im *Shape*/FOURIER-Bereich.
$\tilde{\boldsymbol{w}}_{\mathrm{F}}$	Repräsentation des mittelwertbefreiten Anteils der visuellen Trajektorie im *Shape*/FOURIER-Bereich.
x_i	x-Koordinate des i-ten Sensors.
$x_m(l)$	Signal des m-ten Mikrofons.
$X_m(\iota, k)$	STFT des vom m-ten Mikrofon aufgenommen Signals.
$\boldsymbol{x}$	Nicht näher bezeichnetes Argument.
$\boldsymbol{\mathfrak{x}}$	Nicht näher bezeichneter Vektor.
$\mathbf{x}(\iota, k)$	STFT der Signale aller Mikrofone.
$x^{\mathrm{L}}_{\mathfrak{k}}$	x-Koordinate der $\mathfrak{k}$-ten Landmarke.
$\vec{\boldsymbol{x}}$	Basisvektor des Weltkoordinatensystems.
$\vec{\boldsymbol{x}}_s$	Basisvektor des Sensorkoordinatensystems.
y_i	y-Koordinate des i-ten Sensors.
$\boldsymbol{y}(\iota, k)$	Normierte Darstellung von $\mathbf{x}(\iota, k)$/ Beobachtungen der WATSON-Verteilung.
$\vec{\boldsymbol{y}}$	Basisvektor des Weltkoordinatensystems.
$\vec{\boldsymbol{y}}_s$	Basisvektor des Sensorkoordinatensystems.
$\boldsymbol{\mathcal{y}}$	Nicht näher bezeichnetes Argument.
$\boldsymbol{\mathfrak{y}}$	Nicht näher bezeichneter Vektor.
$y^{\mathrm{L}}_{\mathfrak{k}}$	y-Koordinate der $\mathfrak{k}$-ten Landmarke.
z_i	z-Koordinate des i-ten Sensors.
$\boldsymbol{z}_1,\ \boldsymbol{z}_2$	Einheitsvektor.
$\vec{\boldsymbol{z}}$	Basisvektor des Weltkoordinatensystems.
$\mathbf{1}$	Vektor mit Einsen.
α	Rotation und Skalierung der komplexen RBT.
α	Verhältnis zwischen LOS-Komponente und Reflexionen.

$\boldsymbol{\alpha}(k, \varphi)$	*Steering-Vector* zum Einfallswinkel φ und die Frequenz f_k.
$\boldsymbol{\alpha}(k, \boldsymbol{\tau}_1(\iota))$	*Steering-Vector* zur Signallaufzeitdifferenz $\boldsymbol{\tau}_1(\iota)$ und die Frequenz f_k.
β	Translation der komplexen RBT.
$\gamma(\boldsymbol{l}_1, \boldsymbol{l}_2)$	Volle PROKRUSTES-Distanz der Formen $\boldsymbol{\ell}_1$ und $\boldsymbol{\ell}_2$.
$\boldsymbol{\Gamma}$	Dispersions- bzw. Kovarianzmatrix der Ereignispositionen.
δ_c	Orientierung der c-ten Kamera.
$\boldsymbol{\delta}$	Kameraorientierungen: $[\delta_1 \ \ldots \ \delta_C]$.
$\boldsymbol{\Delta}(k, \varphi)$	*Mode* der komplexen WATSON-Verteilung.
$\tilde{\boldsymbol{\Delta}}(k, \varphi)$	*Mode* der komplexen WATSON-Verteilung (noch nicht normiert).
ε, $\varepsilon(\iota)$	Fehler der Einfallswinkelschätzung (im ι Block).
ε_{P}	Mittlerer Positionierungsfehler (ohne RBT).
$\epsilon_{\mathrm{P,\,Rel}}$	Mittlerer Orientierungsfehler (mit RBT inkl. Skalierung).
ϵ_{P}	Mittlerer Positionierungsfehler (mit RBT ohne Skalierung).
ε_{W}	Mittlerer Orientierungsfehler (ohne RBT).
ϵ_{W}	Mittlerer Orientierungsfehler (mit RBT ohne Skalierung).
ϵ	Entfernung zur Nullstelle.
$\bar{\varepsilon}$	Mittelwert des DOA-Fehlers.
ζ	Schwellwert: Beobachtungen, die mindestens erforderlich sind damit der RANSAC das ermittelte Modell akzeptiert.
ζ_{Fit}	Schwellwert des RANSAC zur Klassifikation, ob eine Beobachtung zum vorliegenden Modell passt oder nicht.
$\eta_{i,d}$	Einfallswinkel (Elevation) des i-ten Sensors zum d-ten Ereignis.
$\vartheta_{c,d}$	Einfallswinkel (Azimuth) der c-ten Kamera zum d-ten Ereignis.
$\boldsymbol{\Theta}$	Einfallswinkel aller Kameras zu allen Sensoren $[[\vartheta_{1,1} \ \ldots \ \vartheta_{C,1}]^{\mathrm{T}} \ \ldots \ [\vartheta_{1,D} \ \ldots \ \vartheta_{C,D}]^{\mathrm{T}}]$.
θ_i	Orientierung des i-ten Sensors.
$\widehat{\theta}_i$	Schätzwert für die Orientierung des i-ten Sensors.
$\widehat{\theta}'_i$	Schätzwert der Orientierung des i-ten Sensors im Referenzkoordinatensystem.
$\boldsymbol{\theta}$	Sensororientierungen: $[\theta_1 \ \ldots \ \theta_I]$.
ι	STFT-Blockindex.
κ	Konzentration der WATSON-Verteilung oder VON MISES-Verteilung.
$\kappa_{i,d}$	Konzentration der VON MISES-Verteilung des d-ten Ereignisse am i-ten Sensor.
$\boldsymbol{\kappa}$	Konzentrationen aller Sensoren und Ereignisse $[\kappa_{1,1} \ \ldots \ \kappa_{I,D}]$.
λ	Zeitlicher Versatz zwischen den Signalen zweier Mikrofone.
$\lambda_{i,d}$	Parameter der Geradenglerichung $\boldsymbol{\varrho}_{i,d}(\lambda_{i,d})$.
$\boldsymbol{\Lambda}$	Unbekannte des Newton-Verfahrens ($\boldsymbol{S}$, $\boldsymbol{\theta}$, $\boldsymbol{E}$).
μ, $\mu_{i,d}$	Mittelwert der VON MISES-Verteilung des d-ten Ereignisse am i-ten Sensor.
$\boldsymbol{\mu}$	Mittelwerte aller Sensoren und Ereignisse $[\mu_{1,1} \ \ldots \ \mu_{I,D}]$.
$\boldsymbol{\mu}_{i,d}$	Mittelwertvektor des i-ten Sensors und des d-ten Ereignisses.
ν	Skalierungsfaktor.
ν_{OK}	Wahrscheinlichkeit, dass eine Beobachtung zu einem tragfähigen Modell führt.
$\xi_m(\varphi)$	Richtcharakteristik des m-ten Mikrofons.

$\boldsymbol{\xi}(\varphi)$	Richtcharakteristik aller Mikrofone: $[\xi_1(\varphi) \ \ldots \ \xi_M(\varphi)]^{\mathrm{T}}$.
$\boldsymbol{\Xi}(\iota, k)$	Eigenvektoren des KLDS $\boldsymbol{\Phi}_{\mathbf{xx}}(\iota, k)$.
$\rho(\boldsymbol{x})$	Gewichtungsfunktion des LYDE.
$\boldsymbol{\varrho}_{i,d}(\lambda_{i,d})$	Gerade, die die möglichen Positionen des d-Ereignisse des i-ten Sensors beschreibt.
$(\sigma^{\mathrm{A}})^2$	Varianz der akustischen Trajektorie.
σ_{Ori}	Standardabweichung der Sensororientierung.
σ_{M}	Standardabweichung der VON MISES-Verteilung ($\sigma_{\mathrm{M}} = \sqrt{1/\kappa}$).
σ_τ	Standardabweichung der TDOA τ.
ς	Dämpfungseigenschaften des Raumes.
$\hat{\tau}_{\mathrm{RIA}}$	TDOA der RIA.
$\boldsymbol{\tau}_1(\iota)$	Signallaufzeitdifferenzen aller Mikrofon im ι-ten STFT-Block bezogen auf das erste Mikrofon: $[\tau_{(1,2)}(\iota) \ \ldots \ \tau_{(1,M)}(\iota)]$.
$\hat{\tau}_{\mathrm{c}}$	Auf das Intervall $[-\tau_{\max}; \tau_{\max}]$ begrenzte TDOA-Schätzung.
$\hat{\tau}$	Schätzung der TDOA.
$\bar{\tau}$	Tatsächliche TDOA.
$\tau_{\max}$	Maximale Signallaufzeitdifferenz.
$\tau_{(m,n),d}$	TDOA für das Mikrofonpaaar m,n und Ereignis d.
$\tilde{\tau}_{(m,n),d}$	Prädiktion der TDOA für das Mikrofonpaar m,n und Ereignis d.
$\upsilon_{r'}$	Verzögerung auf dem r'-ten Pfad.
$\boldsymbol{\Upsilon}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\Psi}$	Matrixen der SVD.
$\boldsymbol{\varphi}_d$	Einfallswinkel (Azimuth) von allen Sensoren zum d-ten Ereignis: $[\varphi_{1,d} \ \varphi_{I,d}]^{\mathrm{T}}$.
φ, $\varphi(\iota)$	Einfallswinkel (für den ι-ten STFT-Block).
$\varphi_{i,d}$	Einfallswinkel (Azimuth) des i-ten Sensors zum d-ten Ereignis.
$\varphi_{i,h,d}$	Einfallswinkel des h-ten Mikrofonpaares des i-ten zum d-ten Ereignis.
$\hat{\varphi}$, $\hat{\varphi}(\iota)$	Schätzwert des Einfallswinkel (für den ι-ten STFT-Block).
$\Phi_{ss}(\iota, k)$	Kovarianzmatrix des Quellsignals $S(\iota, k)$.
$\boldsymbol{\Phi}_{\mathbf{xx}}(\iota, k)$	Kovarianzmatrix der Mikrofonsignale $\mathbf{x}(\iota, k)$.
$\boldsymbol{\Phi}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}}(\iota, k)$	Kovarianzmatrix des Rauschens $\tilde{\mathbf{n}}(\iota, k)$.
$\boldsymbol{\Phi}$	Einfallswinkel (Azmimuth) aller Sensoren zu allen Ereignissen $[\boldsymbol{\varphi}_1 \ \ldots \ \boldsymbol{\varphi}_D]$.
ϕ	Anzahl der Iteration des RANSAC.
$\psi_{i,h}$	Orientierung Mikrofon paar.
ω_m	Orientierung des m-ten Mikrofons.
$\omega(l)$	Fenster-Funktion.
Ω	Menge aller Beochatungen.
Ω_{fit}	Konsens des RANSAC.
Ω_{sel}	Vom RANSAC ausgewählte Beobachtungen, die zur Berechnung der Modellparameter dienen.

Operatoren und spezielle Symbole

$*$	Zeitdiskrete Faltung.
$.^{\mathrm{T}}$	Transpositionsoperator.
$.^{\mathrm{H}}$	Hermitescher Operator.

$\lvert x \rvert$	Betrag von x.
$\lVert \boldsymbol{x} \rVert_2$	L2-Norm des Vektors $\boldsymbol{x}$.
$\{\boldsymbol{x}\}_1$	Erstes Element des Vektors $\boldsymbol{x}$.
$\{\boldsymbol{x}\}_{x:y}$	Elemente von x bis y des Vektors $\boldsymbol{x}$.
$\odot$	Hadamard-Produkt bzw. elementweise Multiplikation.
$\sphericalangle(\boldsymbol{x}, \boldsymbol{y})$	Zwischen den Vektoren $\boldsymbol{x}$ und $\boldsymbol{y}$ eingeschlossener Winkel.
$\arg(\cdot)$	Phase/Winkel des komplexen Arguments.
$\delta(\cdot)$	Dirac-Funktion.
$\mathrm{IDFT}(\cdot)$	IDFT-Operator.
$\#(\cdot)$	Anzahl der Elemente einer Menge.
$\mathcal{O}(\cdot)$	Landau-Symbol für obere Schranke der Ausführungszeit.
$\mathrm{rect}(\cdot)$	Rechteck-Funktion.
$\mathrm{std}(\cdot)$	Standardabweichung.
$\mathrm{svd}(\cdot)$	SVD.
$\mathbb{E}[\cdot]$	Erwartungswertoperator.
$\mathrm{erf}(\cdot)$	Fehlerfunktion.
$\mathrm{Q}(\cdot)$	Q-Funktion. Integral der Standardnormalverteilung von x bis ∞.
$\mathcal{M}(x; \mu_x, \sigma_x)$	VON MISES-Verteilung mit Mittelwert μ_x und Konzentration κ_x.
$\mathcal{N}(x; \mu_x, \sigma_x)$	Normalverteilung mit Mittelwert μ_x und Standardabweichung σ_x.
$\lceil \cdot \rceil$	Runde auf die nächstgrößere ganze Zahl.

Abbildungsverzeichnis

Tabellenverzeichnis

Literatur

[AB79] J. B. Allen und D. A. Berkley. „Image method for efficiently simulating small-room acoustics“. In: *The Journal of the Acoustical Society of America* 65.4 (Apr. 1979), S. 943–950. DOI: 10.1121/1.382599.

[AC07] Y. Avargel und I. Cohen. „On Multiplicative Transfer Function Approximation in the Short-Time Fourier Transform Domain“. In: *IEEE Signal Processing Letters* 14.5 (Mai 2007), S. 337–340. ISSN: 1070-9908. DOI: 10.1109/LSP.2006.888292.

[AHB87] K. Arun, T. Huang und S. Blostein. „Least-Squares Fitting of Two 3-D Point Sets“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9.5 (Sep. 1987), S. 698–700. ISSN: 0162-8828. DOI: 10.1109/TPAMI.1987.4767965.

[Ara+11] S. Araki, H. Sawada, R. Mukai und S. Makino. „DOA Estimation for Multiple Sparse Sources with Arbitrarily Arranged Multiple Sensors“. In: *Journal of Signal Processing Systems* 63.3 (Juni 2011), S. 265–275. ISSN: 1939-8115. DOI: 10.1007/s11265-009-0413-9.

[Ara88] H. Arau-Puchades. „An Improved Reverberation Formula“. In: *Acustica* 65.4 (März 1988), S. 163–180.

[AV07] D. Arthur und S. Vassilvitskii. „K-means++: The Advantages of Careful Seeding“. In: *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA 2007)*. Philadelphia, PA, USA: Society for Industrial und Applied Mathematics, Jan. 2007, S. 1027–1035. ISBN: 978-0-898716-24-5.

[Bar83] A. Barabell. „Improving the resolution performance of eigenstructure-based direction-finding algorithms“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1983)*. Bd. 8. Boston, MA, USA, Apr. 1983, S. 336–339. DOI: 10.1109/ICASSP.1983.1172124.

[BAS95] M. Brandstein, J. Adcock und H. Silverman. „A closed-form method for finding source locations from microphone-array time-decay estimates“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1995)*. Bd. 5. Detroit, MI, USA, Mai 1995, S. 3019–3022. DOI: 10.1109/ICASSP.1995.479481.

[BD10] M. Brückner und J. Denzler. „Active Self-calibration of Multi-camera Systems“. In: *Proceedings of the DAGM Symposium on Pattern Recognition (DAGM 2010)*. Darmstadt, Deutschland: Springer, Sep. 2010, S. 31–40. DOI: 10.1007/978-3-642-15986-2_4.

[Bir03] S. Birchfield. „Geometric microphone array calibration by multidimensional scaling“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*. Bd. 5. Hong Kong, Hong Kong, Apr. 2003, S. 157–160. DOI: 10.1109/ICASSP.2003.1199892.

[BM09] A. Bertrand und M. Moonen. „Robust Distributed Noise Reduction in Hearing Aids with External Acoustic Sensor Nodes“. In: *EURASIP Journal on Advances in Signal Processing* 2009 (Jan. 2009), 12:3–12:3. ISSN: 1110-8657. DOI: 10.1155/2009/530435.

[Bro72a] Brockhaus-Enzyklopädie. *Band 01, A-ATE*. F.A. Brockhaus GmbH, 1972, S. 532.

[Bro72b] Brockhaus-Enzyklopädie. *Band 09, IL-KAS*. F.A. Brockhaus GmbH, 1972, S. 283.

[Bro72c] Brockhaus-Enzyklopädie. *Band 15, POR-RIS*. F.A. Brockhaus GmbH, 1972, S. 173.

[BS05] S. Birchfield und A. Subramanya. „Microphone Array Position Calibration by Basis-Point Classical Multidimensional Scaling“. In: *IEEE Transactions on Speech and Audio Processing* 13.5 (Sep. 2005), S. 1025–1034. ISSN: 1063-6676. DOI: 10.1109/TSA.2005.851893.

[BS06] K. O. Bowman und L. R. Shenton. „Estimation: Method of Moments“. In: *Encyclopedia of Statistical Sciences* 3 (2006). DOI: 10.1002/0471667196.ess1618.pub2.

[BSH07] J. Benesty, M. M. Sondhi und Y. Huang. *Springer handbook of speech processing*. Springer Science & Business Media, 2007. DOI: 10.1007/978-3-540-49127-9.

[Bun14] Bundesverband Informationswirtschaft, Telekommunikation und neue Medien (Bitkom). *Smartphones: Tippst Du noch oder sprichst Du schon?* https://www.bitkom.org/Presse/Presseinformation/Pressemitteilung_3992.html. Juni 2014.

[CDM12] M. Crocco, A. Del Bue und V. Murino. „A Bilinear Approach to the Position Self-Calibration of Multiple Sensors“. In: *IEEE Transactions on Signal Processing* 60.2 (Feb. 2012), S. 660–673. ISSN: 1053-587X. DOI: 10.1109/TSP.2011.2175387.

[CFY76] A. Charnesa, E. L. Fromeb und P. L. Yub. „The Equivalence of Generalized Least Squares and Maximum Likelihood Estimates in the Exponential Family“. In: *Journal of the American Statistical Association* 71.353 (Okt. 1976), S. 169–171. DOI: 10.1080/01621459.1976.10481508.

[Cha95] J. H. Challis. „A procedure for determining rigid body transformation parameters“. In: *Journal of Biomechanics* 28.6 (Juni 1995), S. 733–737. ISSN: 0021-9290. DOI: 10.1016/0021-9290(94)00116-L.

[Che99] P.-C. Chen. „A Non-Line-of-Sight Error Mitigation Algorithm in Location Estimation“. In: *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 1999)*. Bd. 1. New Orleans, LA, USA, Nov. 1999, S. 316–320. DOI: 10.1109/WCNC.1999.797838.

[CM03] L. Cremer und M. Möser. *Technische Akustik*. Springer, 2003.

[CMK03] O. Chum, J. Matas und J. Kittler. „Locally Optimized RANSAC“. In: *Pattern Recognition*. Hrsg. von B. Michaelis und G. Krell. Bd. 2781. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2003, S. 236–243. ISBN: 978-3-540-40861-1. DOI: 10.1007/978-3-540-45243-0_31.

[CMO04] O. Chum, J. Matas und Š. Obdrzálek. „Enhancing RANSAC by generalized model optimization“. In: *Asian Conference on Computer Vision (ACCV 2004)*. Seoul, Süd Korea, Jan. 2004, S. 812–817.

[Con+12] A. Contini, A. Canclini, E. Antonacci, M. Compagnoni, A. Sarti und S. Tubaro. „Self-calibration of microphone arrays from measurement of Times of Arrival of acoustic signals“. In: *Proceedings of International Symposium on Communications Control and Signal Processing (ISCCSP 2012)*. Rom, Italien, Mai 2012, S. 1–6. DOI: 10.1109/ISCCSP.2012.6217837.

[Cro+12] M. Crocco, A. Del Bue, M. Bustreo und V. Murino. „A closed form solution to the microphone position self-calibration problem“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012)*. Kyoto, Japan, März 2012, S. 2597–2600. DOI: 10.1109/ICASSP.2012.6288448.

[Dal06] N. Dalal. „Finding people in images and videos“. Diss. Institut National Polytechnique de Grenoble-INPG, Juli 2006.

[DBA07] J. Dmochowski, J. Benesty und S. Affes. „A Generalized Steered Response Power Method for Computationally Viable Source Localization“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 15.8 (Nov. 2007), S. 2510–2526. ISSN: 1558-7916. DOI: 10.1109/TASL.2007.906694.

[DM98] I. Dryden und K. Mardia. *Statistical shape analysis*. Wiley series in probability and statistics. Chichester [u.a.]: Wiley, 1998. XVII, 347. ISBN: 0471958166.

[Doc+09] S. Doclo, M. Moonen, T. Van den Bogaert und J. Wouters. „Reduced-Bandwidth and Distributed MWF-Based Noise Reduction Algorithms for Binaural Hearing Aids“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 17.1 (Jan. 2009), S. 38–51. ISSN: 1558-7916. DOI: 10.1109/TASL.2008.2004291.

[Dru+14] L. Drude, A. Chinaev, D. H. T. Vu und R. Haeb-Umbach. „Source counting in speech mixtures using a variational EM approach for complex WATSON mixture models“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*. Florenz, Italien, Mai 2014, S. 6834–6838. DOI: 10.1109/ICASSP.2014.6854924.

[DSB01] J. DiBiase, H. Silverman und M. Brandstein. „Robust Localization in Reverberant Rooms“. English. In: *Microphone Arrays*. Hrsg. von M. Brandstein und D. Ward. Digital Signal Processing. Springer Berlin Heidelberg, 2001, S. 157–180. ISBN: 978-3-642-07547-6. DOI: 10.1007/978-3-662-04619-7_8.

[DT05] N. Dalal und B. Triggs. „Histograms of Oriented Gradients for Human Detection“. In: *Proceedings of International Conference on Computer Vision & Pattern Recognition (CVPR 2005)*. Hrsg. von C. Schmid, S. Soatto und C. Tomasi. Bd. 1. San Diego, CA, USA: IEEE Computer Society, Juni 2005, S. 886–893. DOI: 10.1109/CVPR.2005.177.

[El +13] N. El Gemayel, S. Koslowski, F. K. Jondral und J. Tschan. „A low cost tdoa localization system: Setup, challenges and results“. In: *Proceedings of Workshop on Positioning Navigation and Communication (WPNC 2013)*. Dresden, Deutschland, März 2013, S. 1–4. DOI: 10.1109/WPNC.2013.6533293.

[ELF97] D. Eggert, A. Lorusso und R. Fisher. „Estimating 3-D rigid body transformations: a comparison of four major algorithms“. In: *Machine Vision and Applications* 9.5-6 (März 1997), S. 272–290. ISSN: 0932-8092. DOI: 10.1007/s001380050048.

[Esa+12] S. Esaki, K. Niwa, T. Nishino und K. Takeda. „Estimating sound source depth using a small-size array“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012)*. Kyoto, Japan, März 2012, S. 401–404. DOI: 10.1109/ICASSP.2012.6287901.

[Eve+15] C. Evers, A. Moore, P. Naylor, J. Sheaffer und B. Rafaely. „Bearing-only acoustic tracking of moving speakers for robot audition“. In: *Proceedings of IEEE International Conference on Digital Signal Processing (DSP 2015)*. Singapur, Singapur, Juli 2015, S. 1206–1210. DOI: 10.1109/ICDSP.2015.7252071.

[Eyr30] C. F. Eyring. „Reverberation Time in “Dead” Rooms“. In: *The Journal of the Acoustical Society of America* 1.2A (1930), S. 168–168. DOI: 10.1121/1.1901884.

[FB81] M. A. Fischler und R. C. Bolles. „Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography“. In: *Communications of the ACM* 24.6 (Juni 1981), S. 381–395. ISSN: 0001-0782. DOI: 10.1145/358669.358692.

[Fit59] D. Fitzroy. „Reverberation Formula Which Seems to Be More Accurate with Nonuniform Distribution of Absorption“. In: *The Journal of the Acoustical Society of America* 31.7 (1959), S. 893–897. DOI: 10.1121/1.1907814.

[FJ05] M. Frigo und S. Johnson. „The Design and Implementation of FFTW3“. In: *Proceedings of the IEEE* 93.2 (Feb. 2005), S. 216–231. ISSN: 0018-9219. DOI: 10.1109/JPROC.2004.840301.

[FLS63] R. Feynman, R. Leighton und M. Sands. *The Feynman Lectures on Physics*. Second. Bd. 1. Boston: Addison-Wesley, 1963.

[FP03] D. A. Forsyth und J. Ponce. *Computer Vision: A Modern Approach*. 2003.

[Fra+94] J. Fransen, D. Pye, T. Robinson, P. Woodland und S. Young. *WSJCAM0 Corpus and Recording Description*. University of Cambridge, Department of Engineering, 1994.

[G+93] J. S. Garofolo, L. D. Consortium u. a. *TIMIT: acoustic-phonetic continuous speech corpus*. Linguistic Data Consortium, 1993.

[Gau+14] N. D. Gaubitch, J. Martinez, W. B. Kleijn und R. Heusdens. „On near-field beamforming with smartphone-based ad-hoc microphone arrays“. In: *Proceedings of International Workshop on Acoustic Signal Enhacement (IWAENC 2014)*. Antibes - Juan les Pins, Frankreich, Sep. 2014, S. 94–98. DOI: 10.1109/IWAENC.2014.6953345.

[GH10] Y. Guo und M. Hazas. „Acoustic Source Localization of Everyday Sounds Using Wireless Sensor Networks“. In: *Proceedings of the ACM International Conference Adjunct Papers on Ubiquitous Computing (UbiComp 2010)*. Kopenhagen, Dänemark: ACM, Sep. 2010, S. 411–412. ISBN: 978-1-4503-0283-8. DOI: 10.1145/1864431.1864462.

[GKH13] N. Gaubitch, W. Kleijn und R. Heusdens. „Auto-localization in ad-hoc microphone arrays“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*. Vancouver, BC, Kanada, Mai 2013, S. 106–110. DOI: 10.1109/ICASSP.2013.6637618.

[GKH14] N. Gaubitch, W. Kleijn und R. Heusdens. „Calibration of distributed sound acquisition systems using TOA measurements from a moving acoustic source“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*. Florenz, Italien, Mai 2014, S. 7455–7459. DOI: 10.1109/ICASSP.2014.6855049.

[GS08] M. Gillette und H. Silverman. „A Linear Closed-Form Algorithm for Source Localization From Time-Differences of Arrival“. In: *IEEE Signal Processing Letters* 15 (2008), S. 1–4. ISSN: 1070-9908. DOI: 10.1109/LSP.2007.910324.

[Hab10] E. A. Habets. *Room Impulse Response Generator*. 2010. URL: http://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator.

[Had+14] E. Hadad, F. Heese, P. Vary und S. Gannot. „Multichannel audio database in various acoustic environments“. In: *Proceedings of International Workshop on Acoustic Signal Enhacement (IWAENC 2014)*. Antibes - Juan les Pins, Frankreich, Sep. 2014, S. 313–317. DOI: 10.1109/IWAENC.2014.6954309.

[HBE00] Y. Huang, J. Benesty und G. W. Elko. „Acoustic Signal Processing for Telecommunication“. In: Hrsg. von S. L. Gay und J. Benesty. Bd. IV. The Springer International Series in Engineering and Computer Science 551. Springer US, 2000. Kap. Microphone Arrays for Video Camera Steering, S. 239–259. DOI: 10.1007/978-1-4419-8644-3_11.

[Hen+09] M. Hennecke, T. Plotz, G. Fink, J. Schmalenstroer und R. Hab-Umbach. „A hierarchical approach to unsupervised shape calibration of microphone array networks“. In: *Proceedings of IEEE/SP Workshop on Statistical Signal Processing (SSP 2009)*. Cardiff, Wales, Aug. 2009, S. 257–260. DOI: 10.1109/SSP.2009.5278589.

[HF11] M. Hennecke und G. Fink. „Towards acoustic self-localization of ad hoc smartphone arrays“. In: *Proceedings of Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA 2011)*. Edinburgh, Schottland, Mai 2011, S. 127–132. DOI: 10.1109/HSCMA.2011.5942378.

[HG07] E. A. Habets und S. Gannot. „Generating sensor signals in isotropic noise fields“. In: *The Journal of the Acoustical Society of America* 122.6 (Dez. 2007), S. 3464–3470. DOI: 10.1121/1.2799929.

[HMV05] A. Hoffmann, B. Marx und W. Vogt. *Mathematik für Ingenieure*. Bd. 1. Lineare Algebra, Analysis : Theorie und Numerik. Pearson Studium, 2005. ISBN: 3-8273-7113-9.

[HZ04] R. I. Hartley und A. Zisserman. *Multiple View Geometry in Computer Vision*. 2. Aufl. Cambridge University Press, ISBN: 0521540518, 2004.

[Ihl98] F. Ihlenburg. *Finite element analysis of acoustic scattering*. Bd. 132. Applied Mathematical Sciences. Springer-Verlag New York, 1998. DOI: 10.1007/b98828.

[Iva14] M. V. Ivan Dokmanić Laurent Daudet. „How to localize ten microphones in one finger snap“. In: *Proceedings of European Signal Processing Conference (EUSIPCO 2014)*. Lisabon, Portugal, Sep. 2014, S. 2275–2279.

[JCN04] A. Johansson, G. Cook und S. Nordholm. „Acoustic direction of arrival estimation, a comparison between root-music and SRP-PHAT“. In: *Proceedings of IEEE Region 10 Conference (TENCON 2004)*. Bd. B. Adelaide, Australien, Nov. 2004, S. 629–632. DOI: 10.1109/TENCON.2004.1414674.

[Jen+16] J. R. Jensen, J. K. Nielsen, R. Heusdens und M. G. Christensen. „DOA estimation of audio sources in reverberant environments“. In: *(ICASSP 2016)*. Shanghai, China, März 2016, S. 176–180. DOI: 10.1109/ICASSP.2016.7471660.

[Jeu+11] M. Jeub, C. Nelke, C. Beaugeant und P. Vary. „Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals“. In: *Proceedings of European Signal Processing Conference (EUSIPCO 2011)*. Barcelona, Spanien, Aug. 2011, S. 1347–1351.

[JSV09] M. Jeub, M. Schafer und P. Vary. „A binaural room impulse response database for the evaluation of dereverberation algorithms“. In: *Proceedings of IEEE International Conference on Digital Signal Processing (DSP 2009)*. Santorin, Griechenland, Juli 2009, S. 1–5. DOI: 10.1109/ICDSP.2009.5201259.

[JTN14] I. Jafari, R. Togneri und S. Nordholm. „On the use of the Watson mixture model for clustering-based under-determined blind source separation“. In: *Proceedings of Interspeech 2014*. Singapur, Singapur, Sep. 2014, S. 988–992.

[KA13] Y. Kuang und K. Astrom. „Stratified sensor network self-calibration from TDOA measurements“. In: *Proceedings of European Signal Processing Conference (EUSIPCO 2013*. Marrakesch, Marokko, Sep. 2013, S. 1–5.

[KC76] C. Knapp und G. Carter. „The generalized correlation method for estimation of time delay“. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 24.4 (Aug. 1976), S. 320–327. ISSN: 0096-3518. DOI: 10.1109/TASSP.1976.1162830.

[Ken94] J. T. Kent. „The complex Bingham distribution and shape analysis“. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 56.2 (1994), S. 285–299.

[KL08] J. Kemper und H. Linde. „Challenges of passive infrared indoor localization“. In: *Proceedings of Workshop on Positioning Navigation and Communication (WPNC 2008)*. Hannover, Deutschland, März 2008, S. 63–70. DOI: 10.1109/WPNC.2008.4510358.

[KLB08] G. Kurillo, Z. Li und R. Bajcsy. „Wide-area external multi-camera calibration using vision graphs and virtual calibration object“. In: *Proceedings of ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2008)*. Stanford, CA, USA, Sep. 2008, S. 1–9. DOI: 10.1109/ICDSC.2008.4635695.

[KMR12] K. Kumatani, J. McDonough und B. Raj. „Microphone Array Processing for Distant Speech Recognition: From Close-Talking Microphones to Far-Field Sensors“. In: *IEEE Signal Processing Magazine* 29.6 (Nov. 2012), S. 127–140. ISSN: 1053-5888. DOI: 10.1109/MSP.2012.2205285.

[KSS68] A. Krokstad, S. Strom und S. Sørsdal. „Calculating the acoustical room response by the use of a ray tracing technique“. In: *Journal of Sound and Vibration* 8.1 (Juli 1968), S. 118–125. DOI: 10.1016/0022-460X(68)90198-3.

[Kua+13] Y. Kuang, S. Burgess, A. Torstensson und K. Astrom. „A complete characterization and solution to the microphone position self-calibration problem“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*. Vancouver, BC, Canada, Mai 2013, S. 3875–3879. DOI: 10.1109/ICASSP.2013.6638384.

[Kuh+07] T. Kuhnapfel, T. Tan, S. Venkatesh und E. Lehmann. „Calibration of Audio-Video Sensors for Multi-Modal Event Indexing“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2007)*. Bd. 2. Honolulu, HI, USA, Apr. 2007, pages. DOI: 10.1109/ICASSP.2007.366342.

[KWL08] J. Kemper, M. Walter und H. Linde. „Human-Assisted Calibration of an Angulation Based Indoor Location System“. In: *Proceedings of International Conference on Sensor Technologies and Applications (SENSORCOMM 2008)*. Cap Esterel, Frankreich, Aug. 2008, S. 196–201. DOI: 10.1109/SENSORCOMM.2008.17.

[LBD01] K. Lebart, J.-M. Boucher und P. Denbigh. „A new method based on spectral subtraction for speech dereverberation“. In: *Acta Acustica united with Acustica* 87.3 (2001), S. 359–366.

[LD12] T. C. Lawin-Ore und S. Doclo. „Using statistical room acoustics for analysing the output SNR of the MWF in acoustic sensor networks“. In: *Proceedings of European Signal Processing Conference (EUSIPCO 2012)*. Bukarest, Rumänien, Aug. 2012, S. 1259–1263.

[LJ08] E. A. Lehmann und A. M. Johansson. „Prediction of energy decay in room impulse responses simulated with an image-source model“. In: *The Journal of the Acoustical Society of America* 124.1 (Juli 2008), S. 269–277.

[LJ10] E. Lehmann und A. Johansson. „Diffuse Reverberation Model for Efficient Image-Source Simulation of Room Impulse Responses“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 18.6 (Aug. 2010), S. 1429–1439. ISSN: 1558-7916. DOI: 10.1109/TASL.2009.2035038.

[LKH14] V. Leutnant, A. Krueger und R. Haeb-Umbach. „A New Observation Model in the Logarithmic Mel Power Spectral Domain for the Automatic Recognition of Noisy Reverberant Speech“. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.1 (Jan. 2014), S. 95–109. ISSN: 2329-9290. DOI: 10.1109/TASLP.2013.2285480.

[LMC12] K. Lebeda, J. Matas und O. Chum. „Fixing the Locally Optimized RANSAC“. In: *Proceedings of British Machine Vision Conference (BMVC 2012)*. Surrey, England: BMVA Press, Sep. 2012, S. 95.1–95.11. ISBN: 1-901725-46-4. DOI: 10.5244/C.26.95.

[LOG05] G. Lathoud, J.-M. Odobez und D. Gatica-Perez. „AV16. 3: an audio-visual corpus for speaker localization and tracking“. In: *Machine Learning for Multimodal Interaction*. Springer, 2005, S. 182–195.

[LSM06] M. Liebens, T. Sakiyama und J. Miura. „Visual Tracking of Multiple Persons in a Heavy Occluded Space Using Person Model and Joint Probabilistic Data Association“. In: *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*. Heidelberg, Deutschland, Sep. 2006, S. 547–552. DOI: 10.1109/MFI.2006.265635.

[LT99] X. Lai und H. Torp. „Interpolation methods for time-delay estimation using cross-correlation method for blood velocity measurement“. In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 46.2 (März 1999), S. 277–290. ISSN: 0885-3010. DOI: 10.1109/58.753016.

[LV08] H. W. Löllmann und P. Vary. „Estimation of the reverberation time in noisy Environments“. In: *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC 2008)*. Seattle, WA, USA, Sep. 2008, S. 1–4.

[LY10] B. Loesch und B. Yang. „Blind Source Separation Based on Time-Frequency Sparseness in the Presence of Spatial Aliasing“. In: *Latent Variable Analysis and Signal Separation*. Springer, 2010, S. 1–8. DOI: 10.1007/978-3-642-15995-4_1.

[MD99] K. Mardia und I. Dryden. „The complex Watson distribution and shape analysis“. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.4 (1999), S. 913–926.

[MGC13] S. Markovich-Golan, S. Gannot und I. Cohen. „Distributed Multiple Constraints Generalized Sidelobe Canceler for Fully Connected Wireless Acoustic Sensor Networks“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 21.2 (Feb. 2013), S. 343–356. ISSN: 1558-7916. DOI: 10.1109/TASL.2012.2224454.

[Mil32] G. Millington. „A MODIFIED FORMULA FOR REVERBERATION“. In: *The Journal of the Acoustical Society of America* 4.1A (1932), S. 69–82. DOI: 10.1121/1.1915588.

[MJ09] K. V. Mardia und P. E. Jupp. *Directional statistics*. Bd. 494. John Wiley & Sons, 2009.

[MLH08] I. McCowan, M. Lincoln und I. Himawan. „Microphone Array Shape Calibration in Diffuse Noise Fields“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 16.3 (März 2008), S. 666–670. ISSN: 1558-7916. DOI: 10.1109/TASL.2007.911428.

[MP10] I. Marković und I. Petrović. „Speaker localization and tracking with a microphone array on a mobile robot using von Mises distribution and particle filtering“. In: *Robotics and Autonomous Systems* 58.11 (Nov. 2010), S. 1185–1196. ISSN: 0921-8890. DOI: 10.1016/j.robot.2010.08.001.

[Nis04] D. Nistér. „An Efficient Solution to the Five-Point Relative Pose Problem“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26.6 (Juni 2004), S. 756–777. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2004.17.

[NK01] R. Neubauer und B. Kostek. „Prediction of the reverberation time in rectangular rooms with non-uniformly distributed sound absorption“. In: *Archives of Acoustics* 26.3 (2001), S. 183–201. ISSN: 2300-262X. URL: http://acoustics.ippt.pan.pl/index.php/aa/article/view/398.

[NO09] F. Nesta und M. Omologo. „Generalized State Coherence Transform for multidimensional localization of multiple sources“. In: *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009)*. New Paltz, NY, USA, Okt. 2009, S. 237–240. DOI: 10.1109/ASPAA.2009.5346481.

[Oua+12] Y. Oualil, F. Faubel, M. Doss und D. Klakow. „A TDOA Gaussian mixture model for improving acoustic source tracking“. In: *Proceedings of European Signal Processing Conference (EUSIPCO 2012)*. Bukarest, Rumänien, Aug. 2012, S. 1339–1343.

[Pat+05] N. Patwari, J. Ash, S. Kyperountas, A. Hero, R. Moses und N. Correal. „Locating the nodes: cooperative localization in wireless sensor networks“. In: *IEEE Signal Processing Magazine* 22.4 (Juli 2005), S. 54–69. ISSN: 1053-5888. DOI: 10.1109/MSP.2005.1458287.

[Pea05] K. Pearson. „The Problem of the Random Walk“. In: *Nature* 72.1867 (10. Aug. 1905), S. 342. ISSN: 0028-0836.

[PF13] A. Plinge und G. Fink. „Online multi-speaker tracking using multiple microphone arrays informed by auditory scene analysis“. In: *Proceedings of European Signal Processing Conference (EUSIPCO 2013)*. Marrakesch, Marokko, Sep. 2013, S. 1–5.

[PF14a] A. Plinge und G. Fink. „Geometry calibration of distributed microphone arrays exploiting audio-visual correspondences“. In: *Proceedings of European Signal Processing Conference (EUSIPCO 2014)*. Lisabon, Portugal, Sep. 2014, S. 116–120. DOI: 10.1109/MMSP.2010.5661986.

[PF14b] A. Plinge und G. Fink. „Geometry calibration of multiple microphone arrays in highly reverberant environments“. In: *Proceedings of International Workshop on Acoustic Signal Enhacement (IWAENC 2014)*. Antibes - Juan les Pins, Frankreich, Sep. 2014, S. 243–247. DOI: 10.1109/IWAENC.2014.6954295.

[PF14c] A. Plinge und G. Fink. „Multi-speaker tracking using multiple distributed microphone arrays“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*. Florenz, Italien, Mai 2014, S. 614–618. DOI: 10.1109/ICASSP.2014.6853669.

[PHM13] P. Pertilä, M. Hämäläinen und M. Mieskolainen. „Passive Temporal Offset Estimation of Multichannel Recordings of an Ad-Hoc Microphone Array“. In: *IEEE Transactions on Audio, Speech, and Language Processing* 21.11 (Nov. 2013), S. 2393–2402. ISSN: 1558-7916. DOI: 10.1109/TASLP.2013.2286921.

[PMH11] P. Pertilä, M. Mieskolainen und M. Hämäläinen. „Closed-form self-localization of asynchronous microphone arrays“. In: *Proceedings of Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA 2011)*. Edinburgh, Schottland, Mai 2011, S. 139–144. DOI: 10.1109/HSCMA.2011.5942380.

[PN08] M. Pollefeys und D. Nister. „Direct computation of sound and microphone locations from time-difference-of-arrival data“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2008)*. Las Vegas, NV, USA, März 2008, S. 2445–2448. DOI: 10.1109/ICASSP.2008.4518142.

[PPH14] M. Parviainen, P. Pertilä und M. S. Hämäläinen. „Self-localization of wireless acoustic sensors in meeting rooms“. In: *Proceedings of Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA 2014)*. Nancy, Frankreich, Mai 2014, S. 152–156. DOI: 10.1109/HSCMA.2014.6843270.

[PST07] E. Pauwels, A. A. Salah und R. Tavenard. „Sensor Networks for Ambient Intelligence“. In: *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP 2007)*. Chania, Griechenland, Okt. 2007, S. 13–16. DOI: 10.1109/MMSP.2007.4412806.

[RD04] V. Raykar und R. Duraiswami. „Automatic position calibration of multiple microphones“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*. Bd. 4. Montreal, QC, Canada, Mai 2004, S. 69–72. DOI: 10.1109/ICASSP.2004.1326765.

[Red+09] A. Redondi, M. Tagliasacchi, F. Antonacci und A. Sarti. „Geometric calibration of distributed microphone arrays“. In: *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP 2009)*. Rio de Janeiro, Brasilien, Okt. 2009, S. 1–5. DOI: 10.1109/MMSP.2009.5293568.

[RK89] R. Roy und T. Kailath. „ESPRIT-estimation of signal parameters via rotational invariance techniques“. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37.7 (Juli 1989), S. 984–995. ISSN: 0096-3518. DOI: 10.1109/29.32276.

[RKL05] V. Raykar, I. Kozintsev und R. Lienhart. „Position calibration of microphones and loudspeakers in distributed computing platforms“. In: *IEEE Transactions on Speech and Audio Processing* 13.1 (Jan. 2005), S. 70–83. ISSN: 1063-6676. DOI: 10.1109/TSA.2004.838540.

[RKM07] N. Röber, U. Kaminski und M. Masuch. „Ray acoustics using computer graphics technology“. In: *Proceedings of International Conference on Digital Audio Effects (DAFx-07)*. Bordeaux, Frankreich, Sep. 2007, S. 117–124.

[RNL09] N. Raghuvanshi, R. Narain und M. C. Lin. „Efficient and Accurate Sound Propagation Using Adaptive Rectangular Decomposition“. In: *IEEE Transactions on Visualization and Computer Graphics* 15.5 (2009), S. 789–801. DOI: 10.1109/TVCG.2009.28.

[Sab22] W. C. Sabine. *Collected Papers on Acoustics*. Harvard University Press, 1922.

[Sch+09] J. Schmalenstroeer, M. Kelling, V. Leutnant und R. Haeb-Umbach. „Fusing Audio and Video Information for Online Speaker Diarization“. In: *Proceedings of Interspeech 2009*. Brighton, England, Sep. 2009, S. 1163–1166.

[Sch62] M. R. Schroeder. „Frequency-Correlation Functions of Frequency Responses in Rooms“. In: *The Journal of the Acoustical Society of America* 34.12 (1962), S. 1819–1823. DOI: 10.1121/1.1909136.

[Sch65] M. R. Schroeder. „New Method of Measuring Reverberation Time“. In: *The Journal of the Acoustical Society of America* 37.3 (1965), S. 409–412. DOI: 10.1121/1.1909343.

[Sch86] R. Schmidt. „Multiple emitter location and signal parameter estimation“. In: *IEEE Transactions on Antennas and Propagation* 34.3 (März 1986), S. 276–280. ISSN: 0018-926X. DOI: 10.1109/TAP.1986.1143830.

[SE94] W. C. Sabine und M. D. Egan. „Collected Papers on Acoustics“. In: *The Journal of the Acoustical Society of America* 95.6 (1994), S. 3679–3680. DOI: 10.1121/1.409944.

[SH10] J. Schmalenstroeer und R. Haeb-Umbach. „Online Diarization of Streaming Audio-Visual Data for Smart Environments“. In: *IEEE Journal of Selected Topics in Signal Processing* 4.5 (Okt. 2010), S. 845–856. ISSN: 1932-4553. DOI: 10.1109/JSTSP.2010.2050519.

[SJH14] J. Schmalenstroeer, P. Jebramcik und R. Haeb-Umbach. „A combined hardware-software approach for acoustic sensor network synchronization“. In: *Signal Processing* 107 (Feb. 2014). Special Issue on ad hoc microphone arrays and wireless acoustic sensor networks, S. 171–184. ISSN: 0165-1684. DOI: 10.1016/j.sigpro.2014.06.030.

[SSP05] J. Sachar, H. Silverman und W. Patterson. „Microphone position and gain calibration for a large-aperture microphone array“. In: *IEEE Transactions on Speech and Audio Processing* 13.1 (Jan. 2005), S. 42–52. ISSN: 1063-6676. DOI: 10.1109/TSA.2004.834459.

[SZW14] H.-J. Shao, X.-P. Zhang und Z. Wang. „Novel closed-form auxiliary variables based algorithms for sensor node localization using AOA“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*. Florenz, Italien, Mai 2014, S. 1414–1418. DOI: 10.1109/ICASSP.2014.6853830.

[Tag+15] M. J. Taghizadeh, R. Parhizkar, P. N. Garner, H. Bourlard und A. Asaei. „Ad hoc microphone array calibration: Euclidean distance matrix completion algorithm and theoretical guarantees“. In: *Signal Processing* 107 (Feb. 2015). Special Issue on ad hoc microphone arrays and wireless acoustic sensor networks, S. 123–140. ISSN: 0165-1684. DOI: 10.1016/j.sigpro.2014.07.016.

[TH10] D. H. Tran Vu und R. Haeb-Umbach. „Blind speech separation employing directional statistics in an expectation maximization framework“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010)*. Dallas, TX, USA, März 2010, S. 241–244. DOI: 10.1109/ICASSP.2010.5495994.

[Thr05] S. Thrun. „Affine Structure From Sound“. In: *Conference on Neural Information Processing Systems (NIPS 2005)*. Bd. 18. Vancouver, BC, Kanada, Dez. 2005, S. 1353–1360.

[TL08] S. Tervo und T. Lokki. „Interpolation Methods for the SRP-PHAT Algorithm“. In: *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC 2008)*. Seattle, WA, USA, 2008, S. 1–4.

[Tom+15] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi und S. Tubaro. *Deep convolutional neural networks for pedestrian detection*. Okt. 2015. arXiv: 1510.03608 [cs.CV].

[TTH14] O. Thiergart, M. Taseska und E. Habets. „An Informed Parametric Spatial Filter Based on Instantaneous Direction-of-Arrival Estimates“. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.12 (Dez. 2014), S. 2182–2196. ISSN: 2329-9290. DOI: 10.1109/TASLP.2014.2363407.

[Val+10a] S. Valente, E. Antonacci, M. Tagliasacchi, A. Sarti und S. Tubaro. „Self-calibration of two microphone arrays from volumetric acoustic maps in non-reverberant rooms“. In: *Proceedings of International Symposium on Communications Control and Signal Processing (ISCCSP 2010)*. Limassol, Zypern, März 2010, S. 1–4. DOI: 10.1109/ISCCSP.2010.5463318.

[Val+10b] S. Valente, M. Tagliasacchi, E. Antonacci, P. Bestagini, A. Sarti und S. Tubaro. „Geometric calibration of distributed microphone arrays from acoustic source correspondences“. In: *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP 2010)*. Saint-Malo, Frankreich, Okt. 2010, S. 13–18. DOI: 10.1109/MMSP.2010.5661986.

[Vel+15] J. Velasco, M. J. Taghizadeh, A. Asaei, H. Bourlard, C. J. Martin-Arguedas, J. Macias-Guarasa und D. Pizarro. „Novel GCC-PHAT model in diffuse sound field for microphone array pairwise distance based calibration“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015)*. Brisbane, Australien, Apr. 2015, S. 2669–2673. DOI: 10.1109/ICASSP.2015.7178455.

[WC97] H. Wang und P. Chu. „Voice source localization for automatic camera pointing system in videoconferencing“. In: *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 1997)*. New Paltz, NY, USA, Okt. 1997, pages. DOI: 10.1109/ASPAA.1997.625639.

[Wei+07] E. Weinstein, K. Steele, A. Agarwal und J. Glass. „LOUD: A 1020-Node Microphone Array and Acoustic Beamformer“. In: *Proceedings of International Congress on Sound and Vibration (ICSV 2007)*. Cairns, Australia, Juli 2007.

[Wei08] S. Weinzierl. *Handbuch der Audiotechnik*. Springer Science & Business Media, 2008. DOI: 10.1007/978-3-540-34301-1.

[WHP04] E. Warsitz, R. Haeb-Umbach und S. Peschke. „Adaptive Beamforming Combined with Particle Filtering for Acoustic Source Localization“. In: *Proceedings of Interspeech 2004*. Jeju Island, Süd Korea, Okt. 2004, S. 2849–2852.

[WLW03] D. B. Ward, E. Lehmann und R. Williamson. „Particle filtering algorithms for tracking an acoustic source in a reverberant environment“. In: *IEEE Transactions on Speech and Audio Processing* 11.6 (Nov. 2003), S. 826–836. ISSN: 1063-6676. DOI: 10.1109/TSA.2003.818112.

[WY11] G. Wang und K. Yang. „A New Approach to Sensor Node Localization Using RSS Measurements in Wireless Sensor Networks“. In: *IEEE Transactions on Wireless Communications* 10.5 (Mai 2011), S. 1389–1395. ISSN: 1536-1276. DOI: 10.1109/TWC.2011.031611.101585.

[ZBS09] C. Zieger, A. Brutti und P. Svaizer. „Acoustic Based Surveillance System for Intrusion Detection“. In: *Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS 2009)*. Genua, Italien, Sep. 2009, S. 314–319. DOI: 10.1109/AVSS.2009.49.

Eigene Publikationen

[JH16] F. Jacob und R. Haeb-Umbach. „On the Bias of Direction of Arrival Estimation Using Linear Microphone Arrays“. In: *Proceedings of ITG Symposium on Speech Communication (ITG 2016)*. Accepted for Publication. Paderborn, Deutschland, Okt. 2016.

[Pli+16] A. Plinge, F. Jacob, R. Haeb-Umbach und G. A. Fink. „Acoustic Microphone Geometry Calibration - An overview and experimental evaluation of state-of-the-art algorithms“. In: *IEEE Signal Processing Magazine* 33.4 (Juli 2016), S. 14–29. DOI: 10.1109/MSP.2016.2555198.

[DJH15] L. Drude, F. Jacob und R. Hab-Umbach. „DoA-Estimation Based on a Complex Watson Kernel Method“. In: *Proceedings of European Signal Processing Conference (EUSIPCO 2015)*. Nizza, Frankreich, Aug. 2015, S. 255–259. DOI: 10.1109/EUSIPCO.2015.7362384.

[JH15] F. Jacob und R. Haeb-Umbach. *Absolute Geometry Calibration of Distributed Microphone Arrays in an Audio-Visual Sensor Network*. Apr. 2015. arXiv: 1504.03128 [cs.SD].

[JH14] F. Jacob und R. Haeb-Umbach. „Coordinate Mapping Between an Acoustic and Visual Sensor Network in the Shape Domain for a Joint Self-Calibrating Speaker Tracking“. In: *Proceedings of ITG Symposium on Speech Communication (ITG 2014)*. Erlangen, Deutschland, Sep. 2014, S. 1–4.

[JSH13] F. Jacob, J. Schmalenstroeer und R. Haeb-Umbach. „DOA-based Microphone Array Position Self-Calibration Using Circular Statistics“. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*. Vancouver, BC, Kanada, Mai 2013, S. 116–120. DOI: 10.1109/ICASSP.2013.6637620.

[JSH12] F. Jacob, J. Schmalenstroeer und R. Haeb-Umbach. „Microphone Array Position Self-Calibration from Reverberant Speech Input“. In: *Proceedings of International Workshop on Acoustic Signal Enhacement (IWAENC 2012)*. Aachen, Deutschland, Sep. 2012, S. 1–4.

[Sch+11] J. Schmalenstroeer, F. Jacob, R. Haeb-Umbach, M. Hennecke und G. Fink. „Unsupervised Geometry Calibration of Acoustic Sensor Networks Using Source Correspondences“. In: *Proceedings of Interspeech 2011*. Florenz, Italien, Aug. 2011, S. 597–600.

Akronyme

DOA *Direction of Arrival.*

TDOA *Time Difference of Arrival.*

AIR *Aachen Impulse Response Database.*

ASN Akustisches Sensornetz.

BLAS *Basic Linear Algebra Subprograms.*

BMDS Multidimensionale Skalierung mit Basispunkten.

BSS *Blind Source Separation.*

DFT Diskrete FOURIER-Transformation.

DRR *Direct-to-Reverberant Ratio.*

DSB *Delay-and-Sum Beamformer.*

EDC *Energy Decay Curve.*

ESPRIT *Estimation of Signal Parameters via Rotational Invariance Techniques.*

FEM Finite-Elemente-Methode.

FFT *Fast* FOURIER *Transform.*

FFTW *Fastest* FOURIER *Transform in the West.*

FPS *Frames per Second.*

FSB *Filter-and-Sum Beamformer.*

FSBPhat *Steered Response Power with Phase Transform* auf den Impulsantworten des *Filter-and-Sum Beamformer.*

GCCPhat *Generalized Cross Correlation with Phase Transform.*

GSCT *Generalized State Coherrence Transform.*

HMA *Huge Microphone Array.*

HMM *Hidden* MARKOV *Model.*

HOG *Histogram of Oriented Gradient.*

IDFT Inverse diskrete FOURIER-Transformation.

KLDS Kreuzleistungsdichtespektrum.

LDS Leistungsdichtespektrum.

LED *Light-Emitting Diode*.

LORANSAC *Locally Optimized Random Sample Consensus*.

LOS *Line-of-Sight*.

LS *Least-Squares*.

LYDE Loesch und Yang DOA-*Estimator*.

MDS Multidimensionale Skalierung.

MIRD *Multichannel Impulse Response Database*.

ML Maximum-Likelihood.

MTDOA *Maximum Time Difference of Arrival*.

MTF *Multiplicative Transfer Function*.

MUSIC *Multiple Signal Classification*.

NLOS *Non-Line-of-Sight*.

PRANSAC *Partitioned Random Sample Consensus*.

R-MUSIC *Root*-MUSIC.

RAM *Random Access Memory*.

RANSAC *Random Sample Consensus*.

RBT *Rigid Body Transformation*.

RIA Raumimpulsantwort.

RMSE *Root-Mean-Square Error*.

RSS *Received Signal Strength*.

SMS *Short Message Service*.

SNR *Signal-to-Noise Ratio*.

SRP *Steered Response Power*.

SRPPhat *Steered Response Power with Phase Transform*.

STFT *Short-Time* FOURIER *Transform*.

SVD *Singular Value Decomposition.*

SVM *Support Vector Machine.*

TIMIT *Texas Instruments (TI) and Massachusetts Institute of Technology (MIT) Acoustic-Phonetic Continuous Speech Corpus.*

TOA *Time of Arrival.*

TOF *Time of Flight.*

VAD *Voice Activity Detection.*

WKM Watson-Kern-Methode.

WSJCam0 *Wall Street Journal recorded at the University of Cambridge (phase 0).*