

# **Intervall-basierte Kartierung von statischen Soundquellen durch ein mobiles Mikrofonarray**

Von der Fakultät für Elektrotechnik und Informatik  
der Gottfried Wilhelm Leibniz Universität Hannover  
zur Erlangung des akademischen Grades

Doktor-Ingenieur

(abgekürzt Dr.-Ing.)

genehmigte Dissertation

von M. Sc. Axel Rauschenberger  
geboren am 28.10.1988 in Burgwedel

2022

Referent: Prof. Dr.-Ing. Bernardo Wagner

Korreferent: Prof. Dr.-Ing. Holger Blume

Tag der Promotion: 28.07.2022

# Vorwort

Die vorliegende Dissertation habe ich während meiner Zeit als wissenschaftlicher Mitarbeiter am Fachgebiet Echtzeitsysteme (RTS) des Instituts für Systems Engineering an der Leibniz Universität Hannover angefertigt.

Mein besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr.-Ing. Bernardo Wagner, der mich während meines Bachelorstudiums von der mobilen Servicerobotik begeistert hat. Die entfachte Faszination war maßgeblich dafür, dass ich in diesem Bereich mit einer Promotion begonnen habe. Während meiner Zeit als wissenschaftlicher Mitarbeiter hat mich Herr Prof. Dr.-Ing. Bernardo Wagner exzellent betreut, konstruktiv unterstützt und stetig motiviert.

Ein Dank gilt ebenso Herrn Prof. Dr.-Ing. Holger Blume für die Übernahme des Korreferats.

Meine Dissertation wäre nicht ohne die wissenschaftlichen Diskussionen mit meinen ehemaligen Kolleginnen und Kollegen vom RTS und DEI erfolgreich gewesen. An die vertraute Zusammenarbeit und die gemeinsamen Aktivitäten erinnere ich mich gerne zurück.

Nicht zuletzt danke ich meinen Eltern für die finanzielle Unterstützung während meines Studiums und meinen Freunden für die unermüdliche Motivation. Der größte Dank gebührt meiner Frau Veronica, die mir während der anstrengenden Phasen dieser Arbeit stets den Rücken gestärkt hat.



# Kurzfassung

Mobile Serviceroboter nehmen vermehrt Einzug in das tägliche Leben und das industrielle Umfeld. Um ihre Aufgaben umzusetzen, müssen sie die Fähigkeit besitzen, ihre Umgebung ausreichend wahrzunehmen. Viel Aufmerksamkeit wurde bisher der visuellen Erfassung durch Kameras oder Laserscanner gegeben. Im Vergleich dazu konzentriert sich hingegen wenig Forschung auf die akustische Wahrnehmung. In vielen Einsatzgebieten zeigt sich, dass Mikrofone relevante Informationen aus der Umgebung ermitteln können, die durch visuelle Sensoren nicht erfassbar sind.

Ziel dieser Dissertation ist es, ein neues Verfahren zu entwickeln, um die dreidimensionale Kartierung von mehreren statischen Soundquellen innerhalb einer metrischen Karte durchzuführen. Dabei wird eine mobile Plattform verwendet, die mit mehreren Mikrofonen ausgerüstet ist und autonom durch die Umgebung navigiert.

Für die Kartierung wird in dieser Arbeit das neue Verfahren *IB-SSM* (engl. *Interval-based Sound Source Mapping*) vorgestellt. *IB-SSM* basiert auf der Intervall-Arithmetik und führt mathematische Berechnungen auf begrenzten Mengen durch. Im Vergleich zu bekannten Verfahren der Soundquellenkartierung werden einige Einschränkungen aufgehoben. So wird für *IB-SSM* die Gesamtzahl der aktiven Soundquellen nicht als bekannt vorausgesetzt. Bekannte Verfahren heben geometrische Mehrdeutigkeiten durch die Randbedingungen in der Versuchsdurchführung auf. Demgegenüber werden in *IB-SSM* die geometrischen Mehrdeutigkeiten explizit modelliert.

Um die Robustheit des Verfahrens für reale Anwendungen zu steigern, werden in dieser Arbeit die Unsicherheiten der akustischen Merkmale und der Mikrofonpositionen durch Methoden der Intervall-Arithmetik beschrieben. Um die Unsicherheit der akustischen Merkmale zu ermitteln, werden in dieser Dissertation zur vergleichenden Evaluation die zwei alternativen Verfahren *DoATiD* (engl. *Direction-of-Arrival-based Time Difference*) und *InTiD* (engl. *Interval-based Time Difference*) entwickelt. *DoATiD* verwendet für die Berechnung die Ergebnisse von bekannten Ansätzen der Soundquellenlokalisierung. Im Gegensatz dazu basiert *InTiD* direkt auf intervall-basierten Methoden, die auf die Mikrofonsignale angewendet werden.

Die vorgestellten Verfahren werden jeweils zunächst mit simulierten sowie nachfolgend mit realen Daten evaluiert. Weiterhin erfolgt ein systematischer Vergleich mit klassischen Verfahren. Diese Arbeit zeigt, dass das vorgestellte Kartierungsverfahren *IB-SSM* korrekte Bereiche für die Position von mehreren Soundquellen berechnet und in realen Umgebungen anwendbar ist.

## Schlagnworte:

Robot Audition, Soundquellenkartierung, Soundquellenlokalisierung, Intervallarithmetik



# Abstract

Mobile service robots are increasingly finding their way into daily life and the industrial environment. To perform their tasks, they must have the ability to sufficiently perceive their environment. So far, much attention has been given to visual detection by using cameras or laser scanners. In comparison, less research focuses on acoustic perception. However, it has been shown in many applications, that microphones can extract relevant information from the environment which cannot be detected by visual sensors.

The goal of this thesis is to develop a novel method to perform three-dimensional sound source mapping of multiple static sound sources within a metric map. Therefore, a mobile platform is equipped with multiple microphones and navigates autonomously through the environment.

To perform the mapping, the novel method *IB-SSM* is presented in this thesis. *IB-SSM* (*Interval-based Sound Source Mapping*) is based on interval arithmetic and performs mathematical calculations on bounded sets. Compared to known sound source mapping methods, some limitations are removed. For example, *IB-SSM* does not assume the total number of active sound sources to be known. Furthermore, known methods remove geometric ambiguities due to the boundary conditions in the experimental design. In contrast, geometric ambiguities are explicitly modeled in *IB-SSM*.

To increase the robustness of the method for real-world applications, the uncertainties of the acoustic features and microphone positions are described by methods based on the interval analysis. To determine the uncertainty of the acoustic features, two alternative methods *DoATiD* (*Direction-of-Arrival-based Time Difference*) and *InTiD* (*Interval-based Time Difference*) are developed in this dissertation for comparative evaluation. *DoATiD* is based on well-known approaches to estimate the direction of sound sources. In contrast, *InTiD* is based directly on interval-based methods applied to microphone signals.

This thesis shows that the presented mapping method *IB-SSM* computes correct ranges for the position of multiple sound sources and is applicable in real environments.

## **Keywords:**

Robot Audition, Sound Source Localization, Sound Source Mapping, Interval Analysis



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Problemstellung . . . . .	1
1.2.1	Kartierung von Soundquellen . . . . .	2
1.2.2	Wissenschaftlicher Beitrag und Zielsetzung . . . . .	5
1.3	Randbedingungen . . . . .	5
1.4	Lösungsansatz . . . . .	7
1.5	Gliederung der Arbeit . . . . .	8
<b>2</b>	<b>Stand der Wissenschaft</b>	<b>11</b>
2.1	Physikalische Grundlagen der Akustik und Datenverarbeitung . . . . .	11
2.1.1	Akustik . . . . .	11
2.1.2	Dämpfung . . . . .	15
2.1.3	Reflexionen . . . . .	15
2.1.4	Mikrofone und diskrete Audiodaten . . . . .	16
2.1.5	Frequenzanalyse . . . . .	21
2.2	Soundquellenlokalisierung . . . . .	24
2.2.1	Einführung . . . . .	24
2.2.2	Binaurale Verfahren . . . . .	25
2.2.3	Mikrofon-Array-Verfahren . . . . .	28
2.3	Zustandsschätzung und Kartierung . . . . .	40
2.3.1	Einführung . . . . .	40
2.3.2	Soundquellenkartierung - Übersicht . . . . .	41
2.3.3	Probabilistische Zustandsschätzung . . . . .	43
2.3.4	Probabilistische Soundquellenkartierung . . . . .	46
2.4	Intervallarithmetik . . . . .	54
2.4.1	Motivation . . . . .	54
2.4.2	Grundbegriffe . . . . .	55
2.4.3	Inklusionsfunktion . . . . .	56
2.4.4	Subpavings . . . . .	57
2.4.5	SIVIA . . . . .	58
2.4.6	Kontraktoren . . . . .	59
2.5	Bezug zur Arbeit . . . . .	62

<b>3</b>	<b>Intervall-basierte Soundquellenkartierung</b>	<b>65</b>
3.1	Problemdefinition und Notation . . . . .	66
3.2	Erweiterung auf mehrere Paare, Soundquellen und Messpositionen . . . . .	67
3.3	Intervall-basierter Lösungsansatz . . . . .	69
3.4	Behandlung von akustischen Reflexionen . . . . .	71
3.5	Ausblick - Behandlung von Ausreißern . . . . .	72
<b>4</b>	<b>Lokalisierung Mikrofonposition</b>	<b>75</b>
4.1	Relative Mikrofonposition . . . . .	76
4.2	Globale Lokalisierung innerhalb einer Karte . . . . .	77
4.3	Ausblick - Erweiterung auf Multi-Hypothesen . . . . .	80
4.4	Ausblick - Erweiterung auf garantierte Lokalisierung . . . . .	80
<b>5</b>	<b>Zeitliche Differenzschätzung</b>	<b>81</b>
5.1	Richtungs-basierte Zeitdifferenzschätzung (DoATiD) . . . . .	82
5.1.1	Ansatz (1) - Universelles Verfahren (DoATiD) . . . . .	82
5.1.2	Verbesserung der Richtungsschätzung - Ansatz I: M3K . . . . .	88
5.1.3	Verbesserung der Richtungsschätzung - Ansatz II: GM-PHD . . . . .	91
5.1.4	Berechnung der Zeitdifferenz innerhalb eines Kegelsegments . . . . .	92
5.1.5	Ansatz (2) - Konkretisierung auf SRP-PHAT-Verfahren (DoATiD) . . . . .	94
5.2	Mikrofon-basierte Zeitdifferenzschätzung (InTiD) . . . . .	98
5.2.1	Einzelne Soundquelle . . . . .	98
5.2.2	Algorithmus . . . . .	102
5.2.3	Erweiterungen . . . . .	103
5.2.4	Ausblick - Übertragung auf mehrere Soundquellen . . . . .	103
<b>6</b>	<b>Experimentelle Untersuchungen und Ergebnisse</b>	<b>105</b>
6.1	Randbedingungen der Evaluation . . . . .	105
6.1.1	Ziele und Ablauf der Evaluation . . . . .	105
6.1.2	Experimentalumgebung . . . . .	109
6.1.3	Roboterplattform und Sensorik . . . . .	109
6.2	Versuchsdurchführung . . . . .	110
6.3	Evaluation A: Fehlerfreie Messwerte . . . . .	112
6.3.1	Parameter . . . . .	112
6.3.2	Experiment 1 - Einzelne Soundquelle - ohne geometrische Mehrdeutigkeiten . . . . .	113
6.3.3	Experiment 2 - Einzelne Soundquelle - geometrische Mehrdeutigkeiten . . . . .	114
6.3.4	Experiment 3 - Mehrere Soundquellen - ohne geometrische Mehrdeutigkeiten . . . . .	116
6.3.5	Experiment 4 - Mehrere Soundquellen - geometrische Mehrdeutigkeiten . . . . .	118
6.4	Evaluation B: Fehlerbehaftete Messwerte . . . . .	120
6.4.1	Experiment 5 - Einzelne Soundquelle - Abtastfehler . . . . .	120
6.4.2	Experiment 6 - Einzelne Soundquelle - Orientierungsfehler . . . . .	121
6.4.3	Experiment 7 - Einzelne Soundquelle - Mikrofonpositionsfehler . . . . .	122
6.5	Evaluation C: Zeitdifferenzschätzung . . . . .	124
6.5.1	Richtungs-basierte Schätzung (DoATiD) . . . . .	124
6.5.2	Mikrofon-basierte Schätzung (InTiD) . . . . .	132

---

6.6	Evaluation D: Gesamtsystem - Soundquellenkartierung . . . . .	136
6.6.1	Richtungs-basierte Schätzung (DoATiD) . . . . .	136
6.6.2	Mikrofon-basierte Schätzung (InTiD) . . . . .	137
6.6.3	Parameterstudie - Einfluss unterschiedlicher Parameter . . . . .	138
6.6.4	Proof-of-Concept in realer Umgebung . . . . .	143
<b>7</b>	<b>Bewertung und Diskussion</b>	<b>145</b>
7.1	Soundquellenkartierung mit fehlerfreien Messwerten . . . . .	145
7.2	Soundquellenkartierung mit fehlerbehafteten Messwerten . . . . .	147
7.3	Akustische Merkmale - Extraktion der TDoA und Modellierung der Unsicherheit . . . . .	148
7.4	Gesamtsystem Soundquellenkartierung mit Modellierung der Unsicherheit . . . . .	149
7.5	Bezug zur Zielsetzung . . . . .	150
7.6	Zukünftige Forschungsfragen . . . . .	150
<b>8</b>	<b>Zusammenfassung</b>	<b>151</b>
	<b>Literaturverzeichnis</b>	<b>155</b>



# Nomenklatur

## Abkürzungen

<b>2D</b>	zweidimensional
<b>3D</b>	dreidimensional
<b>ASA</b>	engl. <i>Auditory Scene Analysis</i> (Auditorielle Szenenanalyse)
<b>CASA</b>	engl. <i>Computational Auditory Scene Analysis</i> (Computergestützte auditorielle Szenenanalyse)
<b>CC</b>	engl. <i>Cross Correlation</i> (Kreuzkorrelation zur Berechnung der Ähnlichkeit zwischen Signalen)
<b>dB</b>	Dezibel
<b>DoA</b>	engl. <i>Direction of Arrival</i> (Richtung, aus der ein Signal wahrgenommen wird)
<b>DoATiD</b>	engl. <i>Direction-of-Arrival-based Time Difference</i> (In dieser Arbeit vorgestelltes Verfahren zur Ermittlung der TDoA basierend auf DoA-Schätzungen)
<b>EVD</b>	engl. <i>Eigenvalue Decomposition</i> (Eigenwertzerlegung)
<b>ECM</b>	engl. <i>Electret Condenser Microphone</i> (Elektret-Kondensatormikrofon)
<b>EKF</b>	engl. <i>Extended Kalman Filter</i> (Erweiterter Kalman-Filter)
<b>FFT</b>	engl. <i>Fast Fourier transformation</i> (Schnell berechenbare Fouriertransformation)
<b>FT</b>	Fouriertransformation
<b>GCC</b>	engl. <i>Generalized Cross Correlation</i> (Generalisierte Kreuzkorrelation)
<b>GMM</b>	engl. <i>Gaussian Mixture Models</i> (Gaußsche Mischmodelle)
<b>GM-PHD</b>	engl. <i>Gaussian Mixture Probability Hypothesis Density</i> (Gaußsche Mischmodelle zur Ermittlung der PHD)
<b>GEVD</b>	engl. <i>Generalized Eigenvalue Decomposition</i> (Generalisierte Eigenwertzerlegung)
<b>GSVD</b>	engl. <i>Generalized Singular Value Decomposition</i> (Generalisierte Singulärwertzerlegung)
<b>HRTF</b>	engl. <i>Head Related Transfer Function</i> (kopfbezogene Übertragungsfunktion)

---

<b>IB-SSM</b>	engl. <i>Interval-based Sound Source Mapping</i> (In dieser Arbeit vorgestellter intervall-basiertes Verfahren zur Soundquellenkartierung)
<b>ILD</b>	engl. <i>Interaural Level Difference</i> (Interaurale Pegeldifferenz)
<b>InTiD</b>	engl. <i>Interval-based Time Difference</i> (In dieser Arbeit vorgestelltes intervall-basiertes Verfahren zur Ermittlung der TDoA)
<b>ITD</b>	engl. <i>Interaural Time Difference</i> (Interaurale Zeitdifferenz)
<b>LM</b>	Levenberg-Marquardt
<b>M3K</b>	engl. <i>Modified 3D Kalman-Filter</i> (Modifiziertes Kalman-Verfahren zur Ermittlung der Richtung zu einer Soundquelle)
<b>MEMS</b>	engl. <i>Microelectromechanical systems</i> (Mikroelektromechanisches System)
<b>MUSIC</b>	engl. <i>Multiple Signal Classification</i> (Verfahren zur Bestimmung der Richtung zu einer Soundquelle)
<b>PHD-Filter</b>	engl. <i>Probability Hypothesis Density Filter</i> (Rekursives Verfahren zur Ermittlung Posterior-Intensität )
<b>PHAT</b>	engl. <i>Phase Transform</i> (Phasentransformation)
<b>SVD</b>	engl. <i>Singular Value Decomposition</i> (Singularwertzerlegung)
<b>SLAM</b>	engl. <i>Simultaneous Localization and Mapping</i> (Verfahren um eine Karte zu erstellen und eine Lokalisierung in jener durchführen)
<b>SPL</b>	engl. <i>Sound Pressure Level</i> (Schalldruckpegel)
<b>SNR</b>	engl. <i>Signal-to-noise ratio</i> (Signal-Rausch-Verhältnis)
<b>SRP</b>	engl. <i>Steered Response Power</i> (Verfahren zur Bestimmung der Richtung zu einer Soundquelle)
<b>SSL</b>	engl. <i>Sound Source Localization</i> (Soundquellenlokalisierung - Methoden zur Richtungsschätzung zu Soundquellen)
<b>SSM</b>	engl. <i>Sound Source Mapping</i> (Soundquellenkartierung - Verfahren zur Kartierung von Soundquellen)
<b>SSS</b>	engl. <i>Sound Source Separation</i> (Soundquellenseparierung - Verfahren zur Trennung von akustischen Signalen)
<b>STFT</b>	engl. <i>Short-time Fourier transform</i> (Kurzzeit-Fouriertransformation)
<b>TDoA</b>	engl. <i>Time Difference of Arrival</i> (Zeitliche Differenz des Eintreffens eines Signals an unterschiedlichen Positionen)
<b>THD</b>	engl. <i>Total Harmonic Distortion</i> (Gesamte harmonische Verzerrung)

## Allgemeine Notation

$x$	Skalar
$\mathbf{x}$	Vektor
$\mathbf{X}$	Matrix
$X$	Skalar repräsentiert in Fourierraum
$\underline{\mathbf{x}}$	Vektor repräsentiert in Fourierraum
$\underline{\mathbf{X}}$	Matrix repräsentiert in Fourierraum
$[x]$	Intervallgröße von Skalar
$\underline{x}$	Untere Grenze von $[x]$
$\bar{x}$	Obere Grenze von $[x]$
$[\mathbf{x}]$	Intervallvektor (respektive Intervallbox)

## Auswahl von Formelzeichen

${}^{(i)}\tilde{\alpha}_k$	Dämpfung zwischen Soundquelle $s_i$ und Mikrofon $m_k$
${}^{(i)}\alpha_{kl}$	Dämpfung zwischen Mikrofon $m_k$ und $m_l$ für die Soundquelle $i$
$\delta$	Messrauschen für EKF
$\varepsilon$	Prozessrauschen für EKF
$\theta_i$	Azimutwinkel (Horizontaler Winkel) zu Soundquelle $s_i$
$\vartheta$	Temperatur in °C
$\lambda$	Wellenlänge
$\mu$	Erwartungswert von Normalverteilung
$\bar{\mu}$	Prädizierter Erwartungswert von Normalverteilung
$\Sigma$	Kovarianz von Normalverteilung
$\bar{\Sigma}$	Prädizierte Kovarianz von Normalverteilung
$\phi_i$	Elevationwinkel (Vertikaler Winkel) zu Soundquelle $s_i$
$\Phi_0$	Phasenverschiebung
$\Psi(f)$	Gewichtungsfunktion für die Berechnung der Kreuzkorrelation
$\omega$	Kreisfrequenz
${}^h\omega$	Gewicht für Hypothese $h$
$a$	(Schalldruck-)Amplitude
$A^k$	Jacobi-Matrix des Bewegungsmodells für EKF
$\underline{\mathbf{a}}(f, \theta, \phi)$	Ausrichtungsvektor im Fourierraum (engl. <i>Steering Vector</i> )
$\underline{\mathbf{A}}(f, \theta, \phi)$	Matrix von Ausrichtungsvektoren $\underline{\mathbf{a}}(f, \theta, \phi)$ im Fourierraum
$bel$	Belief - Wahrscheinlichkeit
$c$	Schallgeschwindigkeit
$\mathbf{c}$	Steuerungseingabe
$\underline{\mathbf{C}}$	Matrixdarstellung des Kreuzleistungsdichtespektrums
$\mathcal{C}$	Kontraktor für intervall-basierte Berechnung

<b><u>D</u></b>	Diagonalmatrix der Eigenwerte von <b><u>C</u></b> im Fourierraum
<b><u>E</u></b>	Energiefunktion
<b><u>e</u></b>	Eigenvektor von <b><u>C</u></b> im Fourierraum
<b><u>E</u></b>	Matrix mit Eigenvektoren <b><u>e</u></b> im Fourierraum
<b><u>f</u></b>	Frequenz
<b><u>f<sub>s</sub></u></b>	Abtastfrequenz
<b><u>f<sub>l</sub><sup>k</sup></u></b>	Block (engl. <i>Frame</i> ) von Mikrofondaten von Mikrofon $m_l(t)$
<b><u>g</u></b>	Nichtlineares Bewegungsmodell für EKF
<b><u>h</u></b>	Nichtlineares Sensormodell für EKF
<b><u>h(f, θ, φ)</u></b>	Pseudospektrum für MUSIC-Verfahren
<b><u>H</u></b>	Jacobi-Matrix des Sensormodells für EKF
<b><u>J</u></b>	Jacobi-Matrix für Levenberg-Marquardt-Verfahren
<b><u>k<sub>i</sub></u></b>	Repräsentation der Position einer Soundquelle durch eine hyperbolische Kurve
<b><u>K</u></b>	Kreuzleistungsdichtespektrum des Rauschens im Fourierraum
<b><u>L</u></b>	Spatial Likelihood Function
<b><u>L<sub>p</sub></u></b>	Schalldruckpegel
<b><u>m<sub>k</sub>(t)</u></b>	Mikrofonsignal an Index $k$
<b><u>M<sub>k</sub>(f)</u></b>	Fouriertransformierte von $m_k(t)$
<b><u><u>m</u></u>(f)</b>	Vektor von Fouriertransformierten von $M_k(f)$
<b><u>n<sub>e</sub></u></b>	Gesamtanzahl von Fourierpunkten
<b><u>n<sub>f</sub></u></b>	Gesamtanzahl von Frequenzanteilen in einem breitbandigen Signal
<b><u>n<sub>ip</sub></u></b>	Interpolationrate zur Berechnung der Kreuzkorrelation
<b><u>n<sub>k</sub></u></b>	Gesamtanzahl von Kombinationsmöglichkeiten
<b><u>n<sub>l</sub></u></b>	Gesamtanzahl von Messpositionen
<b><u>n<sub>m</sub></u></b>	Gesamtanzahl von Mikrofonen
<b><u>n<sub>s</sub></u></b>	Gesamtanzahl von Soundquellen in Umgebung
<b><u>n<sub>t</sub></u></b>	Gesamtanzahl von Gleichungen für Soundquellenkartierung
<b><u>n<sub>p</sub></u></b>	Gesamtanzahl von Mikrofonpaaren
<b><u>n<sub>z</sub></u></b>	Gesamtanzahl von möglichen Zuordnungen
<b><u>n<sub>λ</sub></u></b>	Gesamtanzahl von Eigenwerten
<b><u><sup>(i)</sup>o(t)</u></b>	Ausgang des Beamforming-Verfahrens von Signal $s_i(t)$
<b><u>p</u></b>	Schalldruck
<b><u>p(·)</u></b>	Wahrscheinlichkeit
<b><u>Q</u></b>	Kovarianz des Messrauschens für EKF
<b><u>r</u></b>	Radius
<b><u>R</u></b>	Kovarianz des Prozessrauschens für EKF
<b><u>r<sub>m<sub>k</sub>,m<sub>l</sub></sub></u></b>	Kreuzkorrelation zwischen Mikrofonsignal $m_k$ und $m_l$
<b><u><math>\hat{r}_{m_k,m_l}</math></u></b>	Generalisierte Kreuzkorrelation zwischen Mikrofonsignal $m_k$ und $m_l$
<b><u>R<sub>m<sub>k</sub>,m<sub>l</sub></sub></u></b>	Kreuzleistungsdichtespektrum zwischen Mikrofonsignal $m_k$ und $m_l$

$\hat{R}_{m_k, m_l}$	Generalisiertes Kreuzleistungsdichtespektrum zwischen Mikrofonensignal $m_k$ und $m_l$
$s(t)$	Akustisches Signal über die Zeit $t$
$S(f)$	Fouriertransformation von $s(t)$ über die Frequenz $f$
$\underline{s}(f)$	Vektor von Fouriertransformationen von $S(f)$
$t$	Zeit
${}^{(i)}t_k$	Ankunftszeit von dem Signal $s_i(t)$ am Mikrofon $k$
${}^{(i)}\Delta t_{kl}^{(n)}$	Zeitliche Differenz TDoA zwischen Mikrofon $m_k$ und $m_l$ für die Soundquelle $i$ an der Messposition $n$
${}^{(i)}\Delta t_j^{(n)}$	Zeitliche Differenz TDoA zwischen dem Mikrofonpaar $j$ für die Soundquelle $i$ an der Messposition $n$
$\mathbf{u}_{k_i}$	Normierte Richtung zu Kandidat $k_i$
$w(t)$	Fensterfunktion über die Zeit $t$
$W_k(f)$	Fouriertransformierter Rauschanteil von Mikrofonensignal $m_k(t)$
$\underline{w}(f)$	Vektor von Fouriertransformierten von $W_k(f)$
$\mathbf{x}$	3D-Punkt
$\mathbf{x}^k$	Zustand zum diskreten Zeitpunkt $k$
$\bar{\mathbf{x}}_i^k$	Prädizierter Zustand zum diskreten Zeitpunkt $k$
$\mathbf{x}_{k_i}$	3D-Position von Kandidat $k_i$
$\mathbf{x}_{s_i}$	3D-Position der Soundquelle $s_i$
$\mathbf{x}_{m_k}^{(n)}$	3D-Position des Mikrofon $m_k$ an der Messposition $n$
$\mathbf{z}^k$	Beobachtung zum diskreten Zeitpunkt $k$



# Kapitel 1

## Einleitung

### 1.1 Motivation

Mobile Roboter sind auf dem Vormarsch. Sie werden sowohl im industriellen Umfeld (z. B. Logistik oder Agrartechnik) als auch für kommerzielle Zwecke (z. B. Staubsauger- und Rasenmäroboter) eingesetzt. Im Jahr 2019 waren schätzungsweise 18,6 Millionen Serviceroboter im privaten Haushalt im Einsatz [Wor20]. Bis zum Jahr 2023 wird damit gerechnet, dass sich die Anzahl der Serviceroboter verdoppelt.

Aufgrund der steigenden Akzeptanz von mobilen Robotern werden weitere Einsatzgebiete erschlossen. Um die wachsenden Anforderungen zu erfüllen, müssen mobile Roboter ein Modell der Umgebung auf Basis ihrer Sensordaten erzeugen können. Eingesetzt werden hierfür vermehrt visuelle Sensoren wie Laserscanner oder Farbkameras. Mikrofone werden hingegen kaum verwendet, um ein Umgebungsmodell zu erstellen. Stattdessen werden akustische Signale hauptsächlich für Sprachassistenten (z. B. Siri von Apple oder Alexa von Amazon) genutzt. Dennoch können durch Mikrofone relevante Informationen über die Umgebung ermittelt werden, die nicht mit visuellen Sensoren erfassbar sind. Beispielsweise können Fehlfunktionen von technischen Anlagen erkannt werden, nachdem ein abnormales Geräusch detektiert wurde. Weiterhin können Informationen von akustischen Signalen für die Suche von vermissten Personen nach einem Disaster-Szenario verwendet werden. Für die genannten Einsatzgebiete ist es erforderlich die Position einer Soundquelle innerhalb einer metrischen Karte zu ermitteln. So könnte technisches Personal zum Fehlerort in einer möglicherweise gestörten Anlage geschickt werden, oder die Suche verschütteter Personen könnte durch ein Bergungsteam in einem konkret identifizierten Bereich aufgenommen werden. Als Resultat können unnötige Risiken vermieden werden, indem die Helfer nur in stark eingeschränkten Bereichen suchen müssen. Für die genannten Einsatzgebiete ist es nicht notwendig und technisch auch aktuell nicht möglich, die Position einer Soundquelle hochpräzise zu ermitteln. Stattdessen ist für technisches Personal ein räumlich begrenzter Bereich ausreichend, in dem sich die fehlerhafte Komponente einer Anlage befindet.

### 1.2 Problemstellung

Das steigende Interesse an der akustischen Wahrnehmung durch mobile Roboter führte erst vor wenigen Jahren zur Entstehung des neuen Forschungsgebietes der *Robot Audition*. Der Begriff wurde

durch Arbeiten von Nakadai et al. geprägt [Nak+00]. Hierbei werden klassische Verfahren der Akustik und Signalverarbeitung im Kontext der mobilen Robotik eingesetzt. Folgende Gebiete bilden die Basis der *Robot Audition* [ON15]:

- Soundquellenlokalisierung (engl. *Sound Source Localization* - SSL),
- Soundquellenseparierung (engl. *Sound Source Separation* - SSS), sowie
- Erkennung und Interpretation (engl. *Automatic Sound Recognition* - ASR).

Die Soundquellenlokalisierung beschäftigt sich mit Verfahren, um die Position von Soundquellen zu ermitteln. Je nach verfügbarem Wissen und eingesetzten Ansätzen beschränkt sich die Lokalisierung jedoch darauf, die relative Richtung zu einer Soundquelle zu schätzen. Als Resultat können akustische Signale aus der ermittelten Richtung verstärkt und aus den anderen Richtungen unterdrückt werden. Weiterhin kann das Ergebnis der Soundquellenlokalisierung genutzt werden, um einzelne Soundsignale (z. B. unterschiedliche Instrumente bei einem Konzert) aus den akustischen Daten zu extrahieren. Dieser Vorgang wird als Soundquellenseparierung bezeichnet. Im Anschluss werden die separierten Signale analysiert. Hierbei kann das Signal charakterisiert werden - z. B. als Sprachsignal, Musik oder Martinshorn - und eine weitergehende Interpretation erfolgen. So können aus einem Sprachsignal die Befehle für ein technisches System erkannt werden.

Die drei Gebiete der *Robot Audition* können durch eine Vielzahl von klassischen akustischen Verfahren adressiert werden. Jedoch ergeben sich für deren Einsatz in der mobilen Robotik aufgrund zusätzlicher Annahmen vorher nicht berücksichtigte Probleme und Einschränkungen. So resultieren akustische Reflexionen in geschlossenen Räumen aufgrund der Wände und der enthaltenen Objekte. Weiterhin können Umgebungsrauschen und die Anwesenheit von mehreren Soundquellen zu Fehlinterpretation bei der Analyse führen. Als Folge können klassische akustische Verfahren fehlerhafte oder unbrauchbare Ergebnisse liefern. Für die eingesetzten Mikrofone gelten geometrische und hardwaretechnische Einschränkungen. So ist die maximale Entfernung zwischen den Mikrofonen durch die Dimension der mobilen Plattform beschränkt. Darüber hinaus sind viele klassische Verfahren der Akustik auf schmalbandige Signale optimiert. Im Gegensatz dazu treten im Kontext der mobilen Robotik vermehrt breitbandige Signale auf, z. B. menschliche Stimmen oder Umgebungsgeräusche. Weiterhin ist die Berechnungsdauer der Verfahren für viele Fragestellungen in der klassischen Akustik nicht von Interesse. Dahingegen sind kurze Berechnungszeiten in der *Robot Audition* essentiell, um aktiv auf akustische Signale in der Umgebung reagieren zu können.

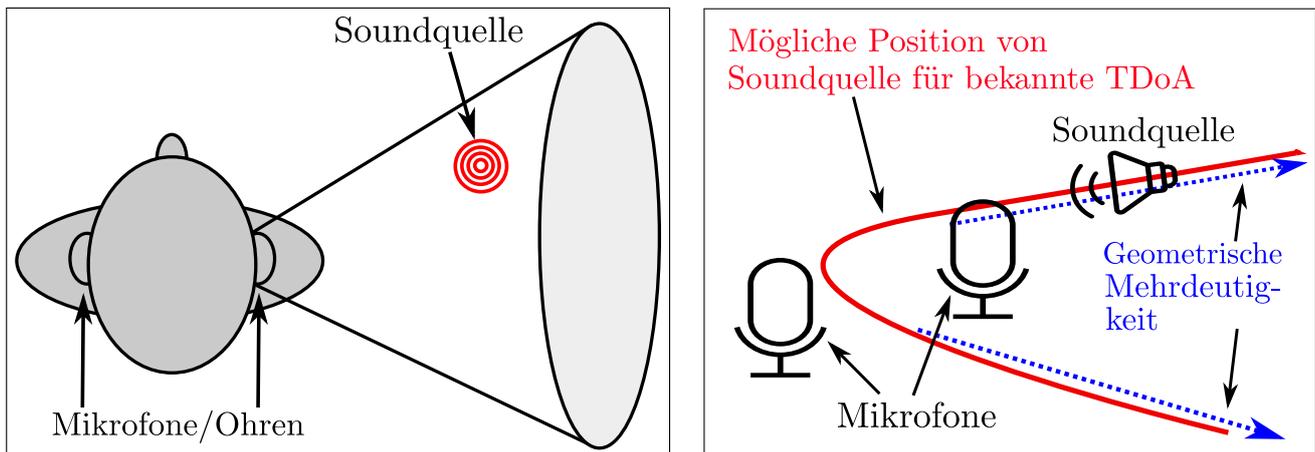
### 1.2.1 Kartierung von Soundquellen

Die Kartierung von Soundquellen (engl. *Sound Source Mapping* - SSM) ist ein wichtiges Ziel in der *Robot Audition*. Aufgabe ist es dabei durch mobile Roboter die zwei- oder dreidimensionale Position von Soundquellen in einer (metrischen oder topologischen) Karte zu ermitteln. Hierfür werden Messungen an unterschiedlichen räumlichen Positionen durchgeführt. Um die Soundquellen zu kartieren, müssen:

- akustische Merkmale aus den Mikrofondaten extrahiert und
- die Positionen der Mikrofone innerhalb der Karte ermittelt werden.

Ein wichtiges akustisches Merkmal ist die *TDoA* (engl. *Time Difference of Arrival*). Sie beschreibt die zeitliche Differenz, die eine akustische Welle benötigt, um ein Mikrofonpaar zu erreichen (die

Definition ist in Abschnitt 2.2.3.1 gegeben). Betrachtet werden akustische Wellen, die von einer Soundquelle emittiert wurden. Falls die Position einer Soundquelle bekannt ist, kann für ein Mikrofonpaar eindeutig die  $TDoA$  berechnet werden. Im Umkehrschluss kann jedoch bei bekannter  $TDoA$  nicht direkt auf die Position der Soundquelle geschlossen werden. Aufgrund von geometrischen Mehrdeutigkeiten befindet sich die Position auf einem Kegel (vgl. Abbildung 1.1a), der in der Psychophysik als *Cone-of-Confusion* bezeichnet wird [WBS54]. Abbildung 1.1a veranschaulicht den resultierenden Kegel, der bei der akustischen Wahrnehmung des Menschen resultiert.



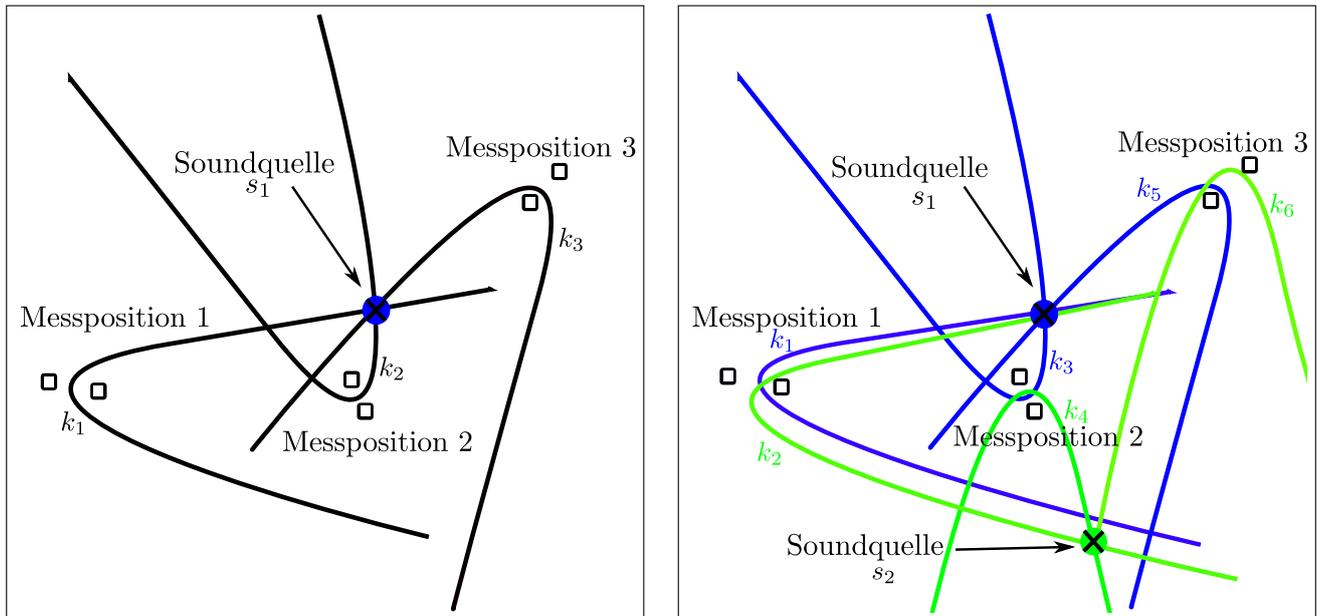
(a) Die Soundquelle befindet sich auf dem *Cone-of-Confusion*.

(b) Im zweidimensionalen Fall befindet sich die Soundquelle auf einer hyperbolischen Kurve.

Abbildung 1.1: Position einer Soundquelle bei bekannter  $TDoA$  für a) den dreidimensionalen und b) den zweidimensionalen Fall.

Viele Arbeiten im Bereich der *Robot Audition* beschränken sich auf den zweidimensionalen Fall. Somit kann die Position einer Soundquelle durch eine hyperbolische Kurve beschrieben werden. Falls angenommen wird, dass die Soundquelle weit von dem Mikrofonpaar entfernt ist, kann die Kurve durch zwei Richtungen (blau markiert in Abbildung 1.1b) approximiert werden. Ohne zusätzliche Information kann nicht entschieden werden, in welcher der beiden Richtungen die Soundquelle angeordnet ist. Diese geometrische Mehrdeutigkeit wird in der Literatur als *Vorne-Hinten-Mehrdeutigkeit* (engl. *Front-Back Ambiguity*) bezeichnet. Um die geometrischen Mehrdeutigkeiten aufzulösen, schränken einige Arbeiten durch gewählte Randbedingungen den betrachteten Bereich ein. So ist es üblich, dass in einigen Experimenten statt den gesamten Winkelbereich von  $360^\circ$  zu betrachten ausschließlich  $180^\circ$  oder weniger berücksichtigt wird [KND06], [KNO15], [LFL13].

Um Soundquellen ausschließlich auf Basis der  $TDoA$  und der Mikrofonpositionen zu kartieren, müssen mehrere Messungen von unterschiedlichen Positionen durchgeführt werden. Im Anschluss kann die Position der Soundquelle ermittelt werden, indem die jeweiligen Repräsentationen (*Cone-of-Confusion*, hyperbolische Kurve oder Richtungsschätzung) kombiniert werden. In Abbildung 1.2a ist ein zweidimensionales Beispiel dargestellt, bei dem eine einzelne Soundquelle  $s_1$  betrachtet und Messungen an drei unterschiedlichen Positionen durchgeführt werden. Die mögliche Position der Soundquelle  $s_1$  ausgehend von der jeweiligen Messposition ist durch die hyperbolischen Kurven  $k_1$ ,  $k_2$  und  $k_3$  repräsentiert. Da eine einzelne Soundquelle betrachtet wird, können die Repräsentationen der Positionen eindeutig kombiniert werden ( $k_1 \cap k_2 \cap k_3$ ). Bei mehreren Soundquellen muss hingegen eine Datenzuordnung durchgeführt werden.



(a) Durchführung von Messungen an drei Positionen für eine Soundquelle.

(b) Durchführung von Messungen an drei Positionen für zwei Soundquellen.

Abbildung 1.2: Kombination mehrerer Messungen im zweidimensionalen Raum.

- (a) Eine Soundquelle  $s_1$  ist aktiv. An drei Messpositionen wird die *TDoA* ermittelt.  
 (b) Zwei Soundquellen  $s_1$  und  $s_2$  sind aktiv. An jeder Messposition werden die beiden *TDoAs* ermittelt, die sich aus  $s_1$  und  $s_2$  ergeben.  $k_1, k_3$  und  $k_5$  (blau) entsprechen der möglichen Position von  $s_1$ . Dies entspricht der Lösung von (a).  $k_2, k_4$  und  $k_6$  (grün) beschreiben die Position von  $s_2$ .

Um die Datenzuordnung für mehrere Soundquellen zu veranschaulichen, ist in Abbildung 1.2b das vorherige Beispiel auf eine weitere Soundquelle  $s_2$  erweitert. Da zwei Soundquellen aktiv sind, resultieren an jeder Messposition zwei *TDoAs*. Somit ergeben sich insgesamt acht mögliche Zuordnungen für die Position von Soundquelle  $s_1$ :

- 1.)  $(k_1 \cap k_3 \cap k_5)$     2.)  $(k_1 \cap k_4 \cap k_5)$     3.)  $(k_1 \cap k_4 \cap k_6)$     4.)  $(k_1 \cap k_3 \cap k_6)$   
 5.)  $(k_2 \cap k_3 \cap k_5)$     6.)  $(k_2 \cap k_4 \cap k_5)$     7.)  $(k_2 \cap k_4 \cap k_6)$     8.)  $(k_2 \cap k_3 \cap k_6)$ .

Ohne zusätzliche Information kann keine Aussage getroffen werden, welche Datenzuordnung die richtige ist. Im vorherigen Beispiel ist die Kombination 1.) korrekt. Mit zunehmender Anzahl von Messungen  $n_l$  und aktiven Soundquellen  $n_s$  nimmt die Anzahl der Kombinationen  $n_k$  exponentiell zu:

$$n_k = (n_s!)^{n_l} \quad (1.1)$$

So ergeben sich bereits für vier Soundquellen und fünf Messpositionen über 7,9 Millionen Kombinationsmöglichkeiten. Um die Positionen der Soundquellen zu schätzen, muss somit einerseits die Anzahl der Soundquellen bekannt sein und andererseits die richtige Datenzuordnung gefunden werden. Weiterhin werden in realen Experimenten akustische Reflexionen beobachtet. Diese ergeben weitere Messungen, die in der Datenzuordnung berücksichtigt werden müssen.

In diesem Abschnitt wurde bisher angenommen, dass die *TDoA* bekannt ist bzw. fehlerfrei ermittelt werden kann. Innerhalb realer Experimente muss jedoch die *TDoA* aus den Mikrofonsignalen geschätzt werden. Aufgrund von u. a. Hintergrundrauschen und akustischen Reflexionen resultiert

eine fehlerhafte Schätzung der *TDoA*. Als Folge ist die Repräsentation der Soundquellenposition inkorrekt. Somit liegt die wahre Soundposition nicht auf der *Cone-of-Confusion* bzw. der hyperbolischen Kurve und die Richtungsapproximation ist fehlerbehaftet.

Schließlich müssen die Mikrofonpositionen bekannt sein, daher muss die mobile Plattform innerhalb der metrischen Karte lokalisiert werden. Die berechnete Lokalisierung kann für die Position oder Orientierung fehlerbehaftet sein. Nachdem mehrere Positionen durch die mobile Plattform angefahren und die jeweiligen fehlerhaften Repräsentationen der Soundquellenposition in einem Modell (z. B. Triangulation oder probabilistisches Modell) kombiniert werden, resultiert ein fehlerbehaftetes Gesamtergebnis der Position der Soundquelle.

## 1.2.2 Wissenschaftlicher Beitrag und Zielsetzung

Ziel dieser Dissertation ist die Entwicklung eines Verfahrens zur dreidimensionalen Kartierung von mehreren statischen Soundquellen innerhalb einer metrischen Karte durch ein mobiles Mikrofonarray. Es soll untersucht werden, wie geometrische Mehrdeutigkeiten und akustische Reflexionen im Kartierungsprozess berücksichtigt werden können. Im Vergleich zu aktuellen Verfahren der *Robot Audition* sollen a priori keine Annahmen über die Anordnung und Anzahl der Soundquellen getroffen werden. Weiterhin sollen die Positionen der Soundquellen ausschließlich über die Mikrofone erfasst werden.

Um den Ansatz in realen Experimenten zu ermöglichen, soll untersucht werden, wie die Unsicherheit der *TDoA* aus den Mikrofondaten und die Unsicherheit der Mikrofonpositionen in einer metrischen Karte beschrieben werden können. Für diesen Zweck sollen neue Verfahren entwickelt werden, die die *TDoA* und ihre Unsicherheit schätzen. Zusammengefasst ergeben sich folgende Ziele für diese Arbeit:

- ( $Z_1$ ) Analyse bestehender Kartierungsverfahren von Soundquellen und auftretender Probleme beim Einsatz in der mobilen Robotik,
- ( $Z_2$ ) Entwicklung von Ansätzen zur Kartierung von Soundquellen unter Berücksichtigung geometrischer Mehrdeutigkeiten und akustischen Reflexionen,
- ( $Z_3$ ) Entwicklung von Fehlermodellen zur Beschreibung der Unsicherheit der Position von Mikrofonen innerhalb einer metrischen Karte,
- ( $Z_4$ ) Entwicklung neuer Verfahren zur Schätzung der *TDoA* zwischen Mikrofonsignalen und Beschreibung der Unsicherheit, sowie
- ( $Z_5$ ) Experimentelle Evaluierung der entwickelten Ansätze sowohl anhand simulierter Daten als auch anhand realer Messwerte

## 1.3 Randbedingungen

Im Folgenden werden die Randbedingungen definiert, die in dieser Arbeit getroffen werden. Sie beziehen sich auf die Umgebung, die untersuchten Soundquellen und die eingesetzte Hardware.

### Anforderungen an die Soundquellen

Die Positionen der Soundquellen werden als statisch angenommen. Die Gesamtzahl der aktiven Soundquellen ist a priori nicht bekannt. Der untersuchte Frequenzbereich liegt zwischen 20 und 20.000 Hz und ist somit auch vom Menschen hörbar. Das ausgesendeten Signale und deren assoziierte Lautstärke sind nicht bekannt. Um Soundquellen zu berücksichtigen, die für kurze Zeiten deaktiviert sind - z. B. Sprachpausen in menschlicher Kommunikation - wird eine ausreichend lange Aufnahmezeit gewählt. Eine Analyse der Mikrofondaten erfolgt für zeitliche Abschnitte innerhalb dieses Zeitfensters.

### Anforderungen an die Umgebung

Die Experimente werden in einem geschlossenen Raum durchgeführt. Die Umgebung wird als statisch angesehen. Das Hintergrundrauschen der untersuchten Umgebung wird im Vergleich zu den emittierten Soundquellen als gering angenommen.

### Anforderungen an die verwendete Hardware

Eine mobile Plattform wird mit unterschiedlichen Sensoren - Mikrofone und Laserscanner - ausgestattet und führt die Soundquellenkartierung durch (vgl. Abbildung 1.3). Es wird angenommen, dass sich die relative Lage der Sensoren nicht verändert. Die akustischen Signale werden durch Mikrofone wahrgenommen. Für diese Arbeit werden handelsübliche Mikrofone verwendet. Hierbei wird der Schwerpunkt auf ECM-Mikrofone gelegt (vgl. Abschnitt 2.1.4) und keine Kalibrierung bezüglich der Amplitude vorausgesetzt. Um die fehlerhafte Detektion von Eigengeräuschen zu minimieren, werden die akustischen Messungen nur im Stillstand durchgeführt. Aus den Mikrofondaten wird für jedes Mikrofonpaar die TDoA ermittelt. Um aus der Gesamtheit aller TDoAs die Soundquellenposition zu schätzen, müssen die betrachteten Mikrofondaten zeitlich zueinander synchronisiert sein. In dieser Arbeit wird angenommen, dass die verwendete Hardware - die Soundkarte - diese Anforderung erfüllt. Im Weiteren wird das Ziel formuliert, dass die Soundquellen ausschließlich durch Mikrofone erfasst und durch keine weitere Sensormodalität wahrgenommen werden. Somit erfolgt die Schätzung der Positionen der Soundquellen ausschließlich auf Basis der Mikrofondaten. Zusätzlich dazu werden die Sensordaten von einem Laserscanner ausschließlich für die Lokalisierung der mobilen Plattform innerhalb einer metrischen Karte eingesetzt.

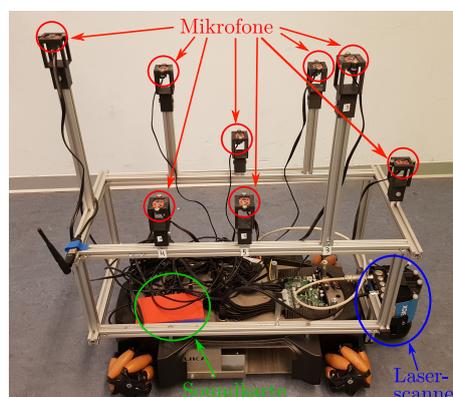


Abbildung 1.3: Experimental-Plattform: Eine KUKA youBot-Plattform ist mit einem Mikrofonarray ausgerüstet.

## 1.4 Lösungsansatz

In dieser Arbeit wird ein neues Verfahren zur Positionsschätzung von statischen Soundquellen in einer bekannten metrischen Karte vorgestellt. Der entwickelte Lösungsansatz besteht aus den Komponenten (1), (2) und (3):

- (1) Der Ermittlung der zeitlichen Differenz ( $TD\text{oA}$ ) und ihrer Unsicherheit aus den Mikrofonsignalen,
- (2) der Lokalisierung der Mikrofonposition und Beschreibung der Unsicherheit in einer metrischen Karte, und schließlich
- (3) der intervall-basierten Soundquellenkartierung

Die Komponenten des gewählten Lösungsansatzes sind in Abbildung 1.4 visualisiert.

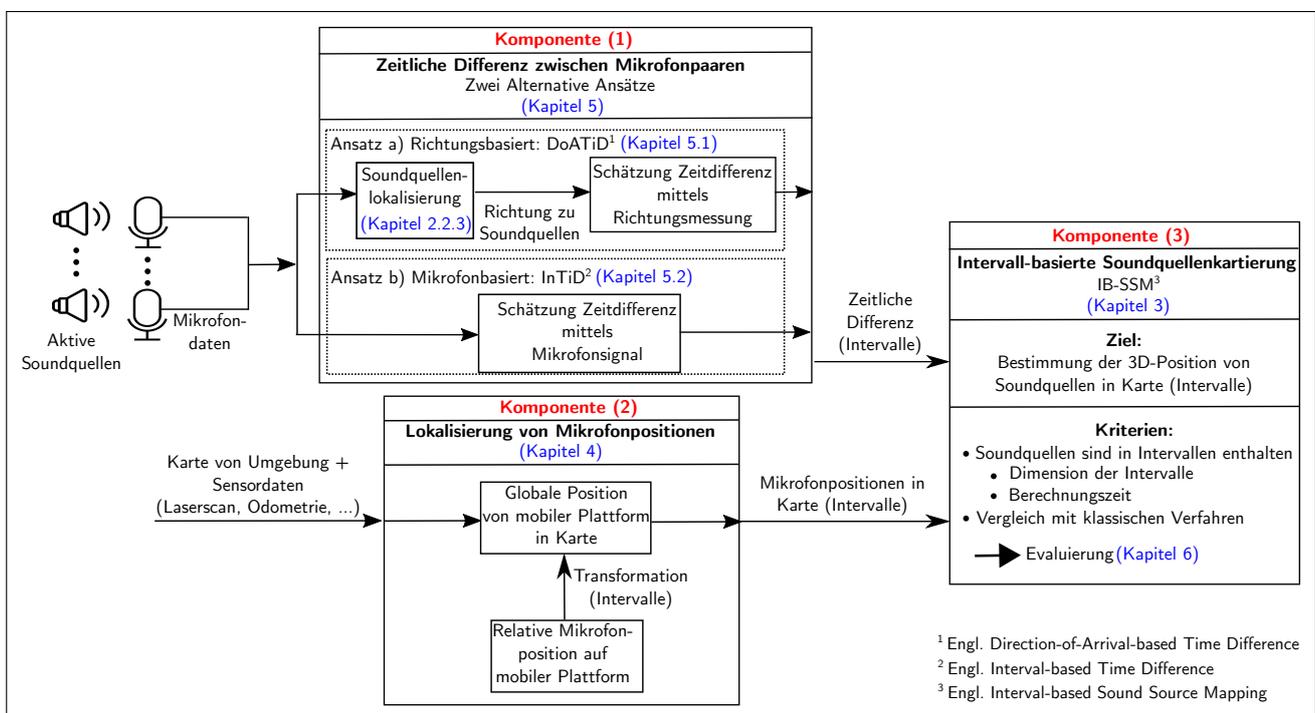


Abbildung 1.4: Systematischer Aufbau der entwickelten Komponenten und des Lösungsansatzes.

In Komponente (3) wird ein geometrischer Zusammenhang zwischen der  $TD\text{oA}$ , der Mikrofon- und der gesuchten Soundquellenposition formuliert. Diese Beschreibung ist für ein Mikrofonpaar und eine einzelne Quelle gegeben und berücksichtigt die *Cone-of-Confusion* (vgl. Abbildung 1.1a). Dieser Zusammenhang wird auf mehrere Mikrofonpaare, unterschiedliche Messpositionen und mehrere gleichzeitig aktiven Soundquellen erweitert. Die resultierenden Gleichungen und deren Abhängigkeiten werden in einem *Constraint-Satisfaction-Problem (CSP)* formuliert und die Lösung durch einen *Branch-and-Bound-Algorithmus* ermittelt. In dieser Arbeit wird ein Algorithmus gewählt, der auf der Intervall-Arithmetik basiert. Dabei wird zunächst der Lösungsbereich durch eine diskrete Menge von dreidimensionalen Intervallen (Intervall-Boxen) repräsentiert. Für jede Intervall-Box wird auf Basis der Restriktionen des *CSPs* überprüft, ob Teile der Lösung enthalten sind. Andernfalls werden inkonsistente Boxen der Lösungsmenge entfernt. Weiterhin werden inkonsistente Lösungsbereiche für Soundquellen verworfen, die aufgrund von akustischen Reflexionen aus weiteren Richtungen wahr-

genommen werden. Somit werden in dem Lösungsansatz akustische Reflexionen berücksichtigt. Die Genauigkeit der resultierenden Lösung der Soundquellenkartierung ist abhängig von der Unterteilung der Lösungsmenge und somit der Dimension der Intervall-Boxen.

In realen Experimenten sind die Messwerte fehlerbehaftet. Daher resultiert kein konsistentes Ergebnis nach Lösung des *CSPs*. Aus diesem Grund wird in dieser Dissertation ein Ansatz entwickelt, der für die Unsicherheit der Mikrofonpositionen und der *TDoA* unbekannte, jedoch begrenzte Fehler berücksichtigt. Dabei werden die Unsicherheiten durch Intervalle modelliert. Im Anschluss wird das *CSP* in Komponente (3) mit den gewählten Unsicherheiten gelöst und somit ermöglicht reale Messwerte zu verwenden.

In Komponente (1) werden auf Basis der Mikrofondaten die *TDoAs* und deren Unsicherheit berechnet. In dieser Dissertation werden hierfür zwei alternative Ansätze *DoATiD*<sup>1</sup> und *InTiD*<sup>2</sup> formuliert. *DoATiD* basiert auf den Richtungsschätzungen von Soundquellen, die durch klassische Verfahren der Soundquellenlokalisierung berechnet werden. *InTiD* ermittelt direkt aus den Daten zweier Mikrofone die Intervalle der *TDoA*. Für die Berechnung werden sogenannte Kontraktoren aus der Intervall-Arithmetik verwendet (vgl. Abschnitt 2.4.6).

In Komponente (2) wird die dreidimensionale Position der Mikrofone innerhalb einer bekannten metrischen Karte ermittelt. Zunächst wird die dreidimensionale Position der jeweiligen Mikrofone in Bezug zu dem mobilen Mikrofonarray angegeben. Hierbei wird sowohl die Dimension der Mikrofonmembran als auch die Genauigkeit der Messmethode berücksichtigt. Insgesamt wird die dreidimensionale Position der Mikrofone durch Intervallboxen repräsentiert. Im Anschluss wird das mobile Mikrofonarray innerhalb einer metrischen Karte lokalisiert. Hierbei werden klassische Lokalisierungsverfahren auf Basis von Rasterkarten verwendet.

## 1.5 Gliederung der Arbeit

Die vorliegende Arbeit gliedert sich wie folgt:

Die Grundlagen im Bereich der akustischen Signalverarbeitung werden in **Kapitel 2** zusammengefasst und auf den Stand der Wissenschaft im Bereich der *Robot Audition* eingegangen. Im Anschluss wird die Intervall-Arithmetik eingeführt, die die Grundlagen für das neue Kartierungsverfahren bildet.

In **Kapitel 3** wird ein neues intervall-basiertes Kartierungsverfahren (*IB-SSM*<sup>3</sup>) vorgestellt, um die dreidimensionale Position von mehreren statischen Soundquellen zu ermitteln. Zunächst wird ein *Constraint-Satisfaction-Problem* (*CSP*) formuliert, um den Zusammenhang zwischen den Mikrofonpositionen, der *TDoA* und der Position der Soundquelle aufzustellen. Das *CSP* wird mittels intervall-basierter Verfahren gelöst. Dabei werden akustische Reflexionen und Ausreißer berücksichtigt. Die Kenntnis der Mikrofonpositionen in der metrischen Karte, sowie die Messwerte der *TDoA* sind in realen Experimenten fehlerbehaftet. Daher wird ein Fehlermodell für die beiden Variablen in **Kapitel 4** und **Kapitel 5** formuliert, das unbekannte aber begrenzte Fehler annimmt.

Die Lokalisierung der Positionen der Mikrofone innerhalb einer metrischen Karte und die Beschreibung möglicher Fehlerarten wird in **Kapitel 4** untersucht.

---

<sup>1</sup> Engl. *Direction-of-Arrival-based Time Difference*

<sup>2</sup> Engl. *Interval-based Time Difference*

<sup>3</sup> Engl. *Interval-based Sound Source Mapping*

In **Kapitel 5** wird die *TDoA* und ihrer Unsicherheit geschätzt. Hierbei werden die alternativen Ansätze *DoATiD* und *InTiD* vorgestellt. *DoATiD* basiert auf Richtungsschätzungen (*DoA*<sup>4</sup>), die durch Verfahren der Soundquellenlokalisierung berechnet werden. Mithilfe der Richtungsschätzungen, einem Filteransatz und geometrischen Restriktionen wird die *TDoA* berechnet. Das alternative Verfahren *InTiD* führt im Gegensatz dazu direkt intervall-basierte Berechnungen auf den Daten der Mikrofon-signale durch. Es werden verschiedene Ansätze vorgestellt, um die zeitlichen Intervalle zu berechnen. Nachdem im ersten Schritt für *InTiD* die akustische Dämpfung vernachlässigt wird, erfolgt im Anschluss eine Erweiterung auf diesen Effekt.

Die vorgestellten Verfahren aus **Kapitel 3** bis **Kapitel 5** werden in **Kapitel 6** zunächst in der Simulation sowie nachfolgend in der realen Umgebung evaluiert und mit klassischen Verfahren verglichen.

In **Kapitel 7** werden die Ergebnisse bewertet und ein Ausblick auf offene Forschungsfragen gegeben.

Mit einer Zusammenfassung schließt diese Arbeit in **Kapitel 8**.

---

<sup>4</sup> Engl. *Direction of Arrival*



# Kapitel 2

## Stand der Wissenschaft

In diesem Kapitel wird ein Überblick über den aktuellen Stand der Wissenschaft gegeben, der die Grundlagen für die vorliegende Arbeit darstellt. Zunächst werden die Grundlagen im Bereich der Akustik und der Datenverarbeitung behandelt. Im Anschluss wird die Entwicklung der akustischen Signalverarbeitung in der mobilen Robotik vorgestellt. Hierfür wird zunächst auf das Teilgebiet der Soundquellenlokalisierung eingegangen und bekannte Verfahren beschrieben. Mithilfe der Ergebnisse der Lokalisierung kann eine Kartierung der Soundquellen durchgeführt werden. Daher werden darauffolgend bekannte Verfahren in diesem Bereich vorgestellt. Die Grundlage für den in dieser Arbeit vorgestellten Ansatz bildet die Intervallarithmetik, die darauffolgend eingeführt wird. Zum Abschluss dieses Kapitels wird der Bezug der vorliegenden Arbeit mit dem präsentierten Stand der Wissenschaft hergestellt und erläutert.

### 2.1 Physikalische Grundlagen der Akustik und Datenverarbeitung

Im Folgenden werden die wesentlichen Begriffe im Bereich der Akustik definiert. Zunächst wird die Ausbreitung von akustischen Wellen mathematisch beschrieben. Darauffolgend wird auf akustische Dämpfung und Reflexionen eingegangen. Um die Kenngrößen von akustischen Wellen zu messen, sind Mikrofone notwendig. Daher wird im Anschluss die Funktionsweise und Einteilung von Mikrofonen, sowie die digitale Repräsentation von akustischen Signalen beschrieben. Zum Abschluss dieses Abschnittes wird auf die Grundlagen der Frequenzanalyse digitaler Signale eingegangen, um eine Basis für die akustische Merkmalsextraktion zu schaffen.

#### 2.1.1 Akustik

Die Akustik (*griechisch akouein*: hören) bezeichnet die Lehre der Ausbreitung mechanischer Wellen in unterschiedlichen Medien wie Luft, Wasser und Festkörpern [Lin73]. Daraus ergeben sich eine Vielzahl von Teildisziplinen, u. a.: Physikalische Akustik, Architektonische Akustik, Unterwasser-Akustik, Physiologische und Psychologische Akustik, Sprache und Musikalische Akustik [Dun+15]. Für die vorliegende Dissertation ist die physikalische Akustik relevant, da in dieser Teildisziplin mathematisch modelliert wird, wie sich akustische Wellen in der Umgebung ausbreiten. Im Folgenden werden wesentliche Begriffe in diesem Bereich eingeführt.

Die Basis bildet ein Körper, der durch Vibration Luftmoleküle in Bewegung versetzt. Durch diesen Vorgang wird eine mechanische Welle erzeugt. In dieser Dissertation werden die Begriffe akustische Welle, Schall und Schallwelle synonym verwendet. Eine akustische Welle breitet sich mit der Schallgeschwindigkeit  $c$  aus. Dabei ist die Schallgeschwindigkeit abhängig vom Medium und der Umgebungstemperatur. So breiten sich akustische Wellen in Festkörpern aufgrund der höheren Materialdichte schneller aus als in Fluiden. Eine Übersicht ist in Tabelle 2.1 gegeben.

Tabelle 2.1: Schallgeschwindigkeit in unterschiedlichen Medien in Längsrichtung. Es wird eine Temperatur von 20 °C angenommen. Die Werte sind [Lid04, Abschnitt 14] entnommen.

Medium	Luft	Wasser	Aluminium
<b>Schallgeschwindigkeit [m/s]</b>	343,4	1483,2	6420

Für die vorliegende Arbeit ist relevant, wie sich akustische Wellen in Luft ausbreiten. In diesem Medium ist die Schallgeschwindigkeit  $c$  neben der Temperatur  $\vartheta$  (in °C) auch von der relativen Luftfeuchtigkeit abhängig [Boh87]. Unter der Annahme einer relativen Luftfeuchtigkeit von 0% kann dieser Zusammenhang durch folgende Gleichung approximiert werden:

$$c = 20,05 \cdot \sqrt{\vartheta + 273,15} \quad m/s. \quad (2.1)$$

Somit ergibt sich für eine Temperatursteigerung von 0 °C auf 30 °C eine Zunahme der Geschwindigkeit von 331,37  $\frac{m}{s}$  auf 349,09  $\frac{m}{s}$ , dies entspricht 17,39  $\frac{m}{s}$ . Es kann ein Modell für die Ausbreitung der akustischen Wellen hergeleitet werden, indem die physikalischen Erhaltungsgleichungen (Massenerhaltung, Impulserhaltung, Energieerhaltung) und die Materialgleichungen für das jeweilige Medium verwendet werden. Hierdurch resultieren Differenzialgleichungen, die unter den jeweiligen Rand- und Anfangsbedingungen gelöst werden müssen. Als Ergebnis kann der Schalldruck  $p$  berechnet werden.

Dabei beschreibt  $p$  die lokale Druckänderung der Luft. Diese führt dazu, dass sich Luftmoleküle in Bewegung versetzen. Neben dem Schalldruck wirkt aufgrund der Gravitation der Luft ein atmosphärischer Druck. Dieser beträgt auf Meereshöhe im Mittel ca. 1 bar und ist im Vergleich zum Schalldruck wesentlich größer. Der Mensch kann durch seine Ohren Änderungen im Schalldruck erfassen, jedoch nicht den atmosphärischen Druck.

Für den Schalldruck ergibt sich ein großer Wertebereich, der vom Menschen wahrgenommen werden kann. Statt einer linearen Repräsentation wird daher eine logarithmische Darstellung gewählt. Diese wird als Pegel in Dezibel (dB) angegeben [Dun+15, Kapitel 6]:

$$L_p = 20 \log_{10} \left( \frac{p}{p_r} \right) dB. \quad (2.2)$$

$L_p$  wird als Schalldruckpegel (SPL, engl. *Sound Pressure Level*) bezeichnet. Hierbei bezeichnet  $p_r$  einer Referenzgröße mit zugehörigem Schalldruck von 20  $\mu\text{Pa}$  (unter der Annahme einer Frequenz von 1 kHz). Diese Referenzgröße ist mit dem kleinsten Schalldruck assoziiert, den der Mensch wahrnehmen kann [Ber19]. Um zu kennzeichnen, dass die dB-Angabe auf diese Referenz bezogen ist, wird die Bezeichnung „dB re 20  $\mu\text{Pa}$ “ oder „dB SPL“ verwendet. Somit ergibt sich, dass der Mensch einen Schallpegel ab 0 dB SPL erfassen kann. Die obere Schranke für den Schallpegel, den der Mensch wahrnehmen kann, ist durch die Schmerzgrenze beschränkt. Diese liegt in Abhängigkeit von der Frequenz und Dauer bei etwa 120 bis 130 dB SPL. Demgegenüber steht der maximal

annehmbare Schalldruckpegel. Dieser ist durch den atmosphärischen Druck nach oben beschränkt und beträgt 194 dB SPL. Um den Wertebereich des Schalldruckpegels zu veranschaulichen, sind einige Beispiele in Tabelle 2.2 angegeben.

Tabelle 2.2: Schalldruckpegel für ausgewählte Geräusche. Die Werte sind [WWS20, Kapitel 3] entnommen.

Geräusch	Schalldruckpegel in dB SPL
Hörschwelle	0
Leises Blätterrauschen	15 - 20
Ruhige Wohnlage	30 - 40
Unterhaltung in normaler Lautstärke	50 - 60
Starker Straßenverkehr	70 - 80
Presslufthammer in 10 m Entfernung	90 - 100
Propellerflugzeug in 3 m Entfernung	120 - 130

Neben dem Bezug auf den Referenzschalldruck werden Messgrößen (wie der Schalldruck oder elektrische Spannungen) in einer logarithmischen Relation zueinander gesetzt. So führt die Verdoppelung (oder Vervierfachung) des Schalldrucks zu einer Erhöhung um gerundet 6 dB (oder 12 dB). Diese Steigerung ist jedoch nicht mit einer Verdoppelung der empfundenen Lautstärke des Menschen assoziiert. Die Lautstärke ist dabei stark vom Individuum und der Frequenz abhängig. Dennoch gilt die grobe Faustregel, dass die Erhöhung der Lautstärke um 10 dB einer wahrgenommenen Verdoppelung der Lautstärke entspricht.

Um den Schalldruckpegel anzugeben, müssen die vorher erwähnten Differenzialgleichungen gelöst werden. Es kann eine Linearisierung der Gleichungen durchgeführt werden, falls der Schalldruck  $p$  als klein gegenüber dem atmosphärischen Druck angenommen wird. Die Linearisierung führt bei fluiden Medien zur nachfolgenden dreidimensionalen Wellengleichung [Dun+15, Kapitel 3.3.3]:

$$\Delta p(\mathbf{x}, t) = \frac{1}{c^2} \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2}. \quad (2.3)$$

Dabei wurde angenommen, dass die Dichte des Fluids unabhängig von der Position ist. Im Gegensatz dazu besitzt der Schalldruck  $p$  eine örtliche und räumliche Abhängigkeit, repräsentiert durch den Ort  $\mathbf{x}$  und der Zeit  $t$ .  $\Delta$  bezeichnet den Laplace-Operator.

Falls die akustischen Größen von der Zeit und nur einer einzigen Raumdimension abhängen, kann der Schalldruck durch eine ebene Welle beschrieben werden. Somit kann die Lösung von Gleichung 2.3 als Sinus-Funktion mit einer konstanten Frequenz dargestellt werden:

$$p(x, t) = a \sin[k(x - ct) + \Phi_0]. \quad (2.4)$$

Hierbei entspricht  $a$  der (Schalldruck-)Amplitude und somit dem betragsmäßigen Spitzenwert des lokalen Schalldrucks.  $\Phi_0$  bezeichnet die Phasenverschiebung und  $k$  der Wellenzahl. Der Zusammenhang zwischen der Wellenlänge  $\lambda$  und der Wellenzahl  $k$  ist folgendermaßen gegeben:

$$k = \frac{2\pi}{\lambda}. \quad (2.5)$$

Weiterhin ist die Wellenlänge  $\lambda$  mit der Frequenz  $f$  durch folgende Beziehung verknüpft:

$$\lambda = \frac{c}{f}. \quad (2.6)$$

In der Realität auftretende akustische Signale bestehen aus mehreren unterschiedlichen Frequenzanteilen. Aufgrund der angenommenen Linearität in 2.3 können die jeweiligen Lösungen summiert werden. Exemplarisch sind in nachfolgender Gleichung  $n$  Frequenzanteile dargestellt:

$$p(x, t) = \sum_{i=1}^n a_i \sin[k_i(x - ct) + \Phi_{0i}]. \quad (2.7)$$

Der Schalldruck  $p$  kann durch ein Mikrofon an einer statischen Position gemessen werden. Daher verschwindet die räumliche Abhängigkeit in Gleichung 2.7. Das resultierende Signal lässt sich folgendermaßen darstellen:

$$p(t) = \sum_{i=1}^N a_i \sin[\omega_i t + \Phi_{0i}]. \quad (2.8)$$

Hierbei bezeichnet  $\omega$  die Kreisfrequenz:

$$\omega = kc = 2\pi f. \quad (2.9)$$

Um Gleichung 2.4 zu veranschaulichen, wird ein sinusförmiges Signal  $s(t)$  mit mehreren Frequenzanteilen betrachtet und in Abbildung 2.1 visualisiert. Die drei Frequenzanteile  $f_1$ ,  $f_2$  und  $f_3$  betragen  $f_1 = 1$  Hz,  $f_2 = 2$  Hz und  $f_3 = 4$  Hz und somit kann  $s(t)$  folgendermaßen angegeben werden:

$$s(t) = \sin(2\pi f_1 t) + 2 \cdot \sin(2\pi f_2 t) + 3 \cdot \sin(2\pi f_3 t). \quad (2.10)$$

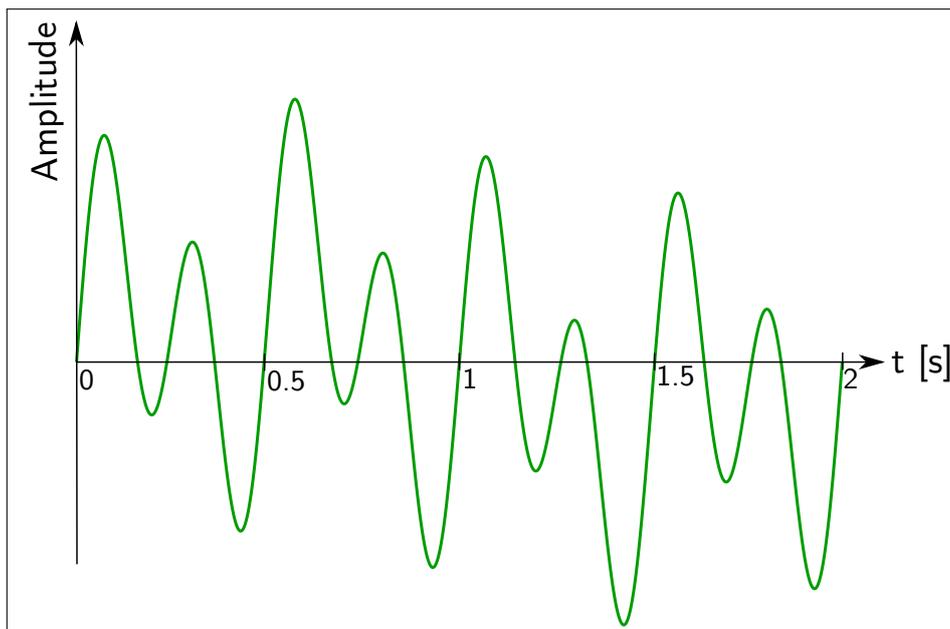


Abbildung 2.1: Exemplarisches Audiosignal  $s(t)$  mit drei Frequenzkomponenten.

## 2.1.2 Dämpfung

Im freien Raumfeld verringert sich der Schalldruck durch zwei Effekte. Der erste Effekt ist unabhängig von der Frequenz der Schallquelle. Dabei nimmt der Schalldruck antiproportional des Ausbreitungsweges  $r$  ab. Wenn die Entfernung verdoppelt wird, resultiert daher eine Abnahme um 6 dB SPL.

Der zweite Effekt ist abhängig von der Luftdämpfung und wird als Dissipation bezeichnet. Hierbei stoßen Luftmoleküle zusammen und ein Teil der Energie wird in Wärme überführt [Bas+95]. Ausgehend von einem initialen Schalldruck  $p(r)$  erfolgt eine Verringerung in Abhängigkeit der Distanz  $r$  zur Schallquelle. Unter der Annahme einer planaren Welle ergibt sich folgender Zusammenhang:

$$p(r + \Delta r) = p(r)e^{-\alpha\Delta r/2}. \quad (2.11)$$

Dabei entspricht  $r$  und  $r + \Delta r$  der Distanz zu einer Schallquelle zweier Bezugsorte, sowie  $p(r)$  und  $p(r + \Delta r)$  dem zugehörigen Schalldruck.

Im Gegensatz zum ersten Effekt ist die Dissipation abhängig von der Frequenz der akustischen Welle, der Lufttemperatur und der relativen Luft-Feuchtigkeit. Dieser Zusammenhang wird in Gleichung (2.11) durch den Dämpfungsfaktor  $\alpha$  beschrieben. In [Bla01] wird vorgestellt, wie  $\alpha$  berechnet werden kann. Im Folgenden ist in Tabelle 2.3 ein Beispiel der Abnahme der Dissipation dargestellt.

Tabelle 2.3: Abnahme des Schalldrucks durch Dissipation in Abhängigkeit der Frequenz. Es wird eine Lufttemperatur von 20 °C angenommen. Die Werte sind entnommen aus [Lid04, Abschnitt 14].

Frequenz [Hz]	100	100	4000	4000	10.000	10.000
Abnahme Schalldruck [dB SPL/km]	1,67	0,5	8,84	38,89	26,28	204,98
Luftfeuchtigkeit [%]	0	30	0	30	0	30

## 2.1.3 Reflexionen

In einem geschlossenen Raum werden akustische Wellen an Wänden oder Hindernissen absorbiert, reflektiert oder gestreut. Dabei kann ein einzelnes Signal aufgrund der Interaktion mit der Umgebung über mehrere Pfade an einem Ort  $x$  beobachtet werden. Als exemplarisches Beispiel ist in Abbildung 2.2 eine Soundquelle gegeben, die durch ein Mikrofon wahrgenommen wird.

Zunächst wird das akustische Signal einer Soundquelle über den direkten Weg zwischen Mikrofon und Soundquelle beobachtet. Dies wird als *Direktschall* (engl. *direct sound*) bezeichnet. Im Anschluss erreichen Reflexionen des Signals das Mikrofon mit einem zeitlichen Versatz. Die frühesten Reflexionen werden als *Erstreflexionen* (engl. *early reflections*) bezeichnet. Mithilfe dieser Erstreflexionen erlangt der Mensch einen Eindruck über die räumlichen Eigenschaften eines Raumes. Aufgrund der großen Vielzahl von möglichen Reflexionswegen erreichen kontinuierlich weitere Reflexionen das Mikrofon. Diese Erscheinung wird als Nachhall (engl. *reverberation*) bezeichnet.

Mathematisch kann der Zusammenhang des wahrgenommenen Signals am Mikrofon mithilfe einer Impulsfunktion hergestellt werden (vgl. Abbildung 2.2). Dabei ist die Impulsfunktion abhängig von den Raumeigenschaften. Falls die Impulsfunktion bekannt ist, kann für eine beliebige Position im Raum das wahrgenommene akustische Signal berechnet werden.

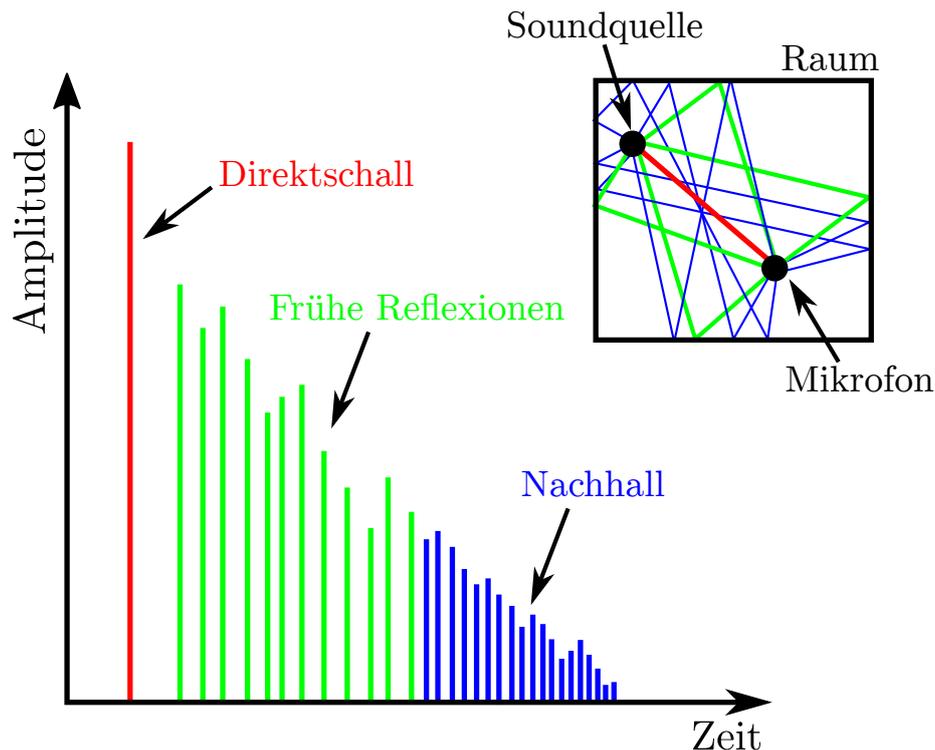


Abbildung 2.2: Gegeben ist die folgende exemplarische Impulsfunktion eines Raumes. Dargestellt ist der Direktschall (rot), die frühen Reflexionen (grün) und der Nachhall (blau).

Impulsfunktionen können mathematisch approximiert oder durch Realmessungen gewonnen werden. Eine vereinfachte Approximation für einen kleinen rechteckigen Raum kann mithilfe der Image Source-Methode ermittelt werden [AB79].

Die Nachhallzeit (engl. *Reverberation Time*) ist eine wichtige Kenngröße, um die Raumeigenschaft zu beschreiben. Sie kennzeichnet die Zeitdauer bis der Schalldruck in einem Raum um 60 dB verringert ist, nachdem die Schallquelle ausgeschaltet wurde [Mös05, Kapitel 7]. Die mittlere Nachhallzeit beträgt in einem durchschnittlichen Wohnzimmer 0,6 s und in einer Kirche 5 s [SF14]. Im Gegensatz dazu besitzt ein echofreier Raum eine Nachhallzeit von näherungsweise 0 s.

#### 2.1.4 Mikrofone und diskrete Audiodaten

Mikrofone werden eingesetzt, um akustische Signale technisch wahrzunehmen. Dabei ist das Einsatzgebiet entscheidend dafür, welche Mikrofonart verwendet wird. Es erfolgt eine Unterteilung in folgende Kategorien:

- Dynamische Mikrofone
  - Tauchspulmikrofone
  - Bändchenmikrofone
- Kondensatormikrofone
  - Kleinmembran-Mikrofone
  - Großmembran-Mikrofone
  - Elektret-Kondensatormikrofon (ECM)
- Mikroelektromechanische System-Mikrofone (MEMS-Mikrofone)



Abbildung 2.3: Unterschiedliche Mikrofonarten. a) Beyerdynamic TG V70 [Bey21] b) Coles 4038 Studio Ribbon [Tho21] c) Schoeps Colette CMC-64 [Sch21] d) Neumann U 87 Ai Studio Set [Neu21] e) CUI Devices CMC-4012-40P [CUI21] f) Infineon IM67D130A [Inf21]

Dynamischen Mikrofone funktionieren nach dem Prinzip der elektrischen Induktion. Sie sind unterteilt in Tauchspulmikrofone und Bändchenmikrofone. Bei Tauchspulmikrofonen ist eine Spule an der Rückseite der Mikrofonmembran montiert und von einem Magneten umgeben. Durch eintreffende Schallwellen wird die Membran mitsamt der Spule in Bewegung versetzt. Folglich wird ein Strom induziert. Im Gegensatz dazu verwenden Bändchenmikrofone anstelle einer Spule ein metallisches Bändchen. Dieses ist zwischen den Polen eines Magneten gespannt. In Abhängigkeit der eintreffenden Schallwellen resultiert eine Auslenkung des Bändchens.

Aufgrund der Konstruktion sind Tauchspulmikrofone robust gegenüber Erschütterungen. Daher eignen sie sich für die Bühne. Bändchenmikrofone sind hingegen empfindlich gegenüber kleinsten Luftbewegungen und sind für die Bühne nicht geeignet. Bändchenmikrofone werden vor allem bei der Aufnahme von E-Gitarren, Blechbläsern und Schlagzeugen eingesetzt.

Die zweite Kategorie bilden die Kondensatormikrofone. Bei ihnen ist eine elektrische Membran in einem geringen Abstand zu einer Metallplatte (Gegenelektrode) angeordnet. Dies entspricht dem Aufbau eines Kondensators. Durch eintreffende Schallwellen wird die Membran in Schwingung versetzt. Hierdurch variiert der Abstand zur Platte. Als Folgerung verändert sich die Kapazität des Kondensators.

sators. Somit kann der zeitliche Verlauf der elektrischen Spannung gemessen werden. Anhand des Membran-Durchmessers erfolgt eine Unterteilung in Kleinmembran- und Großmembran-Mikrofone. Im Vergleich zu den dynamischen Mikrofonen sind Kondensatormikrofone empfindlicher und weisen eine bessere Klangqualität auf. Jedoch benötigen sie eine externe Stromversorgung (P48 Phantomspannung) und neigen zu Rückkopplungen.

Eine weitere Kategorie der Kondensatormikrofone sind Elektret-Kondensatormikrofon (ECM: engl. *electret condenser microphones*). Statt die Ladung des Kondensators durch eine externe Stromversorgung aufzubringen, ist die statische Ladung quasi-permanent gespeichert. Hierbei wird ein Elektret verwendet, das als Film auf der Gegenelektrode angeordnet ist. Als Vorteil können Elektret-Kondensatormikrofone mit geringen Spannungen betrieben werden. Aufgrund ihrer geringen Baugröße werden sie in Mobiltelefonen oder Laptops verwendet. So wird beispielsweise bei der PlayStation-Eye-Kamera mehrere ECM verwendet (vgl. Abbildung 2.10a). Im Vergleich zu den klassischen Kondensatormikrofonen sind ECM kostengünstig, jedoch weisen sie eine schlechtere Klangqualität auf.

Die letzte Kategorie bilden die Mikroelektromechanischen System-Mikrofone (MEMS: engl. *microelectromechanical systems*). Sie bestehen aus einem mechanischen, und somit beweglichen Teil, und einen elektronischen Teil. Die Herstellung von MEMS-Mikrofonen erfolgt wie bei der Produktion von Halbleiter-Chips. In den letzten Jahren verdrängen MEMS-Mikrofone die ECM in vielen Bereichen. Der Grund besteht in der platzsparenden Bauform der MEMS-Mikrofone und der besseren Klangqualität. Daher werden sie heutzutage vermehrt in aktuellen Mobiltelefonen und Laptops integriert. Darüber hinaus hat der Hersteller Infineon das erste zertifizierte MEMS-Mikrofon für Automobilanwendungen entwickelt. Es ist sowohl für den Innenbereich, als auch für den Außenbereich geeignet, und kann bei einer Temperatur von  $-40$  bis  $105$  °C eingesetzt werden [Inf21].

## Richtcharakteristik

Eine wichtige Eigenschaft von Mikrofonen ist ihre Richtcharakteristik. Sie repräsentiert, wie sensitiv ein Mikrofon für die Erfassung des Schalldrucks aus unterschiedlichen Richtungen ist. Diese Eigenschaft wird in Form von Polardiagrammen dargestellt. Dabei werden die Schalldruckmessungen auf den Schalldruck bei  $0^\circ$  bezogen und als Pegel in dB angegeben. Da die Sensitivität der Mikrofone frequenzabhängig ist, wird die Richtcharakteristik für verschiedene Frequenzen angegeben. Eine Auswahl von unterschiedlichen Richtcharakteristiken ist in Abbildung 2.4 dargestellt und wird im Folgenden vorgestellt.

Die Kugelcharakteristik (Omni-direktional) besitzt in alle Richtungen die gleiche Empfindlichkeit. Insbesondere ECM und MEMS-Mikrofone weisen diese Charakteristik auf. Die Nierencharakteristik (engl. *cardioid*) ist am empfindlichsten für einfallende Schallwellen aus Richtungen von Vorne ( $0^\circ$ ). Im Gegensatz dazu unterdrückt sie die Richtungen von der Rückseite ( $180^\circ$ ). Abgeleitete Charakteristiken bilden die Supernierencharakteristik (engl. *supercardioid*) und Hypernierencharakteristik (engl. *hypercardioid*). Die Supernierencharakteristik besitzt eine höhere Empfindlichkeit als die Nierencharakteristik für Schallwellen, die von der Rückseite einfallen. Hierbei liegt die minimale Empfindlichkeit bei etwa  $150^\circ$  und  $210^\circ$ . Die Hypernierencharakteristik besitzt im Vergleich zur Supernierencharakteristik eine stärker Empfindlichkeit für Schallwellen, die von der Rückseite wahrgenommen werden.

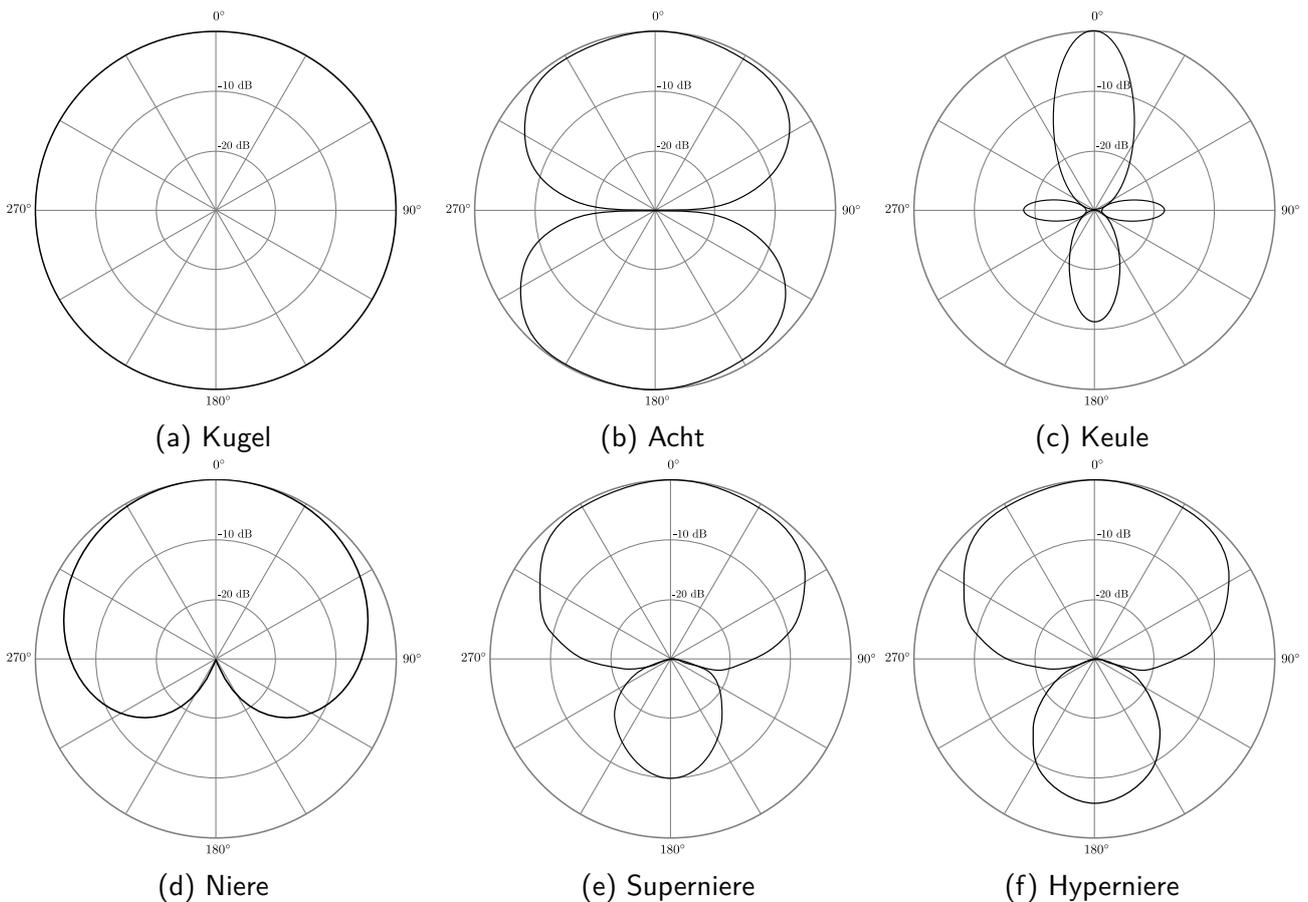


Abbildung 2.4: Exemplarische Darstellung der Richtcharakteristik von Mikrofonen.

Eine weitere Richtcharakteristik ist die Achtcharakteristik. Sie besitzt eine hohe Empfindlichkeit für Schallwellen von sowohl der Vorder- als auch von der Rückseite. Im Gegensatz dazu besitzt sie eine geringe Empfindlichkeit für Schallwellen von der Seite ( $90^\circ$  und  $270^\circ$ ).

Die letzte hier behandelte Richtcharakteristik bildet die Keulencharakteristik (engl. *lobar*). Sie besitzt eine hohe Bündelfähigkeit in gezielten Richtungen. Dabei wird die höchste Empfindlichkeit bei  $0^\circ$  angenommen, gefolgt von  $180^\circ$ . Weiterhin sind die Seiten ( $90^\circ$  und  $270^\circ$ ) empfindlich, jedoch im Vergleich zur Vorder- und Rückseite stark gedämpft.

### Technische Eigenschaften

Im Folgenden wird auf eine Auswahl von technischen Eigenschaften von Mikrofonen eingegangen. Die Vorgehensweise der Messung, als auch die Spezifikation ist nach der IEC 60268-4 (Sound System Equipment - Part 4: Microphones) definiert.

Die Empfindlichkeit eines Mikrofons wird ermittelt, indem eine Schallquelle vor einem Mikrofon platziert wird. Dabei beträgt der Schalldruck der Schallquelle 1 Pa (respektive 140 dB SPL) und es wird eine festgelegte Frequenz (z. B. 1 kHz) verwendet. Im Anschluss wird die resultierende Spannung am Ausgang des Mikrofons bestimmt. Eine hohe Spannung impliziert dabei eine hohe Empfindlichkeit. Das Ergebnis der Messung wird in den Einheiten V/Pa oder dBV angegeben. Dabei ist der Referenzpegel für dBV ein 1 V (vgl. Gleichung 2.2). Um Messunsicherheiten zu berücksichtigen, werden die Schranken in dBV angegeben, z. B.  $\pm 2$  dBV.

Ein weiteres Maß ist die Verzerrung eines Mikrofons. Diese wird durch die gesamte harmonische Verzerrung (THD: engl. *Total Harmonic Distortion*) beschrieben. Die Verzerrung eines reinen Sinussignals wird deutlich, wenn es sich einer Rechtecksfunktion annähert. In diesem Fall werden harmonische Vielfache der Fundamentalfrequenz erzeugt. Dabei setzt die THD den harmonischen Anteil ins Verhältnis zu den reinen Frequenzen. Hierfür werden die zugehörigen Effektivspannungen verwendet. Somit beträgt die THD 0%, falls keine Verzerrung auftritt und 48,3%, falls eine reine Rechtecksfunktion vorliegt. Der Mensch kann eine Verzerrung eines akustischen Signals ab einer THD von 1% erkennen. Aus diesem Grund geben Hersteller den maximalen Schalldruckpegel an, für den eine THD von 1% (bzw. bereits ab 0,5%) angenommen wird. Für die Angabe des Schalldruckpegels wird in der Regel eine Sinusfunktion mit 1 kHz verwendet.

Weiterhin ist das Eigenrauschen eines Mikrofons wichtig. Hierbei werden in einem echofreien Raum die Eigengeräusche des Mikrofons aufgezeichnet. Daraus wird der entsprechende Schalldruckpegel ermittelt. In vielen Fällen wird das Ergebnis mit einer Skala gewichtet, die das menschliche Hörverhalten berücksichtigt. So nimmt der Mensch ein Signal mit 1000 Hz bei 80 dB SPL mit der gleichen Lautstärke wahr, wie ein Signal mit 100 Hz bei 100 dB SPL. Die verwendete Skalierung wird als A-Skala bezeichnet und der gewichtete Schalldruckpegel mit dB(A) notiert. Alternativ wird das Ergebnis durch das Signal-Rausch-Verhalten angegeben. Hierbei wird als Referenzsignal bei 94 dB SPL verwendet und hiervon der Pegel des Eigengeräuschs subtrahiert.

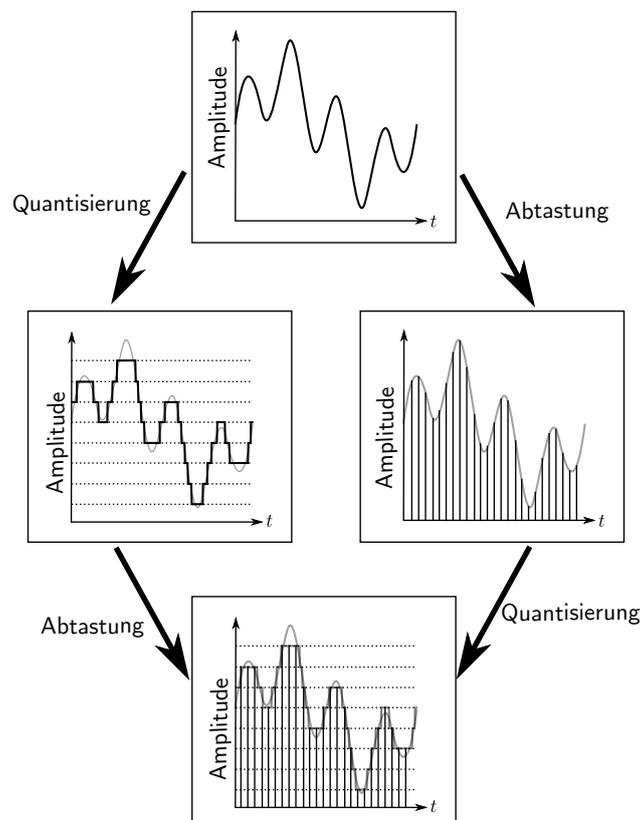


Abbildung 2.5: Vorgang der Digitalisierung eines analogen Signals [Cal09].

## Digitale Darstellung - Mikrofonsignale

Um die Mikrofonsignale zu verarbeiten, müssen die Signale in eine digitale Repräsentation umgewandelt werden. Hierbei wird zunächst das kontinuierliche analoge Amplitudensignal zu diskreten Zeitpunkten abgetastet. Dieser Prozess erfolgt innerhalb der A/D-Umwandlung. Im Anschluss wird der Amplitudenwert diskretisiert. Dieser Prozess wird als Quantisierung bezeichnet. Ein Beispiel des Prozesses ist in Abbildung 2.5 dargestellt. Klassische CDs (engl. *Compact Disc Digital Audio*) verwenden eine Abtastfrequenz von 44,1 kHz und eine Quantisierung von 16 Bit (DIN EN 60908:1999).

### 2.1.5 Frequenzanalyse

Um akustische Signale hinsichtlich der enthaltenen Frequenzanteile zu untersuchen, ist die zeitliche Darstellung nicht geeignet. Stattdessen wird mithilfe der Fourier-Reihenentwicklung das Signal durch wenige Koeffizienten beschrieben. Im Anschluss können durch die Parameter die Amplitude und Phase ermittelt werden. Die Reihenentwicklung kann nur durchgeführt werden, wenn ein periodisches Signal  $s(t)$  mit Periodenzeit  $T$  vorliegt:

$$s(t) = s(t + T). \quad (2.12)$$

Das Signal  $s(t)$  kann durch die Fourier-Reihe in der reellen Darstellung folgendermaßen dargestellt werden [Kid15, Kapitel 2]:

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)). \quad (2.13)$$

Hierbei können nur diskrete Kreisfrequenzen verwendet werden, die ein ganzzahliges Vielfaches von  $\omega_0 = 2\pi f_0$  sind. Somit kann ein monofrequentes sinus-förmiges Signal durch zwei reelle Parameter  $a_n$  und  $b_n$  dargestellt werden. Die restlichen Parameter sind null. Aufgrund der Komplementarität der sinusförmigen Basisfunktionen, kann eine eindeutige Zerlegung von sinusförmigen Funktionen durchgeführt werden. Die Parameter werden folgendermaßen ermittelt [Kid15, Kapitel 2]:

$$a_n = \frac{2}{T} \int_0^T s(t) \cdot \cos(n\omega_0 t) dt \quad n \geq 0, \quad (2.14)$$

$$b_n = \frac{2}{T} \int_0^T s(t) \cdot \sin(n\omega_0 t) dt \quad n \geq 1. \quad (2.15)$$

Zur Verdeutlichung wird die sinusförmige Funktion  $s(t)$  aus Gleichung (2.10) aufgegriffen und die Parameter  $b_n$  ermittelt. Die Ergebnisse sind in Abbildung 2.6 dargestellt.

Weiterhin kann die Amplitude  $d_n$  und die Phase  $\phi_n$  eines Signals aus den Parametern  $a_n$  und  $b_n$  ermittelt werden [Kid15, Kapitel 2]:

$$d_n = \sqrt{a_n^2 + b_n^2} \quad ; \quad \phi_n = \arctan \frac{b_n}{a_n}; \quad a_n \neq 0. \quad (2.16)$$

Hierdurch ist es möglich Gleichung (2.13) folgendermaßen darzustellen:

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} d_k \cos(n\omega_0 t - \phi_n). \quad (2.17)$$

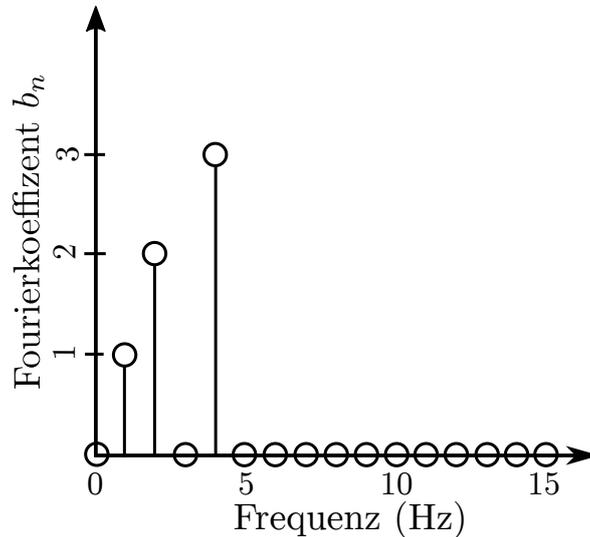


Abbildung 2.6: Fourierkoeffizienten  $b_n$  von  $s(t)$  aus Gleichung (2.10)

Alternativ können die reellen Koeffizienten  $a_n$  und  $b_n$  durch die komplexen Parametern  $c_n$  ersetzt werden:

$$s(t) = \sum_{n=-\infty}^{\infty} (c_n e^{in\omega_0 t}). \quad (2.18)$$

Hierbei entspricht  $i$  der komplexen Zahl  $i^2 = -1$ . Die komplexen Koeffizienten  $c_n$  ergeben sich zu:

$$c_n = \frac{1}{T} \int_0^T s(t) e^{-in\omega_0 t} dt. \quad (2.19)$$

Die reelle Darstellung lässt sich in die komplexe Repräsentation überführen und vice versa. Der Zusammenhang ist folgendermaßen gegeben:

$$c_0 = \frac{a_0}{2}, \quad c_n = \frac{a_n - ib_n}{2}, \quad c_{-n} = \frac{a_n + ib_n}{2}. \quad (2.20)$$

Bisher wurde angenommen, dass ein periodisches Signal betrachtet wird. Jedoch ist diese Annahme für viele Signale in der Realität nicht erfüllt. Um dennoch eine Frequenzanalyse von diesen Signalen durchzuführen, wird die Fourierreihe auf die Fourier-Transformation erweitert. Dabei bezeichnet  $S(w)$  die Fourier-Transformierte von  $s(t)$ . Die Berechnung erfolgt mit [Mer13, Kapitel 3] folgendermaßen:

$$S(w) = \int_{-\infty}^{\infty} s(t) e^{-iwt} dt. \quad (2.21)$$

Im Umkehrschluss kann das Signal  $s(t)$  durch die inverse Fouriertransformation (IFT) aus der Fourier-Transformierten  $S(w)$  ermittelt werden [Mer13, Kapitel 3]:

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(w)e^{iwt} dw. \quad (2.22)$$

Zwischen den komplexen Parametern  $c_n$  der Fourierreihe und der Fouriertransformierten  $S(w)$  besteht folgender Zusammenhang:

$$c_n = \frac{1}{T} S(nw_0) \quad (2.23)$$

Somit kann die Fouriertransformierte als Einhüllende der Spektrallinien von  $c_n$  aufgefasst werden.

Bisher ist die Fouriertransformation ausschließlich für kontinuierliche Signale definiert. Jedoch wird für die Analyse von Mikrofondaten oder anderen digitalen Signalen eine Repräsentation für diskrete Signale benötigt. Aus diesem Grund wird die Diskrete Fourier-Transformation (DFT) definiert.

Sei folgende Zahlenfolge  $\{s_n\}$  der Länge  $n_z$  gegeben:

$$\{s_n\} = \{s_0, s_1, \dots, s_{n_z-1}\} \quad (2.24)$$

Hieraus ergibt sich die DFT  $S_n$  zu [Neu12, Kapitel 3]:

$$S_n = \sum_{k=0}^{n_z-1} s_k e^{-i2\pi \frac{kn}{n_z}} \quad n = 0, 1, \dots, n_z - 1. \quad (2.25)$$

Aus der  $S_n$  lässt sich durch die Umkehrfunktion - der Inversen Diskreten Fourier-Transformation (IDTF) - die ursprüngliche Zahlenfolge  $\{s_n\}$  berechnen [Neu12, Kapitel 3]:

$$s_k = \frac{1}{n_z} \sum_{n=0}^{n_z-1} S_n e^{i2\pi \frac{kn}{n_z}}, \quad k = 0, 1, \dots, n_z - 1. \quad (2.26)$$

Entscheidend für die Berechnung der DFT ist die Länge der betrachteten Zahlenfolge  $\{s_n\}$ . So besitzt z. B. ein Sprachsignal für eine kurze Zeitdauer einen stationären und periodischen Verlauf. Im Gegenzug dazu kann für einen langen Zeitraum eine Aussage über die Zusammensetzung der Frequenzanteile getroffen werden. Bei der Analyse von Sprachsignalen wird daher ein Kompromiss gefunden und ein Zeitabschnitte von 15-20 ms betrachtet [Eul06]. Um den zeitlichen Verlauf der Fouriertransformation über die Zeit darzustellen, wird die Kurzzeit-Fourier-Transformation (STFT: engl. *Short-time-Fourier-Transform*) verwendet. Hierbei wird das gesamte Signal mit einer zeitabhängigen Fenster-Funktion  $w(t)$  multipliziert. Dabei wird die Auswahl der Zahlenfolge  $\{s_n\}$  berechnet. Klassische Fensterfunktionen sind u. a. die Rechteck-Funktion, Hamming-Funktion oder das Von-Hann-Fenster.

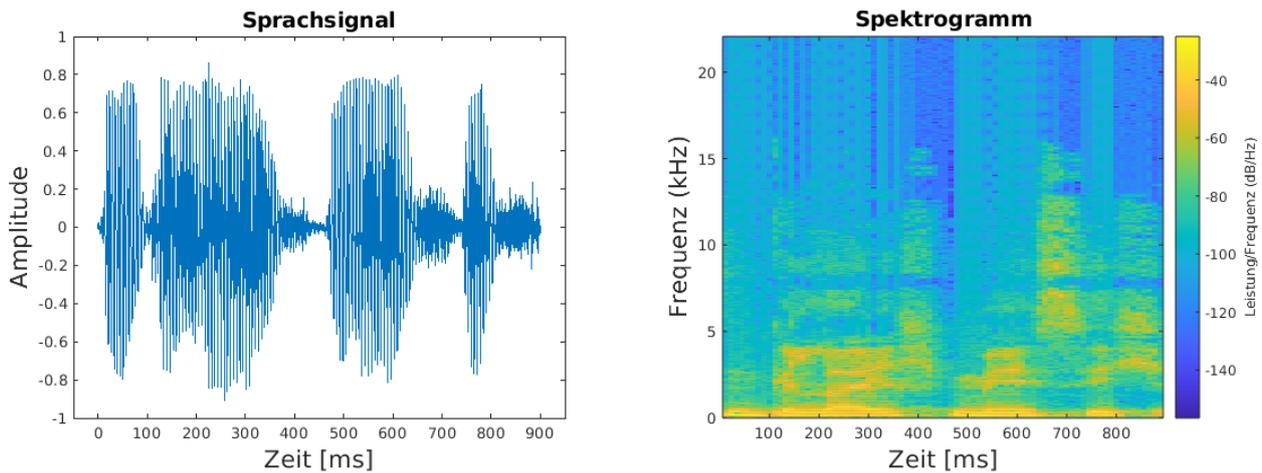
Die diskrete STFT( $t, f$ ) berechnet sich somit folgendermaßen [Mer13, Kapitel 8]:

$$\text{STFT} = \sum_{n=-\infty}^{\infty} s(t)w(t - \tau)e^{-iwt} dt. \quad (2.27)$$

Um die STFT darzustellen, kann das Spektrogramm  $\text{spec}(t, f)$  verwendet werden. Dieses berechnet sich folgendermaßen [Mer13, Kapitel 8]:

$$\text{spec}(t, f) = |\text{STFT}(t, w)|^2. \quad (2.28)$$

In Abbildung 2.7b ist eine Audioaufnahme menschlicher Sprache und das zugehörige Spektrogramm dargestellt.



(a) Audiosignal des Wortes „gleichmäßig“

(b) Spektrogramm von Signal (a).

Abbildung 2.7: (a) Akustisches Signal (b) zugehöriges Spektrogramm.

Im weiteren Verlauf dieser Arbeit werden Merkmale aus den akustischen Signalen extrahiert. In den folgenden Abschnitten werden unterschiedliche Richtungsschätzverfahren vorgestellt.

## 2.2 Soundquellenlokalisierung

### 2.2.1 Einführung

Das Ziel der Soundquellenlokalisierung besteht darin, die Position von Soundquellen zu ermitteln. Je nach verfügbarem Wissen und eingesetzten Verfahren beschränkt sich die Lokalisierung jedoch auf die Schätzung der relativen Richtung zu einer Soundquelle. Im Bereich der *Robot Audition* haben sich für die Soundquellenlokalisierung zwei grundlegend verschiedene Ansätze entwickelt:

- Binaurale Verfahren und
- Mikrofonarray-Verfahren.

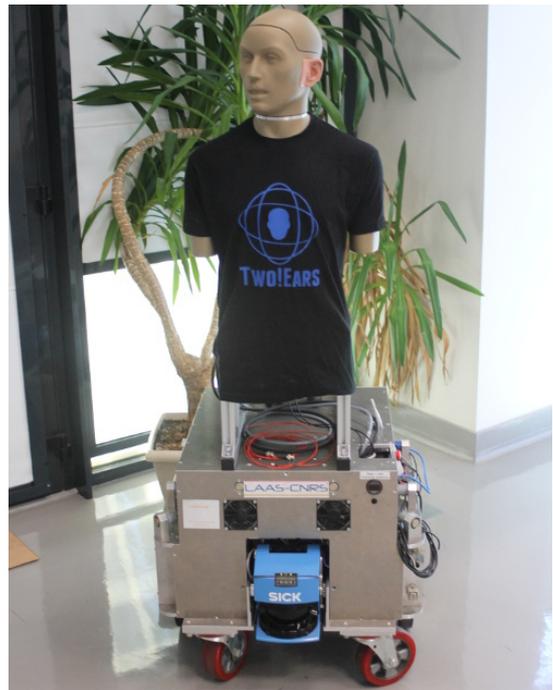
Bei den binauralen Verfahren dient die akustische Verarbeitung und Wahrnehmung des Menschen als Vorbild. Die Ohren des Menschen werden durch zwei Mikrofone repräsentiert und auf einer kopfähnlichen Struktur angeordnet. Hierbei werden die Mikrofone entweder auf einer sphärischen Geometrie angeordnet oder in einem künstlichen Ohrkanal platziert. Als Beispiel ist in Abbildung 2.8b eine künstliche Kopfstruktur (KEMAR) dargestellt, die auf einem mobilen Roboter montiert ist. Berücksichtigt werden für die Richtungsschätzung akustische Merkmale, die der Mensch wahrnimmt.

Demgegenüber stehen Ansätze mittels Mikrofonarray. Hierbei sind mehrere Mikrofone auf einer beliebigen Konfiguration angeordnet. Im Vergleich zu binauralen Verfahren befindet sich in der Regel kein Körper zwischen den Mikrofonen.

Im Folgenden wird auf klassische Verfahren in den Gebieten der Soundquellenlokalisierung eingegangen. Begonnen wird in Abschnitt 2.2.2 mit den binauralen Verfahren und im Anschluss werden in Abschnitt 2.2.3 die Mikrofon-Array-Verfahren vorgestellt.



(a) B1-E Dummy Head [Ent21]



(b) Forschungsroboter Frank

Abbildung 2.8: (a) B1-E-Kopfstruktur (b) Forschungsroboter Frank aus dem Projekt Two!Ears mit KEMAR-Kopf und Torso.

## 2.2.2 Binaurale Verfahren

Binaurale Verfahren orientieren sich an den Ansätzen, die der Mensch zur Lokalisierung von Soundquellen durchführt. Das Ziel ist es, diese Ansätze durch ein rechnergestütztes Modell nachzubilden. Ausgangspunkt ist die auditorielle Szenenanalyse (ASA)<sup>5</sup>, die von Bregman [Bre94] eingeführt wurde. Innerhalb der ASA wird die akustische Wahrnehmung des Menschen beschrieben. Wichtig ist, dass die biologische Plausibilität in den Modellen gegeben ist. Darauf aufbauend wurde die computer-gestützte auditorielle Szenenanalyse (CASA)<sup>6</sup> entwickelt [WB06]. Hierbei werden die akustischen Modelle der ASA auf maschinelle Systeme erweitert.

Klassische Verfahren der Signalverarbeitung analysieren akustische Signale, die in einer Frequenzdarstellung - dem Fourierraum (vgl. Abschnitt 2.1.5) angegeben sind. Dem gegenüberstehend verwenden die binauralen Verfahren eine alternative Darstellung. Dabei orientieren sie sich an der Wahr-

<sup>5</sup> Auditory Scene Analysis

<sup>6</sup> Computational Auditory Scene Analysis

nehmung der Frequenzen des Menschen in der Cochlea des Innenohres: Akustische Signale werden über die Basilarmembran übertragen. Aufgrund der unterschiedlichen Steifigkeit der Basilarmembran, vibriert sie durch Anregung bestimmter Frequenzen an unterschiedlichen räumlichen Positionen. Demnach kann durch den Ort der Vibration auf die Frequenz geschlossen werden. Durch innere Haarzellen wird an der Stelle der Vibration ein Nervenreiz übertragen. Hierdurch kann das menschliche Gehirn eine Analyse durchzuführen. Weitergehende Informationen zur auditoriellen Wahrnehmung des Menschen sind in [Dun+15, Kapitel 12] gegeben.

Der Mensch kann nicht nur unterschiedliche Frequenzen erfassen, sondern ist auch in der Lage die Richtung zu Soundquellen zu schätzen. Dabei unterscheiden sich die Ansätze, um die horizontale und die vertikale Richtung zu ermitteln. Rayleigh konnte Anfang des 20. Jahrhunderts zeigen, dass der Mensch zwei unterschiedliche akustische Merkmale verwendet, und somit die horizontale Richtung zu Soundquellen schätzen kann. Diese Erkenntnisse veröffentlichte er als Duplex-Theorie [Ray07]:

1. Zeitdifferenz (engl. *Interaural Time Difference* - ITD) und
2. Lautstärke-Differenz (engl. *Interaural Level Differenz* - ILD).

Die ITD beschreibt die zeitliche Differenz des Eintreffens eines akustischen Signales zwischen den beiden Ohren. Befindet sich eine akustische Quelle näher am linken als am rechten Ohr, tritt dort das Signal früher auf. Aussagekräftig ist die ITD, falls die einfallenden Wellenlängen größer als der Kopfdurchmesser sind (im Bereich von 1,5 - 2,0 kHz). Falls die Frequenzen höher als 2,0 kHz sind, entstehen Reflexionen und die akustischen Wellen werden gestreut. Die Ursache liegt darin, dass die Wellenlängen kleiner als der Kopfdurchmesser sind. Als Folge nimmt der Schalldruck ab und eine Schalldruckdifferenz kann zwischen den beiden Ohren ermittelt werden. Diese Differenz wird durch die ILD beschrieben. Die ITD und die ILD können mittels verschiedener Verfahren aus den Mikrofondaten ermittelt werden. So werden unterschiedliche Ansätze in [YAZ12] gegenübergestellt.

Aus der ITD und ILD können mithilfe eines mathematischen Modells der horizontale Winkel - Azimut - zu einer Soundquelle ermittelt werden. Hierfür wird die HRTF<sup>7</sup> verwendet. Sie beschreibt den mathematischen Zusammenhang zwischen dem ausgesendeten und dem empfangenen Signal im Ohrkanal [CW99]. Aufgrund individueller Unterschiede der Ohr-, Kopf- und Schultergeometrie unterscheidet sich die HRTF bei jedem Menschen.

Die HRTF kann für Kopfstrukturen analytisch bestimmt werden, die aus einfachen geometrischen Primitiven bestehen, z. B. Kreise oder Kugeln. Hierfür wurden vereinfachte Modelle im Kontext der Robotik entwickelt (vgl. Abbildung 2.9). Ein vereinfachtes Modell beschreibt die Auditory Epipolar Geometry (AEG) [Nak+00]. Hierbei sind die beiden Mikrofone mit dem Abstand  $d$  in einem freien Feld angeordnet. Ein Füllkörper zwischen den Mikrofonen - wie der Kopf des Menschen - wird nicht berücksichtigt. Als Folge wird der Propagationsweg der akustischen Wellen bei der AEG unterschätzt. Im Weiteren wurde das Modell um einen Füllkörper erweitert, und als Revised Epipolar Geometry (RAEG) bezeichnet [NOK01]. Hierbei sind die Mikrofone auf einer Scheibe angeordnet. Die zeitliche Differenz des Erreichens der beiden Mikrofone und der Druckunterschied sind abhängig von dem Füllkörper. Für reale Experimente, bei denen Mikrofone auf einem Kopf angeordnet sind, wurde das Modell innerhalb der Scattering Theory (ST) erweitert [Nak+03]. Die Mikrofone sind bei diesem Ansatz auf einer idealen Kugeloberfläche angeordnet.

---

<sup>7</sup> Head Related Transfer Function

Innerhalb einer numerischen Berechnung - wie durch die Finite-Element-Methode - kann die HRTF unter Kenntnisse der verwendeten Ohr-, Kopf- und Schultergeometrie berechnet werden [Ma+15]. Demgegenüber stehen Ansätze, bei denen die HRTF anhand von Messungen im Feldversuch ermittelt wird [GM95].

Für komplexe Kopfstrukturen, wie sie in der Praxis auftreten, reichen die drei vorgestellten Modelle (vgl. Abbildung 2.9) jedoch nicht aus. Stattdessen ergeben sich folgende zwei Ansätze:

- numerische Berechnung oder
- Messungen im Feldversuch.

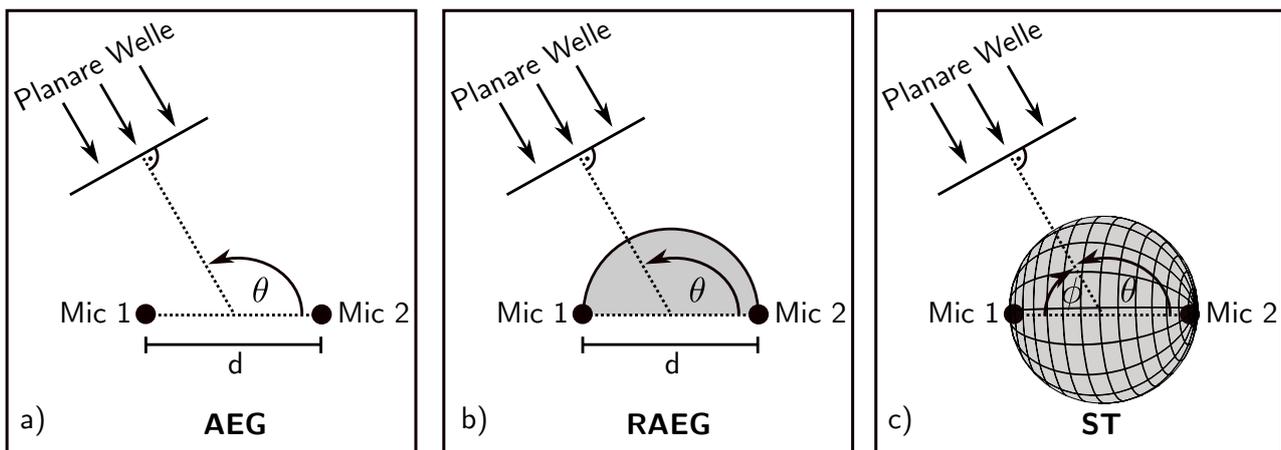


Abbildung 2.9: Unterschiedliche HRTF-Modelle an Anlehnung aus [AD15]: (a) AEG - Zwischen den Mikrofonen ist kein Körper (b) RAEG - Mikrofone sind auf dem Rand einer Scheibe angeordnet (c) ST - Mikrofone sind auf einer Sphäre angeordnet.

Diese zwei Ansätze werden in der Praxis vermehrt eingesetzt. Dabei wird eine Soundquelle an unterschiedlichen Positionen in einem echofreien Raum angeordnet und die Ausgänge im Ohrkanal gemessen. Hierdurch kann eine Beziehung zwischen der Position der Soundquelle und dem gemessenen Mikrofonsignal aufgestellt werden. Mehrere Forschungsinstitute bieten Datenbanken von gemessenen HRTF für ausgewählte Kopfgeometrien an, z. B. die CIPIC-Datenbank [Alg+01] oder die ITA-HRTF-Datenbank [BFF16].

In vielen Fällen genügt es jedoch nicht die HRTF aus den Datenbanken zu verwenden, da sie in realen Umgebungen keine nutzbaren Ergebnisse liefert. Die Ursache liegt darin, dass akustische Reflexionen resultieren und in der HRTF nicht berücksichtigt sind. Zur Lösung des Problems wird die Transformationsfunktion im Allgemeinen am Einsatzort ermittelt. Als Nachteil werden jedoch zusätzlich die Eigenschaften des Raumes erfasst, wodurch die Adaption auf einen anderen Einsatzort häufig ungeeignet ist.

Neben der Schätzung des horizontalen Winkels - Azimut - wird für die Soundlokalisierung der vertikale Winkel - Elevation - benötigt. Im Unterschied zum Azimut wird nur ein einzelnes Signal benötigt, um die Elevation zu schätzen [MG91]. Je nach vertikaler Anordnung einer Soundquelle resultieren Reflexionen an unterschiedlichen Bereichen des Ohrkanals des Menschen. Als Konsequenz überlagern sich die akustischen Wellen und eine stehende Welle resultiert. Dieses Phänomen kann im Frequenzbereich durch einen Knick im Verlauf erkannt werden. Durch einen individuellen Lernprozess kann der Mensch anhand der Position dieser Knicke den Elevation schätzen.

Im binauralen Kontext wurde bisher wenig Forschung bei der Entfernungsbestimmung zu akustischen Quellen geleistet. Dem Mensch fällt es leichter innerhalb einer realen Umgebung die Distanz zu einer akustischen Quelle zu schätzen als in einem echofreien Raum [MG91]. Es besteht dabei ein Zusammenhang zwischen der Signalenergie, die den Hörer direkt erreicht, und der Signalenergie, die durch Reflexionen wahrgenommen wird. Abschließend kann jedoch festgestellt werden, dass der Mensch Schwierigkeiten hat, die Entfernung zu einer Soundquelle akkurat zu schätzen. So unterschätzt der Mensch die Entfernungen oberhalb einen Meter und im Gegenzug überschätzt er Distanzen unterhalb einen Meter [Zah02].

Abschließend sind die binauralen Verfahren in der mobilen Robotik und ihre Entwicklung in [AD15] und [RM17] zusammengefasst.

### 2.2.3 Mikrofon-Array-Verfahren

Neben den binauralen Verfahren, die sich an der auditoriellen Wahrnehmung des Menschen orientieren, existiert ein alternativer Ansatz - die Mikrofon-Array-Verfahren. Bei diesem Ansatz werden mehrere Mikrofone in einem Array angeordnet. Im Gegensatz zu den binauralen Verfahren sind die Mikrofone nicht auf einer Kopfgeometrie angeordnet, sondern zumeist auf einer Struktur ohne Füllkörper. Die Anordnung der Mikrofone innerhalb des Arrays sind essenziell für die Performance der jeweiligen Verfahren. Häufig werden Geometrien für die Mikrofonanordnungen gewählt, die eine vereinfachte mathematische Beschreibung ermöglichen. In Abbildung 2.10 sind unterschiedliche Mikrofonanordnungen dargestellt.

Einerseits existieren Anordnungen, die fest auf einer Struktur angeordnet sind. Die Struktur kann im späteren Verlauf nicht modifiziert werden. In diese Kategorie fallen Konsumgüter, wie die Kamera PlayStation Eye (vgl. Abbildung 2.10a) oder der Amazon Dot 4 (vgl. Abbildung 2.10b). Andererseits existieren fixierte Mikrofonanordnungen, die im Arbeitsbereich eingesetzt werden und somit Video-Konferenzen durchgeführt werden. Beispiele sind in Abbildungen 2.10c und 2.10d dargestellt. Demgegenüber stehen Mikrofonarrays, die flexibel angeordnet werden können. Darunter fallen der Versuchsträger Azimut-3 in Abbildung 2.10e oder der KUKA youBot mit dem Mikrofonarray 8SoundUSB in Abbildung 2.10f. Letzteres Mikrofonarray wird in der Evaluation dieser Dissertation eingesetzt.

Dadurch, dass mehrere Mikrofone verwendet werden, kann die Robustheit von den Ergebnissen der Soundquellenlokalisierung gesteigert werden. Im Vergleich zu den binauralen Verfahren steigt jedoch die Rechendauer, da mehreren Audiosignalen betrachtet werden.

Klassische Verfahren, um die Richtung zu Soundquellen zu schätzen, können in die folgenden drei Kategorien eingeordnet werden [AD15]:

- Beamforming (vgl. Abschnitt 2.2.3.2)
- Kreuzkorrelations-basiert (vgl. Abschnitt 2.2.3.2 bis 2.2.3.3)
- Unterraummethoden, z. B. MUSIC (engl. *MUltiple Signal Classification*) (vgl. Abschnitt 2.2.3.5)

Nachdem im nächsten Abschnitt eine einheitliche Notation definiert wird, folgt im Anschluss die Vorstellung von bekannten Verfahren.

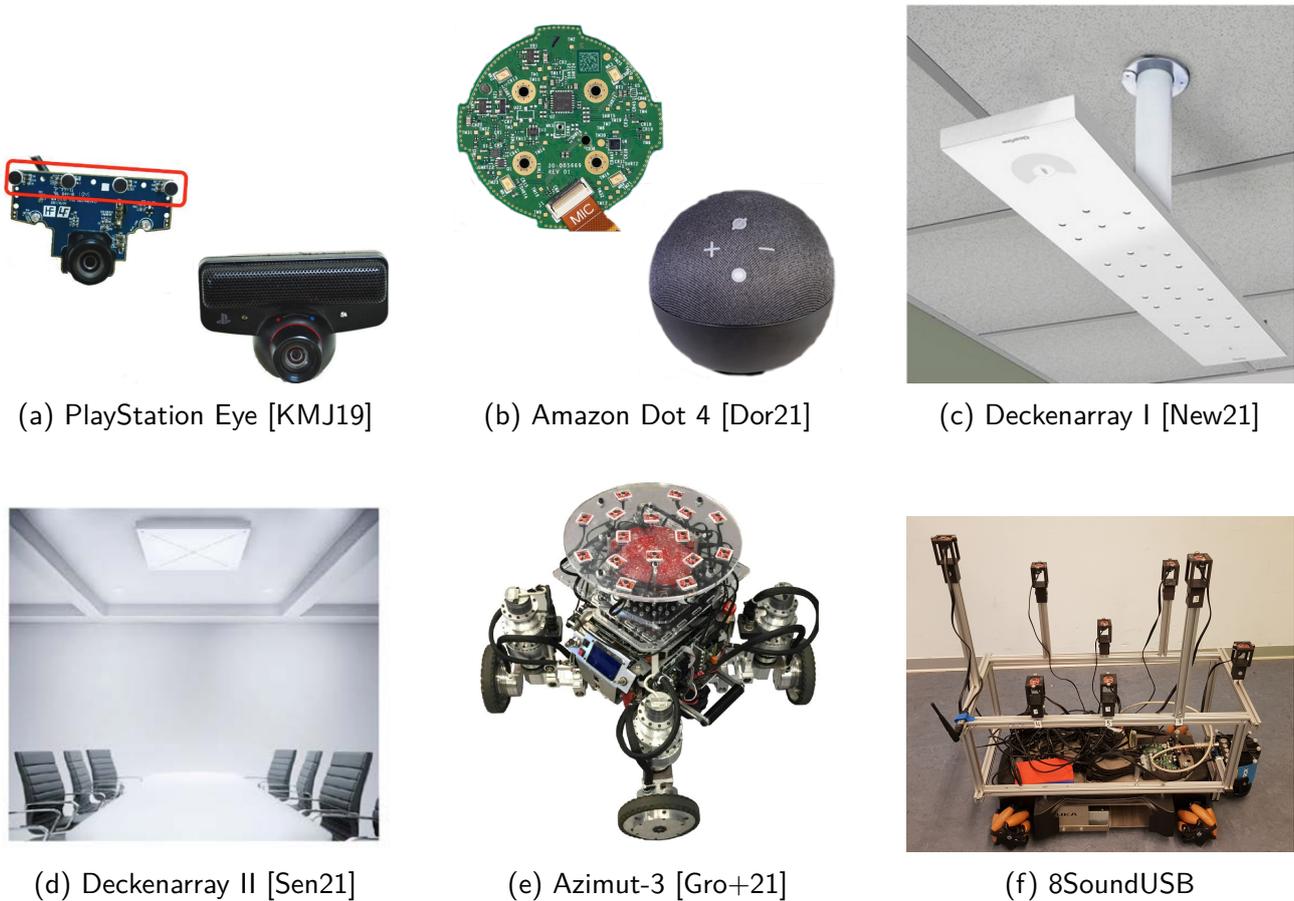


Abbildung 2.10: Darstellung unterschiedlicher Mikrofonarrays. (a) Sony PlayStation Eye Kamera - 4 Mikrofone (b) Amazon Dot 4 - 4 Mikrofone (c) Clear One 910 Beamforming Microphone Array 2 - 24 Mikrofone (d) Sennheiser TeamConnect Ceiling 2 - 28 Mikrofone (e) 16SoundUSB mit Plattform Azimut-3 - 16 Mikrofone (f) 8SoundUSB mit Plattform KUKA youBot - 8 Mikrofone

### 2.2.3.1 Notation und Begriffe

Sei im Folgenden eine aktive Soundquelle  $i$  an der kartesischen Koordinate  $\mathbf{x}_{s_i}$  gegeben, die das Signal  $s_i(t)$  aussendet. Die akustische Welle erreicht  $n_m$  Mikrofone an den kartesischen Koordinaten  $\mathbf{x}_{m_1}, \dots, \mathbf{x}_{m_{n_m}}$  und wird durch die Mikrofonsignale  $m_1(t), \dots, m_{n_m}(t)$  erfasst.

Sei weiterhin angenommen, dass aufgrund des Propagationspfades das Signal  $s_i(t)$  nach der Zeit  ${}^{(i)}t_k$  durch das Mikrofonsignal  $m_k(t)$  wahrgenommen wird. Während sich die akustische Welle räumlich ausbreitet, wird das Signal  $s_i(t)$  um den Faktor  ${}^{(i)}\tilde{\alpha}_k$  gedämpft. Weiterhin wird ein Rauschteil  $v_k(t)$  berücksichtigt, der vom Mikrofon  $k$  und der Umgebung abhängig ist. Es ergibt sich der folgende Zusammenhang:

$$m_k(t) = {}^{(i)}\tilde{\alpha}_k \cdot s_i(t - {}^{(i)}t_k) + v_k(t). \quad (2.29)$$

Es wird angenommen, dass keine Information über das ausgesendete Signal  $s_i(t)$  und dem Zeitpunkt der Aussendung vorliegt. Daher ist es notwendig eine mathematische Beziehung zwischen den unterschiedlichen Mikrofonen aufzustellen. Es wird angenommen, dass sich zwei Mikrofonsignale  $m_k(t)$

und  $m_l(t)$  durch eine zeitliche Differenz  ${}^{(i)}\Delta t_{kl}$ , eine Dämpfung  ${}^{(i)}\alpha_{kl}$  und einen Rauschanteil  $v_{kl}(t)$  unterscheiden:

$$m_k(t) = {}^{(i)}\alpha_{kl} \cdot m_l(t - {}^{(i)}\Delta t_{kl}) + v_{kl}(t). \quad (2.30)$$

Die zeitliche Differenz  ${}^{(i)}\Delta t_{kl}$  wird im weiteren Verlauf dieser Dissertation als TDoA (engl. *Time Difference of Arrival*) bezeichnet. Um den Zusammenhang zwischen den Ankunftszeiten  ${}^{(i)}t_k$  und den Zeitdifferenzen  ${}^{(i)}\Delta t_{kl}$  für eine Soundquelle zu verdeutlichen, ist in Abbildung 2.11a ein exemplarisches Beispiel gegeben.

Sei die Position der Mikrofone zum Messzeitpunkt als statisch angenommen. Durch Variation der TDoA  ${}^{(i)}\Delta t_{kl}$  in Gleichung (2.30) können Mikrofonsignale simuliert werden. Hierdurch ist es möglich eine Soundquellen von beliebigen Positionen (oder Richtungen) wahrzunehmen. Einsatz findet die Simulation von Mikrofonsignalen u. a. bei Videospiele. Dabei wird der Höreindruck des Spielers durch Navigation innerhalb einer virtuellen Welt beeinflusst.

Neben der Simulation von Mikrofonsignalen hat die Variation der TDoA weitere Einsatzmöglichkeiten. So wird in Abschnitt 2.2.3.2 dargestellt, dass eine akustische Ausrichtung der Mikrofone in gewünschte Richtungen durchgeführt werden kann. Als Resultat können dabei unerwünschte Richtungen ausgeblendet werden.

Relevant für die vorliegende Dissertation ist, dass die TDoA genutzt werden kann, um die Position (oder Richtung) einer Soundquelle zu schätzen. Anstatt die TDoAs aus den Mikrofonsignalen zu ermitteln, wird bei bekannten Verfahren der Literatur ein umgekehrter Ansatz verfolgt. Zunächst wird eine diskrete Menge von Hypothesen aufgestellt, die mit potenziellen Positionen (oder Richtung) der Soundquellen assoziiert sind. Für jede Hypothese werden geometrisch die resultierenden TDoAs ermittelt. Im Anschluss wird ein Modell verwendet, um zu charakterisieren wie die theoretischen TDoAs mit den aktuellen Mikrofonsignalen übereinstimmen. Falls die Charakterisierung einen Schwellwert (oder eine Wahrscheinlichkeit) überschreitet, wird die Hypothese als erfüllt angesehen.

Im Folgenden wird beschrieben, wie TDoA geometrisch auf Basis von Hypothesen berechnet werden kann. Zunächst wird eine Hypothese als Kandidaten bezeichnet und in Kugelkoordinaten  $\mathbf{x}_{k_i}(\theta_i, \phi_i, r_i)$  repräsentiert. Hierbei entsprechen  $\theta_i$  dem horizontalen Winkel (Azimut),  $\phi_i$  dem vertikalen Winkel (Elevation) und  $r_i$  dem Abstand zum Ursprung. Abbildung 2.11b dient als exemplarische Darstellung von 16 Kandidaten, die im Kreis um ein Mikrofonarray angeordnet sind.

Im ersten Schritt wird für eine akustische Quelle  $i$  die Zeitdauer  ${}^{(i)}t_l$  berechnet, die sie von Position  $\mathbf{x}_{k_i}$  zum Mikrofon  $m_l$  benötigt:

$${}^{(i)}t_l = \frac{\|\mathbf{x}_{k_i} - \mathbf{x}_{m_l}\|_2}{c}. \quad (2.31)$$

Hierbei entspricht  $c$  der Schallgeschwindigkeit. Darauf folgend wird die TDoA  ${}^{(i)}\Delta t_{kl}$  zwischen den Mikrofonen  $m_k$  und  $m_l$  berechnet:

$${}^{(i)}\Delta t_{kl} = {}^{(i)}t_k - {}^{(i)}t_l. \quad (2.32)$$

Im Weiteren wird angenommen, dass der Kandidat im Vergleich zu dem Abstand der Mikrofone weit entfernt liegt (vgl. Abbildung 2.12). Hierdurch kann die TDoA durch die normierte Richtung  $\mathbf{u}_{k_i}$  des Kandidaten approximiert werden:

$${}^{(i)}\Delta t_{kl} \approx \frac{1}{c}(\mathbf{x}_{m_l} - \mathbf{x}_{m_k}) \cdot \mathbf{u}_{k_i}. \quad (2.33)$$

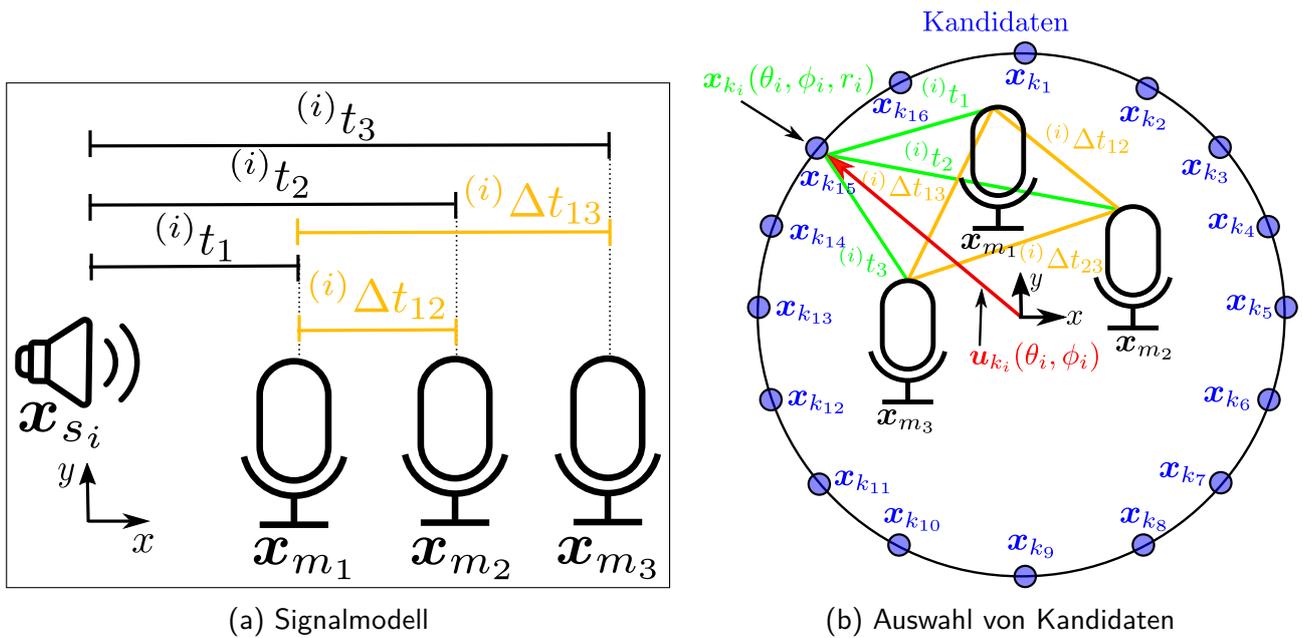


Abbildung 2.11: (a) Darstellung des Signalmodells aus Gleichung (2.30). Eine aktive Signalquelle wird durch drei Mikrofone wahrgenommen. (b) Für Kandidaten (blau) wird geometrisch die zeitliche Differenz (orange) zwischen den Mikrofonpaaren ermittelt. Die Kandidaten können durch ihre 3D-Position  $\mathbf{x}_{k_i}$  oder der normierten Richtung  $\mathbf{u}_{k_i}$  repräsentiert werden.

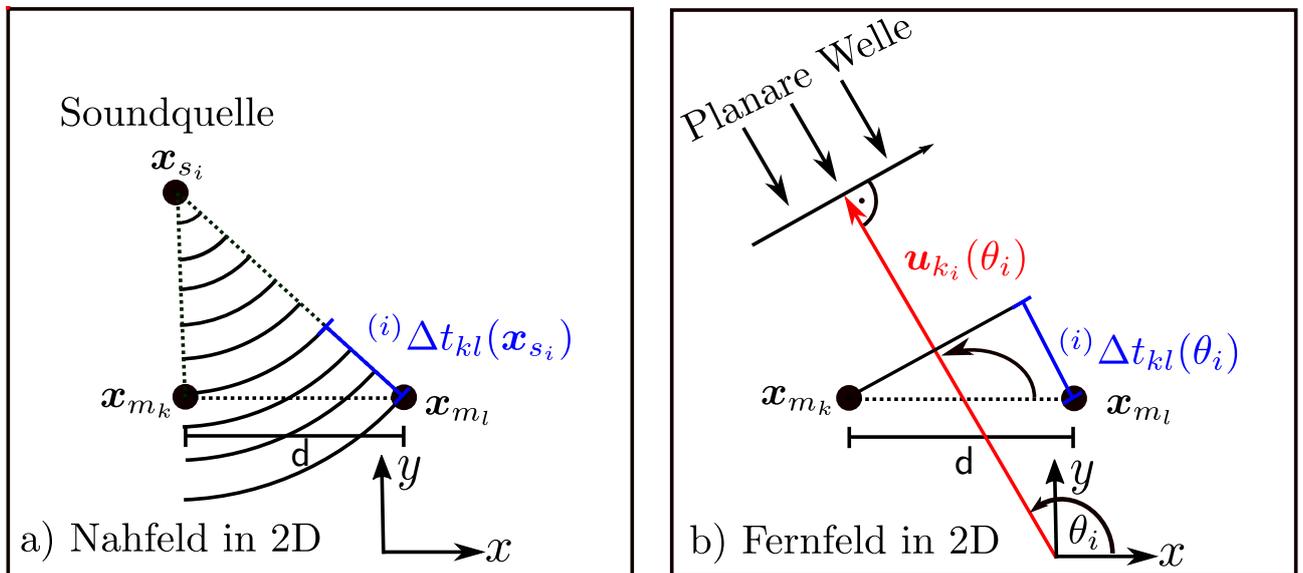


Abbildung 2.12: Unterscheidung zwischen Nah- und Fernfeld.

In den nachfolgenden Abschnitten wird für unterschiedliche Ansätze beschrieben, wie die berechnete TDoA (aus geometrischen Überlegungen) mit den Daten der Mikrofonensignalen in Bezug gestellt werden kann.

Abschließend sei angemerkt, dass im Bereich der Soundquellenlokalisierung die TDoA nicht als Endresultat relevant ist. Stattdessen wird die Position (oder Richtung) einer Soundquelle benötigt. Unter der Annahme einer planaren Welle (vgl. Abbildung 2.12) kann ohne weiteres Wissen - z. B. der Signalcharakteristik und Lautstärke der Soundquelle - nur die Richtung zu einer Soundquelle ermittelt

werden. Diese Richtungsschätzung wird als DoA (engl. *Direction of Arrival*) bezeichnet. Im weiteren Verlauf dieser Arbeit wird die Richtungsschätzung DoA zu einem Kandidaten  $i$  (oder Soundquelle) entweder durch die Winkel  $(\theta_i, \phi_i)$  oder durch den normierten Richtungsvektor  $\mathbf{u}_{k_i}$  repräsentiert.

### 2.2.3.2 Beamforming

Eine bekannte Klasse der Mikrofon-Array-Verfahren ist das Beamforming. Es basiert auf den Arbeiten des deutschen Physikers und Nobelpreisträgers Karl Ferdinand Braun aus dem Jahr 1905. In seinen Arbeiten konnte er durch einen Aufbau von mehreren Antennen die gezielte Abstrahlung eines Signals in eine vorher festgelegte Richtung erzielen [Bra09]. In den weiteren Jahren wurde das Verfahren in unterschiedlichen Disziplinen, wie der Radartechnik [Bar80], Sonar [Tho90], Biomedizin [LZG94] und Akustik [BCH08] eingesetzt.

In Bezug zur Akustik wird Beamforming einerseits eingesetzt, um eine verbesserte Analyse eines Zielsignales zu erreichen. Hierbei werden Störgeräusche aus unerwünschten Richtungen unterdrückt. Dieser Bereich wird als Signalverbesserung (engl. *Signal Enhancement*) bezeichnet. Andererseits wird Beamforming zur Schätzung der Position oder der Richtung eines Zielsignales eingesetzt.

Im Folgenden wird zunächst auf die Signalverbesserung eingegangen und anschließend die Richtungsschätzung erläutert.

In der Literatur haben sich für das Beamforming zwei Ansätze entwickelt:

1. Konventionelle Beamformer
2. Adaptive Beamformer

Konventionelle Beamformer verwenden im Vorfeld feststehende Parameter. Ein klassischer Vertreter ist das sogenannte *Delay-And-Sum-Beamforming* (DS-BF) [Dud77].

Im Gegensatz dazu verwenden adaptive Beamformer dynamische Parameter, die sich in Abhängigkeit der untersuchten Signale ändern. Beispiele sind LCMV (engl. *Linearly constrained minimum variance*) [BM09, Kapitel 3] und MVDR (engl. *Minimum variance distortionless response*) [CBG09, Kapitel 9]

Im Folgenden wird der Ablauf der Signalverbesserung anhand *Delay-And-Sum-Beamforming* an einem vereinfachten Beispiel erläutert. Hierbei ist das Signal  $s_i(t)$  gegeben und es wird von drei

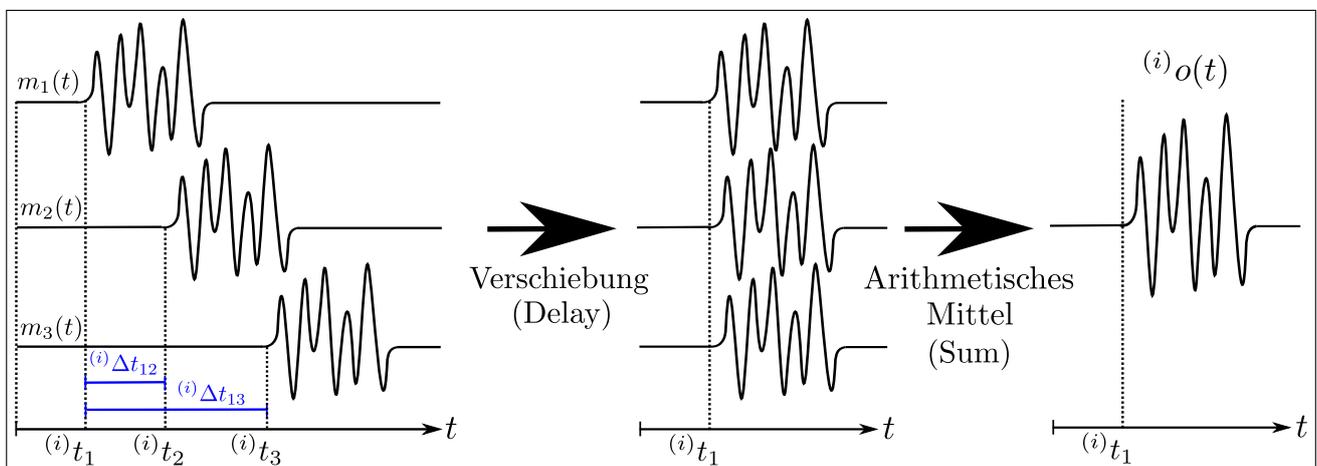


Abbildung 2.13: Exemplarische Darstellung der Funktionsweise des Beamforming-Verfahrens.

Mikrofonen an Positionen  $\mathbf{x}_{m_1}$  bis  $\mathbf{x}_{m_3}$  zu unterschiedlichen Zeitpunkten  ${}^{(i)}t_1$ ,  ${}^{(i)}t_2$  und  ${}^{(i)}t_3$  wahrgenommen (vgl. Abbildung 2.13). Zunächst wird ein Mikrofonsignal  $m_k(t)$  als Referenz ausgewählt und im Anschluss paarweise die zeitliche Differenz  ${}^{(i)}\Delta t_{kl}$  zu einem weiteren Mikrofonsignal  $m_l(t)$  ermittelt.

Die Kernidee des DS-BF besteht darin, alle Mikrofonsignale um die zeitliche Differenz  ${}^{(i)}\Delta t_{kl}$  zu verschieben (engl. *Delay Step*) und im Anschluss das arithmetische Mittel zu berechnen (engl. *Sum Step*):

$${}^{(i)}o(t) = \frac{1}{n_m} \sum_{l=1}^{n_m} m_l(t - {}^{(i)}\Delta t_{kl}) \quad (2.34)$$

Hierbei bezeichnet  ${}^{(i)}o(t)$  den Ausgang (engl. *output*) des Beamforming-Verfahrens,  $n_m$  die Gesamtzahl der Mikrofone und  $k$  notiert den Index des Referenzsignals.

Durch diesen Ablauf wird das Mikrofonarray in eine gewünschte Richtung ausgerichtet. Die Ausrichtung ist dabei von der geometrischen Anordnung der Mikrofone, als auch von der Frequenz des Zielsignals abhängig. Um die Richtcharakteristik (vgl. Abschnitt 2.1.4) einer Mikrofonanordnung zu visualisieren, werden sogenannte Beampattern angefertigt. Ein exemplarisches Beispiel ist in Abbildung 2.14a gegeben. Hierbei sind acht Mikrofone äquidistant mit Abstand 0,2 m auf einer Achse angeordnet. Für die Signalfrequenzen 1 kHz und 1,5 kHz sind die jeweiligen normalisierten Beampattern für eine Ausrichtung in der xy-Ebene (Elevation von  $0^\circ$ ) in Abbildung 2.14b dargestellt.

In den Beampattern sind einerseits die Hauptkeulen (engl. *Main Lobes*), die in  $0^\circ$  und  $180^\circ$  ausgerichtet sind, andererseits auch eine Vielzahl von Nebenkeulen (engl. *Side Lobes*) erkennbar. Durch eine Optimierung der Mikrofonanordnung kann ein Kompromiss zwischen der Stärke von Neben- und Hauptkeule als auch der Breite für die untersuchten Frequenzen gefunden werden.

Die Problematik des DS-BF liegt vor allem darin, dass die räumliche Auflösung für kleine Frequenzen groß ist. In dem behandelten Beispiel kann dies anhand der Nebenkeulen für die Frequenz 1,0 kHz erkannt werden. Weiterhin treten bei hohen Frequenzen Alias-Effekte auf. Als Resultat können Analysen nur in einem kleinen Frequenzbereich von näherungsweise 1 - 2 kHz durchgeführt werden [Tam+05].

Im Folgenden wird beschrieben, wie durch Beamforming-Ansätze die Richtungen von aktiven Soundquellen im Bezug zu einem Mikrofonarray geschätzt werden können. Zunächst werden Kandidaten  $\mathbf{x}_{k_i}$  definiert, die jeweils eine Hypothese der Richtung repräsentieren. Darauffolgend wird die TDoA  ${}^{(i)}\Delta t_{kl}$  für jede Hypothese ermittelt. Dabei ist das Vorgehen wie in vorherigen Abschnitt 2.2.3.1 beschrieben. Mittels der TDoA wird das Beamforming-Verfahren ausgeführt und der Ausgang  ${}^{(i)}o(t)$  berechnet. Schließlich wird die Energie  $E$  von  ${}^{(i)}o(t)$  für den Zeitraum  $t_n$  berechnet:

$$E(i) = \int_{t-t_n}^t |{}^{(i)}o(\tau)|^2 d\tau. \quad (2.35)$$

Abschließend wird die Position der Soundquelle ermittelt, indem der maximale Werte der Energiefunktion geschätzt wird:

$$\mathbf{x}_{k_i} = \arg \max_i E(i) \quad (2.36)$$

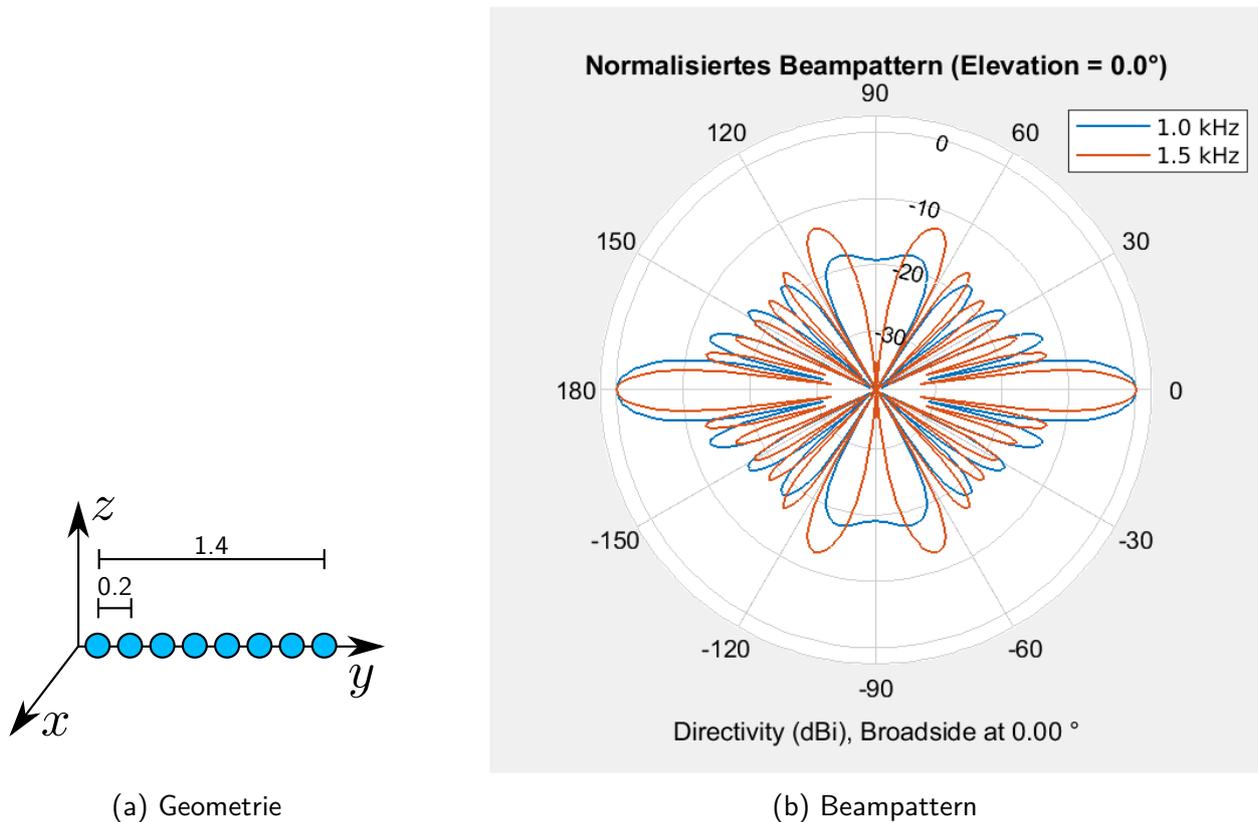


Abbildung 2.14: Ein exemplarisches Mikrofonarray mit zugehörigem Beampattern ist dargestellt.

Aufgrund der Einfachheit und geringen Berechnungsdauer werden Beamforming-Verfahren in vielen Arbeiten der Robotik eingesetzt. Die Güte der Verfahren ist abhängig von der geometrischen Anordnung und der Anzahl der Mikrofone. In [Nak+05] wurde ein Mikrofonarray mit 64 Elementen betrachtet und DS-BF angewendet. In [Nak+09] wurde ein orientierungserweitertes Beamforming-Verfahren vorgestellt und hierfür 96 Mikrofone verwendet. Dabei wurden die Mikrofone innerhalb eines Raumes verteilt. In [Jov+21] wurde untersucht, wie ein Mikrofonarray mit 16 Elementen durch das Beamforming-Verfahren eine Drohe detektieren kann. In weiteren Arbeiten wurde das Mikrofonarray auf eine Drohe verbaut und eine akustische Ortung von menschlichen Stimmen durchgeführt [Aco21].

### 2.2.3.3 Kreuzkorrelation

Im Kapitel 2.2.2 wurden Ansätze vorgestellt, mit denen die interaurale Zeitdifferenz ITD ermittelt werden kann. Im Kontext der Mikrofonarray-Verfahren wird anstelle der Zeitdifferenz eines einzelnen Mikrofonpaares eine Vielzahl von Differenzen verwendet. Dies führt zu denen in Abschnitt 2.2.3.1 definierten TDoA  $^{(i)}\Delta t_{kl}$ . Somit können die gleichen Schätzverfahren aus dem binauralen Kontext verwendet und die TDoA ermittelt werden. Dennoch wird bei Mikrofonarray-Verfahren häufig die Kreuzkorrelation  $r_{m_k, m_l}$  berechnet. Dabei beschreibt die Kreuzkorrelation die Ähnlichkeit zweier Signale  $m_k(t)$  und  $m_l(t)$ . Diese wird ermittelt, indem die Signale zueinander um eine definierte Zeitdifferenz  $\tau$  verschoben und das Ergebnis integriert wird:

$$r_{m_k, m_l}(\tau) = (m_k(t) * m_l(t))(\tau) = \int_{-\infty}^{\infty} m_k(t)^* \cdot m_l(t + \tau) dt. \quad (2.37)$$

Hierbei bezeichnet  $*$  die mathematische Faltungsoperation und  $m_k^*$  die komplex-konjugierte Funktion.

Wenn zwei um  $(i)\Delta t_{kl}$  zeitverschobene Signale so weit verschoben werden bis sie exakt überlappen, liefert Gleichung 2.37 ein Maximum an der Stelle  $(i)\Delta t_{kl}$ . Demnach kann die TDoA durch Argumentenmaximierung von Gleichung (2.37) approximiert werden.

Als Beispiel dient Abbildung 2.15. Hierbei sind die Signale  $m_k(t)$  und  $m_l(t)$  um 50 ms zueinander verschoben. Es kann erkannt werden, dass die Kreuzkorrelationsfunktion ein Maximum bei einer Zeitverschiebung von 50 ms besitzt. Statt die Kreuzkorrelationsfunktion wie in Gleichung (2.37)

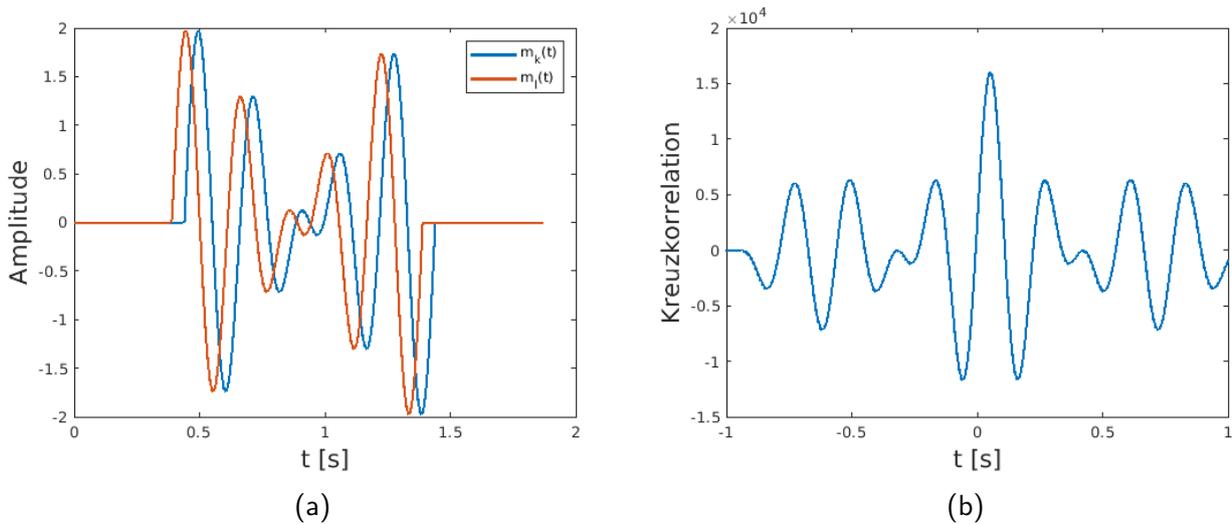


Abbildung 2.15: (a) Zwei Signale  $m_k(t)$  und  $m_l(t)$  sind zeitlich verschoben.  
(b) Das Ergebnis der Kreuzkorrelation  $r_{m_k, m_l}$ .

durch eine Faltungsoperation zu berechnen, kann die Berechnung im Frequenzbereich durchgeführt werden. Zunächst werden die Mikrofonsignale  $m_k(t)$  und  $m_l(t)$  in den Fourierraum transformiert (vgl. Abschnitt 2.1.5). Das Ergebnis wird mit  $M_k(f)$  und  $M_l(f)$  notiert, wobei  $f$  die Frequenz beschreibt. In Anschluss werden  $M_k(f)$  und  $M_l(f)$  komponentenweise konjugiert komplex multipliziert. Das Ergebnis wird als Kreuzleistungsdichtespektrum  $R_{m_k, m_l}$  bezeichnet:

$$R_{m_k, m_l} = M_k(f)^* \cdot M_l(f). \quad (2.38)$$

Die Rücktransformation von  $R_{m_k, m_l}$  in den Zeitbereich liefert schließlich die Kreuzkorrelationsfunktion  $r_{m_k, m_l}$ :

$$r_{m_k, m_l}(\tau) = \int_{-\infty}^{\infty} R_{m_k, m_l} e^{j2\pi f \tau} df. \quad (2.39)$$

Im Bereich der Robotik wurden erste Ansätze in [CH99] vorgestellt, um die TDoA durch die Kreuzkorrelation zu schätzen. In realen Umgebungen, die durch Umgebungsrauschen und akustischer Reflexion beeinflusst sind, kann dieses Verfahren nicht eingesetzt werden. Daher wurde die Kreuzkorrelation  $r_{m_k, m_l}(\tau)$  durch die generalisierte Kreuzkorrelation  $\hat{r}_{m_k, m_l}(\tau)$  erweitert. Dabei wird im Frequenzbereich die Gewichtungsfunktion  $\Psi(f)$  multipliziert:

$$\hat{r}_{m_k, m_l}(\tau) = \int_{-\infty}^{\infty} \Psi(f) \underbrace{M_k(f)^* \cdot M_l(f)}_{R_{m_k, m_l}} e^{j2\pi f \tau} df. \quad (2.40)$$

Unterschiedliche Gewichtungsfunktionen  $\Psi(f)$  wurden in [KC76] vorgestellt. Ein klassischer Vertreter ist die Phasen-Transformation (engl. *Phase Transform* - PHAT):

$$\Psi(f) = \frac{1}{|M_k(f) \cdot M_l(f)| + \epsilon}. \quad (2.41)$$

Hierbei werden die Fouriertransformationen normiert und somit die Amplitude nicht weiter berücksichtigt. Daher besteht nur ein Zusammenhang der Phase der unterschiedlichen Signale. Die Konstante  $\epsilon$  verhindert die Division durch 0 in vorheriger Gleichung. Als problematisch erweist sich dieser Ansatz, falls das Signal viel Umgebungsrauschen enthält. Denn durch die Normierung erhält jeder Frequenzanteil die gleiche Gewichtung. Der Einsatz der GCC-PHAT ist gegeben in [Kim+08] und [Bad+09]. Darüber hinaus wurde in [Fré+12] die GCC-PHAT durch die RWPHAT<sup>8</sup> erweitert. In diesem Verfahren erfolgt eine Gewichtung der Frequenzanteile des Kreuzleistungsdichtespektrum in Abhängigkeit von der Signalenergie. Weiterhin wurde in [Ban+13] ein Verfahren vorgestellt bei dem mehrere Mikrofone und Lautsprecher auf der Oberfläche eines Schlangenroboters montiert wurden.

Analog zu binauralen Verfahren wird anhand der ITD (bzw. TDoA) auf die Position der Soundquelle geschlossen. [MW01] stellen auf Basis von geometrischen Beziehungen die Lösung eines Gleichungssystems vor. Jedoch ist die zugehörige Matrix schlecht konditioniert. Im Weiteren wurde in [AH14] für eine einzelne Quelle eine geometrische Formulierung für die Lokalisierung vorgestellt. Dabei wurde die Annahme getroffen, dass die Laufzeitdifferenz exakt bekannt ist. Dies ist in der Realität jedoch kaum gewährleistet. Schließlich existierten Ansätze auf Basis von neuronalen Netzen, wie in [MEW09].

#### 2.2.3.4 SRP-Verfahren

Ein alternativer Ansatz im Bereich des Mikrofon-Array-Verfahrens ist SRP (engl. *Steered Response Power*). Als Basis dient die generalisierte Kreuzkorrelation (vgl. Abschnitt 2.2.3.3). Im Gegensatz zu vorherigem Abschnitt wird die TDoA nicht direkt aus den Daten der Mikrofone geschätzt. Stattdessen wird eine geometrische Beziehung aufgestellt und durch die Mikrofondaten bewertet.

Zunächst werden potenzielle Kandidaten definiert (vgl. Abbildung 2.11b). Der Vorgang wurde bereits in Abschnitt 2.2.3.2 beschrieben. Im Anschluss wird für jedes Mikrofonpaar  $m_k(t)$  und  $m_l(t)$  geometrisch die resultierende TDoA  ${}^{(i)}\Delta t_{kl}$  berechnet (vgl. Gleichungen (2.31) - (2.31)). Im nächsten Schritt wird ermittelt, wie gut die jeweiligen Kandidat  $\mathbf{x}_{k_i}$  mit den Mikrofondaten übereinstimmen.

Die Kernidee besteht darin, dass die generalisierte Kreuzkorrelation  $\hat{r}_{m_k, m_l}$  ein Maximum annimmt, falls ein potentieller Kandidat mit der korrekten Richtung der Soundquelle übereinstimmt. Das Maximum wird dabei an der geometrisch ermittelten TDoA  ${}^{(i)}\Delta t_{kl}$  angenommen. Dies gilt für alle Mikrofonpaare. Um eine einzelne Kenngröße zu erhalten, wird die Spatial Likelihood Function (SLF)  ${}^{(i)}\mathcal{L}$  definiert. Dabei wird die generalisierten Kreuzkorrelation  $\hat{r}_{m_k, m_l}({}^{(i)}\Delta t_{kl})$  für jedes Mikrofonpaar an der zeitlichen Differenz  ${}^{(i)}\Delta t_{kl}$  ermittelt und im Anschluss summiert:

$${}^{(i)}\mathcal{L} = \sum_{k=1}^{n_m} \sum_{\substack{l=1 \\ k \neq l}}^{n_m} \hat{r}_{m_k, m_l}({}^{(i)}\Delta t_{kl}). \quad (2.42)$$

<sup>8</sup> Relevance-weighted phase transform

Darauffolgend wird  ${}^{(i)}\mathcal{L}$  für jeden Kandidaten  $\mathbf{x}_{k_i}$  berechnet. Schließlich wird die Position (oder Richtung) mit dem größten Wert der Argumentenmaximierung von  ${}^{(i)}\mathcal{L}$  ausgewählt. Der Index  $i^*$  des maximalen Kandidaten  $\mathbf{x}_{k_{i^*}}$  ergibt sich zu:

$$i^* = \arg \max_i {}^{(i)}\mathcal{L}. \quad (2.43)$$

In der Literatur wird die generalisierte Kreuzkorrelation in Gleichung (2.42) häufig durch die Phasentransformation (PHAT) berechnet (vgl. 2.2.3.3). Daher wird das Verfahren als SRP-PHAT bezeichnet. In [Bad+09] werden unterschiedliche Implementierungen des SRP-PHAT-Algorithmus untersucht. Weiterhin ist ein Einsatz des Verfahrens in [Eve+15] dokumentiert.

Einen großen Einfluss auf die Ergebnisse besitzt die Auswahl der zugehörigen Kandidaten. Hierfür wird in [GM19] ein hierarchisches Verfahren vorgestellt. Es wird zunächst eine grobe Abtastung verwendet und im Anschluss die Abtastung in potenziellen Bereichen erhöht. Insgesamt kann durch diesen Prozess sowohl eine hohe Auflösung als auch eine geringe Rechenzeit erzielt werden.

### 2.2.3.5 MUSIC

Zuletzt werden für die Richtungsschätzung eine Unterraum-Methode (engl. *Subspace Method*) vorgestellt. Die Idee basiert darin, dass die Mikrofondaten in zwei disjunkte Unterräume zerlegt werden können:

1. Signalunterraum
2. Rauschunterraum

Dabei enthält der Signalunterraum relevante Informationen von den Zielsignalen und der Rauschunterraum repräsentiert Störgeräusche (Rauschen).

Ein klassisches Verfahren der Unterraum-Methoden ist MUSIC (engl. *MUltiple Signal Classification*). Es wurde erstmals in [Sch86] vorgestellt, und für die Richtungsschätzung von Radarmessungen eingesetzt. Mittlerweile wird MUSIC häufig in der Robotik eingesetzt und zudem für die Richtungsschätzung von akustischen Signalen verwendet. Im Folgenden wird das MUSIC-Verfahren kompakt vorgestellt.

Gegeben sei folgendes Signalmodell im Frequenzbereich:

$$\underbrace{\begin{pmatrix} M_1(f) \\ \vdots \\ M_{n_m}(f) \end{pmatrix}}_{\underline{\mathbf{m}}(f)} = \underbrace{\begin{pmatrix} \underline{\mathbf{a}}(f, \theta_1, \phi_1) & \cdots & \underline{\mathbf{a}}(f, \theta_{n_d}, \phi_{n_d}) \end{pmatrix}}_{\underline{\mathbf{A}}(f, \theta, \phi)} \cdot \underbrace{\begin{pmatrix} S_1(f) \\ \vdots \\ S_{n_d}(f) \end{pmatrix}}_{\underline{\mathbf{s}}(f)} + \underbrace{\begin{pmatrix} W_1(f) \\ \vdots \\ W_{n_m}(f) \end{pmatrix}}_{\underline{\mathbf{w}}(f)}. \quad (2.44)$$

Hierbei entsprechen  $M_k$  und  $S_i$  der Fouriertransformierten der Signale  $m_k(t)$  und  $s_i(t)$ ,  $f$  repräsentiert die Frequenz.  $W_k(f)$  entspricht dem fouriertransformierten Rauschanteil von Mikrofon  $m_k(t)$  assoziiert. Weiterhin ist  $n_s$  mit der Gesamtzahl von Soundquellen, und  $n_m$  mit der Gesamtzahl von Mikrofonen assoziiert.  $\underline{\mathbf{a}}(f, \theta_i, \phi_i)$  entspricht der mathematischen Beziehung zwischen der Fouriertransformierten  $S_i$  des Signals der Soundquelle  $s_i$  und den Fouriertransformationen  $M_1, \dots, M_{n_m}$  der

Mikrofonsignale.  $\underline{a}$  wird als Ausrichtungs-Vektor (engl. *Steering Vector*) bezeichnet und ist abhängig von der Richtung  $(\theta_i, \phi_i)$  zu einer Soundquelle an Position  $\mathbf{x}_{s_i}$ .

Exemplarisch sei das Signal  $s_i(t)$  mit der Zeitdifferenz  ${}^{(i)}t_k(\theta_i, \phi_i)$  an Mikrofon  $m_k$  beobachtbar. Weiterhin sei eine planare Welle vorausgesetzt und keine Dämpfung während der akustischen Ausbreitung angenommen. Somit kann der Ausrichtungsvektor  $\underline{a}(f, \theta_1, \phi_1)$  folgendermaßen dargestellt werden:

$$\underline{a}(f, \theta_1, \phi_1) = \begin{pmatrix} e^{j2\pi f \cdot {}^{(i)}t_1(\theta_1, \phi_1)} \\ \vdots \\ e^{j2\pi f \cdot {}^{(i)}t_{n_m}(\theta_1, \phi_1)} \end{pmatrix}. \quad (2.45)$$

Ausgangspunkt des MUSIC-Verfahrens ist die Ähnlichkeit zwischen den Signalen unterschiedlicher Mikrofonpaare. Für jedes Mikrofonpaar wird für die Mikrofonsignale  $m_k$  und  $m_l$  das Kreuzleistungsdichtespektrum  $R_{m_k, m_l}$  ermittelt (vgl. Gleichung 2.38). Dieses wird innerhalb einer matriziellen Darstellung angeordnet und als  $\underline{C}$  bezeichnet:

$$\underline{C}(f) = \begin{pmatrix} R_{m_1, m_1}(f) & \cdots & R_{m_1, m_{n_m}}(f) \\ & \vdots & \\ R_{m_{n_m}, m_1}(f) & \cdots & R_{m_{n_m}, m_{n_m}}(f) \end{pmatrix}. \quad (2.46)$$

$\underline{C}$  wird folgendermaßen berechnet:

$$\underline{C} = \mathbb{E}[\underline{m} \cdot \underline{m}^H] = \mathbb{E}[\underline{A} \cdot \underline{s} \cdot \underline{s}^H \underline{A}^H] + \mathbb{E}[\underline{w} \cdot \underline{w}^H] + \mathbb{E}[\underline{A} \cdot \underline{s} \cdot \underline{w}^H] + \mathbb{E}[\underline{w} \cdot (\underline{A} \cdot \underline{s})^H]. \quad (2.47)$$

Hierbei entspricht  $\mathbb{E}[\cdot]$  dem Erwartungswert-Operator. Unter der Annahme, dass die Signalquellen unabhängig von dem Rauschen sind, gilt:

$$\mathbb{E}[\underline{w} \cdot (\underline{A} \cdot \underline{s})^H] = \mathbb{O}_{n_m \times n_d}, \quad \mathbb{E}[\underline{A} \cdot \underline{s} \cdot \underline{w}^H] = \mathbb{O}_{n_d \times n_m}. \quad (2.48)$$

Dabei entspricht  $\mathbb{O}$  der Nullmatrix. Falls weiterhin das Rauschen an den jeweiligen Mikrofonen voneinander unabhängig ist gilt:

$$\mathbb{E}[\underline{w} \cdot \underline{w}^H] = \sigma_{\mathcal{N}}^2 \cdot \mathbb{I}_{n_m}. \quad (2.49)$$

$\mathbb{I}_{n_m}$  bezeichnet die Einheitsmatrix mit Dimension  $n_m \times n_m$  und  $\sigma_{\mathcal{N}}^2$  eine Konstante, die von dem Rauschanteil abhängig ist. Somit vereinfacht sich die Berechnung von  $\underline{C}$  aus Gleichung (2.47) zu:

$$\underline{C} = \mathbb{E}[\underline{A} \cdot \underline{s} \cdot \underline{s}^H \cdot \underline{A}^H] + \sigma_{\mathcal{N}}^2 \mathbb{I}_{n_m} \quad (2.50)$$

$\underline{C}$  wird durch eine mathematische Eigenwertzerlegung (engl. Eigen Value Decomposition: EVD) folgendermaßen dargestellt:

$$\underline{C} = \underline{E} \cdot \underline{D} \cdot \underline{E}^{-1} \Leftrightarrow \underline{C} \cdot \underline{E} = \underline{E} \cdot \underline{D} \quad (2.51)$$

Dabei entsprechen  $\underline{\mathbf{E}} = \{\underline{\mathbf{e}}_1, \dots, \underline{\mathbf{e}}_{n_m}\}$  den Eigenvektoren und  $\underline{\mathbf{D}} = \text{diag}(\lambda_1, \dots, \lambda_{n_m})$  der Diagonalmatrix mit den Eigenwerten  $\lambda_i$  von  $\underline{\mathbf{C}}$ . Im Weiteren werden die Eigenwerte der Diagonalmatrix nach ihrer Größe aufsteigend sortiert:

$$\underline{\mathbf{C}} = (\underline{\mathbf{E}}_S | \underline{\mathbf{E}}_N) \left( \begin{array}{ccc|c} \lambda_1 + \sigma_N^2 & & 0 & \\ & \ddots & & \\ 0 & & \lambda_{n_d} + \sigma_N^2 & \\ \hline & \mathbb{O}_{(n_m-n_d) \times n_d} & & \sigma_N^2 \mathbb{I}_{(n_m-n_d)} \end{array} \right) (\underline{\mathbf{E}}_S | \underline{\mathbf{E}}_N)^{-1} \quad (2.52)$$

Im nächsten Schritt wird  $\underline{\mathbf{C}}$  in Anteile zerlegt, die mit dem Signal- und Rausraum assoziiert sind. Hierfür werden zunächst die Eigenwerte betrachtet. Falls diese den Wert  $\sigma_N^2$  übersteigen, enthalten sie Informationen über aktive Soundquellen. Im Folgenden bezeichnet  $n_\lambda$  die Anzahl von Eigenwerten, die diese Schwelle übersteigen. Im Anschluss lässt sich die Gesamtzahl von aktiven Soundquellen folgendermaßen schätzen:

$$n_d = n_m - n_\lambda \quad (2.53)$$

Weiterhin werden die Eigenvektoren  $\underline{\mathbf{E}}_S$ , die einen hohen Eigenwert besitzen, mit dem Signalraum assoziiert. Die restlichen Eigenvektoren  $\underline{\mathbf{E}}_N$  besitzen Eigenwerte mit dem Wert  $\sigma_N^2$  und werden mit dem Rausraum assoziiert.

Der Nutzen dieser Zerlegung liegt darin, dass die Eigenvektoren  $\underline{\mathbf{E}}_N$ , die mit dem Rausraum assoziiert sind, eine geometrische Besonderheit aufweisen. Sie sind orthogonal zu den Steering-Vektoren  $\underline{\mathbf{a}}$ , falls diese in die Richtung einer aktiven Soundquelle ausgerichtet sind. Dabei agiert  $\underline{\mathbf{E}}_N \cdot \underline{\mathbf{E}}_N^H$  als Projektionsoperator in den Rausraum:

$$\underline{\mathbf{A}}(f, \theta, \phi)^H \cdot \underline{\mathbf{E}}_N \cdot \underline{\mathbf{E}}_N^H \cdot \underline{\mathbf{A}}(f, \theta, \phi) = 0. \quad (2.54)$$

Hieraus ergibt sich das sogenannte Pseudospektrum  $h(f, \theta, \phi)$ :

$$h(f, \theta, \phi) = \frac{1}{\underline{\mathbf{A}}(f, \theta, \phi)^H \cdot \underline{\mathbf{E}}_N \cdot \underline{\mathbf{E}}_N^H \cdot \underline{\mathbf{A}}(f, \theta, \phi)}. \quad (2.55)$$

Schließlich kann durch Extraktion von Maximalwerten des Pseudospektrums eine Schätzung der Richtung zu einer Soundquelle durchgeführt werden. Dieser Ansatz wurde im Bereich der Robot Audition zuerst in [AAM99] durchgeführt.

Das Verfahren kann auf breitbandige Signale erweitert werden, indem das durchschnittliche Pseudo-Spektrum  $\bar{h}$  definiert wird. Dieses Spektrum berücksichtigt unterschiedliche Frequenzen:

$$\bar{h}(\theta, \phi) = \sum_{i=1}^{n_f} h(f_i, \theta, \phi). \quad (2.56)$$

Hierbei entspricht  $n_f$  der Gesamtzahl der verwendeten Frequenzen und  $f_i$  die Frequenz. Ein Einsatz ist in [Ish+09] gegeben.

Falls die Signalenergie im Vergleich zur Umgebung gering ist, sind die Ergebnisse des MUSIC-Verfahrens nicht anwendbar. Daher wurde GEVD-MUSIC<sup>9</sup> entwickelt [Nak+11]. Eine Untersuchung

<sup>9</sup> Generalized Eigen Value Decomposition-MUSIC

des GEVD-MUSIC-Verfahrens ist in [TK15] gegeben. Die Idee besteht darin, Informationen über das Rauschen  $\underline{w}$  zu berücksichtigen. Dabei wird entweder auf a-priori-Wissen zurückgegriffen oder Messdaten berücksichtigt, die nur Rauschanteile enthalten. Hierdurch wird das Kreuzleistungsdichtespektrum  $\underline{K}$  des Rauschens geschätzt:

$$\underline{K} = \mathbb{E}[\underline{w} \cdot \underline{w}^H]. \quad (2.57)$$

Mithilfe von  $\underline{K}$  kann der Einfluss des Rauschens verringert (engl. *Noise Whitening*) werden. Dabei wird ein verallgemeinertes Eigenwertproblem (engl. *Generalized Eigenvalue Problem* - GEVP) gelöst:

$$\underline{K}^{-1} \cdot \underline{C} \cdot \underline{E} = \underline{E} \cdot \underline{D} \Leftrightarrow \underline{C} \cdot \underline{E} = \underline{K} \cdot \underline{E} \cdot \underline{D} \quad (2.58)$$

Falls sich das Rauschverhalten ändert - z. B. in dynamischen Szenarien - versagt das GEVD-MUSIC-Verfahren. Daher wurde ein MUSIC-Verfahren (iGEVD-MUSIC) entwickelt, das inkrementell eine Schätzung des Rauschanteils durchführt. Einsatz findet sich in [Oku+12] unter Verwendung eines Quadrocopters.

Neben hochauflösenden Lokalisierungsergebnisse weisen die MUSIC-Verfahren einen hohen rechen-technischen Aufwand auf. Als Lösung wurde eine Zerlegung durch eine generalisierte Singulärwertzerlegung (engl. *Generalized Singular Value Decomposition* - GSVD) durchgeführt. Die Berechnungszeit des GSVD-MUSIC<sup>10</sup> konnte im Vergleich zum GEVD-MUSIC um den Faktor 3 erhöht werden [NNI12]. Einsatz des Verfahrens findet sich in [SBT20].

## 2.3 Zustandsschätzung und Kartierung

Der Schwerpunkt dieser Arbeit liegt darin, die Positionen von Soundquellen innerhalb einer Karte zu bestimmen. Hierfür wird zunächst in Abschnitt 2.3.1 vorgestellt, wie Karten im Kontext der mobilen Robotik repräsentiert und beschrieben werden können. Im Anschluss wird in Abschnitt 2.3.2 auf den Stand der Wissenschaft im Bereich der Soundquellenkartierung eingegangen. In Abschnitt 2.3.4 wird ein probabilistisches Verfahren zur Soundquellenkartierung betrachtet. Dieses dient in der Evaluation als Vergleichsverfahren zu dem in dieser Arbeit entwickelten intervall-basierten Verfahrens (IB-SSM). In Abschnitt 2.3.3 werden die Grundlagen für das Vergleichsverfahren vorgestellt. Dabei werden auf Verfahren der probabilistischen Zustandsschätzung eingegangen. In Kapitel 5 wird die Theorie wieder aufgegriffen, um die Richtungsschätzung von Soundquellen zu verbessern (vgl. Abschnitt 5.1.2).

### 2.3.1 Einführung

Durch verschiedene Sensoren kann die Umgebung erfasst und eine Karte erzeugt werden. Dabei kann die resultierende Karte topologisch oder metrisch repräsentiert werden [Thr+02].

Elemente einer topologischen Karte werden durch ihre Position und weiteren Eigenschaften (z. B. Ausrichtung, Temperatur, Ausdehnung) beschrieben. Dabei hängen die Eigenschaften von der Formulierung und der verwendeten Sensorik ab. Um beispielsweise ein autonomes Fahrzeug zu lokalisieren, sind markante Objekte in der Umgebung relevant.

<sup>10</sup> Generalized Singular Value Decomposition-MUSIC

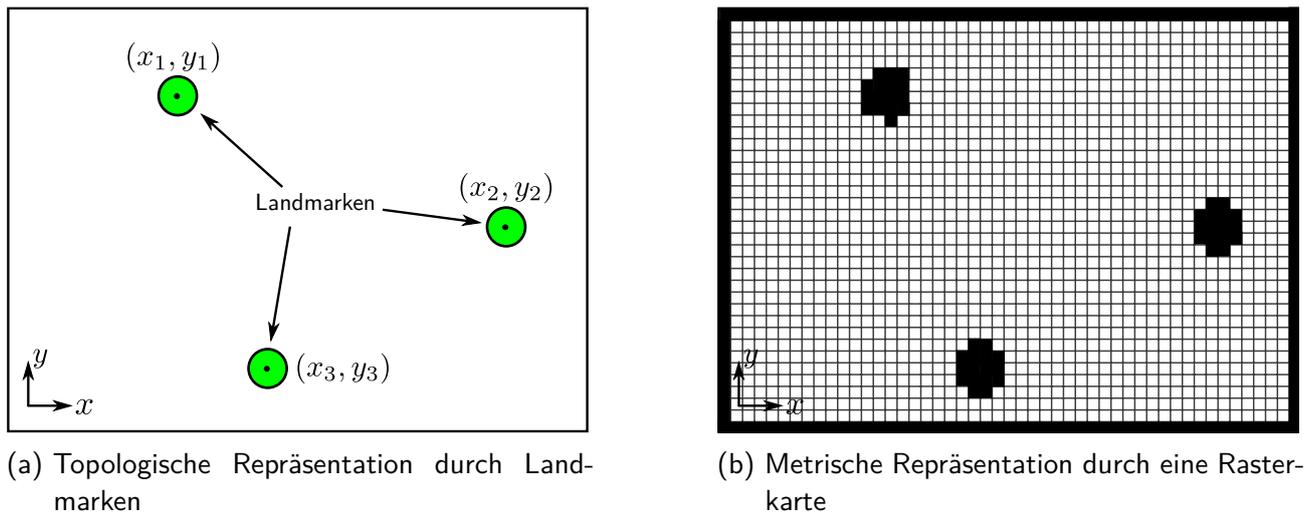


Abbildung 2.16: Repräsentation von Karten

Als Vorteil der topologischen Karte ergibt sich ein geringer Speicherplatz, der für diese Repräsentation notwendig ist. Jedoch können die reinen Sensordaten nicht direkt verwendet werden, um die Eigenschaften der Elemente zu ermitteln. Stattdessen sind Extraktionsverfahren notwendig (z. B. Kreisdetektion von Pfeiler aus den Laserdaten).

Im Gegensatz dazu stehen die metrischen Karten. Hierbei wird ein räumliches Gebiet durch ein Gitter approximiert. Es wird unterschieden in einer äquidistanten Repräsentation und einer Darstellung mit variabler Zellengröße. Jede Zelle des Gitters wird mit einem Skalar belegt. Um unterschiedliche Eigenschaften zu beschreiben, z. B. den Temperaturverlauf und die Gasverteilung, wird das Gitter um mehrere Ebenen erweitert. Jede Ebene (respektive Schicht) repräsentiert eine einzelne Eigenschaft. Im Kontext der Robotik werden Belegungs rasterkarten verwendet. Diese kennzeichnen, ob ein räumlicher Bereich in der Umgebung durch ein Hindernis belegt ist. Die Belegung wird durch eine Wahrscheinlichkeit beschrieben. Diese Repräsentation wurde zuerst von Moravec und Elfes vorgestellt [ME85]. Im Vergleich zur topologischen Darstellung benötigen metrische Karten einen hohen Speicherbedarf. Dieser ist von der Dimension des Raumes, als auch von der Auflösung der Karte abhängig. Als Vorteil sind jedoch keine zusätzlichen Extraktionsverfahren notwendig.

### 2.3.2 Soundquellenkartierung - Übersicht

Nachdem viel Forschung im Bereich der Soundquellenlokalisierung durchgeführt wurde, nimmt in den letzten Jahren das Interesse an der Positionsbestimmung von Soundquellen innerhalb einer Karte (engl. *Sound Source Mapping* - SSM) zu. Generell lassen sich existierende Verfahren des SSMs in zwei Kategorien einteilen:

- Basierend auf Strahlenverfolgung (engl. *ray tracing*)
- Basierend auf Lokalisierungs-Strategien

Bei Strahlenverfolgungsansätzen wird angenommen, dass der Ursprung eines akustischen Signals mit einem visuellen Merkmal korrespondiert. Dieses Merkmal wird von den Sensoren erfasst und ist assoziiert mit Elementen der metrischen Karte. Für die Repräsentation werden Belegungs rasterkarten verwendet (vgl. Abschnitt 2.3). Zunächst wird durch Verfahren der Soundquellenlokalisierung eine

Richtungsschätzung durchgeführt. Im Anschluss werden belegte Zelle ermittelt, die ausgehend von der Messposition einen Schnittpunkt mit der Richtungsschätzung besitzen. Der Wert der belegten Zelle wird mit einer probabilistischen Größe assoziiert. Diese charakterisiert die Wahrscheinlichkeit, dass die Zelle eine Soundquelle enthält. Als Vorteil dieses Ansatzes ergibt sich, dass eine einzelne akustische Messung genügt, um die Position einer Soundquelle in einer metrischen Karte zu ermitteln. Erste Arbeiten finden sich in [Kal+13]. Hierbei wird ein 2D-LiDAR für die Erzeugung der Belegungskarte verwendet (vgl. Abbildung 2.17a). Dieser Ansatz versagt jedoch, falls die Soundquelle nicht innerhalb der wahrnehmbaren Ebene des verwendeten LiDARs liegt. Um dieses Problem zu lösen, erweiterten Even et al. den Ansatz mit einem 3D-LiDAR [Eve+17a] (vgl. Abbildung 2.17b). Hierdurch entsteht jedoch ein größerer Rechenaufwand. Weiterhin muss zusätzliche Hardware integriert werden. Dies kann in vielen Fällen nicht auf bereits existierende Robotik-Systeme angewendet werden. Eine weitere Einschränkung des Verfahrens liegt in der Annahme, dass belegte Zellen keine akustischen Signale passieren lassen. Dies ist in der Realität jedoch nicht gegeben. So existieren akustisch transparente Materialien, die jedoch von LiDAR-Sensoren wahrgenommen werden. Als Folgerung werden zugehörige Zellen mit einer hohen Wahrscheinlichkeit assoziiert eine Soundquelle zu enthalten.

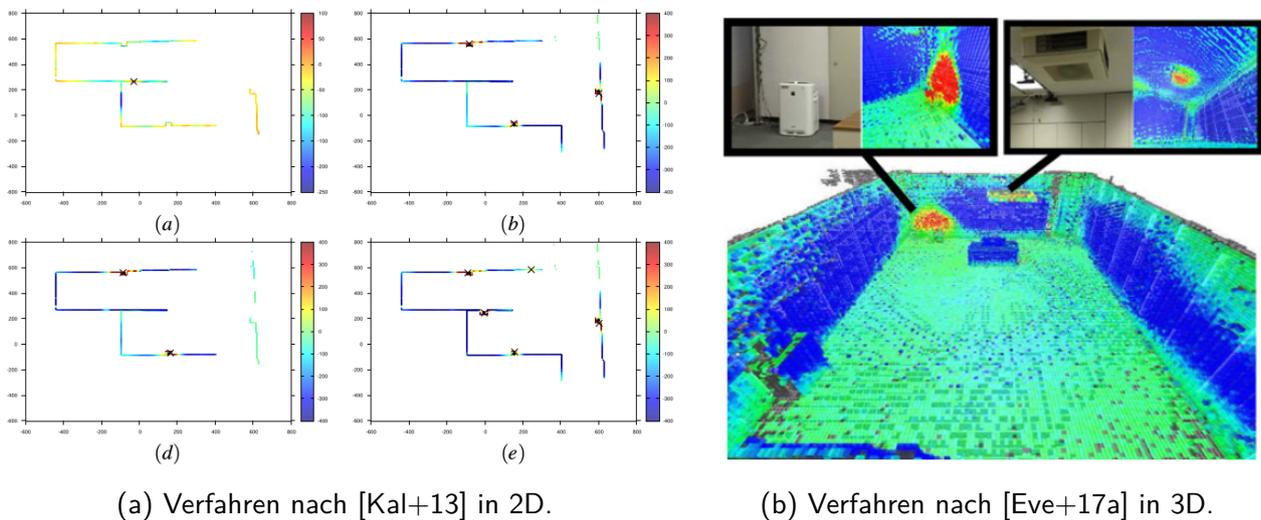
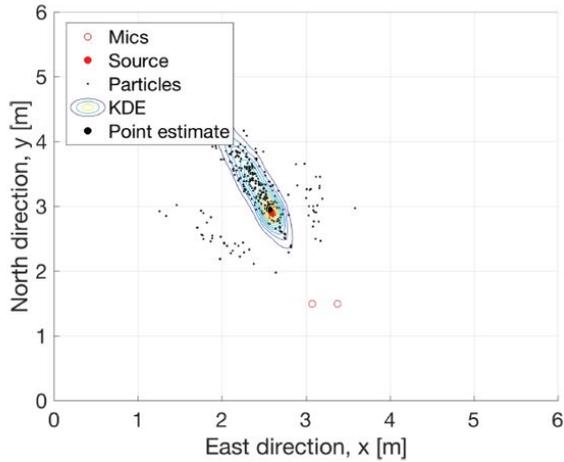


Abbildung 2.17: Soundquellenkartierung basierend auf Strahlenverfolgung.

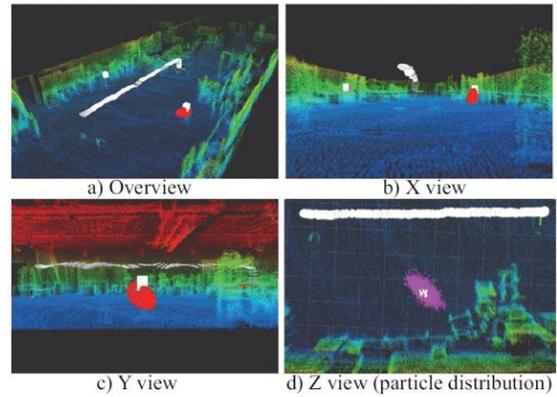
a): zweidimensional b): dreidimensional

Die zweite Kategorie basiert auf Lokalisierungs-Strategien wie Triangulation [Sas+10] und [Mic+20], FastSLAM [Hu+11], Monte-Carlo-Lokalisierung [SRT16], Mixture-Extended-Kalman-Filter [Ngu+16] und Partikel-Filter [Eve+17b]. Im Unterschied zu der ersten Kategorie sind akustische Messungen aus unterschiedlichen Richtungen notwendig. Um dieses Problem zu lösen, verwenden Su et al. einen dreistufigen Ansatz [Su+16], der einen Strahlungsverfolgungsansatz (engl. *ray casting*) mit einem Triangulationsansatz von Sasaki et al. kombiniert [Sas+10]. Für diesen Ansatz wird angenommen, dass die Mehrzahl der Schnittpunkten von den unterschiedlichen Richtungsschätzungen nah an der tatsächlichen Position liegen. Dies ist jedoch in der Praxis nicht gegeben.

Insgesamt adressieren die beiden Kategorien nicht die Problematik der geometrischen Mehrdeutigkeiten (vgl. Abschnitt 1.2). Als Ausnahme sei das Verfahren von [Ngu+16] zu nennen. Meist wird durch eine zusätzliche Annahme eingeschränkt, dass die Soundquelle vor dem mobilen Roboter platziert ist.



(a) Verfahren nach [Eve+17b]



(b) Verfahren nach [SRT16]

Abbildung 2.18: Soundquellenkartierung basierend auf Lokalisierungs-Strategien.

a): Partikelfilter b) Monte-Carlo-Lokalisierung

Weiterhin werden die Unsicherheiten der Mikrofonpositionen und der TDoA nicht vollständig in Betracht gezogen. Beispielsweise modellieren Sasaki et al. die Unsicherheit einer Richtungsmessung durch eine Normalverteilung (mit Erwartungswert 0) [SRT16]. Jedoch treten durch u. a. inakurate Kenntnis über die Mikrofonpositionen systematische Fehler auf. Daher ist die Annahme nicht zutreffend.

### 2.3.3 Probabilistische Zustandsschätzung

Im weiteren Verlauf dieser Arbeit werden Schätzverfahren verwendet, um den Zustand  $\mathbf{x}^k$  eines Systems zu ermitteln. In der Literatur wird das folgende Vorgehen als Bayes-Filter bezeichnet. Der Zustand  $\mathbf{x}^k$  wird zum diskreten Zeitschritt  $k$  beobachtet. Hierfür werden die Beobachtungen  $\mathbf{z}^k$  und die Informationen des Steuerungseingriffs (engl. *control*)  $\mathbf{c}^k$  verwendet. Gesucht ist die Wahrscheinlichkeit  $p(\cdot)$  für den Zustand  $\mathbf{x}^k$  nachdem die Eingriffe  $\mathbf{c}^{1:k}$  durchgeführt und die Beobachtungen  $\mathbf{z}^{1:k}$  erfasst wurden. Diese Wahrscheinlichkeit wird als Belief  $\text{bel}(\mathbf{x}^k)$  des Zustands  $\mathbf{x}^k$  bezeichnet [Thr02]:

$$\text{bel}(\mathbf{x}^k) = p(\mathbf{x}^k | \mathbf{z}^{1:k}, \mathbf{c}^{1:k}). \quad (2.59)$$

Es wird angenommen, dass der Zustand  $\mathbf{x}^k$  sämtliche Informationen der Vergangenheit enthält (Markov-Annahme). Somit kann der Belief  $\text{bel}(\mathbf{x}^k)$  durch die folgende Rechenvorschrift beschrieben werden:

$$\text{bel}(\mathbf{x}^k) = \eta \cdot \overbrace{p(\mathbf{z}^k | \mathbf{x}^k)}^{\text{Sensormodell}} \cdot \underbrace{\int \overbrace{p(\mathbf{x}^k | \mathbf{x}^{k-1}, \mathbf{u}^k)}^{\text{Bewegungsmodell}} \cdot \text{bel}(\mathbf{x}^{k-1}) d\mathbf{x}^{k-1}}_{\text{bel}(\mathbf{x}^k)}. \quad (2.60)$$

Dabei wird der Belief  $\text{bel}(\mathbf{x}^{k-1})$  zum vorherigen Zeitschritt  $k - 1$  verwendet. Dies impliziert, dass eine initiale Schätzung  $\text{bel}(\mathbf{x}^0)$  für den Startzustand vorliegen muss.

Um den Belief berechnen zu können, wird ein Sensor- und ein Bewegungsmodell benötigt. Das Bewegungsmodell (engl. *motion model*)  $p(\mathbf{x}^k | \mathbf{x}^{k-1}, \mathbf{c}^k)$  beschreibt den Übergang des Zustands  $\mathbf{x}^{k-1}$

in den Zustand  $\mathbf{x}^k$  nachdem die Steuereingabe  $\mathbf{c}^k$  durchgeführt wurde. Das Sensormodell - respektive Beobachtungsmodell -  $p(\mathbf{z}^k|\mathbf{x}^k)$  beschreibt den probabilistischen Zusammenhang der Erfassung einer Messung  $\mathbf{z}^k$  unter der Annahme, dass der Zustand  $\mathbf{x}^k$  gegeben ist.  $\eta$  bezeichnet eine Skalierungskonstante zur Gewährleistung der Normierungseigenschaft einer Wahrscheinlichkeitsfunktion. Die Gleichung (2.60) wird in zwei Schritten berechnet:

$$\text{Prädiktion: } \overline{\text{bel}}(\mathbf{x}^k) = \int p(\mathbf{x}^k|\mathbf{x}^{k-1}, \mathbf{c}^k) \cdot \text{bel}(\mathbf{x}^{k-1}) d\mathbf{x}^{k-1}. \quad (2.61)$$

$$\text{Korrektur: } \text{bel}(\mathbf{x}^k) = \eta \cdot p(\mathbf{z}^k|\mathbf{x}^k) \cdot \overline{\text{bel}}(\mathbf{x}^k). \quad (2.62)$$

Im ersten Schritt - der Prädiktion - wird in der Chapman-Kolmogorov-Gleichung (2.61) der prädiizierte Zustand  $\overline{\text{bel}}(\mathbf{x}^k)$  berechnet, indem das Bewegungsmodell und der vorherige Belief  $\text{bel}(\mathbf{x}^{k-1})$  verwendet wird.

Im zweiten Schritt - der Korrektur - wird in Gleichung (2.62) der Belief  $\text{bel}(\mathbf{x}^k)$  zum aktuellen Zeitschritt  $k$  mithilfe des prädiizierten Zustands  $\overline{\text{bel}}(\mathbf{x}^k)$  und dem Sensormodell berechnet.

Um die beiden Schritte durchführen zu können, müssen für das Beobachtungsmodell, das Sensormodell und für den initialen Zustand  $\text{bel}(\mathbf{x}^0)$  eine geeignete Repräsentation gewählt werden. In einer Vielzahl von Problemstellungen wird als Repräsentation die Normalverteilung gewählt. Dies führt zu dem sogenannten Kalman-Filter [Kal60]. Dieser Ansatz setzt eine lineare Darstellung des Beobachtungs- und des Sensormodells voraus. In der Realität sind die Zusammenhänge jedoch in den meisten Fällen nichtlinear. Zur Lösung wird daher entweder eine Linearisierung durchgeführt (EKF - engl. *Extended Kalman Filter*), oder alternativ wird der nichtlineare Zusammenhang betrachtet und die Ergebnisse durch eine Normalverteilung approximiert. Dies führt zum *Unscented Kalman-Filter* (UKF) [JUD95].

Im weiteren Verlauf dieser Dissertation wird der Belief  $\text{bel}(\mathbf{x}^k)$  an verschiedenen Stellen berechnet. Einerseits wird in Abschnitt 2.3.4 die Position von Soundquellen durch ein probabilistisches Verfahren geschätzt. Dieses Verfahren dient im weiteren Verlauf als Vergleich zu dem in der Dissertation entwickelten intervall-basierten Verfahren. Andererseits wird die rekursive Zustandsschätzung in Abschnitt 5.1.2 verwendet, um die Richtung zu unterschiedlichen Soundquellen zu ermitteln.

Im Folgenden wird in Abschnitt 2.3.3.1 der erweiterte Kalman-Filter betrachtet und im Anschluss auf die Soundquellenkartierung angewendet.

### 2.3.3.1 Erweiterter Kalman-Filter

Das Ziel ist den Belief  $\text{bel}(\mathbf{x}^k)$  aus Gleichung (2.60) durch ein rekursives Verfahren zu schätzen. Hierfür wird im Folgenden der erweiterte Kalman-Filter vorgestellt.

In dieser Arbeit werden mehrere zeitgleich aktive Soundquellen berücksichtigt. Daher wird die Notation der jeweiligen Zustände  $\mathbf{x}_i^k$  um den Index  $i$  erweitert und mit der jeweiligen Soundquelle  $s_i$  assoziiert.

Für das Bewegungsmodell, das Sensormodell und den Belief  $\text{bel}(\mathbf{x}_i^k)$  aus Gleichung (2.60) wird die Normalverteilung als Repräsentation gewählt. Dabei wird der aktuelle Zustand  $\mathbf{x}_i^k$  durch den Erwartungswert  $\boldsymbol{\mu}_i^k$  und der Kovarianz  $\boldsymbol{\Sigma}_i^k$  repräsentiert. Bevor der Belief berechnet wird, muss eine initiale Schätzung für  $\boldsymbol{\mu}_i^0$  und  $\boldsymbol{\Sigma}_i^0$  vorliegen. Im Anschluss wird für jeden Zeitschritt  $k$  die Prädiktion und Korrektur berechnet.

### Prädiktion

Das nichtlineare Bewegungsmodell  $\mathbf{g}(\mathbf{c}^k, \mathbf{x}_i^{k-1})$  berechnet aus dem vorherigen Zustand  $\mathbf{x}_i^{k-1}$  und der Steuerungseingabe  $\mathbf{c}^k$  einen prädizierten Zustand  $\bar{\mathbf{x}}_i^k$ :

$$\bar{\mathbf{x}}_i^k = \mathbf{g}(\mathbf{c}^k, \mathbf{x}_i^{k-1}) + \boldsymbol{\varepsilon}^k \quad \text{mit} \quad \boldsymbol{\varepsilon}^k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}^k). \quad (2.63)$$

Hierbei entspricht  $\boldsymbol{\varepsilon}^k$  einem gaußförmigen Fehlermodell - bezeichnet als Prozessrauschen - mit Erwartungswert null und der Kovarianz  $\mathbf{R}^k$ . In dieser Arbeit wird die Steuerungseingabe  $\mathbf{c}^k$  als unbekannt angenommen, wodurch sich die vorherige Gleichung folgendermaßen vereinfacht:

$$\bar{\mathbf{x}}_i^k = \mathbf{g}(\mathbf{x}_i^{k-1}) + \boldsymbol{\varepsilon}^k. \quad (2.64)$$

Da ein nichtlinearer Zusammenhang angenommen wird, ist der prädizierte Zustand nicht normalverteilt. Dies widerspricht der vorherigen Annahme. Daher wird Gleichung (2.64) an der Stelle des Erwartungswertes  $\boldsymbol{\mu}_i^{k-1}$  durch die ersten zwei Summanden der Taylor-Reihenentwicklung linearisiert:

$$\mathbf{g}(\mathbf{x}_i^{k-1}) \approx \mathbf{g}(\boldsymbol{\mu}_i^{k-1}) + \underbrace{\frac{\partial \mathbf{g}(\mathbf{x}_i^{k-1})}{\partial \mathbf{x}_i^{k-1}} \Big|_{\boldsymbol{\mu}_i^{k-1}}}_{A^k} \cdot (\mathbf{x}_i^{k-1} - \boldsymbol{\mu}_i^{k-1}). \quad (2.65)$$

Hierbei entspricht  $A^k$  der Jacobi-Matrix des Bewegungsmodells. Im Folgenden sei angenommen, dass  $A^k$  für alle Zustände  $s_i$  identisch ist. Der Erwartungswert des prädizierten Zustands ergibt sich zu:

$$\bar{\boldsymbol{\mu}}_i^k = \mathbf{g}(\boldsymbol{\mu}_i^{k-1}). \quad (2.66)$$

Weiterhin wird die prädizierte Kovarianz  $\bar{\boldsymbol{\Sigma}}_i^k$  zum Zeitpunkt  $k$  geschätzt, indem die Varianzfortpflanzung durchgeführt wird:

$$\bar{\boldsymbol{\Sigma}}_i^k = \mathbf{A}^k \boldsymbol{\Sigma}_i^{k-1} (\mathbf{A}^k)^T + \mathbf{R}^k. \quad (2.67)$$

### Korrektur

Im zweiten Schritt wird eine Korrektur des prädizierten Zustands durchgeführt, indem ein Vergleich mit der Beobachtung  $\mathbf{z}_i^k$  erfolgt (vgl. Gleichung (2.62)). Hierbei wird ein nichtlineares Sensormodell  $\mathbf{h}(\mathbf{x}_i^{k-1})$  betrachtet, das den aktuellen Zustand  $\mathbf{x}_i^k$  in den Beobachtungsraum abbildet. Es gilt:

$$\mathbf{z}_i^k = \mathbf{h}(\mathbf{x}_i^k) + \boldsymbol{\delta}^k \quad \text{mit} \quad \boldsymbol{\delta}^k \sim \mathcal{N}(0, \mathbf{Q}^k). \quad (2.68)$$

$\boldsymbol{\delta}^k$  entspricht einem gaußförmigen Fehlermodell (bezeichnet als Messrauschen) mit Erwartungswert null und der Kovarianz  $\mathbf{Q}^k$  zum Zeitpunkt  $k$ .

Wie für das Bewegungsmodell im Schritt der Prädiktion wird das Sensormodell durch eine Taylor-Reihenentwicklung linearisiert. Die Linearisierung erfolgt an der Stelle des Erwartungswertes des prädizierten Zustands:

$$\mathbf{h}(\mathbf{x}_i^k) \approx \mathbf{h}(\bar{\boldsymbol{\mu}}_i^k) + \underbrace{\frac{\partial \mathbf{h}(\mathbf{x}_i^k)}{\partial \mathbf{x}_i^k} \Big|_{\bar{\boldsymbol{\mu}}_i^k}}_{H^k} \cdot (\mathbf{x}_i^k - \bar{\boldsymbol{\mu}}_i^k). \quad (2.69)$$

$H^k$  entspricht der Jacobi-Matrix des Sensormodells. Es wird angenommen, dass  $H^k$  für alle Soundquellen  $s_i$  identisch ist.

Zunächst wird der sogenannte Kalman-Gain  $\mathbf{K}_i^k$  berechnet. Dieser beschreibt, wie sich der Einfluss vom dem prädizierten Zustand und der Beobachtung auf das Ergebnis auswirkt. Je nach Wahl des Kalman-Gains kann das Ergebnis entweder ausschließlich von dem prädizierten Zustand, ausschließlich von der Beobachtung oder von der Kombination beider Größen abhängen.  $\mathbf{K}_i^k$  wird folgendermaßen berechnet:

$$\mathbf{K}_i^k = \bar{\Sigma}_i^k (\mathbf{H}^k)^T \left( \mathbf{H}^k \bar{\Sigma}_i^k (\mathbf{H}^k)^T + \mathbf{Q}^k \right)^{-1}. \quad (2.70)$$

Darauffolgend werden der Mittelwert  $\mu_i^k$  und die Kovarianz  $\Sigma_i^k$  zum aktuellen Zeitpunkt  $k$  berechnet:

$$\mu_i^k = \bar{\mu}_i^k + \mathbf{K}_i^k \left( z_i^k - h(\bar{\mu}_i^k) \right) \quad (2.71)$$

$$\Sigma_i^k = \bar{\Sigma}_i^k - \mathbf{K}_i^k \mathbf{H}^k \bar{\Sigma}_i^k \quad (2.72)$$

Abschließend ist das gesamte Schema in Abbildung 2.19 dargestellt.

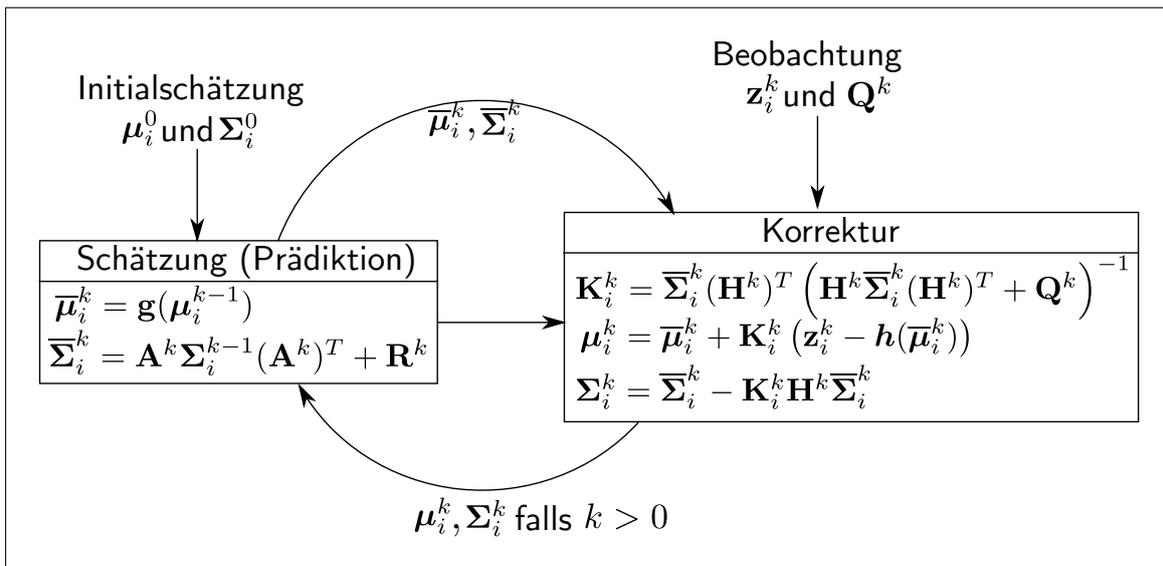


Abbildung 2.19: Schematische Darstellung des erweiterten Kalman-Filters ohne Behandlung von Steuerungseingaben.

### 2.3.4 Probabilistische Soundquellenkartierung

Im Folgenden wird ein probabilistisches Verfahren vorgestellt, das den erweiterten Kalman-Filter für die Soundquellenkartierung verwendet. Das vorgestellte Verfahren erweitert die Arbeiten von [Ngu+16]. Die Autoren beschränken sich auf die zweidimensionale Positionsschätzung für maximal zwei Soundquellen. In dieser Dissertation wird das Verfahren auf die dreidimensionale Schätzung erweitert. Dabei wird eine beliebige Anzahl von Soundquellen berücksichtigt. Das erweiterte Verfahren dient in Kapitel 6 zum Vergleich mit dem intervall-basierten Ansatz dieser Dissertation.

Zunächst werden keine geometrischen Mehrdeutigkeiten (vgl. Abschnitt 1.2) bei der Richtungsschätzung berücksichtigt. Im Anschluss wird der Verfahren auf die Mehrdeutigkeiten in Abschnitt 2.3.4.4 erweitert.

### 2.3.4.1 Modellierung

Im Weiteren wird der erweiterte Kalman-Filter auf die Schätzung der dreidimensionalen Position von mehreren statischen Soundquellen angewendet. Die Positionen der Soundquellen werden durch den Zustandsraum  $\mathbf{x}_i^k \in \mathbb{R}^3$  beschrieben. Hierbei bezeichnet  $i$  den Index einer einzelnen Soundquelle  $s_i$  und  $k$  den diskreten Zeitpunkt.  $\mathbf{x}_i^k$  ist in Bezug zu dem globalen Koordinatensystem  $\text{KS}_G$  formuliert (vgl. Abbildung 2.20):

$$\mathbf{x}_i^k = [x_i^k \quad y_i^k \quad z_i^k]^T. \quad (2.73)$$

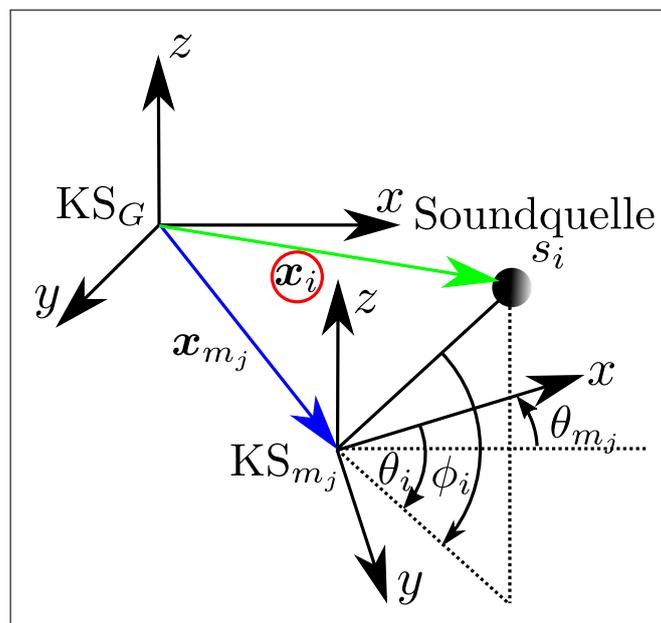


Abbildung 2.20: Bezogen auf ein globales Koordinatensystem  $\text{KS}_G$  wird die Position  $\mathbf{x}_i^k$  einer Soundquelle  $s_i$  gesucht. An Messposition  $\mathbf{x}_{m_j}^k$  werden die Winkel  $(\theta_i^k, \phi_i^k)$  beobachtet. Diese sind auf das Koordinatensystem  $\text{KS}_{m_j}$  bezogen. Zur vereinfachten Darstellung wird der Index  $k$  nicht berücksichtigt.

Für die Zustandsschätzung werden die Prädiktion und Korrektur erläutert.

#### Prädiktion

Die jeweiligen Soundquellen werden als statisch angenommen, wodurch die Funktion  $g$  des Bewegungsmodells durch die Identitätsabbildung dargestellt werden kann. Somit gilt für den Erwartungswert  $\bar{\boldsymbol{\mu}}_i^k$  des prädierten Zustands:

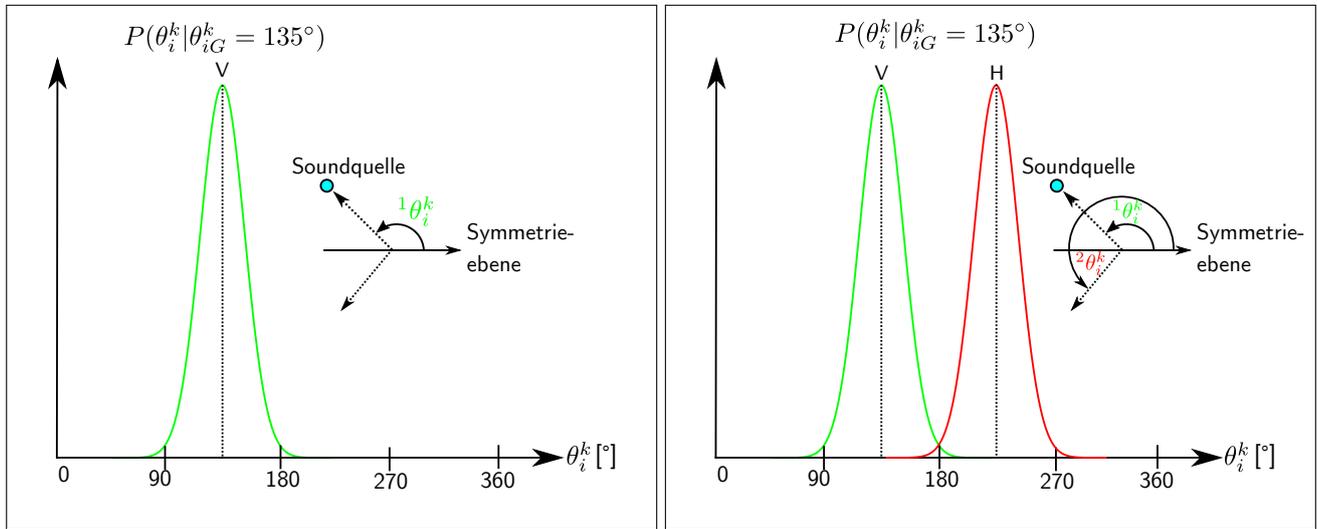
$$\bar{\boldsymbol{\mu}}_i^k = \boldsymbol{\mu}_i^{k-1}. \quad (2.74)$$

Die Kovarianz  $\bar{\boldsymbol{\Sigma}}_i^k$  zum Zeitpunkt  $k$  wird mittels Gleichung (2.67) berechnet.

### Korrektur

Der Zustandsraum  $\mathbf{x}_i^k$  kann nicht direkt erfasst werden. Stattdessen wird die Richtung  $\mathbf{z}_i^k = (\theta_i^k, \phi_i^k)$  von der Messposition  $\mathbf{x}_{m_j}^k = [x_{m_j}^k \ y_{m_j}^k \ z_{m_j}^k]^T$  beobachtet. Dabei bezeichnet  $\theta_i^k$  den horizontalen Winkel (Azimut) und  $\phi_i^k$  den vertikalen Winkel (Elevation) zum Zeitpunkt  $k$  zur Soundquelle  $s_i$  (vgl. Abbildung 2.20). Die Richtung ist bezogen auf das Koordinatensystem  $\text{KS}_{m_j}$  und mit der jeweiligen Messposition assoziiert. Weiterhin wird angenommen, dass sich die Orientierung der beiden Koordinatensysteme nur durch eine Rotation um die z-Achse unterscheidet. Die Rotation wird mit dem Winkel  $\theta_{m_j}$  beschrieben.

In diesem Abschnitt wird angenommen, dass die Richtung durch eine unimodale Normalverteilung dargestellt wird (vgl. Abbildung 2.21a). Unter dieser Annahme können jedoch keine geometrischen Mehrdeutigkeiten (vgl. Vorne-Hinten-Mehrdeutigkeit) berücksichtigt werden. Aus diesem Grund wird in Abschnitt 2.3.4.2 eine multimodale Gaußverteilung vorgestellt (vgl. Abbildung 2.21b).



(a) Keine geometrischen Mehrdeutigkeiten angenommen. (b) Geometrische Mehrdeutigkeit der Richtungsmessung angenommen.

Abbildung 2.21: Beobachtungsmodelle exemplarisch dargestellt für den horizontalen Winkel  $\theta_i$  mit dem tatsächlichen Winkel  $\theta_{iG}^k$ . a) Keine Berücksichtigung b) Berücksichtigung von geometrischen Mehrdeutigkeiten (vgl. Abschnitt 2.3.4.2)

Die Beziehung zwischen dem Zustandsraum  $\mathbf{x}_i^k$  im globalen Koordinatensystem und der Beobachtung  $\mathbf{z}_i^k$  im relativen Koordinatensystem wird durch die nichtlineare Funktion  $\mathbf{h}^k$  beschrieben:

$$\mathbf{z}_i^k = \mathbf{h}^k(\mathbf{x}_i^k) = [h_1^k(\mathbf{x}_i^k) \ h_2^k(\mathbf{x}_i^k)]^T \quad (2.75)$$

Hierbei wird der Zustandsraum auf die jeweiligen Winkel  $\theta_i^k$  und  $\phi_i^k$  projiziert.

$$\theta_i^k = h_1^k(\mathbf{x}_i^k) = \arctan\left(\frac{y_i^k - y_{m_j}^k}{x_i^k - x_{m_j}^k}\right) + \theta_{m_j}^k \quad (2.76)$$

$$\phi_i^k = h_2^k(\mathbf{x}_i^k) = \arctan\left(\frac{\sqrt{(x_i^k - x_{m_j}^k)^2 + (y_i^k - y_{m_j}^k)^2}}{z_i^k - z_{m_j}^k}\right). \quad (2.77)$$

Wie in Abschnitt 2.3.3.1 geschildert, muss die Funktion  $\mathbf{h}^k$  linearisiert werden. Hierfür wird die Jacobi-Matrix  $\mathbf{H}_k$  aufgestellt:

$$\mathbf{H}_k = \begin{pmatrix} \frac{\partial h_1^k}{\partial x_i} & \frac{\partial h_1^k}{\partial y_i} & \frac{\partial h_1^k}{\partial z_i} \\ \frac{\partial h_2^k}{\partial x_i} & \frac{\partial h_2^k}{\partial y_i} & \frac{\partial h_2^k}{\partial z_i} \end{pmatrix}. \quad (2.78)$$

Die jeweiligen Anteile ergeben sich folgendermaßen:

$$\begin{aligned} \frac{\partial h_1^k}{\partial x_i} &= -\frac{y_i - y_{m_j}}{(x_i - x_{m_j})^2 + (y_i - y_{m_j})^2} \\ \frac{\partial h_1^k}{\partial y_i} &= \frac{x_i - x_{m_j}}{(x_i - x_{m_j})^2 + (y_i - y_{m_j})^2} \\ \frac{\partial h_1^k}{\partial z_i} &= 0 \\ \frac{\partial h_2^k}{\partial x_i} &= -\frac{(x_{m_j} - x_i) \cdot (z_i - z_{m_j})}{\sqrt{(x_i - x_{m_j})^2 + (y_i - y_{m_j})^2} \cdot \sqrt{(x_i - x_{m_j})^2 + (y_i - y_{m_j})^2 + (z_i - z_{m_j})^2}} \\ \frac{\partial h_2^k}{\partial y_i} &= -\frac{(y_{m_j} - y_i) \cdot (z_i - z_{m_j})}{\sqrt{(x_i - x_{m_j})^2 + (y_i - y_{m_j})^2} \cdot \sqrt{(x_i - x_{m_j})^2 + (y_i - y_{m_j})^2 + (z_i - z_{m_j})^2}} \\ \frac{\partial h_2^k}{\partial z_i} &= -\frac{\sqrt{(x_i - x_{m_j})^2 + (y_i - y_{m_j})^2}}{\sqrt{(x_i - x_{m_j})^2 + (y_i - y_{m_j})^2 + (z_i - z_{m_j})^2}}. \end{aligned} \quad (2.79)$$

Es sei angenommen, dass die Beobachtungen voneinander unkorreliert sind und die Varianz jeder Komponente  $\sigma_Q^2$  entspricht. Somit ergibt sich die Kovarianzmatrix  $\mathbf{Q}^k$  zu:

$$\mathbf{Q}^k = \begin{pmatrix} \sigma_Q^2 & 0 \\ 0 & \sigma_Q^2 \end{pmatrix}. \quad (2.80)$$

Im Anschluss wird der Erwartungswert  $\boldsymbol{\mu}_i^k$  und die Kovarianz  $\boldsymbol{\Sigma}_i^k$  geschätzt, indem die Gleichungen (2.71) und (2.72) verwendet werden.

### 2.3.4.2 Berücksichtigung von geometrischer Mehrdeutigkeit

Bisher werden in dem Kalman-Verfahren keine geometrischen Mehrdeutigkeiten berücksichtigt. Daher wird das Verfahren im Weiteren erweitert. Es wird angenommen, dass die Symmetrie-Ebene der Mikrofonanordnung bekannt ist. Die Messung  $\mathbf{z}_i^k$  wird daher entweder oberhalb oder unterhalb dieser Ebene wahrgenommen (vgl. Abbildung 2.21b).

Um diesen Effekt zu modellieren, wird das Sensormodell  $P(\mathbf{z}_i^k | \mathbf{x}_i^k)$  angepasst. Es ergibt sich, indem zwei Normalverteilungen linear überlagert werden:

$$P(\mathbf{z}_i^k | \mathbf{x}_i^k) = \sum_{j=1}^2 \frac{1}{2} \mathcal{N}({}^j \mathbf{h}(\mathbf{x}_i^k), \mathbf{Q}^k). \quad (2.81)$$

Hierbei bezeichnet der Index  $j = 1$  die Projektion in die Richtung der Vorderseite und  $j=2$  in die Richtung der Rückseite.

Es sei zunächst eine einzelne initiale Zustandsschätzung gegeben. Nachdem eine Messung durchgeführt wurde, ergeben sich jeweils zwei Hypothesen. Entweder ist die Messung der Soundquelle mit der Vorderrichtung (V) oder der Hinterrichtung (H) assoziiert. Im darauffolgenden Zeitschritt ergeben sich für diese zwei Hypothesen insgesamt vier Hypothesen.

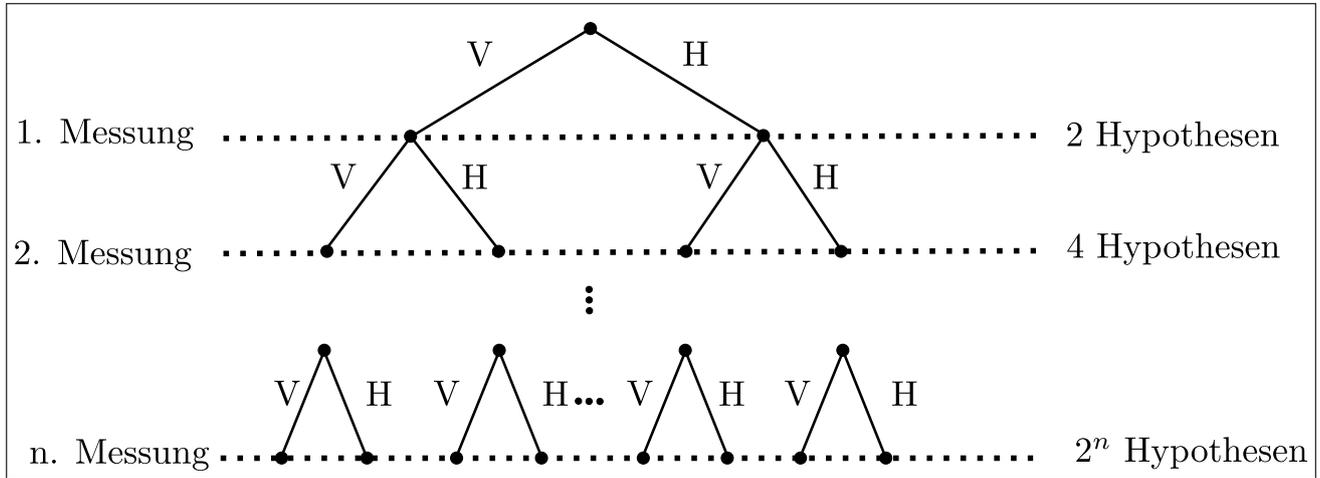


Abbildung 2.22: Kombination der unterschiedlichen Hypothesen für eine Soundquelle. V: Vorne, H: Hinten

Es wird angekommen, dass im Zeitschritt  $k$  insgesamt  $n_h^k$  Hypothesen vorliegen. Somit wird der Belief  $\text{bel}(\mathbf{x}_i^k)$  an der Position der Soundquelle  $s_i$  aus der gewichtigen Summe mehrerer Normalverteilungen dargestellt. Dieses Modell wird als Gaußsches Mischmodell (engl. *Gaussian Mixture Model*, kurz GMM) bezeichnet [Rey09]:

$$\text{bel}(\mathbf{x}_i^k) = p(\mathbf{x}_i^k | \mathbf{z}^{1:k}) = \sum_{h=1}^{n_h^k} h \omega_i^k \mathcal{N}(h \boldsymbol{\mu}_i^k, h \boldsymbol{\Sigma}_i^k). \tag{2.82}$$

Hierbei bezeichnen  $h \boldsymbol{\mu}^k$  den Mittelwert und  $h \boldsymbol{\Sigma}^k$  die Kovarianz der  $h$ -ten Hypothese zum Zeitschritt  $k$ .  $h \omega^k$  beschreibt das Gewicht der jeweiligen Hypothese  $h$  in Bezug auf die Gesamtheit aller Hypothesen zum Zeitpunkt  $k$ . Insgesamt summieren sich die Gewichte  $h \omega_i^k$  zu eins:

$$\sum_{h=1}^{n_h^k} h \omega_i^k = 1. \tag{2.83}$$

Im Folgenden werden die Prädiktion und Korrektur betrachtet. Zunächst muss eine initiale Schätzung  $\text{bel}(\mathbf{x}_i^0)$  vorliegen. Im Gegensatz zu dem vorgestellten EKF-Ansatz in Abschnitt 2.3.4.1 können mehrere initiale Hypothesen  $n_h^0$  vorgegeben und durch ein Gaußsches Mischmodell beschrieben werden:

$$\text{bel}(\mathbf{x}_i^0) = \sum_{h=1}^{n_h^0} h \omega_i^0 \mathcal{N}(h \boldsymbol{\mu}_i^0, h \boldsymbol{\Sigma}_i^0). \tag{2.84}$$

**Prädiktion**

Zunächst wird der prädizierte Erwartungswert  $h \bar{\boldsymbol{\mu}}_i^k$  und die Kovarianz  $h \bar{\boldsymbol{\Sigma}}_i^k$  für den prädizierten Zustand für jede Hypothese  $h$  separat wie im Prädiktionsschritt des EKF-Ansatzes berechnet. Es

wird angenommen, dass der vorherige Zustand  $n_h^{k-1}$  Hypothesen besitzt. Die prädierten Gewichte  $h\bar{\omega}_i^k$  bleiben unverändert:

$$h\bar{\omega}_i^k = h\omega_i^k. \quad (2.85)$$

Somit gilt für den prädierten Zustand:

$$\text{bel}(\bar{\mathbf{x}}_i^k) = \sum_{h=1}^{n_h^{k-1}} h\bar{\omega}_i^k \mathcal{N}(h\bar{\boldsymbol{\mu}}_i^k, h\bar{\boldsymbol{\Sigma}}_i^k). \quad (2.86)$$

### Korrektur

Das Beobachtungsmodell besteht aus einem Gaußschen Mischmodell mit zwei Komponenten. Daraus resultieren für jede Hypothese  $h$  jeweils zwei weitere Hypothesen. Entweder wird die Soundquelle von der Vorder- oder der Rückseite erfasst. Für jede prädierte Hypothese  $h$  wird zunächst separat die Wahrscheinlichkeit  $h,j\lambda_i$  berechnet, dass diese von der jeweiligen Messung  $\mathbf{z}_i^k$  erzeugt wurde. Dabei wird der prädierte Zustand der Hypothese  $h$  mit Hilfe des nichtlinearen Beobachtungsmodells  ${}^j\mathbf{h}$  in den Beobachtungsraum projiziert ( $j = 1$ : Vorne und  $j=2$ : Hinten). Die projizierte Richtung wird im Anschluss mit der jeweiligen Messung  $\mathbf{z}_i^k$  verglichen. Dabei wird die resultierende Kovarianz durch die Varianzfortpflanzung und dem Messrauschen  $\mathbf{Q}^k$  beschrieben. Die Wahrscheinlichkeit  $h,j\lambda_i$  ergibt sich zu:

$$\begin{aligned} h,j\lambda_i &= \mathcal{N}(\mathbf{z}_i^k | \mathbf{h}(h\bar{\boldsymbol{\mu}}_i^k), \mathbf{H}^k \cdot h\bar{\boldsymbol{\Sigma}}_i^k \cdot (\mathbf{H}^k)^T + \mathbf{Q}^k) \\ &= \frac{1}{(2\pi)^d \sqrt{|\mathbf{H}^k \cdot h\bar{\boldsymbol{\Sigma}}_i^k \cdot (\mathbf{H}^k)^T + \mathbf{Q}^k|}} e^{-\frac{1}{2} [\mathbf{z}_i^k - {}^j\mathbf{h}(h\bar{\boldsymbol{\mu}}_i^k)]^T [\mathbf{H}^k \cdot h\bar{\boldsymbol{\Sigma}}_i^k \cdot (\mathbf{H}^k)^T + \mathbf{Q}^k]^{-1} [\mathbf{z}_i^k - {}^j\mathbf{h}(h\bar{\boldsymbol{\mu}}_i^k)]}. \end{aligned} \quad (2.87)$$

Hierbei entspricht  $d$  der Dimension des Zustandsraum  $\mathbf{x}_i^k$ . Für jede Hypothese  $h$  wird der Erwartungswert  $h,j\boldsymbol{\mu}_i^k$  und die Kovarianz  $h,j\boldsymbol{\Sigma}_i^k$  durch die Gleichungen (2.71) und (2.72) geschätzt. Dabei wird die Richtung  $j$  berücksichtigt. Schließlich resultiert ein Gaußsches Mischmodell mit  $n_h^k = 2 \cdot n_h^{k-1}$  Komponenten:

$$\text{bel}(\mathbf{x}_i^k) = \sum_{h=1}^{n_h^{k-1}} \sum_{j=1}^2 h,j\omega_i^k \mathcal{N}(h,j\boldsymbol{\mu}_i^k, h,j\boldsymbol{\Sigma}_i^k) = \sum_{h=1}^{n_h^k} h\omega_i^k \mathcal{N}(h\boldsymbol{\mu}_i^k, h\boldsymbol{\Sigma}_i^k). \quad (2.88)$$

Die Gewichte  $h,j\omega_i^k$  der jeweiligen Komponenten ergeben sich folgendermaßen:

$$h,j\omega_i^k = \frac{1}{2} h\bar{\omega}_i^k \cdot h,j\lambda_i \cdot \eta. \quad (2.89)$$

$\eta$  ist hierbei ein Skalierungsparameter.

### 2.3.4.3 Reduktion von Hypothesen

In Abschnitt 2.3.4.2 wurde dargestellt, dass sich durch die Modellierung der geometrischen Mehrdeutigkeiten die Anzahl der Hypothesen in jedem Zeitschritt verdoppelt. Um die Anzahl der Hypothesen für einen einzelnen Zustand zu beschränken, existieren unterschiedliche Ansätze. Im Folgenden werden Verfahren beschrieben, die ursprünglich im Bereich des Verfolgens von einzelnen Objekten verwendet werden. Sie können jedoch auch für die Hypothesenreduktion eingesetzt werden:

1. Nächster-Nachbar-Standard-Filter (engl. *nearest neighbour standard filter* - NNSF-Filter) [BFC90]
2. Probabilistischer Datenassoziiierungsfilter (engl. *probabilistic data association filter* - PDAF) [Bar75]
3. Gauß'scher-Summen-Filter (engl. *gaussian sum filter* - GSF) [AS72]

Der NNSF-Filter approximiert die gesamte Anzahl von Hypothesen durch eine einzelne Hypothese  $h$ , die mit dem höchsten Gewicht  ${}^h\omega_i^k$  assoziiert ist. Die anderen Hypothesen werden verworfen.

Im Gegensatz dazu approximiert PDAF alle Hypothesen durch eine einzelne Normalverteilung. Die Parameter der Verteilung setzen sich aus allen Hypothesen zusammen. Dieser Vorgang wird als *Merging* bezeichnet.

Schließlich verwirft der GSF zunächst alle Hypothesen, dessen Gewichte  $\omega_i^k$  kleiner als ein vorgegebener Schwellwert sind. Im Anschluss werden ähnliche Hypothesen zusammengeführt (*Merging*) und jeweils durch Parameter einer einzelnen Normalverteilung beschrieben. Im letzten Schritt wird eine Teilmenge der verbliebenen Hypothesen ausgewählt, die mit den höchsten Gewichten  ${}^h\omega_i^k$  assoziiert sind. Dabei wird die maximale Anzahl von Hypothesen auf  $n_{max}$  Elemente beschränkt.

### 2.3.4.4 Erweiterung auf mehrere Soundquellen

In den vorherigen Abschnitten wurde zunächst eine einzelne Soundquelle betrachtet. Im Folgenden wird das Zustandsschätzverfahren auf mehrere zur gleichen Zeit aktive Soundquellen erweitert. Exemplarisch wird die Zustandsschätzung auf zwei Soundquellen  $\mathbf{x}_1^k$  und  $\mathbf{x}_2^k$  übertragen. Ohne Beschränkung der Allgemeinheit ist es möglich diesen Ansatz auf weitere Soundquellen zu verallgemeinern. Es sei angenommen, dass die jeweiligen Soundquellen unabhängig voneinander sind. Gesucht ist der Belief  $\text{bel}(\mathbf{x}_1^k, \mathbf{x}_2^k)$  für die Zustände  $\mathbf{x}_1^k$  und  $\mathbf{x}_2^k$ . Dabei wurden die Beobachtungen  $z^1$  bis  $z^k$  durchgeführt:

$$\text{bel}(\mathbf{x}_1^k, \mathbf{x}_2^k) = p(\mathbf{x}_1^k, \mathbf{x}_2^k | \mathbf{z}^{1:k}). \quad (2.90)$$

Aufgrund der Aktivität von mehreren Soundquellen zum Zeitpunkt  $k$  setzt sich eine Beobachtung  $z^k$  aus mehreren Richtungsmessungen zusammen:

$$\mathbf{z}^k = (z_i^k, z_j^k) \quad i \in \{1, 2\} \text{ mit } i \neq j. \quad (2.91)$$

A priori ist die Datenassoziiierung zwischen den jeweiligen Messungen und den Zuständen der Soundquellen nicht bekannt. Weiterhin können für jede Messung geometrische Mehrdeutigkeiten auftreten (vgl. Abschnitt 2.3.4.2). Um die möglichen Datenassoziiierung von zwei Soundquellen zu visualisieren, ist ein Beispiel in Abbildung 2.4 gegeben. Es resultieren für zwei Richtungsmessungen bereits zwölf Hypothesen (vgl. Tabelle 2.4).

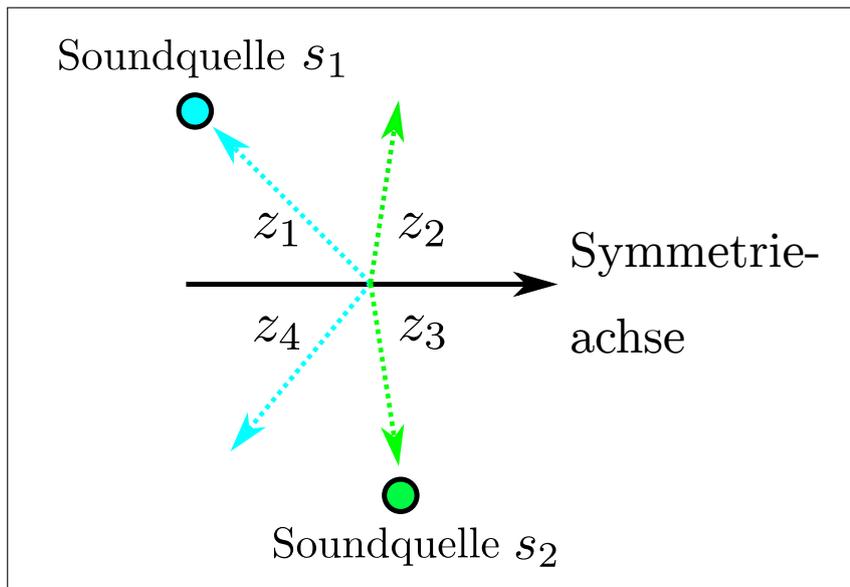


Abbildung 2.23: Darstellung der Geometrie.

Zuordnung	$s_1$	$s_2$
1	$z_1$	$z_2$
2	$z_1$	$z_3$
3	$z_1$	$z_4$
4	$z_2$	$z_1$
5	$z_2$	$z_3$
6	$z_2$	$z_4$
7	$z_3$	$z_1$
8	$z_3$	$z_2$
9	$z_3$	$z_4$
10	$z_4$	$z_1$
11	$z_4$	$z_2$
12	$z_4$	$z_3$

Tabelle 2.4: Zuordnung

Abbildung 2.24: rechts: In Abbildung 2.23 sind zwei aktive Soundquellen dargestellt. Aufgrund geometrischer Mehrdeutigkeiten können vier unterschiedliche Beobachtungen wahrgenommen werden.

links: In Tabelle 2.4 sind die zugehörigen Datenassoziiierung zwischen den Soundquellen und den Messungen dargestellt.

Die Anzahl von Hypothesen steigt stark an bei zunehmender Anzahl von aktiven Soundquellen. Daher ist es essentiell die Anzahl der Hypothesen zu reduzieren. Um den Belief  $\text{bel}(x_1^k, x_2^k)$  zu berechnen, werden die drei Verfahren aus Abschnitt 2.3.4.3 auf mehrere Zustände erweitert [BL95]:

1. Globaler Nächster-Nachbar-Standard-Filter (engl. *global nearest neighbour filter* - GNN-Filter)
2. Gemeinsamer Probabilistischer Datenassoziiierungsfiler (engl. *joint probabilistic data association filter* - JPDAF)
3. Multihypothesen Tracker (engl. *multi-hypothesis tracker* - MHT)

Diese Ansätze finden vor allem Einsatz in der Verfolgung von mehreren Objekten, falls die Anzahl der Zustände bekannt ist. Weiterhin wird die Anzahl der Hypothesen reduziert.

### 2.3.4.5 Erweiterung auf Ausreißer und Fehlmessungen

In der Realität treten Fehlmessungen aufgrund von Messunsicherheiten und akustischen Effekten auf. Diese können Reflexions-, Streuungs- oder Rauscheffekte beinhalten. Weiterhin ist nicht gewährleistet, dass alle aktiven Soundquellen zu jedem Messzeitpunkt beobachtet werden.

Eine weitere Schwierigkeit entsteht, wenn die Anzahl der aktiven Soundquellen im Vorfeld nicht bekannt ist. Hierfür wurden Verfahren entwickelt, deren Elemente durch Mengen repräsentiert werden - bezeichnet als *Random Finite Sets* (RFS). Dabei entspricht jedes Element einer Zufallsvariablen. Um mathematische Operationen auf den Mengen durchzuführen, wurde die *Finite Set Statistics* (FISST) entwickelt [Mah03].

Die RFS werden für das Verfolgen (engl. *tracking*) von mehreren Objekten verwendet [VSD05]. Weiterhin ist der Einsatz der RFS auf die Lokalisierung und Kartierung (SLAM) in [Mul+11a] und [Mul+11b] gegeben.

Die Kernidee besteht darin, dass die Fehlmessungen (engl. *clutter*) durch einen Punktprozess (engl. *Point Process* - PP) beschrieben wird. Dieser Prozess modelliert sowohl die Wahrscheinlichkeit der Anzahl der Elemente der Menge als auch die Wahrscheinlichkeit der Zuweisung der Zufallsvariablen. Ein klassischer Vertreter um Fehlmessungen zu modellieren ist der *Poisson-Punkt-Prozess* (PPP). Dieser Prozess ist durch Vorgabe einer Intensitätsfunktion vollständig beschrieben.

Die Posterior-Verteilung aus Gleichung (2.90) kann mithilfe der FISST ermittelt werden. Dabei kann die Lösung durch den PHD-Filter (engl. *probability hypothesis density* Filter) geschätzt werden. Die Kernidee liegt darin, dass die RFS durch die PHD-Funktion approximiert wird. Diese entspricht dem ersten statistischen Moment der RFS. Die PHD-Funktion wird verwendet, um den Erwartungswert der Anzahl von Elementen in einem vorgegebenen Bereich zu schätzen. Für einen PPP ergibt sich die PHD-Funktion aus der Intensitätsfunktion. Um die Intensitätsfunktion zu approximieren, wird häufig eine Repräsentation gewählt, bei der mehrere Normalverteilungen (engl. *Gaussian Mixture*) linear überlagert werden. Dies liefert den sogenannten GM-PHD-Filter (engl. *Gaussian Mixture PHD*) [VM06]. In Abschnitt 5.1.3 wird das Verfahren dargestellt und die Zustandsschätzung von Richtungen zu aktiven Soundquellen wird durchgeführt.

## 2.4 Intervallarithmetik

In dieser Arbeit bildet die Intervallarithmetik die Grundlage für das Kartierungsverfahren. Zunächst wird motiviert, wozu die Intervallarithmetik im Kontext dieser Arbeit verwendet wird. Darauf folgend werden Grundbegriffe definiert, die notwendig für das Verständnis der entwickelten Ansätze in dieser Arbeit sind.

### 2.4.1 Motivation

In Abschnitt 1.2 wurde auf die Herausforderungen der Soundquellenkartierung eingegangen. Im Weiteren wird argumentiert, auf welche Weise Intervalle verwendet werden können, um die Probleme zu adressieren.

Betrachtet wird es vereinfachtes Beispiel bei dem die Orientierung zwischen einem Mikrofonarray und einer mobilen Plattform inkorrekt angenommen wird. Es resultiert ein systematischer Fehler  $e_s$ :

$$e_s = a_{model} - a_{ist}. \quad (2.92)$$

$a_{ist}$  entspricht der tatsächlichen Orientierung und  $a_{model}$  der modellierten Orientierung. Ein klassisches Vorgehen besteht darin, den Fehler  $e_s$  durch einen probabilistischen Ansatz zu modellieren.

In einer Vielzahl von Ansätzen wird angenommen, dass der Erwartungswert  $\mu$  null beträgt und somit im Mittel kein Fehler auftritt. Diese Annahme trifft für das beschriebene Beispiel nicht zu. Aufgrund des systematischen Fehlers in der Modellierung ist der Orientierungsfehler im Durchschnitt ungleich null. Dies ist in Abbildung 2.25a veranschaulicht.

Falls der systematische Fehler bekannt ist, kann der Erwartungswert  $\mu$  entsprechend gewählt werden. Somit würde die Fehlermodellierung durch die Normalverteilung im Mittel mit dem tatsächlichen Fehler übereinstimmen. Jedoch ist diese Information in der Realität nicht gegeben. Weiterhin kann die Anordnung der Mikrofone durch Interaktion mit der Umgebung beeinflusst werden.

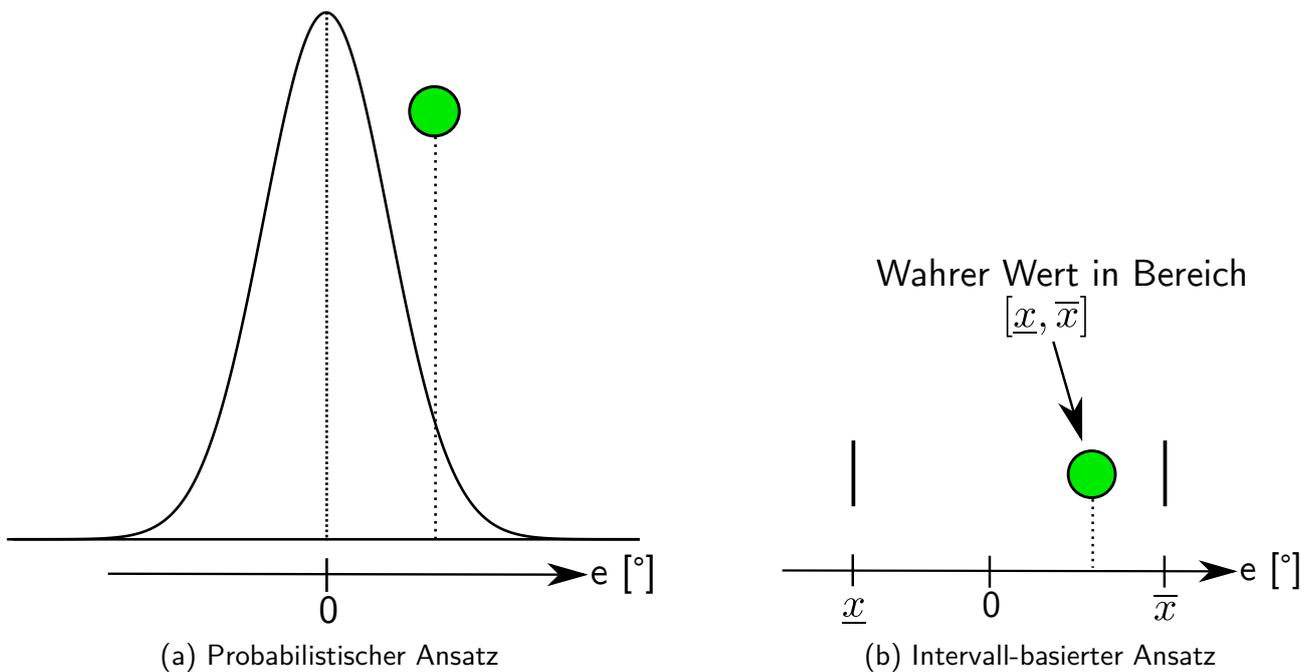


Abbildung 2.25: Die korrekte Orientierung (grüner Punkt) ist dargestellt und das Fehlermodell  $e$  ist durch einen probabilistischen Ansatz (a) und einen intervall-basierenden Ansatz (b) gegeben.

So könnte beispielsweise die Plattform mit einem Hindernis zusammenstoßen und somit eine Verdrehung der Montierung resultieren. Auch in diesem Fall würde der gewählte Erwartungswert  $\mu$  nicht mit der Realität übereinstimmen.

Statt den Fehler durch einen Skalar oder eine probabilistische Formulierung zu beschreiben, kann ein intervall-basierender Ansatz gewählt werden. Dabei werden eine untere  $\underline{x}$  und eine obere Schranke  $\bar{x}$  definiert [Jau+01]. Es wird angenommen, dass der wahre Wert in dem Intervall  $[x] = [\underline{x}, \bar{x}]$  enthalten ist. Jedoch wird keine Aussage über die Wahrscheinlichkeitsverteilung innerhalb des Intervalls getroffen (vgl. Abbildung 2.25b). Falls die Grenzen für  $\underline{x}$  und  $\bar{x}$  die korrekten Werte einschließen, kann für intervall-basierte Operation mathematisch garantiert werden, dass die Ergebnisse innerhalb der berechneten Intervalle liegen.

## 2.4.2 Grundbegriffe

Es sei im Folgenden die Distanzmessung eines Zollstocks gegeben und angenommen, dass die Genauigkeit auf  $\pm 1$  mm limitiert ist. Als Beispiel werden ausgehend von einem Referenzpunkt die Distanzmessung  $a = 9$  mm und  $b = 3$  mm ermittelt. Nachdem die Unsicherheiten der Messungen berücksichtigt wurden, kann garantiert werden, dass die korrekten Distanzwerte in den Intervallen  $A = [8, 10]$  und  $B = [2, 4]$  liegen. Die Distanz zwischen den Messungen wird folgendermaßen ermittelt:

$$[8, 10] - [2, 4] = [8 - 4, 10 - 2] = [4, 8]. \quad (2.93)$$

Somit ist es garantiert, dass die Distanz zwischen 4 und 8 mm liegt.

Für die Intervalle können unterschiedliche Eigenschaften definiert werden [Jau+01]. Wichtige Eigenschaften sind die Breite  $w([x])$  eines Intervalls:

$$w([x]) = \bar{x} - \underline{x} \quad (2.94)$$

und der Mittelpunkt:

$$\text{mid}([x]) = \frac{1}{2}(\bar{x} + \underline{x}). \quad (2.95)$$

Um mehrdimensionale Größen durch Intervalle zu beschreiben, werden Intervallvektoren  $[\mathbf{x}]$  verwendet. Somit kann ein dreidimensionaler Messpunkt im kartesischen Raum folgendermaßen angegeben werden:

$$[\mathbf{x}] = \begin{pmatrix} [\underline{x}_1, \bar{x}_1] \\ [\underline{x}_2, \bar{x}_2] \\ [\underline{x}_3, \bar{x}_3] \end{pmatrix}. \quad (2.96)$$

Hierbei entsprechen  $\underline{x}_1, \bar{x}_1, \underline{x}_2, \bar{x}_2, \underline{x}_3, \bar{x}_3 \in \mathbb{R}$  den Grenzen in den jeweiligen Dimensionen  $x_1, x_2$  und  $x_3$ .

### 2.4.3 Inklusionsfunktion

Um mathematische Zusammenhänge zu formulieren, müssen Funktionen auf die Intervallarithmetik erweitert werden. Im Folgenden sei  $\mathbf{f}$  eine Funktion die von  $\mathbb{R}^n$  nach  $\mathbb{R}^m$  abbildet. Zur Veranschaulichung ist ein exemplarisches Beispiel gegeben, das in Abbildung 2.26 visualisiert ist.

Die zweidimensionale Intervallmenge  $[\mathbf{x}] = ([\underline{x}_1, \bar{x}_1] \quad [\underline{x}_2, \bar{x}_2])^T$  wird durch Auswertung von  $\mathbf{f}([x])$  in den Bildraum transformiert werden. Diese Abbildung kann sowohl nicht konvex als auch unverbunden sein.

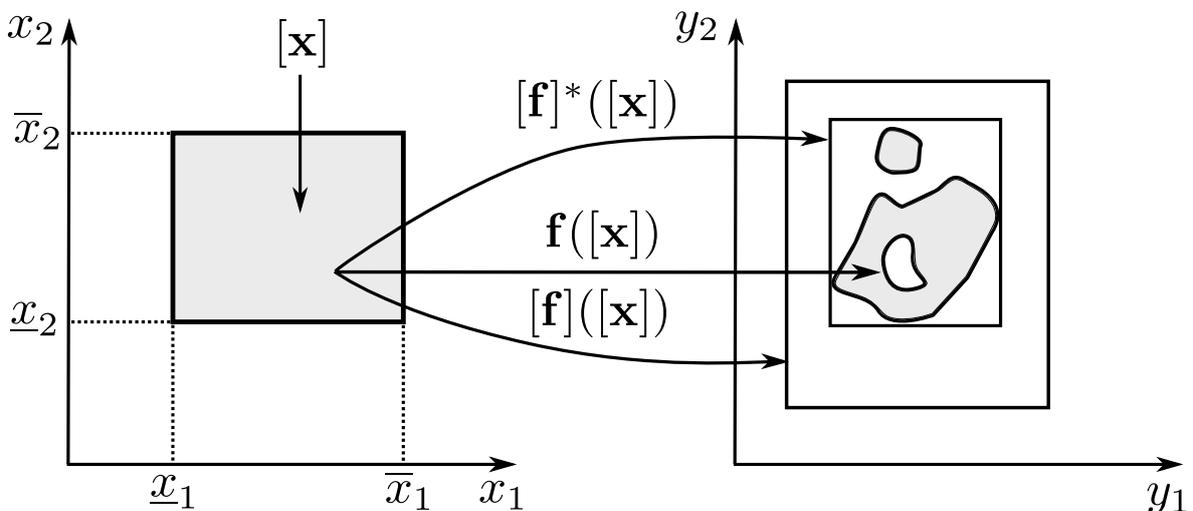


Abbildung 2.26: Inklusionsfunktion. [Jau+01]

Die Funktion  $f$  wird auf die Intervallararithmetik erweitert, indem eine Inklusionsfunktion  $[f] : \mathbb{IR}^n \rightarrow \mathbb{IR}^m$  gebildet wird:

$$\forall [\mathbf{x}] \in \mathbb{IR}^n, f([\mathbf{x}]) \subset [f]([\mathbf{x}]). \quad (2.97)$$

Die Bildmenge der Inklusionsfunktion bildet eine Obermenge der Funktion  $[f]$ . Die Angabe der Inklusionsfunktion ist nicht eindeutig. Daher ist nicht garantiert, dass eine optimale Darstellung vorliegt. Im Folgenden bezeichnet  $[f]^*([\mathbf{x}])$  eine optimale Inklusionsfunktion. Für konkrete Einsatzgebiete erweist sich jedoch die Ermittlung von  $[f]^*([\mathbf{x}])$  als Herausforderung.

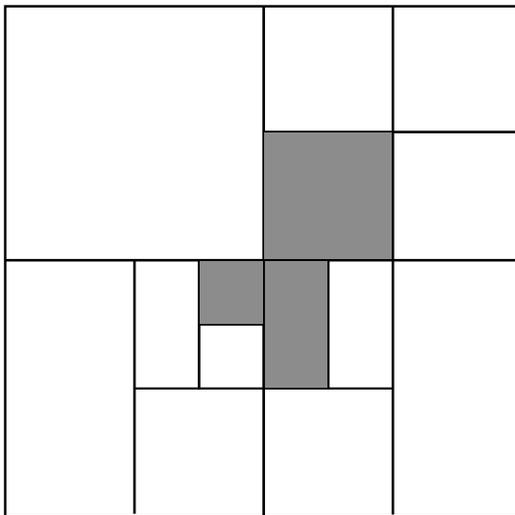
### 2.4.4 Subpavings

Um die Lösungsmenge präzise zu beschreiben, ist eine einzelne Intervall-Box nicht ausreichend. So ist für die Soundquellenkartierung von mehreren Soundquellen eine Lösungsmenge notwendig, die aus einer Menge von mehreren Boxen bestehen.

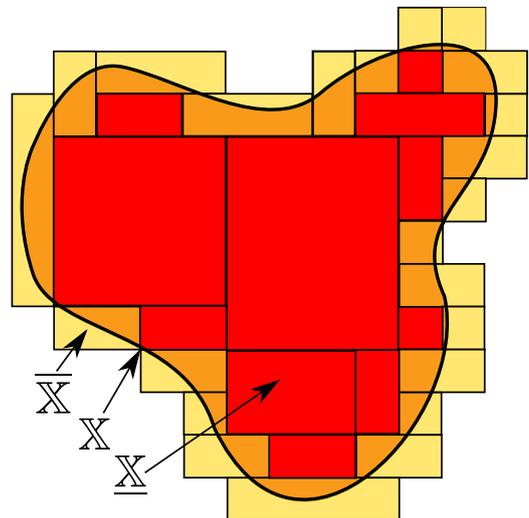
Im Folgenden sei  $\mathbb{P}$  eine Menge von  $m$  überlappungsfreien Boxen:

$$\mathbb{P} = \{[\mathbf{x}_1], [\mathbf{x}_2], \dots, [\mathbf{x}_m]\}. \quad (2.98)$$

$\mathbb{P}$  wird als *Subpaving* bezeichnet. Ausgehend von einer initialen Box kann durch rekursive Unterteilung ein Subpaving erzeugt werden. Falls die Unterteilung gleichmäßig ist, wird das Subpaving als regulär bezeichnet. Ein Beispiel eines regulären Subpavings ist in Abbildung 2.27a gegeben.



(a) Darstellung eines regulären Subpaving, das durch rekursive Halbierung (Bisektion) entsteht.



(b) Die Lösung  $\mathbb{X}$  wird durch ein inneres Subpaving  $\underline{\mathbb{X}}$  (rot) und ein äußeres Subpaving  $\overline{\mathbb{X}}$  (gelb) dargestellt.

Abbildung 2.27

Für die Approximation einer Lösung  $\mathbb{X}$  sind zwei Untermengen des Subpavings  $\mathbb{P}$  relevant. Dabei beinhaltet das innere Subpaving  $\underline{\mathbb{X}}$  ausschließlich Boxen, die komplett in der Lösungsmenge liegen.

Im Gegensatz dazu, enthält das äußere Subpaving  $\overline{\mathbb{X}}$  sowohl Boxen des inneren Subpavings als auch Boxen, die Teile der Lösung enthalten. Es gilt somit:

$$\underline{\mathbb{X}} \subseteq \mathbb{X} \subseteq \overline{\mathbb{X}}. \quad (2.99)$$

## 2.4.5 SIVIA

In Kapitel 2.4.4 wurde die Approximation des Lösungsraums  $\mathbb{X}$  durch Subpavings dargestellt. Im Folgenden wird die Ermittlung dieser Mengen beschrieben. Hierfür wird ein Zusammenhang zwischen einer Messung  $\mathbb{Y} \in \mathbb{R}^m$  (z. B. einem akustischen Merkmal einer Mikrofonaufnahme) und der Lösung  $\mathbb{X}$  durch eine nichtlineare Funktion  $\mathbf{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  angenommen:

$$\mathbb{X} = \{x \in \mathbb{R}^n \mid \mathbf{f}(x) \in \mathbb{Y}\} = \mathbf{f}^{-1}(\mathbb{Y}). \quad (2.100)$$

Die unbekannte Menge  $\mathbb{X}$  kann mit einem *Branch-and-Bound*-Algorithmus ermittelt werden. Im Kontext der Intervall-Arithmetik wird der Algorithmus *Set Inversion Via Interval Analysis* (SIVIA) verwendet [Jau+01]. Das Vorgehen ist in Algorithmus 1 gegeben.

---

### Algorithm 1 SIVIA nach [Jau+01]

---

**Eingabe:**  $[\mathbf{f}], [\mathbf{x}_0], \mathbb{Y}, \varepsilon$

**Ausgabe:**  $\underline{\mathbb{X}}, \overline{\mathbb{X}}$

```

1:  $\mathcal{S} := \{[\mathbf{x}_0]\}$ 
2: while  $\mathcal{S} \neq \emptyset$  do
3:    $[\mathbf{x}] := \text{top}(\mathcal{S})$ 
4:   if  $[\mathbf{f}]([\mathbf{x}]) \cap \mathbb{Y} = \emptyset$  then
5:     continue
6:   else if  $[\mathbf{f}]([\mathbf{x}]) \subset \mathbb{Y}$  then
7:      $\underline{\mathbb{X}} := \underline{\mathbb{X}} \cup [\mathbf{x}]$ 
8:      $\overline{\mathbb{X}} := \overline{\mathbb{X}} \cup [\mathbf{x}]$ 
9:   else if  $\omega([\mathbf{x}]) < \varepsilon$  then
10:     $\overline{\mathbb{X}} := \overline{\mathbb{X}} \cup [\mathbf{x}]$ 
11:   else
12:      $([\mathbf{x}_1], [\mathbf{x}_2]) = \text{bisect}([\mathbf{x}])$ 
13:      $\mathcal{S} := \mathcal{S} \cup \{[\mathbf{x}_1]\} \cup \{[\mathbf{x}_2]\}$ 
14:   end if
15: end while

```

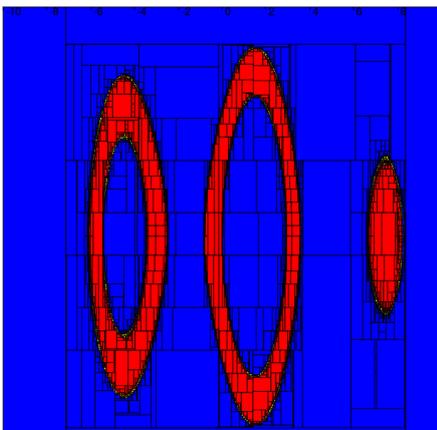
---

Als Eingang für das Verfahren wird der initiale Suchbereich  $[\mathbf{x}_0]$ , die Funktion  $[\mathbf{f}]$ , die Messung  $\mathbb{Y}$  und der Parameter  $\varepsilon$  benötigt.  $\varepsilon$  charakterisiert die Genauigkeit der Lösung und wird im weiteren Verlauf näher beschrieben.

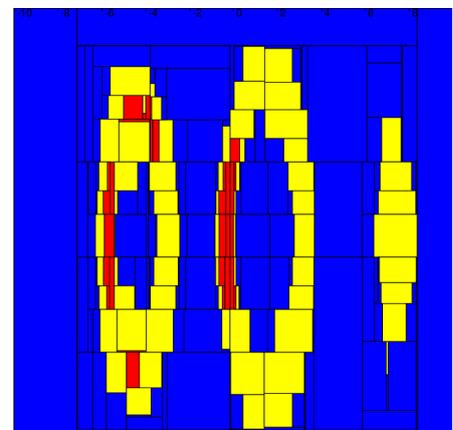
Als Ergebnis wird das innere Subpaving  $\underline{\mathbb{X}}$  und das äußere Subpaving  $\overline{\mathbb{X}}$  ermittelt. Im ersten Schritt wird die Menge  $\mathcal{S}$  definiert und als Stapelspeicher (engl. *stack*) assoziiert. Dieser wird zu Beginn mit dem initialen Suchbereich  $[\mathbf{x}_0]$  initialisiert. Solange Elemente auf dem Stapel  $\mathcal{S}$  enthalten sind, wird das Verfahren weiter ausgeführt. Zunächst wird das oberste Element  $[\mathbf{x}]$  vom Stapel  $\mathcal{S}$  entnommen.

Dabei werden vier Fälle unterschieden.

- 1.) Die Schnittmenge von  $[f]([x])$  und der Messung  $\mathbb{Y}$  sind disjunkt. Somit beinhaltet  $[x]$  keinen Teil der Lösung und ist für die weitere Betrachtung nicht relevant.
- 2.)  $[f]([x])$  ist eine Teilmenge von  $\mathbb{Y}$ . Somit ist  $[x]$  vollständig in der Lösungsmenge  $\mathbb{X}$  enthalten. Als Folge werden das innere und äußere Subpaving um diese Menge erweitert.
- 3.) Die längste Dimension der Intervall-Box ( $[x]$ ) liegt unter dem festgelegten Schwellwert  $\varepsilon$ . In diesem Fall enthält ( $[x]$ ) einen Teil der Lösung. Somit wird das äußere Subpaving um ( $[x]$ ) erweitert. Dabei hat der Schwellwert  $\varepsilon$  einen Einfluss auf die Genauigkeit der Lösung. Zur Veranschaulichung ist in Abbildung 2.28 die Lösungsmenge für unterschiedliche Schwellwerte visualisiert.
- 4.)  $[x]$  enthält Teile der Lösung, jedoch ist die Dimensionen der Intervall-Boxen noch nicht ausreichend fein. Daher wird die Intervall-Box ( $[x]$ ) an der längsten Achse halbiert und die resultierende Untermenge wird dem Stapel  $\mathcal{S}$  hinzugefügt.



(a)  $\varepsilon = 0.1$



(b)  $\varepsilon = 2.0$

Abbildung 2.28: Darstellung der Genauigkeit des SIVIA-Verfahrens durch Charakterisierung des Parameters  $\varepsilon$ . Als Beispiel wird die Lösung folgender Funktion approximiert:  $f = x^2 + y^2 - 10 \cdot \sin(x) \cdot 4$  mit  $[0, 30]$

Zur Veranschaulichung ist das Verfahren in Abbildung 2.29 für ein zweidimensionales Beispiel dargestellt. Hierbei wurden die jeweiligen Boxen in vier Teile unterteilt, falls Bedingung 4.) erfüllt wurde. Blau markierte Boxen enthalten keinen Teil der Lösung (Fall 1), gelbe Boxen enthalten Teile der Lösung, jedoch ist die geforderte Genauigkeit noch nicht erfüllt (Fall 3). Schließlich enthalten rot markierte Boxen Teile der Lösung und erfüllen die geforderte Genauigkeit.

## 2.4.6 Kontraktoren

Ein Nachteil des in Kapitel 2.4.5 vorgestellten SIVIA-Verfahrens besteht in dem hohen Rechenaufwand. Dieser ist aufgrund des rekursiven Ansatzes exponentiell. Ein alternativer Ansatz liegt darin, sogenannte Kontraktoren zu verwenden. Diese führen eine einzelne Reduktion durch, anstelle den Suchbereich iterativ zu teilen. Die Funktion  $f$  aus Gleichung (2.100) wird als sogenanntes *Constraint-Satisfaction-Problem* (CSP) definiert [Apt03, Kapitel 2]. Ein Kontraktor muss die mathematische Eigenschaft aufweisen, dass inkonsistente Bereiche der Lösungsmenge reduziert werden. Jedoch dürfen keine Teile der Lösung entfernt werden. Der Berechnungsaufwand der Kontraktion ist polynomial und somit geringer im Vergleich zum SIVIA-Verfahren.

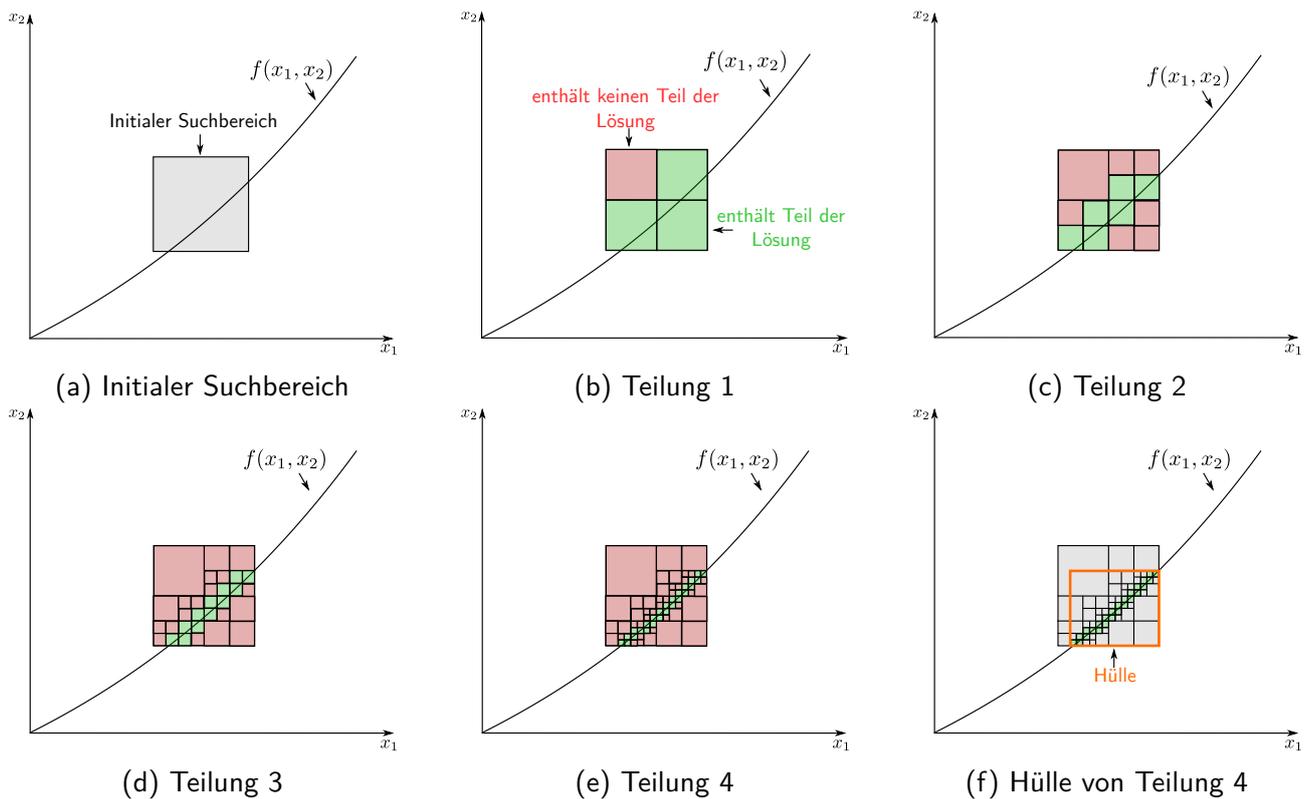


Abbildung 2.29: Exemplarische Durchführung des SIVIA-Verfahrens. Initialer Suchbereich (grau), Box enthält einen Teil der Lösung (grün), Box enthält keinen Teil der Lösung (rot)

In Abbildung 2.30 ist ein exemplarisches Beispiel einer Kontraktion gegeben. Die Herausforderung liegt darin einen optimalen Kontraktor zu finden, der den gesamten inkonsistenten Bereich entfernt.

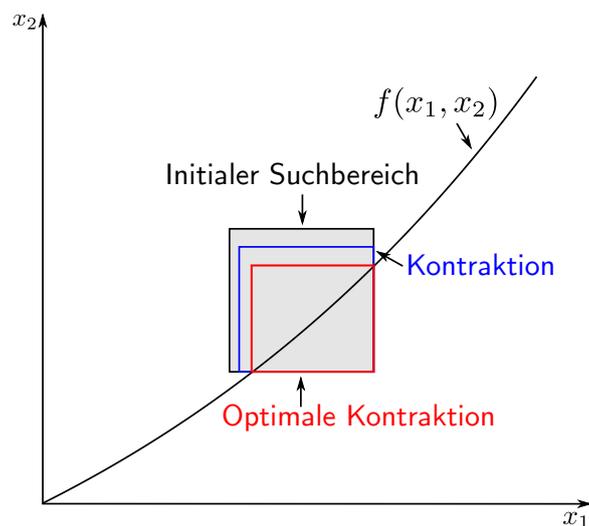


Abbildung 2.30: Kontraktion eines initialen Suchbereiches um die Lösungsfunktion  $f(x_1, x_2)$  zu approximieren. Verwendung eines nicht optimalen (blau) und optimalen (rot) Kontraktors

In der Literatur wurden unterschiedliche Kontraktoren entwickelt, die u. a. auf dem Gauss-Eliminations-Verfahren, der linearen Programmierung, der Krawczyk-Methode und der Forward-

Backward-Propagierung [Jau+01] basieren. In dieser Arbeit wird der Forward-Backward-Kontraktor (FB-Kontraktor) verwendet und im Folgenden anhand eines vereinfachten Beispiels vorgestellt. Die Kontraktion basiert auf der Constraint-Propagation. Hierbei werden die Restriktion in Isolation betrachtet und damit die Wertebereiche kontrahiert.

Als Beispiel sei folgendes Constraint-Satisfaction-Problem  $\mathcal{H}$  gegeben:

$$\mathcal{H} : \begin{cases} \text{Variablen: } x_1, x_2 \\ \text{Constraints:} \\ 1. \quad 4x_1 + x_2^2 - 5 = 0 \\ 2. \quad x_1 + 2x_2 - 3 = 0 \\ \text{Wertebereich: } [x_1] \in [0, 10], [x_2] \in [-100, 100] \end{cases}$$

Zunächst wird die erste Restriktion betrachtet. Hierfür werden Hilfsvariablen  $a_1, a_2, a_3, a_4$  benötigt, die den gesamten Wertebereich abbilden:  $[a_1] = [a_2] = [a_3] = [a_4] = [-\infty, \infty]$ . Zunächst wird ein Vorwärtsschritt durchgeführt:

$$\begin{aligned} 1. \quad [a_1] &:= 4 \cdot [x_1] \\ 2. \quad [a_2] &:= [x_2] \\ 3. \quad [a_3] &:= [a_2] \cdot [a_2] \\ 4. \quad [a_4] &:= [a_1] + [a_3] - 5 \end{aligned} \tag{2.101}$$

Durch die vorgegebene Restriktion muss  $a_4$  mit dem Skalar null übereinstimmen. Hieraus ergibt sich folgende Bedingung:

$$5. \quad [a_4] := [a_4] \cap \{0\}. \tag{2.102}$$

Im Anschluss erfolgt der Rückwärtsschritt:

$$\begin{aligned} 6. \quad [a_1] &:= ([a_4] - [a_3] + 5) \cap [a_1] \\ 7. \quad [a_3] &:= ([a_4] - [a_1] + 5) \cap [a_3] \\ 8. \quad [a_2] &:= \sqrt{[a_3]} \cap [a_2] \\ 9. \quad [x_2] &:= [a_2] \cap [x_2] \\ 10. \quad [x_1] &:= 0,25 \cdot [a_1] \cap [x_1]. \end{aligned} \tag{2.103}$$

Tabelle 2.5: Iteration über Forward-Backward-Kontraktor

Iteration k	$[x_1] (k)$	$[x_2] (k)$
0	[0;10]	[-100;100]
1	[0;10]	[-3,5;1,5]
2	[0;2,5625]	[0,21875;1,5]
3	[0,6875;1,23804]	[0,880981;1,15625]
⋮	⋮	⋮
12	[0,999999;1]	[1;1]
13	[1;1]	[1;1]

Analog erfolgen der Vorwärts- und Rückwärtsschritt für die zweite Restriktion von  $\mathcal{H}$ . Im Anschluss werden beide Kontraktoren nacheinander ausgeführt. Tabelle 2.5 stellt die kontrahierten Wertebereiche für mehrere Iterationsschritte dar.

## 2.5 Bezug zur Arbeit

Dieser Abschnitt stellt den Bezug der vorliegenden Arbeit mit dem Stand der Wissenschaft her.

Die physikalischen Grundlagen in Abschnitt 2.1.1 sind relevant für das grundlegende Verständnis der Ausbreitung von akustischen Wellen. Die akustische Dämpfung (vgl. Abschnitt 2.1.2) und Reflexion (vgl. Abschnitt 2.1.3) beeinflussen, wie die akustischen Signale wahrgenommen werden. Um akustische Signale zu interpretieren, müssen diese zunächst aufgezeichnet werden. Hierfür wurden in Abschnitt 2.1.4 unterschiedliche Mikrofontypen dargestellt. In dieser Arbeit werden mehrerer Mikrofone auf einer mobilen Plattform montiert. Dabei ist zu beachten, dass der verfügbare Raum auf der Plattform beschränkt ist und die Mikrofone robust gegenüber Erschütterungen sein müssen. Daher eignen sich Elektret-Kondensator oder MEMS-Mikrofone. Die Analyse der akustischen Signale erfolgt im Frequenzbereich. Dazu wurde in Abschnitt 2.1.5 die Frequenzanalyse durch die Kurzzeitfouriertransformation beschrieben.

Um die Position von Soundquellen zu bestimmen, ist es zunächst notwendig auditorielle Merkmale aus den Mikrofonsignalen zu extrahieren. In dieser Arbeit wird die TDoA zwischen unterschiedlichen Mikrofonpaaren verwendet. Um die TDoA zu ermitteln, werden zwei neue Ansätze - *DoATiD*<sup>11</sup> und *InTiD*<sup>12</sup> - entwickelt. *DoATiD* basiert auf Richtungsschätzungen, die aus Verfahren im Bereich der Soundquellenlokalisierung gewonnen werden können. Hierfür wurden in Abschnitt 2.2.3 unterschiedliche Verfahren vorgestellt. Da keine Kalibrierung der eingesetzten Mikrofone vorausgesetzt wird, müssen Verfahren ausgewählt werden, die akustische Dämpfung berücksichtigen. Hierfür wird in dieser Arbeit das SRP-PHAT-Verfahren verwendet. Das Verfahren wurde in Abschnitt 2.2.3.4 vorgestellt und ist unabhängig von der Amplitude der Mikrofonsignale. Für reale Experimente ist es nicht ausreichend die reinen Richtungsschätzverfahren zu verwenden. Stattdessen müssen die Daten zuvor gefiltert werden. Hierfür werden in dieser Arbeit zwei Ansätze - *M3K*<sup>13</sup> und *GM-PHD*<sup>14</sup> - vorgestellt. *M3K* beruht auf einem erweiterten Kalman-Filter (vgl. Abschnitt 2.3.3) und *GM-PHD* auf einen PHD-Filter (vgl. Abschnitt 2.3.4.5).

Neben dem richtungsbasierten Ansatz *DoATiD* wird der mikrofon-basierte Ansatz *InTiD* in dieser Dissertation vorgestellt, um die TDoA zu schätzen. *InTiD* basiert auf einem neuen intervall-basierten Verfahren, das direkt die Mikrofondaten berücksichtigt. Ein bestehender Kontraktor (vgl. 2.4.6) wird dabei um die akustische Dämpfung erweitert. Hiermit wird die Annahme berücksichtigt, dass keine Kalibrierung der Mikrofone vorliegt.

Mit den Informationen der auditoriellen Merkmale und der Position der Mikrofone wird die Position der aktiven Soundquellen in einer metrischen Karte ermittelt. In Abschnitt 2.3.1 wurden die Grundlagen für die Kartierung vorgestellt und in Abschnitt 2.3.2 der Stand der Wissenschaft im Bereich der Soundquellenkartierung gegeben. In Abschnitt 1.2 wurden Probleme dargestellt, die durch

<sup>11</sup> Engl. *Direction-of-Arrival-based Time Difference*

<sup>12</sup> Engl. *Interval-based Time Difference*

<sup>13</sup> Engl. *Modified 3D Kalman Method*

<sup>14</sup> Engl. *Gaussian Mixture Probability Hypothesis Density*

aktuelle Verfahren nicht adressiert werden. Zur Lösung wird in dieser Arbeit ein neues Verfahren vorgestellt und somit die Kartierung von Soundquellen durchgeführt. Dabei werden Verfahren der Intervallarithmetik eingesetzt, die in Abschnitt 2.4 vorgestellt wurden.

Schließlich werden zwei klassische Verfahren zum Vergleich der entwickelten intervall-basierten Kartierung verwendet. Einerseits wird der in Abschnitt 2.3.4 vorgestellte probabilistische Ansatz verwendet. Andererseits wird ein Ansatz durch Lösung eines Optimierungsproblems untersucht. Dieser Ansatz wird in Abschnitt 3.2 vorgestellt.



# Kapitel 3

## Intervall-basierte Soundquellenkartierung

Dieses Kapitel stellt die Kernidee des entwickelten Lösungsansatzes zur Kartierung von Soundquellen vor. Dabei dient Abbildung 3.1 der Einordnung des Ansatzes in das Gesamtsystem. Zunächst wird in Abschnitt 3.1 die verwendete Notation eingeführt. Anschließend wird der Zusammenhang zwischen den Mikrofonpositionen, den extrahierten akustischen Merkmalen und den gesuchten Positionen der Soundquellen aufgestellt. Dieser Zusammenhang gilt für ein einzelnes Mikrofonpaar und wird in Abschnitt 3.2 auf mehrere Soundquellen, Mikrofonpaare und Messpositionen erweitert. Als Resultat entsteht ein nichtlineares Gleichungssystem. Zur Lösung dieses Gleichungssystems wird in Abschnitt 3.3 ein intervall-basierter Ansatz vorgestellt. Dieser Ansatz wird in Abschnitt 3.4 auf Reflexionen und in Abschnitt 3.5 auf Ausreißer erweitert.

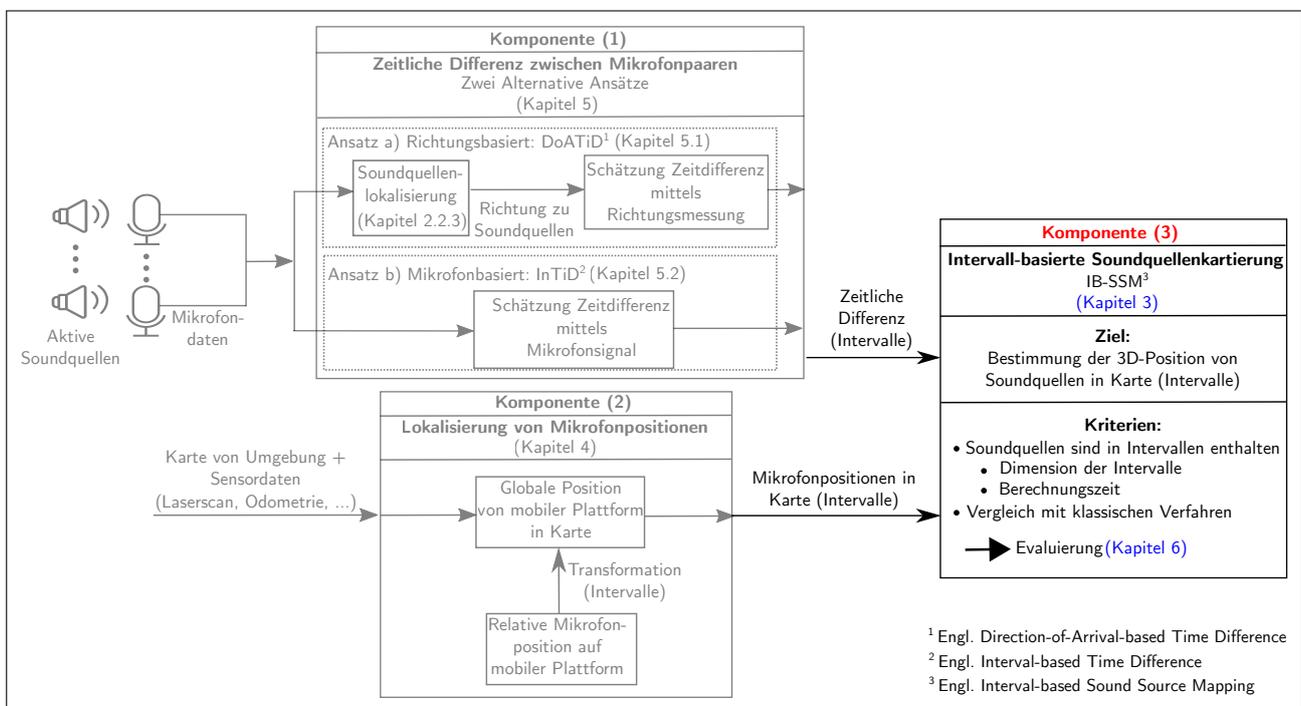
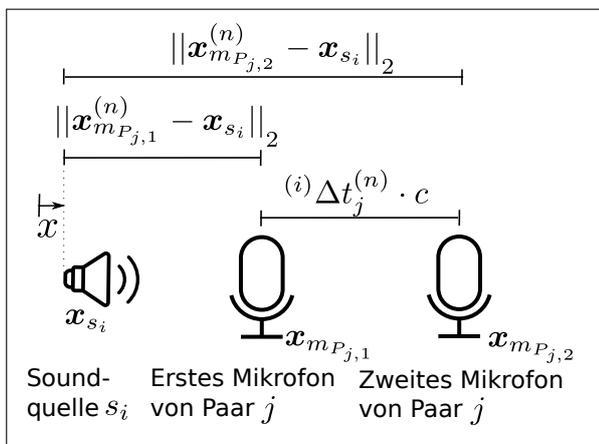


Abbildung 3.1: Systematischer Aufbau des entwickelten Gesamt-Konzeptes. Der relevante Teil für dieses Kapitel ist hervorgehoben.

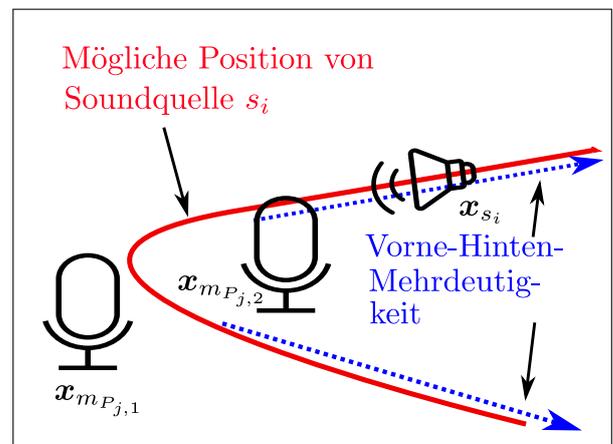
### 3.1 Problemdefinition und Notation

Es seien in der untersuchten Umgebung verschiedene Soundquellen  $s_i$  mit  $i \in \{1, \dots, n_s\}$  vorhanden, die akustische Signale mit der Schallgeschwindigkeit  $c$  aussenden.  $n_s$  bezeichnet die Gesamtzahl der Soundquellen. In dieser Arbeit wird angenommen, dass  $n_s$  im Vorfeld nicht bekannt ist. Die Positionen der Soundquellen werden durch den Zustandsraum  $\mathbf{x}_{s_i} \in \mathbb{R}^3$  beschrieben. Weiterhin nimmt ein mobiler Roboter die akustischen Signale durch ein Mikrofonarray mit  $n_m$  Mikrofonen wahr. Im Folgenden wird eine mathematische Beziehung zwischen den Mikrofonpaaren benötigt. Daher wird die erste und zweite Position des Mikrofonpaares  $j \in \{1, \dots, n_p\}$  mit  $\mathbf{x}_{m_{P_{j,1}}}^{(n)}$  und  $\mathbf{x}_{m_{P_{j,2}}}^{(n)} \in \mathbb{R}^3$  beschrieben.  $n_p$  ist die Gesamtzahl der verwendeten Mikrofonpaare. Aufgrund der Bewegung des Roboters, verändern sich die Mikrofonpositionen innerhalb einer globalen Karte. Hierfür bezeichnet der Exponent  $n \in \{1, \dots, n_l\}$  den Index der Position und  $n_l$  die Gesamtzahl der räumlichen Positionen. Weiterhin resultiert die Zeitdifferenz  ${}^{(i)}\Delta t_j^{(n)}$  zwischen dem ersten und zweiten Mikrofonpaar.  ${}^{(i)}\Delta$  wird im Weiteren als TDoA bezeichnet (vgl. Abschnitt 2.2.3.1). Der Exponent  $i$  bezeichnet die TDoA, die von der Soundquelle  $s_i$  resultiert. Weiterhin hängt die TDoA von der Position des Mikrofonpaares  $j$  am Positionsindex  $n$  ab. Schließlich wird der Zusammenhang zwischen der Position eines Mikrofonpaares und der TDoA für eine einzelne Soundquelle  $s$  folgendermaßen definiert:

$$\|\mathbf{x}_{m_{P_{j,1}}}^{(n)} - \mathbf{x}_{s_i}\|_2 - \|\mathbf{x}_{m_{P_{j,2}}}^{(n)} - \mathbf{x}_{s_i}\|_2 = {}^{(i)}\Delta t_j^{(n)} \cdot c. \quad (3.1)$$



(a) Beziehung zwischen der Position eines Mikrofonpaares und der TDoA aus Gleichung (3.1).



(b) Die Position einer Soundquelle ist für eine fixe TDoA auf einer hyperbolischen Kurve (rot) verortet.

Abbildung 3.2: Zusammenhang zwischen den Merkmalen der Mikrofonensignale, der Mikrofonpositionen und der Position einer Soundquelle in 2D.

Für eine vereinfachte Darstellung des Zusammenhanges wird in Abbildung 3.2a der Exponent  $n$  vernachlässigt.

Die linke Seite der Gleichung (3.1) enthält die geometrischen Eigenschaften der Mikrofonkonfiguration. Dahingegen beschreibt die rechte Seite die Messungen des Mikrofonpaares in Form der TDoA. Wie in Abbildung 3.2b dargestellt, liegt die Lösung von Gleichung (3.1) für eine fixierte TDoA auf einer hyperbolischen Kurve. Falls die Annahme getroffen wird, dass die Soundquelle weit entfernt ist (Fernfeld-Annahme vgl. Abbildung 2.12 in Abschnitt 2.2.3.1), resultieren im zweidimensionalen zwei

mögliche Richtungen zu einer Soundquelle. Diese geometrischen Mehrdeutigkeiten werden in der Literatur als Vorne-Hinten-Mehrdeutigkeiten (engl. *Front-Back-Ambiguity*) bezeichnet und sind im Folgenden in Abbildung 3.2b visualisiert. Im Gegensatz zu den präsentierten Methoden in Abschnitt 2.3.2 wird in dieser Arbeit die Lösung der Kartierung nicht auf die beiden Richtungen beschränkt und stellt somit einen wichtigen Beitrag dieser Arbeit dar. Stattdessen wird die gesamte Information des *Cone-of-Confusion* (vgl. Abbildung 1.1a) verwendet, um auf die Position der Soundquellen zu inferieren.

## 3.2 Erweiterung auf mehrere Paare, Soundquellen und Messpositionen

Die Gleichung (3.1) beschreibt ausschließlich den Zusammenhang zwischen einem Mikrofonpaar für eine Soundquelle an einer Messposition. Jedoch nimmt jedes Mikrofonpaar verschiedene Soundquellen an unterschiedlichen Positionen wahr. Die Gesamtzahl der resultierenden Gleichungen  $n_t$  wird folgendermaßen berechnet:

$$n_t = n_s \cdot n_p \cdot n_l. \quad (3.2)$$

Insgesamt müssen die verschiedenen Gleichungen in einem mathematischen System formuliert und schließlich gelöst werden. Aufgrund der hohen Nichtlinearität und den Restriktionen zwischen den Gleichungen, ist die Lösung des Systems eine Herausforderung. Um dies zu verdeutlichen, wird im Folgenden ein klassischer Ansatz beschrieben.

Zunächst sei vorausgesetzt, dass die zeitliche Differenz  ${}^{(i)}\Delta t_j^{(n)}$  bekannt ist. Somit kann die rechte Seite von Gleichung (3.1) angegeben werden. Im Weiteren wird die linke Seite mit der Funktion  $f_k(\mathbf{x}_{s_i})$  assoziiert. Hierbei entspricht der Index  $k \in \{1, n_p \cdot n_l\}$  der Nummerierung der Gleichungen zugehörig zur Soundquelle  $s_i$ . Die rechte Seite wird mit  $z_k$  assoziiert. Somit gilt:

$$f_k(\mathbf{x}_{s_i}) = z_k \quad \forall k \in \{1, 2, \dots, n_p \cdot n_l\}. \quad (3.3)$$

Die Differenz zwischen der Messung  $z_k$  und dem Modell  $f_k(\mathbf{x}_{s_i})$  wird als Fehler  $e_k$  bezeichnet:

$$e_k = z_k - f_k(\mathbf{x}_{s_i}). \quad (3.4)$$

Das Ziel ist die Position  $\mathbf{x}_{s_i}$  so zu wählen, dass die Summe der Fehlerquadrate über alle Gleichungen minimal wird. Dieses Vorgehen wird als Methode der kleinsten Quadrate (engl. *least-square approach*) bezeichnet und führt zu folgendes Optimierungsproblem:

$$x^* = \underset{x_{s_i}}{\operatorname{argmin}} \sum_{k=1}^{n_p \cdot n_l} e_k^T \Omega_k e_k. \quad (3.5)$$

Hierbei bezeichnet  $\Omega_k$  die Informationsmatrix. Sie entspricht der Inversen der Kovarianz zugehörig zu der Messung  $z_k$ .  $\Omega_k^{-1}$  berücksichtigt somit die Unsicherheit der Messung. Im Folgenden sind  $\Omega_k$  und  $\Omega_k^{-1}$  jeweils skalare Größen.

Das Optimierungsproblem kann beispielsweise mit dem Levenberg-Marquardt-Verfahren gelöst werden [Lev44]. Hierfür wird Gleichung  $f_k$  durch eine Taylor-Reihenentwicklung linearisiert, ein linea-

res Gleichungssystem aufgestellt (siehe Gleichung (3.6)) und anschließend eine rekursive Berechnung durchgeführt [Foy76]:

$$(\mathbf{H} - \lambda \mathbf{I}) \Delta x = -\mathbf{b} \text{ mit } \mathbf{H} = \sum_k \mathbf{J}_k^T \Omega_k \mathbf{J} \text{ und } \mathbf{b} = \sum_k e_k^T \Omega_k \mathbf{J} \quad (3.6)$$

$$x \leftarrow x + \Delta x. \quad (3.7)$$

Hierbei bezeichnet  $\lambda$  einen Dämpfungsfaktor. Zur Berechnung wird die Jacobi-Matrix  $\mathbf{J}$  benötigt. Sie beinhaltet die Ableitungen von  $f_k$  für die Unbekannten  $(x_{s_i}, y_{s_i}, z_{s_i})$ :

$$\mathbf{J}_k(x) = \begin{pmatrix} \frac{\partial f_k}{\partial x_{s_i}} & \frac{\partial f_k}{\partial y_{s_i}} & \frac{\partial f_k}{\partial z_{s_i}} \end{pmatrix}. \quad (3.8)$$

Zur vereinfachten Darstellung der Ableitung wird im Folgenden die Bezeichnung der Messpositionen ( $n$ ) vernachlässigt und die Koordinaten der beiden Mikrofone mit  $\mathbf{m}_{P_{j,1}} = (m_{1x}, m_{1y}, m_{1z})^T$  und  $\mathbf{m}_{P_{j,2}} = (m_{2x}, m_{2y}, m_{2z})^T$  bezeichnet:

$$f_k = \sqrt{(x_{s_i} - m_{1x})^2 + (y_{s_i} - m_{1y})^2 + (z_{s_i} - m_{1z})^2} - \sqrt{(x_{s_i} - m_{2x})^2 + (y_{s_i} - m_{2y})^2 + (z_{s_i} - m_{2z})^2}. \quad (3.9)$$

Somit gilt für die Jacobi-Matrix  $\mathbf{J}_k$ :

$$\mathbf{J}_k(x) = \begin{pmatrix} \frac{x_{s_i} - m_{1x}}{\sqrt{(x_{s_i} - m_{1x})^2 + (y_{s_i} - m_{1y})^2 + (z_{s_i} - m_{1z})^2}} - \frac{x_{s_i} - m_{2x}}{\sqrt{(x_{s_i} - m_{2x})^2 + (y_{s_i} - m_{2y})^2 + (z_{s_i} - m_{2z})^2}} \\ \frac{y_{s_i} - m_{1y}}{\sqrt{(x_{s_i} - m_{1x})^2 + (y_{s_i} - m_{1y})^2 + (z_{s_i} - m_{1z})^2}} - \frac{y_{s_i} - m_{2y}}{\sqrt{(x_{s_i} - m_{2x})^2 + (y_{s_i} - m_{2y})^2 + (z_{s_i} - m_{2z})^2}} \\ \frac{z_{s_i} - m_{1z}}{\sqrt{(x_{s_i} - m_{1x})^2 + (y_{s_i} - m_{1y})^2 + (z_{s_i} - m_{1z})^2}} - \frac{z_{s_i} - m_{2z}}{\sqrt{(x_{s_i} - m_{2x})^2 + (y_{s_i} - m_{2y})^2 + (z_{s_i} - m_{2z})^2}} \end{pmatrix}^T. \quad (3.10)$$

Bisher ist das Verfahren für eine einzelne Soundquelle  $s_i$  definiert. Falls mehrere Soundquellen betrachtet werden, muss das Verfahren erweitert werden. Zunächst muss die Anzahl der aktiven Soundquellen bekannt sein. Für jede Messposition muss eine Zuordnung zwischen der Messungen und der Soundquelle erfolgen. Für  $n_s$  aktiven Soundquellen und  $n_l$  verschiedenen Messpositionen ergeben sich  $n_z$  mögliche Zuordnungen:

$$n_z = (n_s!)^{n_l}. \quad (3.11)$$

Somit resultieren aus zwei Soundquellen und zehn Messpositionen insgesamt 1024 mögliche Zuordnungen.

Schließlich resultieren hohe Berechnungszeiten für die Lösung des Optimierungsproblems. Weiterhin wird eine gute Startlösung vorausgesetzt. Andernfalls ist nicht gewährleistet, dass eine korrekte Lösung ermittelt wird. Im Experimentaltel in Kapitel 6 wird auf dieses Verfahren eingegangen.

Im Gegensatz zu dieser Methode wurde ein intervall-basierter Ansatz in [RJC09] vorgestellt. Dort wird Gleichung (3.1) in einem anderen Kontext gelöst, bei dem der Ursprung eines statischen Transmitters gesucht wird. Dieser Transmitter sendet im Gegensatz zu der vorliegenden Dissertation ein elektromagnetisches Signal aus. Es werden drei statische Receiver eingesetzt und eine bekannte TDoA an den Receivern angenommen. Dieser Ansatz kann auf das in dieser Arbeit vorliegende Problem angewendet werden. Daher wird im nächsten Abschnitt die Methodologie definiert.

### 3.3 Intervall-basierter Lösungsansatz

Im Folgenden wird die Position der gesuchten Soundquellen in der Umgebung durch eine diskrete Menge von Intervallboxen repräsentiert. In diesem Abschnitt wird beschrieben, wie diese Boxen ermittelt werden. Zunächst wird Gleichung (3.1) als CSP formuliert:

$$\mathcal{H} : \begin{cases} \text{Variables: } \mathbf{x} \\ \text{Constants: } \mathbf{x}_{m_{P_{j,1}}}^{(n)}, \mathbf{x}_{m_{P_{j,2}}}^{(n)}, {}^{(i)}\Delta t_j^{(n)}, c \\ \text{Constraints:} \\ 1. \quad \|\mathbf{x}_{m_{P_{j,1}}}^{(n)} - \mathbf{x}\|_2 - \|\mathbf{x}_{m_{P_{j,2}}}^{(n)} - \mathbf{x}\|_2 = {}^{(i)}\Delta t_j^{(n)} \cdot c \\ \text{Domains: } [\mathbf{x}], [\mathbf{x}_{m_{P_{j,1}}}^{(n)}], [\mathbf{x}_{m_{P_{j,2}}}^{(n)}], [{}^{(i)}\Delta t_j^{(n)}], [c] \end{cases}$$

Durch die Wahl des Wertbereiches (engl. *Domain*) kann die Unsicherheit der jeweiligen Größen modelliert werden. Weiterhin kann garantiert werden, dass die korrekte Position in der Lösungsmenge enthalten ist, falls die Grenzen der Wertebereiche richtig ausgewählt werden. Jedoch ist die Zuweisung dieser Grenzen eine Herausforderung.

Die Grenzen der Schallgeschwindigkeit  $c$  hängen von dem Temperaturbereich in der untersuchten Umgebung ab. In Kapitel 4 wird ein Ansatz vorgestellt, um die Grenzen für die Mikrofonposition zu ermitteln. Weiterhin wird darauffolgend in Kapitel 5 beschrieben, wie die Grenzen der TDoA ermittelt werden können. Hierfür werden zwei alternative Ansätze verfolgt.

Um das CSP  $\mathcal{H}$  zu lösen, wird der Branch-and-Bound-Algorithmus *Set Inversion Via Interval Analysis* (SIVIA) aus Kapitel 2.4.5 mit einem Forward-Backward Kontraktor aus Kapitel 2.4.6 verwendet und mit  ${}^{(i)}\mathcal{C}_j^{(n)}$  bezeichnet. Die Idee besteht darin, mit einer initialen Box  $[\mathbf{x}]$  zu starten, die die gesamte Dimension der betrachteten Umgebung beinhaltet. Im Anschluss wird eine Berechnung auf  $[\mathbf{x}]$  mit dem Kontraktor  ${}^{(i)}\mathcal{C}_j^{(n)}$  durchgeführt, der inkonsistente Bereiche von  $[\mathbf{x}]$  entfernt.

Das formulierte CSP beschreibt die Position einer einzelnen Soundquelle  $s_i$  bezogen auf ein Mikrofonpaar  $j$  an der Messposition mit Index  $n$ . Jedoch muss die Position der Soundquellen  $s_i$  basierend auf den verbleibenden Gleichungen der anderen Mikrofonpaare einbezogen werden. Hierfür müssen im ersten Schritt alle Gleichungen identifiziert werden, die einen Zusammenhang mit der Soundquelle  $s_i$  und von der Messposition  $n$  besitzen. Diese Identifizierung wird in Kapitel 5 untersucht.

Im Anschluss wird ein einzelner Kontraktor  ${}^{(i)}\mathcal{C}^{(n)}$  für alle Mikrofonpaare definiert, indem die Schnittmenge der korrespondierenden Kontraktoren  ${}^{(i)}\mathcal{C}_j^{(n)}$  berechnet wird:

$$\text{Kontraktor Quelle } s_i: {}^{(i)}\mathcal{C}^{(n)} = \bigcap_{j=1}^{n_p} {}^{(i)}\mathcal{C}_j^{(n)}. \quad (3.12)$$

Um die entwickelte Methode zu vereinfachen, wird die Kartierung der Soundquelle in einer schematischen Skizze in Abbildung 3.3 dargestellt. Die Signale von zwei Soundquelle  $s_1$  und  $s_2$  werden durch einen mobilen Roboter an den Positionen  $n$  und  $n+1$  erfasst. Für jede Position und für jede Soundquelle wird ein Kontraktor  ${}^{(i)}\mathcal{C}^{(n)}$  ermittelt. Bei einer separaten Nutzung jedes Kontraktors auf  $[\mathbf{x}]$  entstehen vier Bereiche. Diese stehen zunächst in keiner mathematischen Beziehung zueinander.

Um dies zu lösen, werden zunächst alle Kontraktoren kombiniert, die mit der gleichen Messposition korrespondieren, indem ihre Vereinigung berechnet wird:

$$\text{Kontraktor an Position } n: \mathcal{C}^{(n)} = \bigcup_{i=1}^{n_s} {}^{(i)}\mathcal{C}^{(n)}. \quad (3.13)$$

Schließlich wird die Schnittmenge der beiden Kontraktoren an den Positionen  $n$  und  $n+1$  ermittelt, die zu einem finalen Kontraktor  $\mathcal{C}$  führt. Dieser Kontraktor wird verwendet, um die Lösung der Soundquellenkartierung zu berechnen, nachdem alle Messungen durchgeführt wurden:

$$\text{Finaler Kontraktor: } \mathcal{C} = \bigcap_{n=1}^{n_l} \mathcal{C}^{(n)}. \quad (3.14)$$

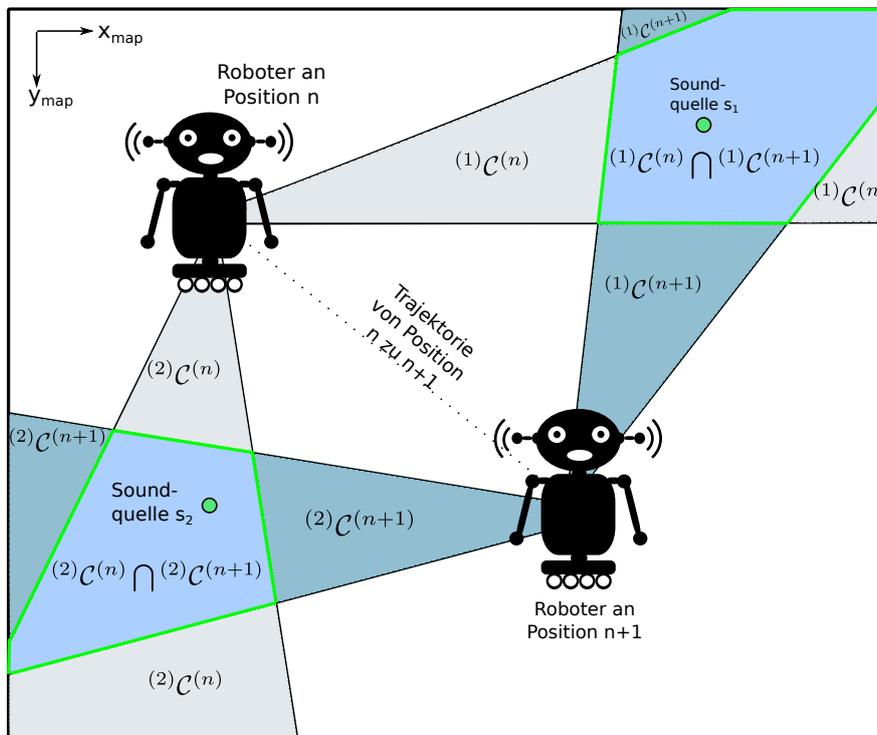


Abbildung 3.3: Vereinfachte Darstellung der Durchführung von Kontraktoren.

Das entwickelte Kartierungsverfahren (**IB-SSM**: engl. *Interval-Based Sound Source Mapping*) ist in Algorithmus 2 dargestellt.

Zuerst wird eine Messung an der Position  $n$  durchgeführt (vgl. Zeile 4) und das Intervall für die Mikrofonposition (vgl. Zeile 5) und für die TDoA (vgl. Zeile 6) geschätzt. Im Anschluss wird der Kontraktor aus Gleichung (3.13) ermittelt (vgl. Zeile 7). Danach werden aus der vorherigen Lösungsmenge (oder die Startlösung, falls die erste Messung durchgeführt wurde) inkonsistente Anteile entfernt, indem der Kontraktor  $\mathcal{C}^{(n)}$  mit SIVIA verwendet wird (vgl. Zeile 8). Im Gegensatz zur Bestimmung des finalen Kontraktors  $\mathcal{C}$  am Ende aller Messungen, hat dieser vorgestellte Algorithmus einen Vorteil. Es können direkt neue Informationen gewonnen werden, nachdem eine Messung durchgeführt wurde. Dies ermöglicht es, optimale Anfahrpositionen während der Laufzeit der Messungen zu bestimmen. Somit kann eine höhere Genauigkeit der Soundquellenkartierung realisiert werden.

**Algorithm 2** Pseudo-Code: IB-SSM**Eingabe:** Initialer Bereich:  $[x]$ , Anzahl Messpositionen:  $n_l$ **Ausgabe:** Reduzierter Bereich  $[x_c]$ 

```

1: for  $n = 1$  to  $n_l$  do
2:    $\mathbf{p} \leftarrow \text{getNextMeasurementPosition}(n)$ 
3:    $\text{driveToPosition}(\mathbf{p})$ 
4:    $\mathbf{z} \leftarrow \text{takeMeasurement}(\mathbf{p})$ 
5:    $[\mathbf{m}] \leftarrow \text{calcIntMicPos}(\mathbf{p})$  ▷ vgl. Kapitel 4
6:    $[\mathbf{t}] \leftarrow \text{calcIntTDoA}(\mathbf{z})$  ▷ vgl. Kapitel 5
7:    $\mathcal{C} \leftarrow \text{buildContractor}([\mathbf{m}], [\mathbf{t}])$  ▷ vgl. Kapitel 3.3
8:    $[x_c] \leftarrow \text{contractBoxes}(\mathcal{C}, [x])$  ▷ vgl. Kapitel 3.3
9:    $[x] \leftarrow [x_c]$ 
10: end for

```

### 3.4 Behandlung von akustischen Reflexionen

Im vorherigen Abschnitt wurde der Ablauf des Soundkartierungsverfahren vorgestellt. Im aktuellen Abschnitt wird das Verfahren auf akustische Reflexionen erweitert. Wie in Kapitel 2.1.3 dargestellt, erreicht das ausgesendete Signal  $s$  einen Raumpunkt  $x$  nicht nur über den direkten Propagationpfad. Stattdessen resultieren Reflexionen in Abhängigkeit der Geometrie des Raumes an Wänden und Objekten. Nach Durchführung von Verfahren zur Richtungsschätzung aus Kapitel 2.2.3 erscheint die Anzahl von aktiven Quellen aufgrund der Reflexionen größer als die tatsächliche Anzahl. Zur Illustrierung des Soundkartierungsverfahren dient ein vereinfachtes Beispiel in Abbildung 3.4.

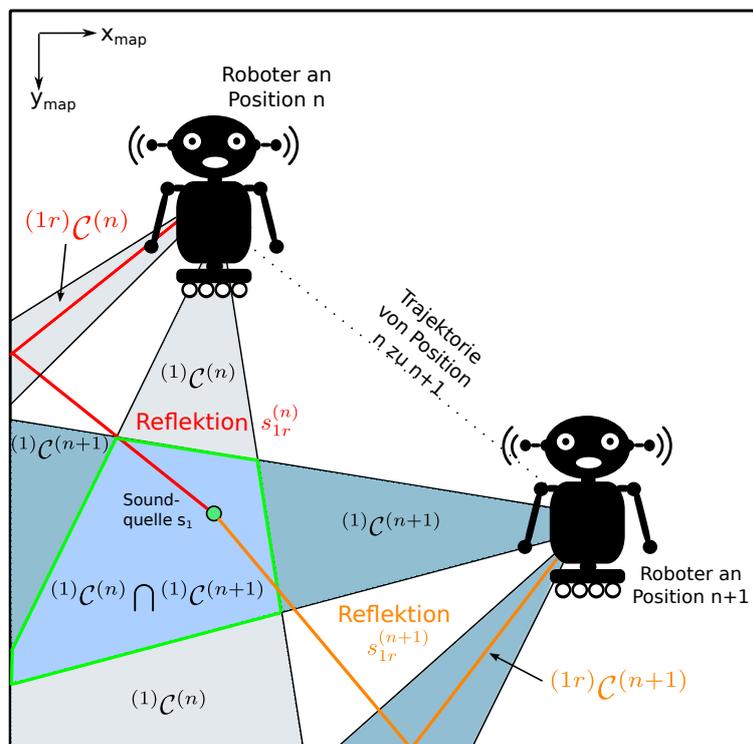


Abbildung 3.4: Behandlung von Reflexionen.

Hierbei wird die Soundquelle  $s_1$  an zwei Messpositionen  $n$  und  $n+1$  beobachtet. Zusätzlich wird an Position  $n$  eine Richtungsschätzung aufgrund einer Reflexion  $s_{1r}^{(n)}$ , und an Position  $n+1$  eine Reflexion  $s_{1r}^{(n+1)}$  wahrgenommen. Der Subskript  $r$  bezeichnet die Charakterisierung als Reflexion. Als Resultat werden an jeder Position zwei potenzielle Soundquellen wahrgenommen. Daher resultierten nach Gleichung (3.12) insgesamt vier Kontraktoren  $(1)\mathcal{C}^{(n)}$ ,  $(1r)\mathcal{C}^{(n)}$ ,  $(1)\mathcal{C}^{(n+1)}$  und  $(1r)\mathcal{C}^{(n+1)}$ . Der finale Kontraktor nach Gleichung (3.14) ergibt sich folgendermaßen:

$$\mathcal{C} = \underbrace{(1)\mathcal{C}^{(n)} \cap (1)\mathcal{C}^{(n+1)}}_{\emptyset} \cup \underbrace{(1)\mathcal{C}^{(n)} \cap (1r)\mathcal{C}^{(n+1)}}_{\emptyset} \cup \underbrace{(1r)\mathcal{C}^{(n)} \cap (1)\mathcal{C}^{(n+1)}}_{\emptyset} \cup \underbrace{(1r)\mathcal{C}^{(n)} \cap (1r)\mathcal{C}^{(n+1)}}_{\emptyset}. \quad (3.15)$$

Aufgrund der Abhängigkeit der Position der Reflexion von der beobachteten Messposition resultieren für die Schnittmengenberechnung jeweils leere Mengen.

### 3.5 Ausblick - Behandlung von Ausreißern

Als Randbedingung für das in Kapitel 3.3 entwickelte Soundkartierungsverfahren ist es notwendig, dass die Soundquellen während der Messungen aktiv sind. Falls die Quellen in diesem Zeitraum kein Signal aussenden oder die wahrgenommene Lautstärke der Signalquelle zu gering ist, kann keine erfolgreiche Richtungsschätzung durchgeführt werden. Abbildung 3.5 stellt ein Beispiel dar, bei dem die Soundquelle  $s_1$  von der Position  $n$  nicht beobachtet wurde. Nach Berechnung der Intersektion der Beobachtung beider Positionen  $n$  und  $n+1$  resultiert ein gemeinsamer Bereich, der mit der Soundquelle  $s_2$  assoziiert ist. Somit wird die Beobachtung von Position  $n$  zur Soundquelle  $s_1$  als Reflexion aufgefasst. Im Weiteren wird das Soundkartierungsverfahren für Soundquellen berücksichtigt,

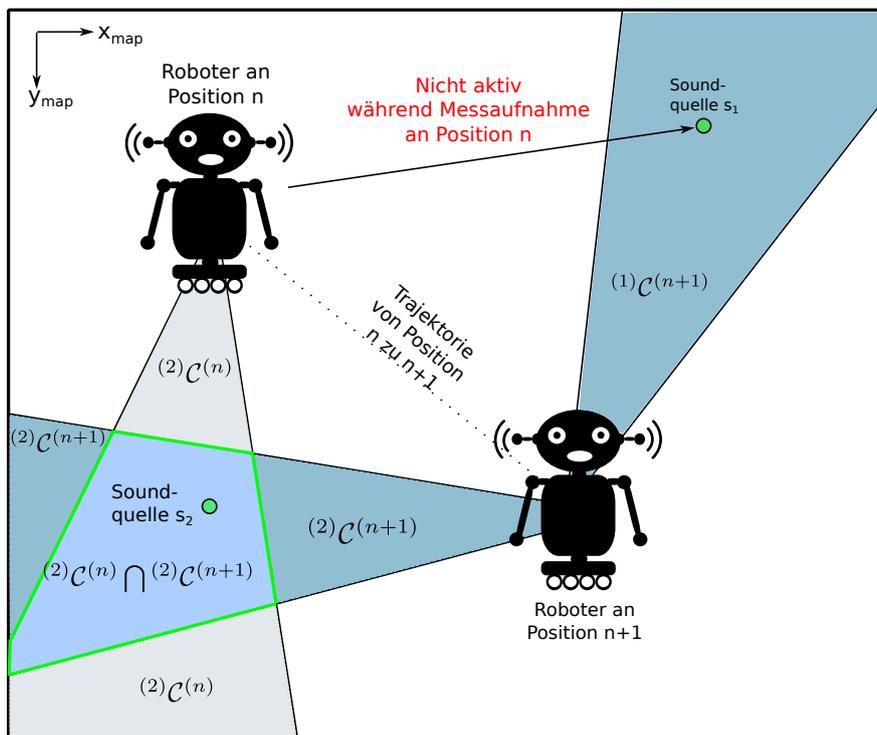


Abbildung 3.5: Behandlung von Ausreißern.

die nicht zu jedem Messzeitpunkt aktiv sind. Dabei wird im Folgenden eine fehlende Messung als Ausreißer interpretiert. Ein Ansatz mit Ausreißern umzugehen ist in [Jau09] dargestellt. Die Grundidee besteht darin, die reguläre Schnittmenge durch die sogenannte  $q$ -relaxte Schnittmenge zu ersetzen. Dabei ergeben sich bei einer Menge von  $n$  Elementen  $n - q$  Schnittmengen. Abbildung 3.6 veranschaulicht die  $q$ -relaxte Schnittmenge für die Mengen  $\mathbb{X}_1$ ,  $\mathbb{X}_2$ ,  $\mathbb{X}_3$  und  $\mathbb{X}_4$ . Diese Mengen besitzen keine gemeinsame Schnittmenge:

$$\mathbb{X}_1 \cap \mathbb{X}_2 \cap \mathbb{X}_3 \cap \mathbb{X}_4 = \emptyset \quad (3.16)$$

Die Bildung der 1-relaxte Schnittmenge  $\cap^1$  repräsentiert die Schnittmenge von mindestens drei Mengen:

$$\cap^1 = (\mathbb{X}_1 \cap \mathbb{X}_2 \cap \mathbb{X}_3) \cup (\mathbb{X}_1 \cap \mathbb{X}_2 \cap \mathbb{X}_4) \cup (\mathbb{X}_1 \cap \mathbb{X}_3 \cap \mathbb{X}_4) \cup (\mathbb{X}_2 \cap \mathbb{X}_3 \cap \mathbb{X}_4). \quad (3.17)$$

Dahingegen ist die 3-relaxte Schnittmenge  $\cap^3$  assoziiert mit der Vereinigung aller 4 Mengen.

Die Erweiterung des Ansatzes auf die  $q$ -relaxte Schnittmenge wirkt sich darauf aus, wie Reflexionen behandelt werden. Daher muss für den Einsatz in einer realen Umgebungen ein Kompromiss zwischen Ausreißer- und Reflexionsbehandlung getroffen werden.

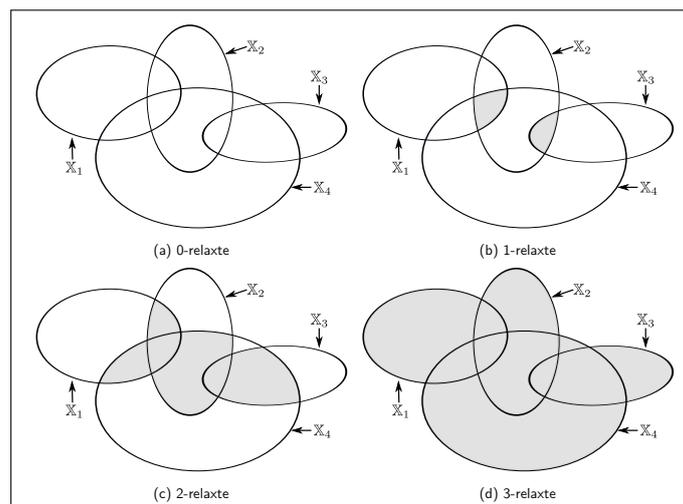


Abbildung 3.6:  $q$ -Relaxte Schnittmenge



# Kapitel 4

## Lokalisierung Mikrofonposition

Um im späteren Verlauf die Positionen der Soundquellen zu kartieren, müssen zunächst die Positionen der Mikrofone innerhalb einer metrischen Karte ermittelt werden. Hierbei werden zwei Aspekte modelliert:

1. Relative Mikrofonposition auf mobiler Plattform
2. Globale Lokalisierung innerhalb einer Karte

Im ersten Schritt wird in Abschnitt 4.1 die relative Mikrofonposition im Koordinatensystem der mobilen Plattform beschrieben. Im Anschluss wird in Abschnitt 4.2 die Position der Plattform innerhalb einer globalen metrischen Karte modelliert.

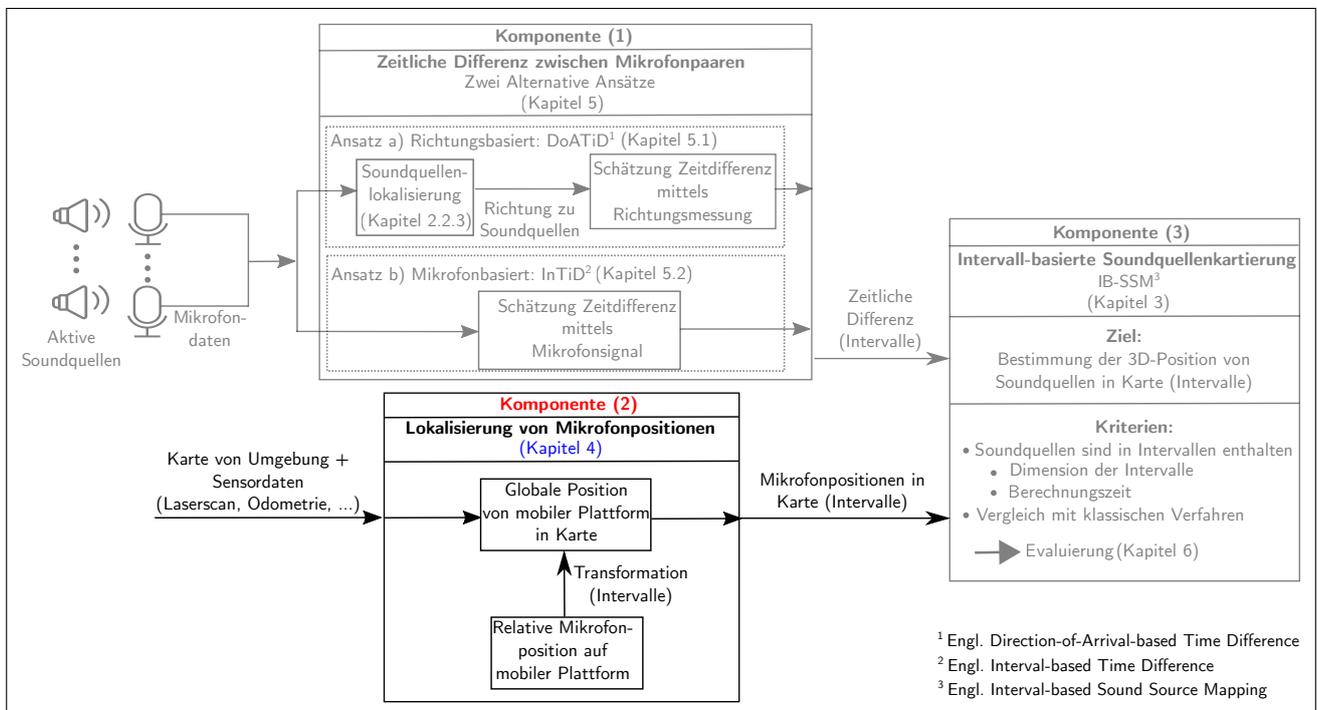


Abbildung 4.1: Systematischer Aufbau des entwickelten Gesamt-Konzeptes. Der relevante Teil für dieses Kapitel ist hervorgehoben.

## 4.1 Relative Mikrofonposition

Zunächst wird die Position  $\mathbf{x}_{m_i} \in \mathbb{R}^3$  von Mikrofon  $m_i$  bezogen auf das Koordinatensystem  $\text{KS}^{\text{mob}}$  der mobilen Plattform beschrieben (vgl. Abbildung 4.2). Dabei beschreibt  $T_{m_i}^{\text{mob}}$  die Transformation zwischen Koordinatensystem  $\text{KS}^{\text{mob}}$  und dem Koordinatensystem  $\text{KS}^{m_i}$  zum zugehörigen Mikrofon  $m_i$ . Durch Verwendung einer homogenen Transformation wird der Zusammenhang zwischen den Koordinatensystemen folgendermaßen angegeben:

$$\mathbf{x}_{m_i}^{\text{mob}} = T_{m_i}^{\text{mob}} \cdot \mathbf{x}_{m_i} \quad (4.1)$$

mit:

$$T_{m_i}^{\text{mob}} = \begin{bmatrix} \mathbf{R}_{m_i}^{\text{mob}} & \mathbf{t}_{m_i}^{\text{mob}} \\ 0 & 1 \end{bmatrix}. \quad (4.2)$$

Hierbei bezeichnet  $\mathbf{t}_{m_i}^{\text{mob}}$  die Translation zwischen den beiden Koordinatensystemen und  $\mathbf{R}_{m_i}^{\text{mob}}$  die Rotation. Die Elemente  $\mathbf{x}_{m_i}^{\text{mob}} \in \mathbb{R}^4$  sind in homogenen Koordinaten angegeben:

$$\mathbf{x}_{m_i}^{\text{mob}} = (x_{m_i}, y_{m_i}, z_{m_i}, 1)^T. \quad (4.3)$$

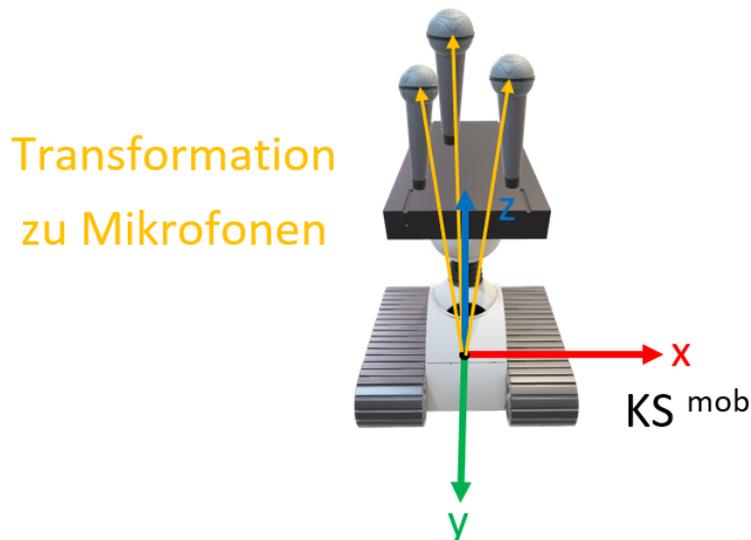


Abbildung 4.2: Relative Position von Mikrofonen innerhalb des Koordinatensystems einer mobilen Plattform.

Die Transformation  $T_{m_i}^{\text{mob}}$  kann innerhalb einer Kalibrierung oder manuell ermittelt werden. Die Orientierung  $\mathbf{R}_{m_i}^{\text{mob}}$  kann vernachlässigt werden, falls omnidirektionale Mikrofon verwendet werden. Um die Translation  $\mathbf{t}_{m_i}^{\text{mob}}$  zu ermitteln, muss ein Referenzpunkt am Mikrofon angegeben werden. Da eine Mikrofonmembran eine räumliche Ausdehnung besitzt, ist dieser Referenzpunkt nicht eindeutig festlegbar. Abbildung 4.3 veranschaulicht diesen Effekt. Weiterhin enthält ein Mikrofonsignal keine Information über den Auftreffpunkt der Schallwellen auf der Mikrofonmembran. Diese Eigenschaft unterscheidet Mikrofone von Kamerasensoren, die Informationen über den Einfallsort der Lichtwellen erhalten.



Abbildung 4.3: Die Membran eines Mikrofons besitzt eine räumliche Ausdehnung. Die Modellierung durch einen Punkt-Receiver führt zu Fehlern.

Um diese Herausforderung in der vorliegenden Arbeit zu lösen, wird zunächst die Transformation zu einem beliebigen Referenzpunkt auf der Mikrofonmembran ermittelt. Im Anschluss wird die Mikrofonposition durch eine Intervallbox repräsentiert. Diese besitzt ihren Ursprung auf der Referenzmarkierung. Die Dimension der Intervallbox ist abhängig von der ermittelten Ausdehnung der Mikrofonmembran  $[\mathbf{x}_{mem}]$  und der Genauigkeit der Kalibrierung  $[\mathbf{x}_g]$ :

$$\begin{aligned} [\mathbf{x}_{mem}] &= ([\underline{x}_{mem}, \bar{x}_{mem}], [\underline{y}_{mem}, \bar{y}_{mem}], [\underline{z}_{mem}, \bar{z}_{mem}]) \\ [\mathbf{x}_g] &= ([\underline{x}_g, \bar{x}_g], [\underline{y}_g, \bar{y}_g], [\underline{z}_g, \bar{z}_g]). \end{aligned} \quad (4.4)$$

Besondere Aufmerksamkeit muss den ermittelten Intervallgrenzen  $[\mathbf{x}_{mem}]$  und  $[\mathbf{x}_g]$  zukommen. Auch für eine schlechte Kalibrierung kann eine garantierte Schätzung der Soundquellenkartierung angegeben werden. Diese Schätzung beinhaltet die wahre Position der Soundquellen, wenn korrekte Intervallgrenzen ausgewählt wurden.

Somit ergibt sich die Beschreibung der erweiterten Translation  $[\mathbf{t}_{m_i}^{mob}]$  eines Mikrofons  $m_i$  folgendermaßen:

$$[\mathbf{t}_{m_i}^{mob}] = \mathbf{t}_{m_i}^{mob} + [\mathbf{x}_{mem}] + [\mathbf{x}_g]. \quad (4.5)$$

## 4.2 Globale Lokalisierung innerhalb einer Karte

Nachdem in Abschnitt 4.1 die Mikrofonposition in einem relativen Koordinatensystem der mobilen Plattform beschrieben wurde, erfolgt im Folgenden die Beschreibung in einem globalen Koordinatensystem. Dabei bezeichnet  $T_{mob}^{map}$  die Transformation zwischen der mobilen Plattform  $KS^{mob}$  und dem globalen Koordinatensystem  $KS^{map}$  einer metrischen Karte (vgl. Abbildung 4.4).

Die Transformation wird analog zu Gleichung (4.1) definiert:

$$\mathbf{x}_{m_i}^{map} = T_{mob}^{map} \cdot \mathbf{x}_{m_i}^{(mob)} \quad (4.6)$$

mit

$$T_{m_i}^{mob} = \begin{bmatrix} \mathbf{R}_{mob}^{map} & \mathbf{t}_{mob}^{map} \\ 0 & 1 \end{bmatrix}. \quad (4.7)$$

$\mathbf{t}_{mob}^{map}$  beschreibt die Translation des mobilen Koordinatensystems  $KS^{mob}$  zum globalen Koordinatensystem der metrischen Karte  $KS^{map}$ .  $\mathbf{R}_{mob}^{map} \in SO(3)$  beschreibt die Rotation.

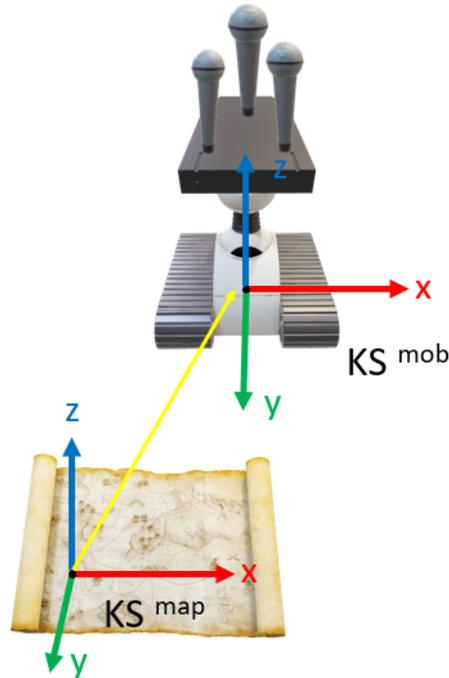


Abbildung 4.4: Position einer mobilen Plattform innerhalb einer Karte

In folgender Arbeit ist die Bewegung der mobilen Plattform auf die XY-Ebene beschränkt, wodurch eine Rotation um die z-Achse resultiert. Somit wird die Rotation  $R_{rob}^{map}$  folgendermaßen angegeben:

$$R_{mob}^{map} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4.8)$$

In Abhängigkeit des verwendeten Lokalisierungsalgorithmus und der Repräsentation der metrischen Karte ist die Position der mobilen Plattform nicht fehlerfrei ermittelbar. In dieser Arbeit wird eine metrische Repräsentation der Karte in Form einer Belegungskarte angenommen (vgl. Kapitel 2.3). Dabei stellt Abbildung 4.5 eine Belegungskarte mit einer Auflösung von  $dx \cdot dy$  dar.

Zur Lösung der fehlerbehafteten Lokalisierung wird in der Arbeit die geschätzte Position der mobilen Plattform mit einer Intervallbox erweitert. Dabei sind die Grenzen der Intervallbox abhängig von der Dimension der Gitterelemente der Rasterkarte  $[\mathbf{x}_{rkt}]$  und der Unsicherheit des Lokalisierungsverfahrens  $[\mathbf{x}_{lt}]$ :

$$\begin{aligned} [\mathbf{x}_{rkt}] &= ([x_{rkt}, \bar{x}_{rkt}], [y_{rkt}, \bar{y}_{rkt}], [z_{rkt}, \bar{z}_{rkt}]) \\ [\mathbf{x}_{lt}] &= ([x_{lt}, \bar{x}_{lt}], [y_{lt}, \bar{y}_{lt}], [z_{lt}, \bar{z}_{lt}]). \end{aligned} \quad (4.9)$$

Es ergibt sich somit die Beschreibung der erweiterten Translation  $[\mathbf{t}_{mob}^{map}]$  der mobilen Plattform folgendermaßen:

$$[\mathbf{t}_{mob}^{map}] = \mathbf{t}_{mob}^{map} + [\mathbf{x}_{lt}] + [\mathbf{x}_{rkt}]. \quad (4.10)$$

Im letzten Schritt wird die Unsicherheit der Orientierung der mobilen Plattform innerhalb der Karte betrachtet.

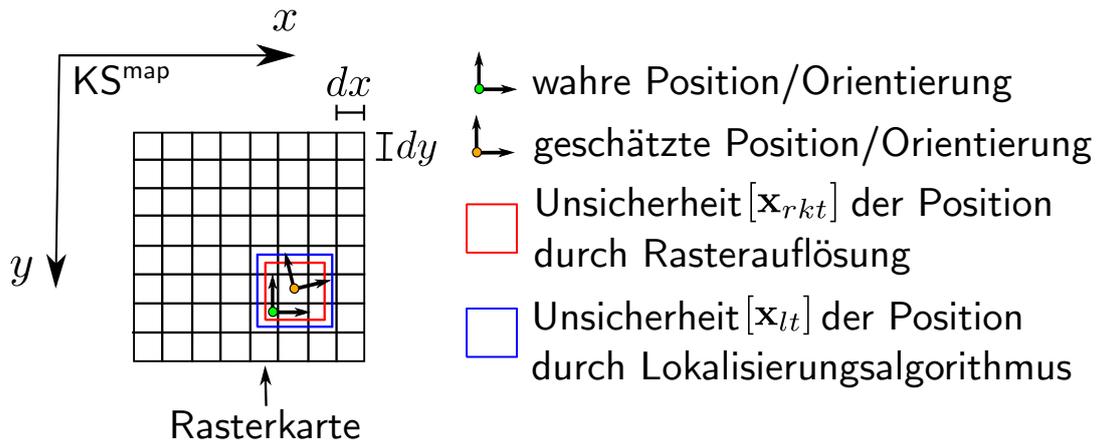


Abbildung 4.5: Vergleich der Positionsschätzung (orange) und wahren Position (grün) innerhalb einer globalen Rasterkarte und Darstellung der Unsicherheiten der Position durch die Rasterauflösung und den Lokalisierungsalgorithmus.

Hierbei wird angenommen, dass der Winkel  $\alpha$  aus Gleichung (4.8) nur mit einer Unsicherheit bestimmt werden kann. Diese ist, wie die Translation aus Gleichung (4.10), von dem Lokalisierungsalgorithmus  $[\mathbf{x}_{l\alpha}]$  und der Rasterauflösung der Karte  $[\mathbf{x}_{rk\alpha}]$  abhängig:

$$\begin{aligned} [\mathbf{x}_{rk\alpha}] &= ([x_{rk\alpha}, \bar{x}_{rk\alpha}], [y_{rk\alpha}, \bar{y}_{rk\alpha}], [z_{rk\alpha}, \bar{z}_{rk\alpha}]) \\ [\mathbf{x}_{l\alpha}] &= ([x_{l\alpha}, \bar{x}_{l\alpha}], [y_{l\alpha}, \bar{y}_{l\alpha}], [z_{l\alpha}, \bar{z}_{l\alpha}]). \end{aligned} \quad (4.11)$$

Somit wird  $\alpha$  folgendermaßen erweitert:

$$[\alpha, \bar{\alpha}] = \alpha + [\mathbf{x}_{l\alpha}] + [\mathbf{x}_{rk\alpha}]. \quad (4.12)$$

Hieraus ergibt sich somit eine intervall-erweiterte Rotation  $[R]_{mob}^{map}$ :

$$[R]_{mob}^{map} = \begin{pmatrix} \cos([\alpha]) & -\sin([\alpha]) & 0 \\ \sin([\alpha]) & \cos([\alpha]) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4.13)$$

Schließlich werden die Ergebnisse zusammengefasst. Die Position eines Mikrofons lässt sich daher folgendermaßen darstellen:

$$[\mathbf{x}_{m_i}^{(map)}] = [T_{mob}^{map}] \cdot [T_{m_i}^{mob}] \cdot \mathbf{x}_{m_i}. \quad (4.14)$$

### 4.3 Ausblick - Erweiterung auf Multi-Hypothesen

In realen Experimenten ist nicht gewährleistet, dass eine eindeutige Schätzung bei der Lokalisierung der Plattform resultiert. Insbesondere in symmetrischen Umgebungen können mehrere Hypothesen auftreten.

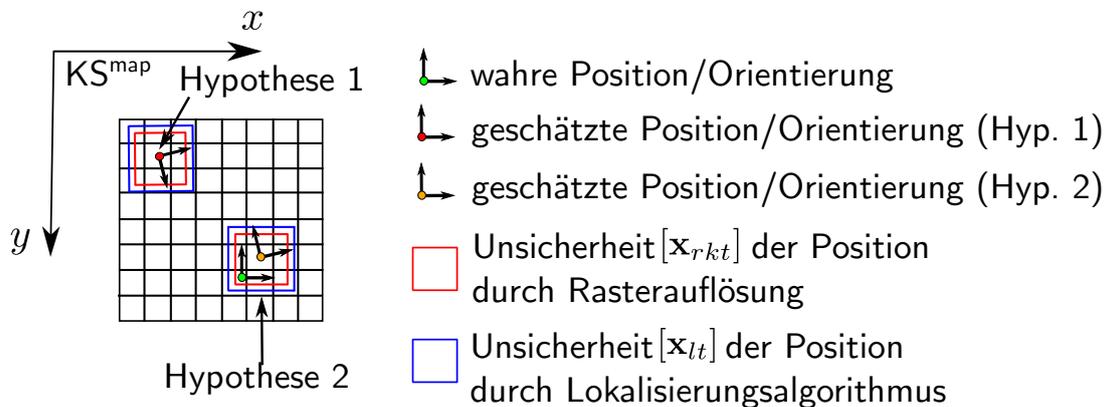


Abbildung 4.6: Vergleich der Positionsschätzung (orange) und wahren Position (grün) innerhalb einer globalen Rasterkarte und Darstellung der Unsicherheiten der Position durch die Rasterauflösung und den Lokalisierungsalgorithmus.

Daher kann statt einer einzelnen Hypothese mehrere verwendet werden. Ein Beispiel für zwei Hypothesen ist in Abbildung 4.6 gegeben.

### 4.4 Ausblick - Erweiterung auf garantierte Lokalisierung

Das bestehende Verfahren kann auf eine garantierte Lokalisierung erweitert werden. Hierfür können Arbeiten von Langerwisch herangezogen werden, in denen ein intervall-basiertes Verfahren zur Kartierung und Lokalisierung mobiler Serviceroboter unter der Annahme begrenzter Sensorfehler vorgestellt wurde [LW12]. Hierfür wurde eine raster-basierte Karte zur Darstellung der Umgebung verwendet. Kernpunkte sind die Einführung eines intervall-basierten Bewegungsmodells basierend auf Odometriedaten und die Einführung eines inversen Sensormodells für die Beschreibung von Laserscan-Messungen.

Langerwisch konnte zeigen, dass eine garantierte Kartierung und Lokalisierung erzielt werden kann. Jedoch sind die Ergebnisse konservativer als klassische probabilistische Verfahren.

# Kapitel 5

## Zeitliche Differenzschätzung

Die zeitliche Differenz - TDoA (vgl. Abschnitt 2.2.3.1) - zwischen Mikrofonpaaren wird benötigt, um auf die räumliche Position einer Soundquelle zu schließen. In diesem Kapitel wird dargestellt, wie die TDoA für jedes Mikrofonpaar und jede aktive Soundquelle  $s_i$  geschätzt wird. Dabei werden Diskretisierungsfehler und Messrauschen der Mikrofon-signale berücksichtigt.

In dieser Arbeit werden zwei alternative Ansätze *DoATiD* und *InTiD* vorgestellt, um die TDoA und ihre Unsicherheit zu schätzen:

- (a) Richtungs-basierte Methode - *DoATiD* (Kapitel 5.1)
  - (1) Universelles Verfahren - unabhängig von SSL-Algorithmus und Signalcharakteristik (Kapitel 5.1.1)
  - (2) Konkretisierung: Einsatz des SRP-PHAT-Verfahrens (Kapitel 5.1.5)
- (b) Mikrofon-signal-basierte Methode - *InTiD* (Kapitel 5.2)

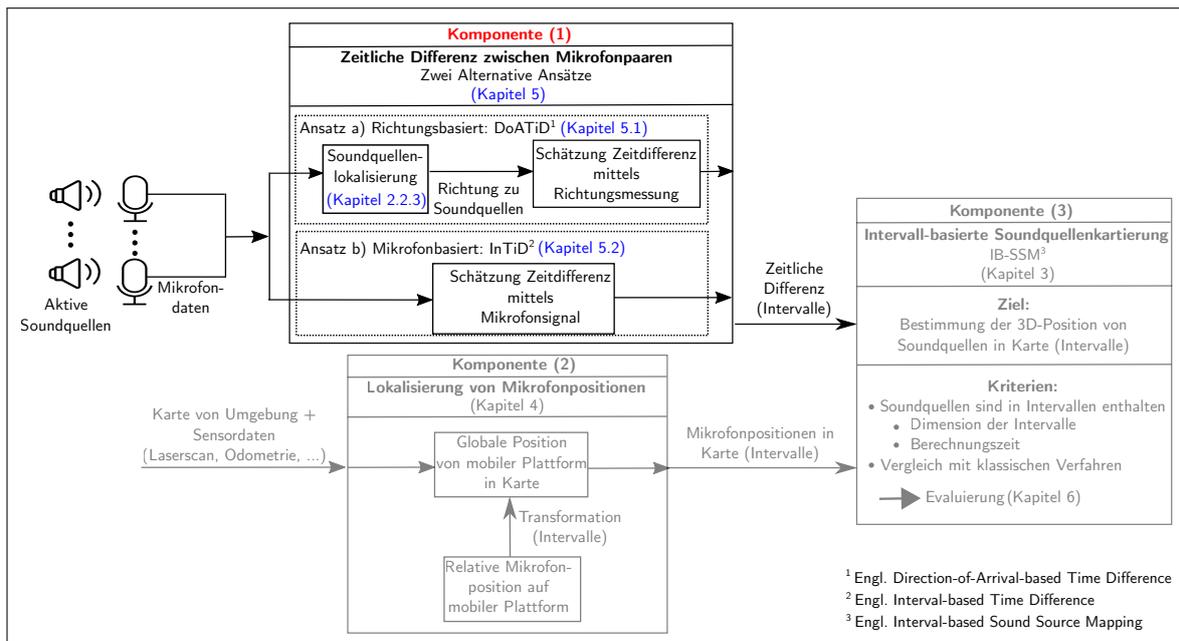


Abbildung 5.1: Systematischer Aufbau des entwickelten Gesamt-Konzeptes. Der relevante Teil für dieses Kapitel ist hervorgehoben.

*DoATiD* basiert auf Ergebnissen der Richtungsschätzung von klassischen Verfahren aus dem Bereich der Soundquellenlokalisierung (vgl. Abschnitt 2.2.3). Im Weiteren werden die Begriffe Richtung und DoA (vgl. Abschnitt 2.2.3.1) synonym verwendet. Derzeit sind Software-Bibliotheken frei zur Verfügung, um DoA-Schätzungen von Soundquelle zu ermitteln, z. B. HARK [Miz+11] oder ODAS [Gro+21]. Darüber hinaus ist es denkbar, dass in naher Zukunft kommerzielle Mikrofonarrays dem Nutzer die DoA-Schätzungen als Sensorausgabe zur Verfügung stellen.

Zunächst wird in Abschnitt 5.1.1 für *DoATiD* ein universelles Verfahren vorgestellt (1). Dieses ist unabhängig von dem verwendeten SSL-Algorithmus und der Signalcharakteristik der Soundquellen. Somit kann der entwickelte Ansatz direkt auf den Richtungsschätzungen von bestehenden Software-Bibliotheken angewendet werden.

Da in den DoA-Schätzverfahren direkt oder indirekt die TDoA ermittelt wird, widmet sich im Anschluss der Ansatz (2) in Abschnitt 5.1.5 mit der Konkretisierung von *DoATiD* für ein ausgewähltes SSL-Verfahren. Indem Kenntnisse über das bestehende Verfahren vorliegen, kann die Genauigkeit der DoA geschätzt werden. In dieser Arbeit wird das SRP-PHAT-Verfahren (vgl. Abschnitt 2.2.3.4) verwendet. Es wurde ausgewählt, da ein direkter Zusammenhang zwischen der TDoA und DoA vorliegt.

Um die Robustheit gegenüber Messrauschen und Fehlmessungen in realen Experimenten zu steigern, wird die DoA-Schätzung von *DoATiD* als Beobachtung in einem rekursiven Zustandsschätzverfahren integriert (vgl. Abschnitt 2.3.3.1). Die resultierenden Zustände (oder gefilterten Beobachtungen) werden schließlich für die Schätzung der Intervalle der TDoAs verwendet.

Die beiden Verfahren (1) und (2) von *DoATiD* verwenden für die DoA-Schätzung mathematische Methoden außerhalb der Intervall-Arithmetik. So wird beim SRP-PHAT- und MUSIC-Verfahren die Kreuzkorrelation berechnet. Unter mathematischen Kriterien ist jedoch ein rein intervall-basierter Ansatz notwendig, um mathematisch garantierte Intervallgrenzen zu ermitteln. Dieser Ansatz muss direkt Methoden der Intervallarithmetik auf die Daten der Mikrofone anwenden. Daher wird im zweiten Teil dieses Kapitels der mikrofonsignal-basierte Ansatz *InTiD* vorgestellt.

## 5.1 Richtungs-basierte Zeitdifferenzschätzung (DoATiD)

Im Folgenden wird das neue richtungsbasiertes Verfahren DoATiD (engl. *Direction-of-Arrival-based Time Difference*) vorgestellt, um die TDoA zu schätzen.

Zunächst wird in Abschnitt 5.1.1 ein Verfahren vorgestellt, das universell eingesetzt werden kann und sowohl unabhängig von dem Algorithmus der DoA-Schätzung als auch von der Charakteristik der Soundsignale ist. Hierfür wird zunächst das Gesamtsystem und im Anschluss die jeweiligen Komponenten vorgestellt.

Im weiteren Verlauf wird in Abschnitt 5.1.5 der Einsatz eines ausgewähltes DoA-Schätzverfahrens betrachtet und eine optimierte Schätzung der TDoA durchgeführt.

### 5.1.1 Ansatz (1) - Universelles Verfahren (DoATiD)

In diesem Abschnitt wird ein universeller Ansatz beschrieben, der TDoA-Intervalle für jedes Mikrofonpaar berechnet. Dieser Ansatz basiert auf den Ergebnissen der DoA-Schätzung von Verfahren aus dem Bereich der Soundquellenlokalisierung (SSL).

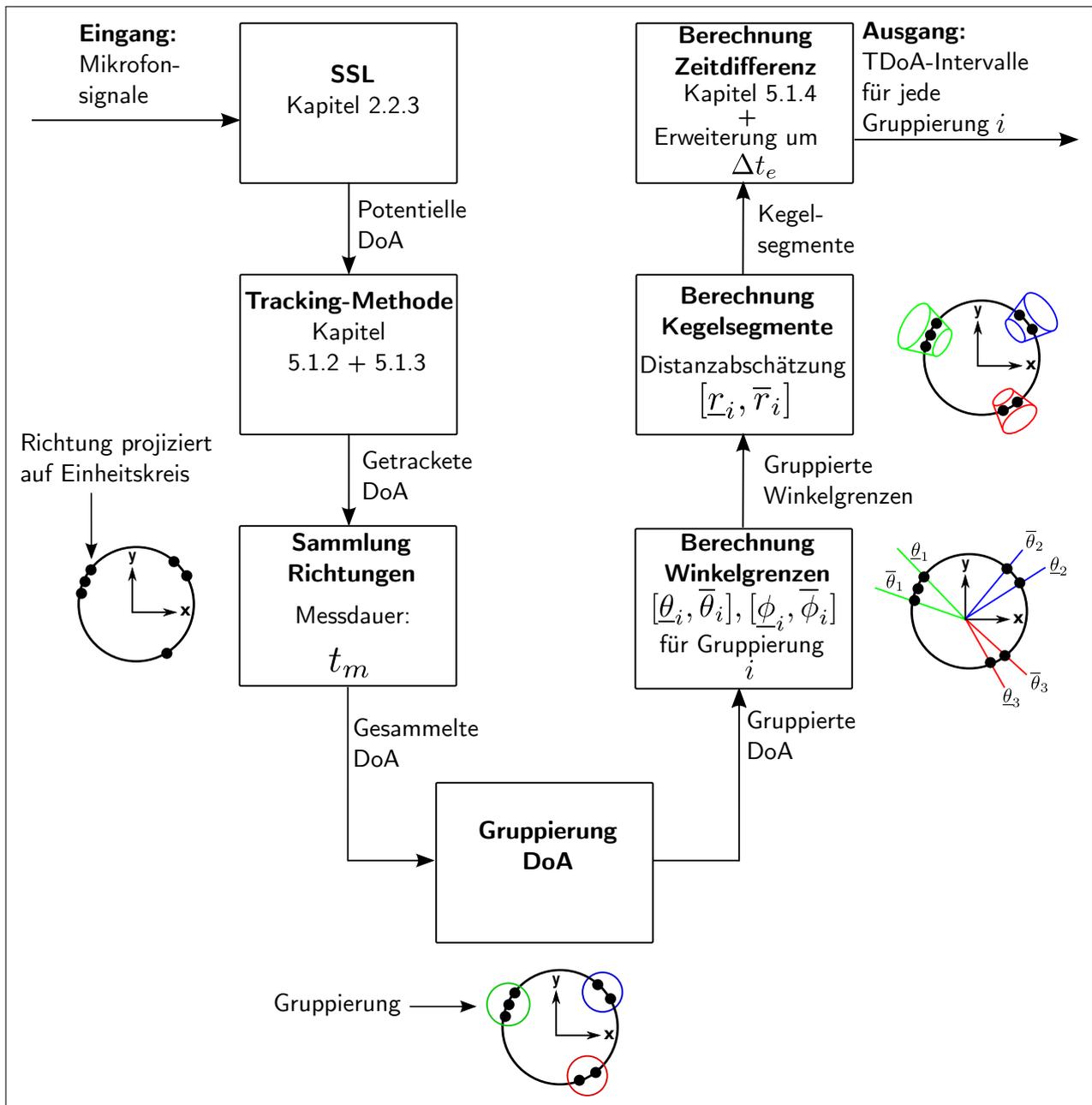


Abbildung 5.2: Universelles Verfahren aus Abschnitt 5.1.1: Berechnungsschritte für die Intervalle der TDoA auf Basis von DoA-Schätzungen.

Im Folgenden wird keine Annahme über die Signalcharakteristik der Soundquelle und den SSL-Algorithmus getroffen. Somit können ohne Beschränkung der Allgemeinheit sowohl binaurale Verfahren (vgl. Abschnitt 2.2.2) als auch Mikrofonarray-Verfahren (vgl. 2.2.3) verwendet werden. Weiterhin können existierende SSL-Systeme auf diesen Ansatz erweitert werden.

Die einzelnen Schritte des universellen Verfahrens sind in Abbildung 5.2 visualisiert und werden im Folgenden vorgestellt.

Zunächst wird auf Basis der Mikrofonssignale und des SSL-Verfahrens zum Zeitpunkt  $k$  eine Menge von DoAs berechnet. Diese wird im Folgenden als potenzielle DoAs (oder potenzielle Richtungen) bezeichnet und durch  ${}^p DoA^k$  notiert:

$${}^p DoA^k = \{{}^p DoA_1^k, {}^p DoA_2^k, \dots\}. \quad (5.1)$$

Dabei kennzeichnet  $p$ , dass die Beobachtung als potentiell angesehen wird. Das  $i$ -te Element  ${}^p DoA_i$  ist durch die Winkel  $(\theta_i, \phi_i)$  charakterisiert. Die verwendete Winkelkonvention ist in Abbildung 5.3 dargestellt. In Abhängigkeit des verwendeten SSL-Verfahrens wird für jede Richtung  ${}^p DoA_i$  die wahrgenommene Energie  $\Lambda_i^k$  ermittelt. So ist die Zuordnung der Energie beim Beamforming-Verfahren durch Gleichung (2.35), beim SRP-PHAT-Verfahren durch Gleichung (2.42) und beim MUSIC-Verfahren durch Gleichung (2.55) gegeben. Somit setzt sich das Ergebnis  $\psi^k$  des SSL-Verfahrens zum Zeitpunkt  $k$  aus folgenden Tupeln zusammen:

$$\psi^k = \left\{ \underbrace{({}^p DoA_1^k, \Lambda_1^k)}_{\psi_1^k}, \underbrace{({}^p DoA_2^k, \Lambda_2^k)}_{\psi_2^k}, \dots \right\}. \quad (5.2)$$

Für das universelle Verfahren werden die Details der SSL-Verfahren nicht weiter betrachtet, sondern  $\psi^k$  als Beobachtungen eines Sensors interpretiert. Es wird jedoch eine Abschätzung der Genauigkeit dieser Beobachtung verwendet.

Für die weitere Verarbeitung sind nicht alle Elemente von  $\psi^k$  relevant, da diese Fehlmessungen enthalten und mit Messrauschen versehen sein können. Daher werden die potenziellen DoAs und die Energiewerte als Beobachtung einer rekursiven Zustandsschätzung aufgefasst. Im Anschluss wird eine Klassifizierung durchgeführt, ob die Elemente von  $\psi^k$  mit einer bekannten Soundquelle, einer neuen Soundquelle oder einer Fehlmessung assoziiert sind. Die Details der Zustandsschätzung und der Assoziation sind in Abschnitt 5.1.2 dargestellt.

Das Ergebnis des vorherigen Schrittes beinhaltet DoAs, die nicht mit einer Fehlmessung oder Rauschen assoziiert sind. Sie werden im Folgenden als getrackte DoAs bezeichnet und mit  ${}^t DoA^k$  notiert:

$${}^t DoA^k = \{{}^t DoA_1^k, {}^t DoA_2^k, \dots\}. \quad (5.3)$$

Dabei notiert  $t$  die Charakterisierung als getrackte (oder gefilterte) Beobachtung.

In realen Experimenten wird die Robustheit der SSL-Algorithmen gegenüber Rauschen verringert, indem verschiedene Messungen in einer definierten Zeitspanne  $t_m$  gesammelt werden. In dieser Arbeit wird angenommen, dass die jeweiligen Soundquellen für mindestens 20% dieser Zeitspanne aktiv sind. Somit resultiert für Messungen in dieser Zeitspanne eine Menge von DoAs. Diese werden als  ${}^g DoA$  bezeichnet:

$${}^g DoA = \left\{ \underbrace{{}^t DoA_1^k, {}^t DoA_2^k, \dots}_{\text{Zeitpunkt: } k}, \underbrace{{}^t DoA_1^{k+1}, {}^t DoA_2^{k+1}, \dots, \dots}_{\text{Zeitpunkt: } k+1} \right\}. \quad (5.4)$$

Dabei beschreibt  $g$  die Charakterisierung als gesammelte Beobachtungen.

Es wird angenommen, dass mehrere Soundquellen in dieser Zeit aktiv sein können. Daher muss im Weiteren die Menge  ${}^g DoA$  in unterschiedliche Cluster eingeteilt werden. Für das Clustering werden die DoA zunächst auf einen Einheitskreis projiziert. Im Anschluss erfolgt die Einteilung durch klassische

Verfahren wie das K-Means-Clustering [Mac+67]. Im Folgenden bezeichnet  ${}^cDoA_i$  die Menge der gesammelten DoAs, die dem Cluster  $i$  zugeordnet wurden.

Für jede Clustermenge  ${}^cDoA_i$  werden die Grenzen der Winkel  $[\theta_i, \bar{\theta}_i]$  und  $[\phi_i, \bar{\phi}_i]$  ermittelt. Dabei werden die minimalen und maximalen Werte verwendet. Besondere Aufmerksamkeit muss der Auflösung der verwendeten SSL-Algorithmen gegeben werden. Diese Information muss den Intervallen für  $\theta_i$  und  $\phi_i$  hinzugefügt werden.

Um Kollisionen mit Soundquellen zu vermeiden und den Einfluss der Geräusentwicklung der verwendeten mobilen Plattform zu reduzieren, wird ein minimaler Abstand  $\underline{r}_i$  definiert. Dieser gibt an, ab welcher Minimaldistanz eine Soundquelle lokalisiert werden kann.

Weiterhin wird angenommen, dass die Messungen in einem abgeschlossenen Raum durchgeführt werden und sich dort alle akustisch wahrnehmbaren Quellen befinden. Somit kann die maximale Distanz  $\bar{r}_i$  durch die Struktur des Raumes abgeschätzt werden. Eine exemplarische Darstellung der Grenzen der Distanz  $[\underline{r}_i, \bar{r}_i]$  ist in Abbildung 5.4a gegeben.

Als Resultat ergibt sich, dass die Position jeder Soundquelle in folgendem Kegelsegment liegt:  $\text{con}_i([\phi_i, \bar{\phi}_i], [\theta_i, \bar{\theta}_i], [\underline{r}_i, \bar{r}_i]) \in \mathbb{R}^3$  visualisiert in Abbildung 5.3.

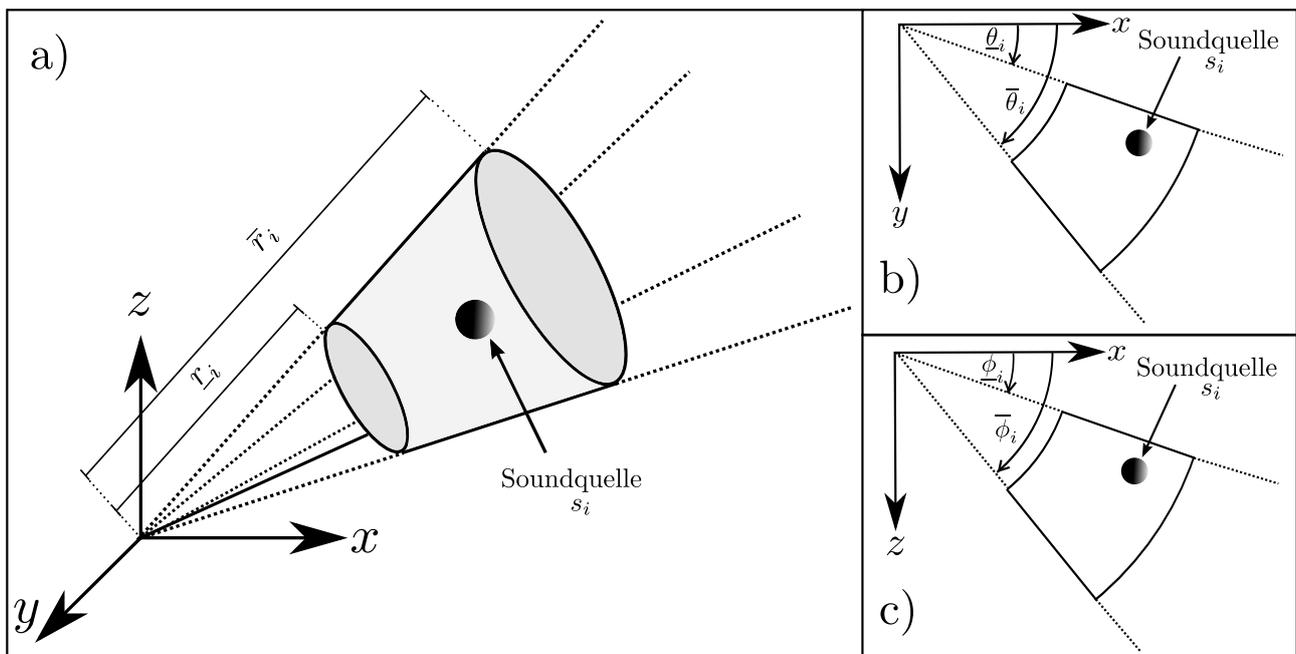


Abbildung 5.3: Kegelsegment a) 3D-Ansicht: Grenzen  $[\underline{r}_i, \bar{r}_i]$  für Radius  $r_i$

b) XY-Ebene: Grenze  $[\theta_i, \bar{\theta}_i]$  für Azimut  $\theta$

c) XZ-Ebene: Grenze  $[\phi_i, \bar{\phi}_i]$  für Elevation  $\phi$

Es ist wichtig hervorzuheben, dass durch geometrische Mehrdeutigkeiten die wahre Position der Soundquelle nicht in allen Fällen in diesem Kegelsegment liegt (vgl. Abbildung 5.4b). Als Folgerung kann der räumliche Bereich des Kegelsegments nicht verwendet werden, um auf die Position der Soundquelle zu schließen.

Stattdessen wird die zeitliche Differenz  ${}^{(i)}\Delta t_j^{(n)}$  verwendet. Diese wird aus dem Kegelsegment  $\mathbf{x}_c \in \text{con}_i$  und der Positionen der Mikrofonpaare  $j$  ermittelt.

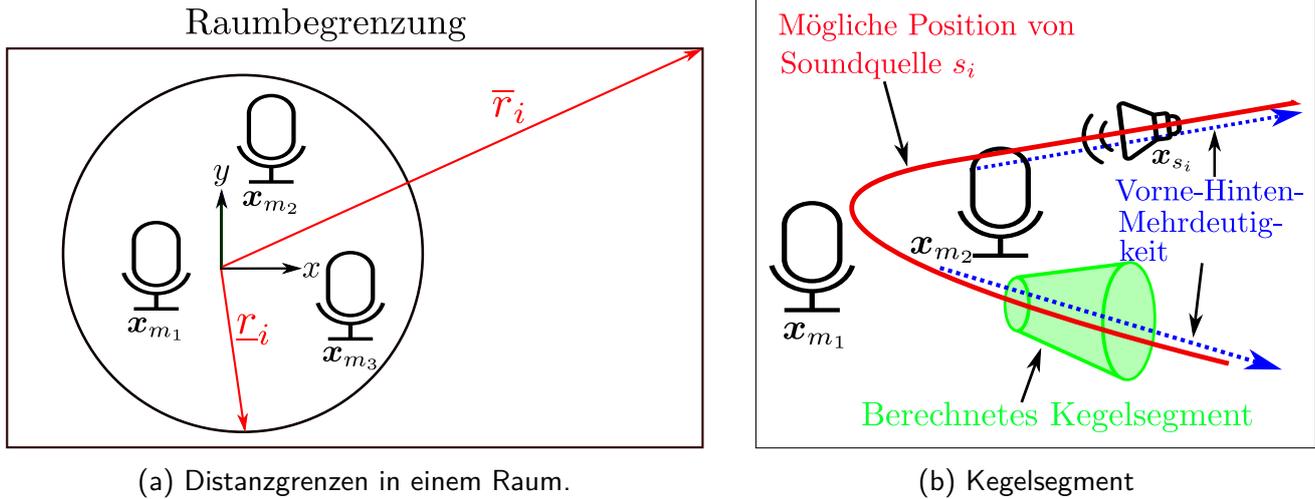


Abbildung 5.4: (a) Darstellung der Distanzgrenzen in Abhängigkeit der Raumstruktur.  
 (b) Das Kegelsegment enthält nicht die wahre Position der Soundquelle.

Dabei ergeben sich für die einzelnen Elemente des Kegelsegments unterschiedliche TDoAs. Im Anschluss werden die minimalen und maximalen TDoAs ermittelt und durch das Intervall  $[(^{(i)}\Delta t_j^{(n)}, (^{i)}\Delta \bar{t}_j^{(n)})]$  charakterisiert:

$$\begin{aligned} (^{i)}\Delta t_j^{(n)} &= \frac{1}{c} (\|\mathbf{x}_{m_{Pj,1}}^{(n)} - \mathbf{x}_{\text{con}}\|_2 - \|\mathbf{x}_{m_{Pj,2}}^{(n)} - \mathbf{x}_{\text{con}}\|_2), \\ (^{i)}\Delta \bar{t}_j^{(n)} &= \min_{\mathbf{x}_{\text{con}}} (^{i)}t_j^{(n)}, \quad (^{i)}\Delta \bar{t}_j^{(n)} = \max_{\mathbf{x}_{\text{con}}} (^{i)}t_j^{(n)}. \end{aligned} \quad (5.5)$$

Die Berechnung der zeitlichen Grenzen erfolgt durch Lösung eines nichtlinearen Optimierungsproblems. Eine genaue Beschreibung ist in Abschnitt 5.1.4 dargestellt.

Weiterhin ist nicht garantiert, dass die wahre TDoA in den berechneten Intervallen enthalten ist. Die Ursache liegt darin, dass bisher keine Diskretisierungsfehler der Mikrofon-signale betrachtet wurden.

Zur Verdeutlichung ist im Folgenden ein exemplarisches Beispiel gegeben. Das Signal  $s_i(t)$  erreicht zum Zeitpunkt  $(^{i)}t_1 = 0,95$  ms das erste und zum Zeitpunkt  $(^{i)}t_2 = 0,3$  ms das zweite Mikrofon eines Paares. Weiterhin sei die Abtastrate des akustischen Signals  $t_s = 0,2$  ms. Als Folgerung können beide Zeitpunkte nicht akkurat gemessen werden. Zur Beschreibung der Unsicherheit wird folgendes Intervall bestimmt:

$$\begin{aligned} [a] &= [(^{i)}t_1 - t_s, (^{i)}t_1 + t_s] = [0,75; 1,15] \\ [b] &= [(^{i)}t_2 - t_s, (^{i)}t_2 + t_s] = [0,1; 0,5]. \end{aligned}$$

Damit ergibt sich die TDoA folgendermaßen:

$$\begin{aligned} [a] - [b] &= \\ [(^{i)}t_1 - (^{i)}t_2 - 2 \cdot \frac{1}{f_s}, (^{i)}t_1 - (^{i)}t_2 + 2 \cdot \frac{1}{f_s}] &= \\ [0,25; 1,05]. \end{aligned}$$

Somit ist garantiert, dass die TDoA einen Wert zwischen 0,25 ms und 1,05 ms annimmt.

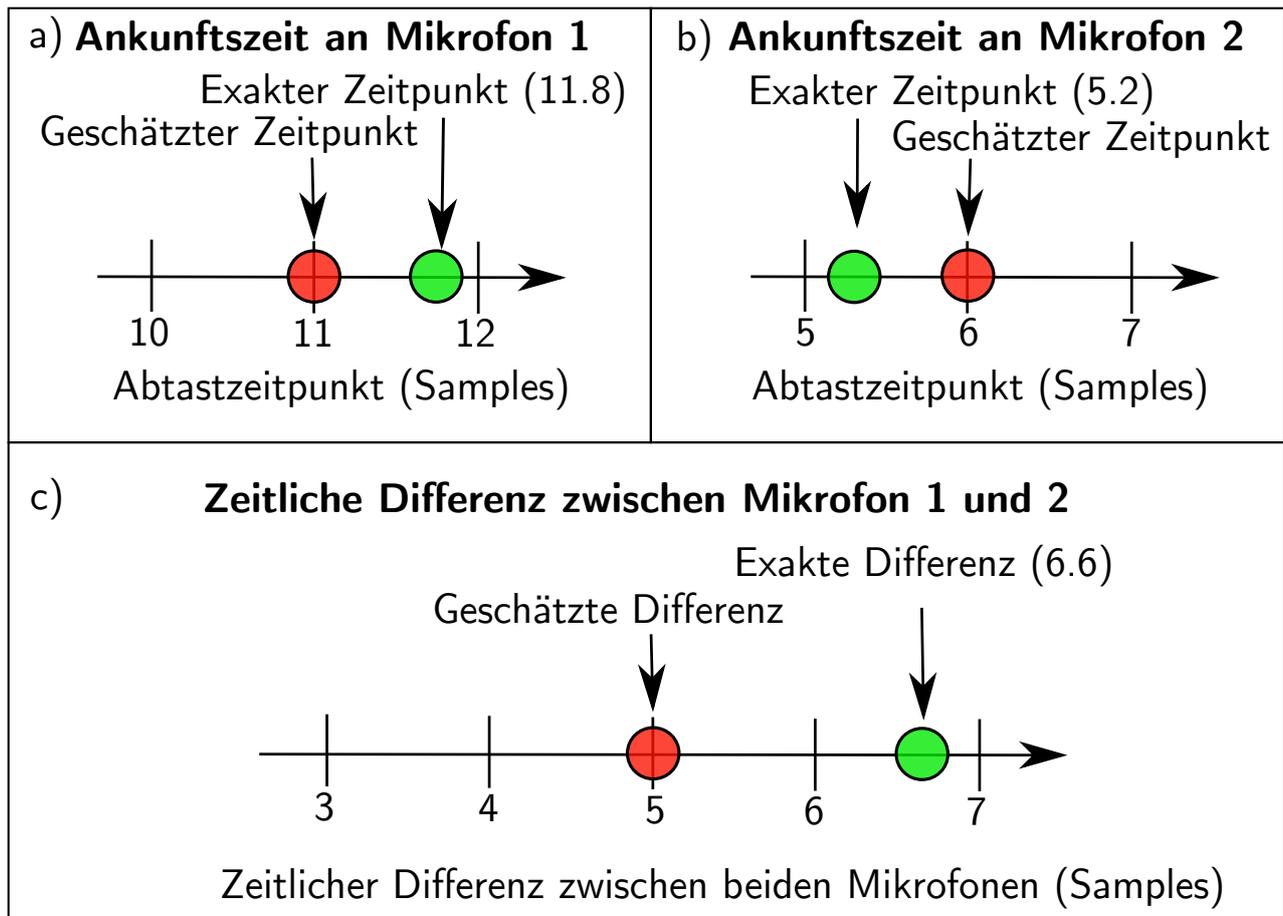


Abbildung 5.5: Ein Signal erreicht Mikrofon 1 in a) und Mikrofon 2 in b) zu unterschiedlichen Zeitpunkten. Aufgrund der diskreten Abtastung kann der exakte Zeitpunkt nicht erfasst werden, daher wird in Abhängigkeit des Signals eine Zuweisung zu einem diskreten Zeitpunkt ermittelt. Die zeitliche Differenz c) (Samples) unterscheidet sich von der exakten Differenz.

Dieses Ergebnis wird verwendet, um die Unsicherheit des Diskretierungsfehler durch Erweiterung des vorherigen TDoA-Intervalls aus Gleichung (5.5) zu modellieren:

$$\left[ \begin{aligned} & \left[ {}^{(i)}\Delta \underline{t}_j^{(n)} - \frac{2}{f_s}, {}^{(i)}\Delta \bar{t}_j^{(n)} + \frac{2}{f_s} \right] = \\ & \left[ {}^{(i)}\Delta \underline{t}_j^{(n)} - \Delta t_e, {}^{(i)}\Delta \bar{t}_j^{(n)} + \Delta t_e \right]. \end{aligned} \right. \quad (5.6)$$

Hierbei bezeichnet  $\Delta t_e$  die Erweiterung des Intervalls und wird im weiteren Verlauf als Sampling-Erweiterung gekennzeichnet. Die Größenordnung von  $\Delta t_e$  liegt in der Regel für reale Experimente zwischen 0,02 ms bis 0,125 ms. Dabei wurde eine Abtastfrequenz von 16.000 bis 96.000 Hz angenommen.

## 5.1.2 Verbesserung der Richtungsschätzung - Ansatz I: M3K

Im vorherigen Abschnitt wurden die Schritte des universellen Verfahrens zur Ermittlung der TDoA-Intervalle vorgestellt. Dieser Abschnitt stellt eine Strategie vor, um die Menge von potenziellen Richtungen  ${}^pDoA$  zu analysieren. Ziel ist es, Fehlmessungen zu erkennen und das Messrauschen der Richtungsschätzungen zu reduzieren. Hierfür wird auf die Methoden von Abschnitt 2.3.3 zurückgegriffen und eine rekursive Zustandsschätzung durchgeführt.

Der Zustand  $\mathbf{x}_i^k$  repräsentiert eine Richtungsschätzung mit den Winkeln  $(\theta_i^k, \phi_i^k)$ . Für das weitere Vorgehen wird die Richtung auf die Einheitskugel projiziert und in kartesischer Darstellung angegeben:

$$\mathbf{x}_i^k = \begin{bmatrix} x_i^k & y_i^k & z_i^k \end{bmatrix}^T \quad (5.7)$$

Das beschriebene Verfahren basiert auf der modifizierten 3D-Kalman-Methode (M3K) von [GM19]. Inklusiv der Prädiktion und Korrektur (vgl. Abschnitt 2.3.3) setzt sich das Verfahren aus folgenden - für diese Arbeit relevanten - Schritten zusammen:

1. Prädiktion
2. Probabilistische Zuordnung der Messungen
3. Hinzufügen und Entfernen von Zuständen
4. Korrektur

Im Weiteren werden die einzelnen Schritte beschrieben.

### Prädiktion

Es wird keine Steuerungseingabe vorgegeben, da die Richtungsschätzung nicht beeinflusst werden kann. Weiterhin werden die Soundquellen als statisch angenommen, wodurch eine statische Richtungsschätzung folgt. Somit resultiert für die Funktion  $\mathbf{g}$  des Bewegungsmodells die Identitätsabbildung (vgl. Abschnitt 2.3.4.1). Für den Erwartungswert  $\bar{\boldsymbol{\mu}}_i^k$  des prädizierten Zustands ergibt sich somit:

$$\bar{\boldsymbol{\mu}}_i^k = \boldsymbol{\mu}_i^{k-1}. \quad (5.8)$$

Die Schätzung der Kovarianz  $\bar{\boldsymbol{\Sigma}}_i^k$  zum Zeitpunkt  $k$  wird mittels Gleichung (2.67) berechnet.

### Probabilistische Zuordnung der Messungen

Die Messungen zum Zeitpunkt  $k$  sind durch  $\boldsymbol{\psi}^k$  gegeben. Dabei ist  $\boldsymbol{\psi}^k$  durch ein Tupel aus Richtungen  ${}^pDoA_i^k$  und Energien  $\Lambda_i^k$  beschrieben (vgl. Abschnitt 5.1.1). Für den Kalman-Ansatz werden ausschließlich die Richtungen  ${}^pDoA_i^k$  als Beobachtung aufgefasst. Dabei ist die  $i$ -te Beobachtung  $\mathbf{z}_i^k \in \mathbb{R}^3$  durch folgende Darstellung gegeben:

$$\mathbf{z}_i^k = \begin{bmatrix} x_i^k & y_i^k & z_i^k \end{bmatrix}^T \quad (5.9)$$

Die Richtung  ${}^pDoA_i^k$  ist durch Projektion auf die Einheitskugel in kartesischen Koordinaten dargestellt. Somit liegt für die Repräsentation des Zustands  $\mathbf{x}_i^k$  und der Beobachtung  $\mathbf{z}_i^k$  die gleiche Darstellung vor. Somit entspricht der Funktion  $\mathbf{h}$  des erweiterten Kalman-Filters der Identitätsabbil-

und die Jacobi-Matrix  $\mathbf{H}_k$  ist mit der Einheitsmatrix assoziiert. Weiterhin bezeichnet  $\mathbf{z}^k$  die Gesamtheit aller Einzelbeobachtungen  $z_i^k$ .

Es sei angenommen, dass die jeweiligen Beobachtungen unkorreliert voneinander sind und die Varianz für jede Komponente  $\sigma_Q^2$  entspricht. Somit ist die Kovarianzmatrix  $\mathbf{Q}^k$  gegeben durch:

$$\mathbf{Q}^k = \begin{pmatrix} \sigma_Q^2 & 0 & 0 \\ 0 & \sigma_Q^2 & 0 \\ 0 & 0 & \sigma_Q^2 \end{pmatrix}. \quad (5.10)$$

Im ersten Schritt wird auf Basis der Energie  $\Lambda_i^k$  geschätzt, ob eine Soundquelle in der Richtung  ${}^pDoA_i^k$  aktiv ist. Hierfür wird die Annahme getroffen, dass die Energie einer aktiven Soundquelle durch eine Normalverteilung approximiert werden kann:

$$p(\Lambda_i^k | a) = \mathcal{N}(\Lambda_i^k | \mu_a, \sigma_a). \quad (5.11)$$

$a$  charakterisiert, dass die Soundquelle aktiv ist.  $\mu_a$  und  $\sigma_a$  bezeichnen den Erwartungswert und die Standardabweichung der Normalverteilung. Weiterhin wird der gleiche Ansatz für die Beschreibung der Inaktivität der Soundquelle gewählt:

$$p(\Lambda_i^k | e) = \mathcal{N}(\Lambda_i^k | \mu_e, \sigma_e). \quad (5.12)$$

$e$  charakterisiert, dass die Soundquelle inaktiv ist.  $\mu_e$  und  $\sigma_e$  bezeichnen den Erwartungswert und die Standardabweichung der Normalverteilung.

Die Bestimmung der Parameter muss in realen Experimenten durchgeführt werden. Mit der Information über die Aktivität oder Inaktivität kann für jede Beobachtung  $j$  eine Einteilung in die folgenden drei Kategorien durchgeführt:

1. Fehldetektion: Die Soundquelle ist inaktiv und die Beobachtung entstammt einer Fehlmessung
2. Neue Soundquelle: Die Soundquelle ist aktiv und die Beobachtung entstammt einer bisher unbekanntem Soundquelle.
3. Verfolgte Soundquelle  $s_i$ : Die Soundquelle ist aktiv und die Beobachtung entstammt einer bekannten Soundquelle.

Im Weiteren bezeichnet die Funktion  $f(\boldsymbol{\psi}_j^k)$  die Zuordnung der Messung  $\boldsymbol{\psi}_j^k$  zu einer der drei Hypothesen:

$$f(\boldsymbol{\psi}_j^k) = \begin{cases} -2 & \text{Fehlmessung} \\ -1 & \text{Neue Soundquelle} \\ i \geq 1 & \text{Zugehörig zu Soundquelle } i \end{cases}$$

Für alle Elemente einer Messung zum Zeitpunkt  $k$  repräsentiert  $\mathbf{f}(\boldsymbol{\psi}^k)$  die Gesamtheit der Zuordnungen:

$$\mathbf{f}(\boldsymbol{\psi}^k) = (f(\boldsymbol{\psi}_1^k), f(\boldsymbol{\psi}_2^k), \dots). \quad (5.13)$$

Im Anschluss wird die stochastische Funktion  $p(\mathbf{z}_j^k | \bar{\mathbf{x}}_i^k)$  (vgl.  $\lambda_i$  in Abschnitt 2.3.4.2) berechnet. Sie beschreibt die Wahrscheinlichkeit, mit der eine Beobachtung  $\mathbf{z}_j^k$  von einer bekannten Soundquelle  $s_i$

resultiert. Berücksichtigt wird der prädizierte Zustand  $\bar{\mathbf{x}}_i^k$  der Soundquelle und die Berechnung wird folgendermaßen durchgeführt:

$$p(\mathbf{z}_j^k | \bar{\mathbf{x}}_i^k) = \mathcal{N}(\mathbf{z}_i^k | \mathbf{h}(\bar{\boldsymbol{\mu}}_i^k), \mathbf{H}^k \cdot \boldsymbol{\Sigma}_i^k \cdot (\mathbf{H}^k)^T + \mathbf{Q}^k). \quad (5.14)$$

Dabei wird der Erwartungswert des prädizierten Zustands  $\bar{\boldsymbol{\mu}}_i^k$  (vgl. Prädiktionsschritt) in den Zustandsraum projiziert. Anschließend wird die Kovarianz berechnet, indem die Varianzfortpflanzung durchgeführt und das Messrauschen berücksichtigt wird.

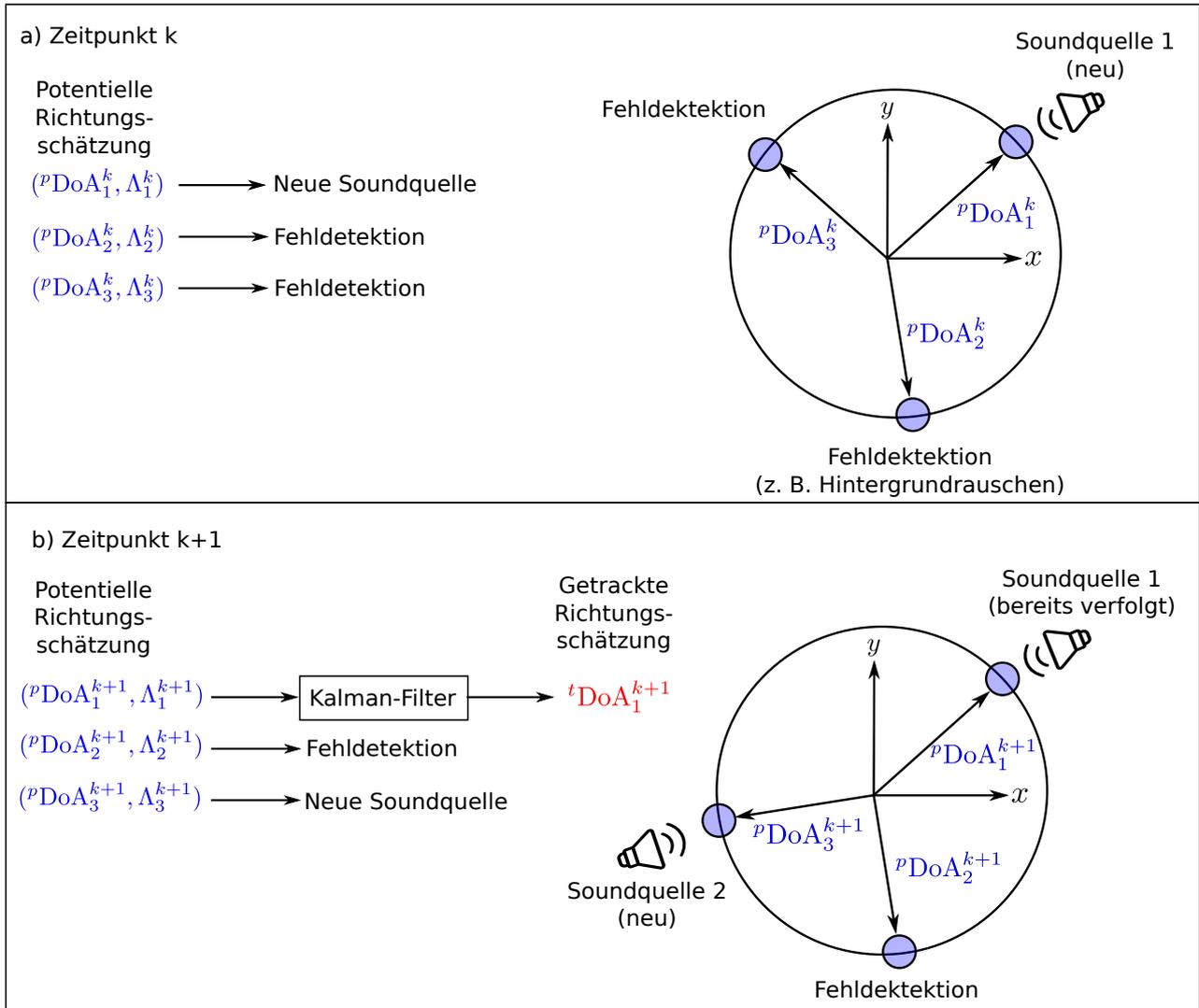


Abbildung 5.6: Darstellung des M3K-Verfahrens nach [GM19]. a) Zum Zeitpunkt  $k$  werden 3 potentielle DoA's beobachtet.  $pDoA_1^k$  ist zugehörig zu Soundquelle 1,  $pDoA_2^k$  und  $pDoA_3^k$  entsprechen Fehldetektionen. b) Zum Zeitpunkt  $k + 1$  werden 3 DoAs beobachtet.  $pDoA_1^{k+1}$  entspricht der im vorherigen Schritt detektierten Soundquelle 1,  $pDoA_2^{k+1}$  einer neuen Soundquelle und  $pDoA_3^{k+1}$  einer Fehldetektion.

Im nächsten Schritt wird für jede Beobachtung  $\boldsymbol{\psi}^k$  und möglicher Zuordnung  $\mathbf{f}$  die zugehörige Wahrscheinlichkeit  $p(\mathbf{f}(\boldsymbol{\psi}^k))$  ermittelt. Als Resultat können für den nächsten Schritt weitere stochastische Größen gefolgert werden. Einerseits wird die Wahrscheinlichkeit  $p(i|\boldsymbol{\psi}_j^k)$  berechnet. Sie beschreibt die Wahrscheinlichkeit, dass die Beobachtung  $\boldsymbol{\psi}_j^k$  von der Soundquelle  $s_i$  entammt.

Weiterhin bezeichnet  $p(i|\psi^k)$  die Wahrscheinlichkeit, dafür dass mindestens ein Element der Beobachtung  $\psi^k$  von der Soundquelle  $s_i$  resultiert. Andererseits wird die Wahrscheinlichkeit  $p(\text{neu}|\psi_j^k)$  ermittelt, dafür dass die Beobachtung  $\psi_j^k$  einer neuen Soundquelle entspricht. Die Details der Berechnung der Wahrscheinlichkeiten sind in [GM19] dargestellt.

### Hinzufügen und Entfernen von Zuständen

Bisher wurde ausschließlich ein einzelner Zustand betrachtet. Im Folgenden wird beschrieben, wie weitere Zustände hinzugefügt und entfernt werden können.

Ein neuer Zustand  $x_i$  wird dem Kalman-Ansatz hinzugefügt, falls die Wahrscheinlichkeit  $p(\text{neu}|\psi_j^k)$  über den Schwellwert  $\theta_{\text{neu}}$  liegt. Um kurzzeitige Detektionen von Soundquellen zu verhindern, wird dieser Zustand zunächst als potenzieller Kandidat betrachtet. Erst nachdem dieser Kandidat mindestens für  $n_{\text{neu}}$  Zeitschritte in einem festgelegten Zeitraum erkannt wurde, wird der Zustand als getrackter (oder verfolgter) Zustand berücksichtigt. Die resultierende Richtungsschätzung entspricht einem Element von  ${}^t\text{DoA}^k$  (vgl. Abschnitt 5.1.1), das für die weiteren Schritte berücksichtigt wird.

Weiterhin wird der Zustand  $x_i$  aus dem Kalman-Ansatz entfernt, falls die Wahrscheinlichkeit  $p(i|\psi^k)$  unter den Schwellwert  $\theta_{\text{tot}}$  für  $n_{\text{tot}}$  nachfolgende Zeitschritte fällt.

### Korrektur

Um die Korrektur des prädizierten Zustandes  $\bar{x}_i$  durchzuführen, wird eine Beobachtung  $z_j^k$  benötigt. Die Auswahl erfolgt durch Suche des Argumentes, das die Wahrscheinlichkeit  $p(i|\psi_j^k)$  maximiert:

$$j^*(i) = \arg \max_j p(i|\psi_j^k). \quad (5.15)$$

$j^*(i)$  entspricht dem Index und  $z_{j^*}^k$  der zugehörigen Beobachtung. Weiterhin wird der Kalman-Gain  $\mathbf{K}_i^k$  mit der Wahrscheinlichkeit  $p(i|\psi^k)$  gewichtet. Hierdurch wird sichergestellt, dass ein Zustand  $x_i$  nicht modifiziert wird, falls keine zugehörige Messung vorliegt. Somit können der Erwartungswerte  $\mu_i^k$  und der Kovarianz  $\Sigma_i^k$  folgendermaßen berechnet werden:

$$\mu_i^k = \bar{\mu}_i^k + p(i|\psi^k) \mathbf{K}_i^k \left( z_{j^*(i)}^k - \mathbf{h}(\bar{\mu}_i^k) \right) \quad (5.16)$$

$$\Sigma_i^k = \bar{\Sigma}_i^k - p(i|\psi^k) \mathbf{K}_i^k \mathbf{H}^k \bar{\Sigma}_i^k. \quad (5.17)$$

## 5.1.3 Verbesserung der Richtungsschätzung - Ansatz II: GM-PHD

Im vorherigen Abschnitt wurde dargestellt, dass es von unterschiedlichen Parametern abhängig ist, ob Zustände hinzugefügt oder entfernt werden. Diese Parameter sind die Wahrscheinlichkeiten  $p(\text{neu}|\psi_j^k)$  und  $p(i|\psi^k)$ , der Schwellwert  $\theta_{\text{neu}}$  und die Anzahl der Zeitschritte  $\theta_{\text{neu}}$  und  $\theta_{\text{tot}}$ .

Im Kontext der DoA-Schätzung zu Soundquellen hat sich ein weiterer Ansatz durchgesetzt. Dieser basiert auf den PHD-Filter (engl. *Probability Hypothesis Density*) [Mah03] und wird im Folgenden kompakt vorgestellt. Die Grundlagen wurden bereits in Abschnitt 2.3.4.5 vorgestellt.

Beim PHD-Ansatz wird durch die Funktion  $\lambda_b(\mathbf{x})$  beschrieben, wann ein neuer Zustand hinzugefügt wird. Der Prozess wird als spontane Geburt (engl. *birth*) eines Zustandes interpretiert. Die Funktion beschreibt die Intensität, mit der ein neuer Zustand an jeweiliger Position  $\mathbf{x}$  erscheint. In Bezug zur DoA-Schätzung beschreibt  $\lambda_b(\mathbf{x})$  die Richtung aus der eine Soundquelle wahrgenommen

wird. Falls a priori keine Information vorliegt, wird jede Richtung mit der gleichen Intensität assoziiert. Weiterhin können Bereiche definiert werden, die für die Richtungswahrnehmung nicht berücksichtigt werden sollen, z. B. Richtungsmessungen ausgehend vom Boden oder der Decke.

Die Funktion  $\lambda_s(\mathbf{x})$  beschreibt die Wahrscheinlichkeit (oder Intensität), dass ein bereits beobachteter Zustand  $\mathbf{x}_i$  weiterhin im nächsten Zeitschritt existiert. Dieser Vorgang wird als Überleben (engl. *surviving*) bezeichnet. Hierdurch kann modelliert werden, dass Zustände entfernt werden. Weiterhin beschreibt  $p_d$  die Wahrscheinlichkeit, dass eine Beobachtung durch einen Sensor detektiert wird.

Für die Implementierung des PHD-Filters muss eine Repräsentation des Bewegungsmodells und Sensormodells (vgl. Abschnitt 2.3.3) gewählt werden. Dabei wird in vielen Arbeiten ein Gaußsches Mischmodell (GM) (vgl. Abschnitt 2.3.4.2) verwendet. Hieraus resultiert der GM-PHD-Filter, dessen Zustandsschätzung durch einen Prädiktions- und Korrekturschritt des EKF (vgl. Abschnitt 2.3.3.1) durchgeführt werden kann. Details zur Durchführung ist in [VM06] gegeben. Dabei wird die Integration von neuen Zuständen, sowie die Ermittlung der Gewichtung der jeweiligen Gauß-Komponenten beschrieben.

#### 5.1.4 Berechnung der Zeitdifferenz innerhalb eines Kegelsegments

In diesem Abschnitt wird darauf eingegangen, wie die minimale und maximale TDoA für jedes Kegelsegment berechnet werden. Zunächst wird der zweidimensionale Fall behandelt. Hierfür ist in Abbildung 5.7 ein exemplarisches Beispiel gegeben.

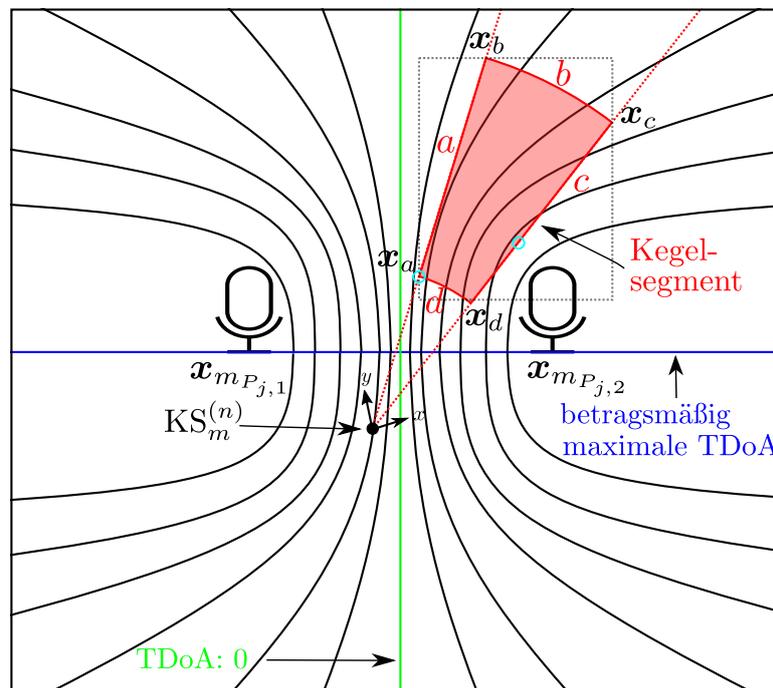


Abbildung 5.7: Exemplarische Darstellung der Isolinien für unterschiedliche Zeitdifferenzen und gegebenes Kegelsegment in 2D.

Für ein Mikrofonpaar an Position  $\mathbf{x}_{mP_{j,1}}$  und  $\mathbf{x}_{mP_{j,2}}$  wird der Verlauf der TDoA betrachtet. Für eine fixierte TDoA liegt die Position einer Soundquelle auf einer Isolinie mit hyperbolischer Form. Die betragsmäßig größte TDoA resultiert für eine Soundquelle, die auf der Mikrofonachse des Paares angeordnet ist. Dabei wird angenommen, dass die Soundquelle nicht zwischen dem Paar liegt.

Orthogonal zur Mikrofonachse resultiert ausgehend vom Mittelpunkt eine TDoA von null. Hier ist der Propagationspfad zu beiden Mikrofonen identisch.

In Abschnitt 5.1.1 wurde erläutert, wie auf Basis von potenziellen Richtungen  ${}^pDoA^k$  eine Menge von Kegelsegmenten berechnet wird. Im Folgenden wird die minimale und maximale TDoA für jedes Kegelsegment benötigt. Zu beachten ist, dass im Allgemeinen das Koordinatensystem  $KS_m^{(n)}$  der  $n$ -ten Messung nicht mit dem Mittelpunkt des Mikrofonpaares übereinstimmt.

Ein konservativer Ansatz, um die Grenzen für die TDoA zu ermitteln, kann mithilfe der Intervall-Arithmetik erfolgen. Hierbei wird zunächst der Bereich des Kegelsegments mit einer umhüllenden Intervallbox  $[\mathbf{x}_{\text{con}}]$  approximiert (vgl. den grau-markierten Bereich in Abbildung 5.7). Daraus ergibt sich:

$$[{}^{(i)}\Delta t_j^{(n)}] = \frac{1}{c} (\|\mathbf{x}_{m_{P_{j,1}}}^{(n)} - [\mathbf{x}_{\text{con}}]\|_2 - \|\mathbf{x}_{m_{P_{j,2}}}^{(n)} - [\mathbf{x}_{\text{con}}]\|_2). \quad (5.18)$$

Diese Approximation ist nicht optimal, wie in Abbildung 5.7 entnommen werden kann.

Aufgrund der Geometrie der Lösung kann gezeigt werden, dass das Minimum und Maximum der TDoA auf dem Rand des Kegelsegments angenommen wird. Dabei ist der Rand durch die Seiten  $a, b, c$  und  $d$  repräsentiert. Die zugehörigen Extrema sind im Beispiel durch blaue Kreise markiert.

Die Seiten  $a$  und  $b$  können durch begrenzte Geraden  $\mathbf{v}_1$  und  $\mathbf{v}_2$  in folgender parametrisierten Form angegeben werden:

$$\text{Seite } a: \quad \mathbf{v}_1 = \mathbf{x}_a + \lambda_a \cdot (\mathbf{x}_b - \mathbf{x}_a) \quad , \quad \lambda_a \in [0, 1] \quad (5.19)$$

$$\text{Seite } c: \quad \mathbf{v}_2 = \mathbf{x}_d + \lambda_c \cdot (\mathbf{x}_c - \mathbf{x}_d) \quad , \quad \lambda_c \in [0, 1]. \quad (5.20)$$

Für die Seiten  $b$  und  $d$  ergeben sich Kreissegmente mit folgender Parametrisierung:

$$\text{Seite } b: \quad \mathbf{v}_3 = \bar{r} \cdot [\cos(\underline{\theta} + \lambda_b \cdot (\bar{\theta} - \underline{\theta})), \sin(\underline{\theta} + \lambda_b \cdot (\bar{\theta} - \underline{\theta}))]^T, \lambda_b \in [0, 1] \quad (5.21)$$

$$\text{Seite } d: \quad \mathbf{v}_4 = \underline{r} \cdot [\cos(\underline{\theta} + \lambda_d \cdot (\bar{\theta} - \underline{\theta})), \sin(\underline{\theta} + \lambda_d \cdot (\bar{\theta} - \underline{\theta}))]^T, \lambda_d \in [0, 1] \quad (5.22)$$

Hierbei bezeichnen  $[\underline{r}, \bar{r}]$  die Schranken der Distanz und  $[\underline{\theta}, \bar{\theta}]$  die Schranken des Azimut-Winkels (in 2D) des betrachteten Kegelsegments (vgl. Abschnitt 5.1.1).

Um die Extremwerte zu ermitteln, wird zunächst für jede Funktion  $\mathbf{v}_k$  eine Menge von Samples  $\mathbf{x}_{v_k}$  erzeugt. Im Anschluss werden für alle Samples die maximalen und minimalen Werte ermittelt:

$$\begin{aligned} {}^{(i)}\Delta t_{j,k}^{(n)} &= \frac{1}{c} (\|\mathbf{x}_{m_{P_{j,1}}}^{(n)} - \mathbf{x}_{v_k}\|_2 - \|\mathbf{x}_{m_{P_{j,2}}}^{(n)} - \mathbf{x}_{v_k}\|_2), \\ {}^{(i)}\Delta \underline{t}_{j,k}^{(n)} &= \min_{\mathbf{x}_{v_k}} {}^{(i)}t_{j,k}^{(n)} \quad , \quad {}^{(i)}\Delta \bar{t}_{j,k}^{(n)} = \max_{\mathbf{x}_{v_k}} {}^{(i)}t_{j,k}^{(n)}. \end{aligned} \quad (5.23)$$

Darauffolgend werden die Grenzen der TDoA als Extremwerte aller Randbereiche ermittelt:

$${}^{(i)}\Delta \underline{t}_j^{(n)} = \min_k {}^{(i)}t_{j,k}^{(n)} \quad , \quad {}^{(i)}\Delta \bar{t}_j^{(n)} = \max_k {}^{(i)}t_{j,k}^{(n)}. \quad (5.24)$$

Das beschriebene Vorgehen ist bisher beschränkt auf 2D. Analog zum zweidimensionalen Fall kann eine konservative Abschätzung durchgeführt werden. Dabei wird das Kegelsegment  $[\mathbf{x}_{\text{con}}]$  durch eine umhüllende Intervallbox approximiert und anschließend Gleichung (5.18) gelöst.

Alternativ kann für den zweidimensionalen Randbereich des dreidimensionalen Kegelsegments ein Sampling mit den Elementen  $x_{v_k}$  durchgeführt werden. Für diese Elemente werden die Grenzen der TDoA durch Lösung der Gleichungen (5.23) und (5.24) ermittelt.

### 5.1.5 Ansatz (2) - Konkretisierung auf SRP-PHAT-Verfahren (DoATiD)

In Abschnitt 5.1.1 wurde dargestellt, wie die potenziellen Richtungsschätzungen  ${}^p DoA^k$  durch ein beliebiges SSL-Verfahren berechnet werden können. Dabei wurden die Details des Verfahrens nicht betrachtet.  ${}^p DoA^k$  wurde als Sensoreingang in ein rekursives Zustandsschätzverfahren aufgefasst. Im Anschluss wurden aus den resultierenden Richtungen  ${}^t DoA^k$  die TDoA-Intervalle für jede Gruppierung ermittelt. Hierbei wurden geometrische Überlegungen miteinbezogen. Nicht berücksichtigt wurde, dass  ${}^p DoA^k$  nicht für alle Richtungen die gleiche Genauigkeit besitzt. In Abhängigkeit der Geometrie des Mikrofonarrays kann die Genauigkeit stark variieren. Weiterhin wird während der DoA-Schätzung die TDoA entweder direkt (z. B. SRP-Verfahren) oder indirekt (z. B. MUSIC-Verfahren) verwendet. Unsicherheiten der TDoA können direkt im SSL-Verfahren berücksichtigt werden.

Im folgenden Abschnitt wird die Unsicherheit der TDoA direkt im Schätzverfahren berücksichtigt. Dabei wird exemplarisch das SRP-PHAT-Verfahren ausgewählt (vgl. Abschnitt 2.2.3.4). Die Berechnungsschritte des erweiterten Ansatzes sind in Abbildung 5.8 dargestellt.

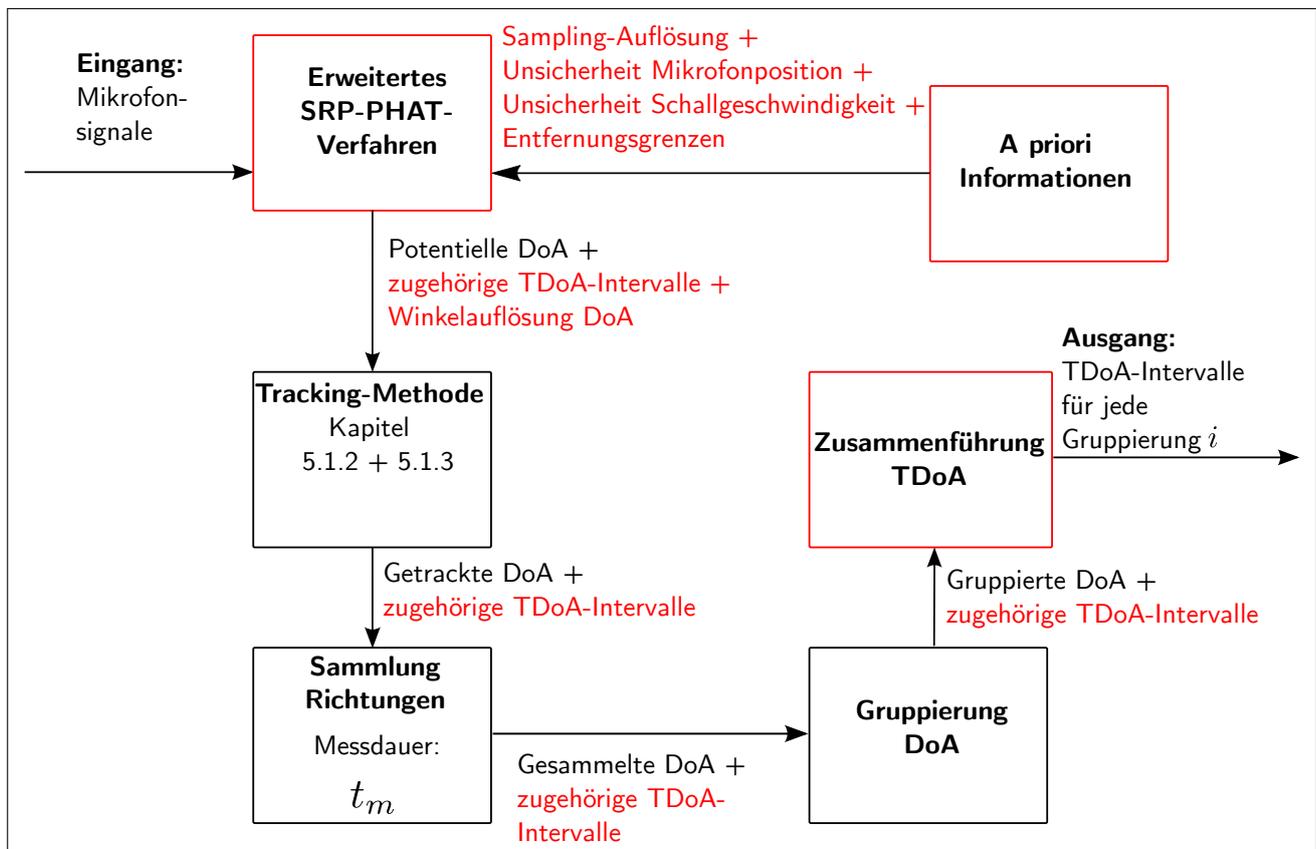


Abbildung 5.8: Erweitertes Verfahren unter Verwendung des SRP-PHAT-Verfahrens: Berechnungsschritte für die Intervalle der TDoA auf Basis von DoA-Schätzungen. Veränderte Schritte im Vergleich zum universellen Verfahren aus Abschnitt 5.1.1 sind rot markiert.

Es seien zunächst die diskreten Mikrofonsignale  $m_1(i), \dots, m_{n_m}(i)$  mit der Abtastfrequenz  $f_s$  gegeben.  $i \in \mathbb{N}$  bezeichnet den diskreten Index des jeweiligen Mikrofonsignals und  $n_m$  die Gesamtzahl der Mikrofone. Der zugehörige Zeitschritt  $t$  berechnet sich folgendermaßen:

$$t = \frac{i}{f_s}. \quad (5.25)$$

Im ersten Schritt wird zum aktuellen Messzeitpunkt  $k$  für jedes Mikrofonsignal  $m_l(i)$  ein Block von  $n_f$  nachfolgenden Daten betrachtet und als Frame  $\mathbf{f}_l^k$  bezeichnet:

$$\mathbf{f}_l^k = [m_l(0), m_l(1), \dots, m_l(n_f - 1)]. \quad (5.26)$$

Hierbei wird angenommen, dass  $m_l(i)$  den  $i$ -ten Mikrofonamplitudenwert ab dem Messzeitpunkt  $k$  beschreibt. Da die Länge der jeweiligen Frames  $\mathbf{f}_l^k$  nicht mit der vollständigen Periode von beliebigen Signalen übereinstimmt, werden die Daten der Frames mit einer Fensterfunktion  $w(i)$  gefiltert. Dabei sind in [Pra14, Kapitel 3] unterschiedliche Fensterfunktionen ausführlich dargestellt. Im Kontext der Signalverarbeitung von akustischen Signalen wird das Von-Hann-Fenster vermehrt eingesetzt:

$$w(i) = \frac{1}{2} \left[ 1 - \cos \left( \frac{2\pi i}{n_f - 1} \right) \right]. \quad (5.27)$$

Der gefilterte Frame wird mit  $\hat{\mathbf{f}}_l^k$  bezeichnet und folgendermaßen berechnet:

$$\hat{\mathbf{f}}_l^k(i) = \mathbf{f}_l^k(i) \cdot w(i) \quad \forall i \in \{0, \dots, n_f - 1\}. \quad (5.28)$$

Im nächsten Schritt wird eine diskrete Fouriertransformation der gefilterten Frames  $\hat{\mathbf{f}}_l^k$  durchgeführt (vgl. Abschnitt 2.1.4). Da die Mikrofonsignale durch reelle Zahlen repräsentiert sind, genügt es, die Hälfte der Werte  $n_h = \frac{n_f}{2} + 1$  zu berechnen.

Darauffolgend wird die generalisierte Kreuzkorrelation für jedes Mikrofonpaar  $j$  im Fourierraum ermittelt und die Phasen-Transformation (vgl. Kapitel 2.2.3.4) verwendet. Das Resultat wird mit  $X_j^k(i)$  bezeichnet. Hierbei bezeichnet  $i \in \{0, \dots, \frac{n_f}{2}\}$  den Index des zugehörigen Frequenzbereichs  $f$ :

$$f = \frac{i \cdot f_s}{n_f}. \quad (5.29)$$

Im späteren Verlauf wird die Kreuzkorrelation im Zeitbereich verwendet, um die TDoA zwischen zwei Mikrofonsignalen zu ermitteln. Dabei ist die kleinste darstellbare TDoA durch den Kehrwert der Abtastfrequenz  $f_s$  festgelegt. Um die Genauigkeit zu erhöhen, kann somit  $f_s$  erhöht werden. Jedoch ist die Erhöhung durch die eingesetzte Hardware beschränkt. Weiterhin resultiert eine erhöhte Rechenzeit, da weitere Samples berücksichtigt werden müssen. Stattdessen wird eine Interpolation im Zeitbereich durchgeführt werden. Unter Annahme von sinusförmigen Signalen ergibt sich dabei eine Sinc-Interpolation. Statt die Interpolation im Zeitbereich durchzuführen, erfolgen die Rechenschritte im Fourierbereich. Beim sogenannten *Zero-Padding* wird dem Signal  $X_j^k(i)$  eine Menge von Nullelementen hinzugefügt. Dabei charakterisiert die Anzahl der hinzugefügten Elemente die Größenordnung der Interpolation. Im Folgenden bezeichnet  $n_{ip}$  die Interpolationsrate. Hierbei entspricht  $n_{ip} = 1$  der ursprünglichen und  $n_{ip} = 2$  der Verdopplung der resultierenden Auflösung. Schließlich

bezeichnet  ${}^{n_{ip}}X_j^k(f)$  das erweiterte Signal mit der Interpolationrate  $n_{ip}$  und wird folgendermaßen dargestellt:

$${}^{n_{ip}}X_j^k(f) = [X_j^k(0), X_j^k(1), \dots, X_j^k(n_h - 1), \underbrace{0, 0, 0, 0, 0, \dots, 0, 0, 0, 0, 0}_{n_h \cdot (n_{ip} - 1) \text{ Elemente}}]. \quad (5.30)$$

Im Anschluss wird  ${}^{n_{ip}}X_j^k(f)$  in den Zeitbereich zurücktransformiert, indem die Inverse Diskrete Fourier-Transformation (vgl. Abschnitt 2.1.4) berechnet wird. Dabei resultiert die Kreuzkorrelationsfunktion  $r_{m_k, m_l}$  (vgl. Abschnitt 2.2.3.3).

Für eine Menge von normierten Richtungen  $\mathbf{u}_{k_i}$  (oder Kandidaten) wird im Anschluss aufgrund geometrischer Beziehungen die TDoA  ${}^{(i)}\Delta t_j^{(n)}$  für das jeweilige Mikrofonpaar  $j$  ermittelt. Wie in Abschnitt 2.2.3.3 dargestellt, besitzt die Kreuzkorrelationsfunktion  $r_{m_k, m_l}$  an der Stelle  ${}^{(i)}\Delta t_j^{(n)}$  ein theoretisches Maximum. Nachdem die Kreuzkorrelationsfunktion an der Stelle der entsprechenden Zeitdifferenz ausgewertet wird, kann beurteilt werden, ob eine aktive Soundquelle in der Richtung  $\mathbf{u}_{k_i}$  vorliegt. In realen Experimenten wird das theoretische Maximum nicht zwingend an der zu erwartenden Stelle angenommen. Die Ursache liegt in der diskretisierten Darstellung der Mikrofonsignale, sowie der Unsicherheit der wahren Mikrofonpositionen und der Schallgeschwindigkeit  $c$ .

Zur Lösung dieses Problems wurde in [GM19] ein probabilistischer Ansatz vorgestellt, bei dem der Fehler der TDoA durch eine Normalverteilung modelliert wird. Hierfür werden sowohl die Schallgeschwindigkeit  $c$  als auch und die Position der Mikrofone  $\mathbf{x}_{m_k}$  und  $\mathbf{x}_{m_l}$  als gauß-verteilt mit einem mittleren Fehler von 0 angenommen. Diese Annahme ist insbesondere für die Schallgeschwindigkeit  $c$  nicht zutreffend.

Daher wird in dieser Arbeit ein intervall-basierter Ansatz gewählt, um die TDoA zu beschreiben und im Folgenden vorgestellt.

Wie in Abschnitt 2.2.3.1 dargestellt, berechnet sich die TDoA  ${}^{(i)}\Delta t_{kl}$  für eine Richtung  $\mathbf{u}_{k_i}$  folgendermaßen:

$${}^{(i)}\Delta t_{kl} = \frac{1}{c}(\mathbf{x}_{m_l} - \mathbf{x}_{m_k}) \cdot \mathbf{u}_{k_i}. \quad (5.31)$$

Um die Kreuzkorrelation auszuwerten, wird für die TDoA eine diskrete Repräsentation benötigt. Im Folgenden bezeichnet  $\Delta t_{kl} = 1$ , dass die Mikrofonsignale  $m_k(i)$  und  $m_l(i)$  um ein Sample verschoben sind. Hierdurch ergibt sich eine zeitliche Verschiebung von  $1 \cdot \frac{1}{f_s \cdot n_{ip}}$ . Die Umrechnung wird folgendermaßen durchgeführt:

$${}^{(i)}\Delta t_{kl} = \frac{f_s \cdot n_{ip}}{c}(\mathbf{x}_{m_l} - \mathbf{x}_{m_k}) \cdot \mathbf{u}_{k_i}. \quad (5.32)$$

Um die Unsicherheit der TDoA zu repräsentieren, werden die Parameter  $c$ ,  $\mathbf{x}_{m_k}$  und  $\mathbf{x}_{m_l}$  durch die Intervalle  $[c]$ ,  $[\mathbf{x}_{m_k}]$  und  $[\mathbf{x}_{m_l}]$  ersetzt. Somit wird Gleichung 5.32 folgendermaßen angepasst:

$$[{}^{(i)}\Delta t_{kl}] = \frac{f_s \cdot n_{ip}}{[c]}([\mathbf{x}_{m_l}] - [\mathbf{x}_{m_k}]) \cdot \mathbf{u}_{k_i}. \quad (5.33)$$

In Kapitel 4.1 wurde beschrieben, wie die Grenzen für die lokale Mikrofonposition ermittelt werden. Für die Schallgeschwindigkeit  $c$  werden die Grenzen  $[\underline{c}, \bar{c}]$  in Abhängigkeit des erwarteten Temperaturbereichs am Einsatzort berechnet.

Seien im Folgenden  ${}^{(i)}a_{kl}, b_{kl} \in \mathbb{R}$  die oberen und unteren Grenzen von  $[{}^{(i)}\Delta t_{kl}]$ . Im nächsten Schritt werden  ${}^{(i)}a_{kl}$  und  ${}^{(i)}b_{kl}$  auf die nächste benachbarte natürliche Zahl gerundet, sodass das ursprüngliche Intervall in dem resultierenden Intervall  $[{}^{(i)}\Delta \hat{t}_{kl}]$  enthalten ist:

$$[{}^{(i)}\Delta \hat{t}_{kl}] = [{}^{(i)}\hat{a}_{kl}, {}^{(i)}\hat{b}_{kl}] \quad \text{mit} \quad {}^{(i)}\hat{a}_{kl} = \lceil {}^{(i)}a_{kl} \rceil \quad \text{und} \quad {}^{(i)}\hat{b}_{kl} = \lfloor {}^{(i)}b_{kl} \rfloor. \quad (5.34)$$

Um alle ganzzahligen Werte in  $[{}^{(i)}\Delta \hat{t}_{kl}]$  zu beschreiben, wird die Menge  ${}^{(i)}S_{kl}$  definiert:

$${}^{(i)}S_{kl} = \{{}^{(i)}\hat{a}_{kl}, {}^{(i)}\hat{a}_{kl} + 1, \dots, {}^{(i)}\hat{b}_{kl} - 1, {}^{(i)}\hat{b}_{kl}\} \quad (5.35)$$

Die ermittelte Kreuzkorrelation  $r_{m_k, m_l}$  wird für alle Elemente aus  ${}^{(i)}S_{kl}$  untersucht. Anschließend wird das Maximum berechnet und mit  ${}^{(i)}\text{msw}_{kl}$  bezeichnet:

$${}^{(i)}\text{msw}_{kl} = \max_j r_{m_k, m_l}(j) \quad \forall j \in {}^{(i)}S_{kl}. \quad (5.36)$$

Bisher wurde ausschließlich ein Mikrofonpaar  $j$  berücksichtigt. Im nächsten Schritt wird für eine gegebene Richtung  $\mathbf{u}_{k_i}$  die Summe für alle Paare betrachtet:

$${}^{(i)}\text{msw} = \sum_{(k,l) \in P} \text{msw}_{kl}. \quad (5.37)$$

Hierbei bezeichnet  $P$  die Menge, die alle Kombinationen  $(k, l)$  von unterschiedlichen Mikrofonpaarungen enthält. Als Ergebnis wird jeder Richtung  $\mathbf{u}_{k_i}$  ein Skalar zugeordnet. Je größer dieser Wert ist, desto wahrscheinlicher befindet sich eine aktive Soundquelle in dieser Richtung.

Um potenzielle Richtungen zu extrahieren, werden über den gesamten Suchraum lokale Maxima ermittelt. Hierfür kann im Vorfeld die Annahme getroffen werden, dass maximal  $n_{max}$  unterschiedliche Soundquellen aktiv sind. Im nächsten Schritt werden  $n_{max}$  lokale Maxima aus  ${}^{(i)}\text{msw}$  extrahiert:

$$\mathbf{i}^* = \arg \max_i {}^{(i)}\text{msw}. \quad (5.38)$$

Die resultierenden Richtungen  $\mathbf{u}_{k_{i^*}}$  werden als potenzielle Richtungen  ${}^pDoA$  betrachtet. Wie in Abschnitt 5.1.2 und 5.1.3 beschrieben, wird durch eine rekursive Zustandsschätzung die Richtungs-schätzung verbessert. Weiterhin wird dabei der Einfluss vom Messrauschen verringert.

Im Gegensatz zu dem universellen Verfahren in Abschnitt 5.1.1 liegen die zugehörigen Intervalle der TDoAs bereits vor. Sie werden im nachfolgenden Schritt verwendet.

Im Anschluss wird für die Richtungsschätzung  ${}^pDoA$  mithilfe der in Abschnitt 5.1.2 oder 5.1.3 behandelten Verfahren gefiltert und als Ergebnis resultiert die Menge  ${}^tDoA$  der getrackten DoAs und die zugehörigen TDoA-Intervalle. Im nächsten Schritt werden für die Messdauer  $t_m$  getrackte DoAs gesammelt und im Anschluss eine Gruppierung durchgeführt. Darauffolgend werden die Intervalle der jeweiligen TDoAs für jede Gruppierung zusammengeführt, indem eine Vereinigung aller Elemente berechnet wird.

Im letzten Schritt können in Abhängigkeit der vorhandenen Distanzinformationen die Intervalle der TDoA weiter angepasst werden.

## 5.2 Mikrofon-basierte Zeitdifferenzschätzung (InTiD)

Der in Abschnitt 5.1 vorgestellte Ansatz DoATiD verwendet Methoden aus dem Bereich der SSL, um die Richtung zu einer Soundquelle zu schätzen. Jedoch kann nicht ohne weitere Annahmen garantiert werden, dass diese Methoden die korrekte Richtungen enthalten. Daher wird in diesem Abschnitt ein alternativer Ansatz - Intervall-basierte Zeitdifferenz (engl. *Interval-Based Time Differenz - InTiD*) - vorgestellt. Dieser Ansatz verwendet die akustischen Signale und führt eine intervall-basierte Berechnungen durch. Zu Beginn wird in Abschnitt 5.2.1 zunächst eine einzelne Soundquelle betrachtet. Zum Ende wird der Ansatz in Abschnitt 5.2.4 auf mehrere aktive Soundquellen erweitert.

### 5.2.1 Einzelne Soundquelle

Es sei zunächst eine einzelne Soundquelle  $s_i(t)$  angenommen, die durch die Mikrofon-signale  $m_k(t)$  und  $m_l(t)$  wahrgenommen wird. Die akustische Welle benötigt die Zeit  ${}^{(i)}t_k$  um das Mikrofon  $m_k$  und die Zeit  ${}^{(i)}t_l$  um das Mikrofon  $m_l$  zu erreichen (vgl. Abbildung 5.9).

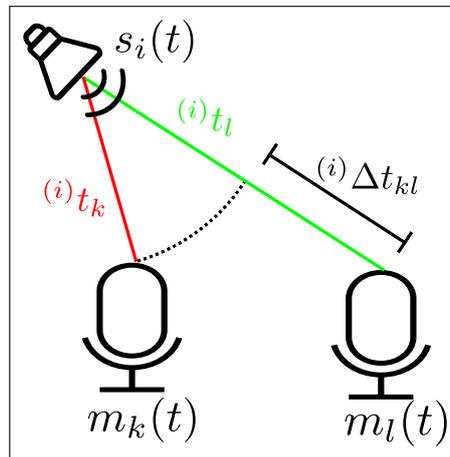


Abbildung 5.9: Das akustische Signal  $s_i(t)$  wird durch zwei Mikrofone empfangen.

Während der Propagation der akustischen Welle sei angenommen, dass die Amplitude der Mikrofon-signale um einen konstanten Parameter  ${}^{(i)}\tilde{\alpha}_k$  oder  ${}^{(i)}\tilde{\alpha}_l$  gedämpft wird. Diese Dämpfung resultiert aufgrund der Propagation der akustischen Welle innerhalb der Umgebung und der unterschiedlichen Sensitivitäten der jeweiligen Mikrofone (vgl. Kapitel 2.1.2).

Der Zusammenhang zwischen der Soundquelle und den Mikrofonen wird folgendermaßen formuliert:

$$\begin{aligned} m_k(t) &= {}^{(i)}\tilde{\alpha}_k \cdot s_i(t - {}^{(i)}t_k) \\ m_l(t) &= {}^{(i)}\tilde{\alpha}_l \cdot s_i(t - {}^{(i)}t_l). \end{aligned} \quad (5.39)$$

Es sei weiterhin angenommen, dass das Signal  $s_i(t)$  a priori unbekannt ist. Weiterhin liegen keine Informationen vor, wann die akustische Welle ausgesendet wird. Als Schlussfolgerungen können die zeitlichen Differenzen  ${}^{(i)}t_k$  und  ${}^{(i)}t_l$  nicht direkt ermittelt werden. Aufgrund des zeitinvarianten Verhalten des Systems kann jedoch eine Beziehung zwischen den Mikrofon-signalen hergestellt werden. Es gilt:

$$s_i(t - {}^{(i)}t_l) = s_i\left(t - \underbrace{({}^{(i)}t_l - {}^{(i)}t_k)}_{{}^{(i)}\Delta t_{kl}} - {}^{(i)}t_k\right) = s_i\left((t - {}^{(i)}\Delta t_{kl}) - {}^{(i)}t_k\right). \quad (5.40)$$

Durch Einsetzen der ersten Gleichung von (5.39) in (5.40) folgt:

$$s_i(t - {}^{(i)}t_l) = {}^{(i)}\tilde{\alpha}_k \cdot s_i(t - {}^{(i)}t_k) = \underbrace{{}^{(i)}\tilde{\alpha}_l \frac{1}{{}^{(i)}\tilde{\alpha}_k}}_{{}^{(i)}\alpha_{kl}} m_k(t - {}^{(i)}\Delta t_{kl}). \quad (5.41)$$

Somit kann ein direkter Zusammenhang zwischen den beiden Mikrofonsignalen hergestellt werden:

$$m_k(t) = {}^{(i)}\alpha_{kl} \cdot m_l(t - {}^{(i)}\Delta t_{kl}) \quad (5.42)$$

Das Mikrofonsignal  $m_k(t)$  unterscheidet sich von dem Signal  $m_l(t)$  durch eine zeitliche Differenz (TDoA)  ${}^{(i)}\Delta t_{kl}$  und einer relativen Dämpfung  ${}^{(i)}\alpha_{kl}$ . Es ist ausreichend die Mikrofonsignale zu messen und somit auf die Parameter  ${}^{(i)}\Delta t_{kl}$  und  ${}^{(i)}\alpha_{kl}$  zu schließen. Daher sind keine Kenntnisse über das Signal  $s_i$  notwendig.

In realen Experimenten sind die Mikrofonsignale mit Sensorrauschen behaftet, beschrieben durch  $n_k$  und  $n_l$ . Das Rauschen ist von dem verwendeten Mikrofon und den emittierten Signalen abhängig. Im Folgenden wird angenommen, dass  $n_k \sim \mathcal{N}(0, \sigma_k^2)$  und  $n_l \sim \mathcal{N}(0, \sigma_l^2)$  normalverteilt sind mit einem Mittelwert von 0 und einer Varianz von  $\sigma_k^2$  oder  $\sigma_l^2$ .

Darüber hinaus wird die Amplitude der Mikrofonsignale durch die Digitalisierung quantisiert (vgl. Abbildung 2.5 von Kapitel 2.1.4). Dabei bezeichnet  $q_e$  den maximalen Diskretisierungsfehler. Dieser ist von der Anzahl der Bits abhängig, die für die Darstellung des Signals gewählt wurde.

Weiterhin liegen die Mikrofonsignale in einer diskretisierten Beschreibung vor, die im Folgenden mit  $m_k(n)$  und  $m_l(n)$  bezeichnet wird. Dabei charakterisiert der Index  $n \in \{1, \dots, n_t\}$  den Zeitpunkt  $t = n \cdot \frac{1}{f_s}$  der Messung mit der Abtastfrequenz  $f_s$  und Gesamtzahl  $n_t$  der Signaldaten. Es wird angenommen, dass beide Mikrofonsignale Informationen im gemeinsamen Zeitraum  $[t_g] = [t_a, t_b]$  besitzen. Dabei bezeichnet  $t_a$  den frühesten und  $t_b$  den spätesten Zeitpunkt. Darüber hinaus wird  $[t_g]$  in  $n_f$  nachfolgende Blöcke aufgeteilt. Hierfür wird der zeitliche Bereich von Block  $r \in \{1, \dots, n_f\}$  durch  $[t_a^r, t_b^r]$  bezeichnet. Um dies zu veranschaulichen, wird folgendes Beispiel verwendet. Die Messungen der Mikrofonpaare wurden im Zeitraum von 1 bis 2 Sekunden durchgeführt. Dieser zeitliche Bereich wird in 4 Blöcke mit einer Länge von 250 ms aufgeteilt:

$$[t_g] = [1; 2] = \underbrace{[1; 1, 25]}_{[t_a^1; t_b^1]} \cup \underbrace{[1, 25; 1, 5]}_{[t_a^2; t_b^2]} \cup \underbrace{[1, 5; 1, 75]}_{[t_a^3; t_b^3]} \cup \underbrace{[1, 75; 2, 0]}_{[t_a^4; t_b^4]}. \quad (5.43)$$

Statt einen einzelnen Wert für die Amplitude  $m_k(n)$  (oder  $m_l(n)$ ) zu berücksichtigen, wird ein Intervall betrachtet. Dieses ist abhängig von dem Sensorrauschen der Mikrofone:

$$[m_k(n)] = [m_k(n) - \Delta m_k, m_k(n) + \Delta m_k] \quad (5.44)$$

$$[m_l(n)] = [m_l(n) - \Delta m_l, m_l(n) + \Delta m_l]. \quad (5.45)$$

Mittels unterschiedlicher Ansätze können die Schranken  $\Delta m_k$  und  $\Delta m_l$  ermittelt werden. Ein Ansatz besteht darin, Messungen durchzuführen, wenn keine Soundquellen aktiv sind. Im Anschluss kann die Standardabweichung der jeweiligen Amplitudenwerte ermittelt werden. Durch die  $3\sigma$ -Umgebung wer-

den 99,7% aller Messwerte beschrieben und somit können die Schranken folgendermaßen festgelegt werden:

$$[m_k(n)] = [m_k(n) - 3\sigma_k; m_k(n) + 3\sigma_k] \quad (5.46)$$

$$[m_l(n)] = [m_l(n) - 3\sigma_l; m_l(n) + 3\sigma_l]. \quad (5.47)$$

Im letzten Schritt werden diese Schranken um den Quantisierungsfehler  $q_e$  ergänzt:

$$[m_k(n)] = [m_k(n) - 3\sigma_k - q_e; m_k(n) + 3\sigma_k + q_e] \quad (5.48)$$

$$[m_l(n)] = [m_l(n) - 3\sigma_l - q_e; m_l(n) + 3\sigma_l + q_e]. \quad (5.49)$$

Schließlich resultieren Schläuche (engl. *Tubes*)  $m_k(\cdot)$  und  $m_l(\cdot)$ , die die akustischen Signale beinhalten (vgl. Abbildung 5.10). Weitere Details zu Tubes sind in [BJ14] gegeben.

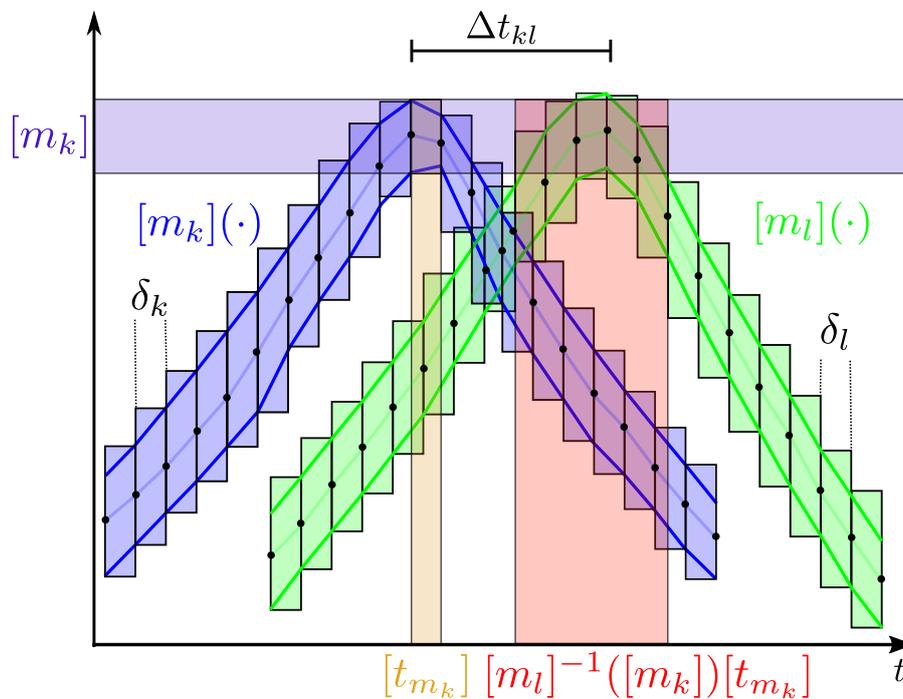


Abbildung 5.10: Tubes  $m_k(\cdot)$  und  $m_l(\cdot)$  sind dargestellt. Sie umhüllen die akustischen Signale.

Weiterhin wird ein linearer Verlauf zwischen benachbarten Messwerten  $m_k(n)$  und  $m_k(n+1)$  angenommen. Als Folgerung können die Mikrofondaten im Bereich  $[n \cdot \frac{1}{f_s}; (n+1) \cdot \frac{1}{f_s}]$  durch zwei-dimensionale Intervallboxen repräsentiert werden (vgl. Abbildung 5.10).

Insgesamt liegt das Ziel darin, die Schranken für die Intervalle  $[a_{kl}]$  und  $[\Delta t_{kl}]$  zu ermitteln. Dabei soll der Zusammenhang von Gleichung (5.42) für die Tubes  $m_k(\cdot)$  und  $m_l(\cdot)$  für alle Messungen im zeitlichen Bereich  $[t_g]$  gelten. Weiterhin muss die Grenze der Blöcke  $n_f$  in Abhängigkeit des Experimentes, der Mikrofonkonfiguration und weiterem Vorwissen der Signalcharakteristik angegeben werden.

Im ersten Schritt wird zunächst der Einfluss der Dämpfung  $a_{kl}$  vernachlässigt. Diese Annahme ist erfüllt, falls die betrachteten Mikrofone in unmittelbarer Nähe angeordnet sind oder die Mikrofone

hinsichtlich der Amplitude kalibriert sind. Somit wird ein CSP  $\mathcal{H}_2$  formuliert, das ein einzelnes Mikrofonpaar im Block  $r$  im zugehörigen Zeitbereich  $[t_a^r; t_b^r]$  berücksichtigt:

$$\mathcal{H}_2 : \begin{cases} \textbf{Variablen: } \Delta t_{kl}, m_k(\cdot), m_l(\cdot) \\ \textbf{Constraints:} \\ 1. \forall t \in [t_a^r, t_b^r] : m_k(t) = m_l(t + \Delta t_{kl}) \\ \textbf{Wertebereiche: } [\Delta t_{kl}]; [m_k](\cdot); [m_l](\cdot) \end{cases}$$

Der initiale Bereich  $[\Delta t_{kl}]$  beinhaltet die minimale und maximale zeitliche Differenz der beiden Mikrofone. Dieser Bereich wird durch die Euklidische Distanz der Mikrofone und der Schallgeschwindigkeit  $c$  beschrieben. Um das CSP  $\mathcal{H}_2$  zu lösen, wird der Kontraktor  $C_2^r$  aus [Vog20] verwendet:

$$\begin{pmatrix} [\Delta t_{kl}] \\ [m_k](t) \\ [m_l](t) \end{pmatrix} \xrightarrow{C_2^r} \begin{pmatrix} [\Delta t_{kl}] \cap \bigcap_{t_1=t_0}^{t_f} ([m_l]^{-1}([m_k](t_1)) - t_1) \cap \bigcap_{t_1=t_0}^{t_f} (t_1 - [m_k]^{-1}([m_l](t_1))) \\ [m_k](t) \cap [m_l](t + [\Delta t_{kl}]) \\ [m_l](t) \cap [m_k](t - [\Delta t_{kl}]) \end{pmatrix} \quad (5.50)$$

Weiterhin wird angenommen, dass die Soundquelle für den gesamten Zeitbereich  $[t_g]$  aktiv ist. Es gilt, dass für alle Blöcke  $r$  die korrekte TDoA  $\Delta t_{kl}$  eine Lösung darstellt. Daher wird ein globaler Kontraktor  $C_2$  berechnet, indem die Schnittmenge aller  $n_f$  Kontraktoren  $C_2^r$  gebildet wird:

$$C_2 = \bigcap_{r=1}^{n_f} C_2^r. \quad (5.51)$$

Im nächsten Schritt wird die Dämpfung berücksichtigt und das folgende CSP  $\mathcal{H}_3$  formuliert:

$$\mathcal{H}_3 : \begin{cases} \textbf{Variablen: } \Delta t_{kl}, \alpha_{kl}, m_k(\cdot), m_l(\cdot) \\ \textbf{Constraints:} \\ 1. \forall t \in [t_a^r, t_b^r] : m_k(t) = \alpha_{kl} \cdot m_l(t + \Delta t_{kl}) \\ \textbf{Wertebereiche: } [\Delta t_{kl}], [\alpha_{kl}], [m_k](\cdot), [m_l](\cdot) \end{cases}$$

Der initiale Bereich  $[\alpha_{kl}]$  wird auf Basis der euklidischen Distanz zwischen dem Mikrofonenpaar und einer Abschätzung der Sensitivität der jeweiligen Mikrofone angegeben. Um das CSP  $\mathcal{H}_3$  zu lösen, wird im Folgenden der neu-entwickelte Kontraktor  $C_3^r$  formuliert:

$$\begin{pmatrix} [\Delta t_{kl}] \\ [\alpha_{kl}] \\ [m_k](t) \\ [m_l](t) \end{pmatrix} \xrightarrow{C_3^r} \begin{pmatrix} [\Delta t_{kl}] \cap \bigcap_{t_1=t_0}^{t_f} ([m_l]^{-1}(\frac{1}{[\alpha_{kl}]} \cdot [m_k](t_1)) - t_1) \cap \bigcap_{t_1=t_0}^{t_f} (t_1 - [m_k]^{-1}([\alpha_{kl}] \cdot [m_l](t_1))) \\ [\alpha_{kl}] \cap \bigcap_{t_1=t_0}^{t_f} (\frac{[m_k](t_1)}{[m_l](t_1 + [\Delta t_{kl}])}) \cap \bigcap_{t_1=t_0}^{t_f} (\frac{[m_k](t_1 - [\Delta t_{kl}])}{[m_l](t_1)}) \\ [m_k](t) \cap ([\alpha_{kl}] \cdot [m_l](t + [\Delta t_{kl}])) \\ [m_l](t) \cap \frac{1}{[\alpha_{kl}]} \cdot ([m_k](t - [\Delta t_{kl}])) \end{pmatrix} \quad (5.52)$$

Es gilt, dass für alle Blöcke  $r$  die korrekte zeitliche Differenz  $\Delta t_{kl}$  eine Lösung darstellt. Daher wird ein globaler Kontraktor  $C_3$  berechnet, indem die Schnittmenge aller  $n_f$  Kontraktoren  $C_3^r$  gebildet wird:

$$C_3 = \bigcap_{r=1}^{n_f} C_3^r. \quad (5.53)$$

## 5.2.2 Algorithmus

Im vorherigen Abschnitt wurden die Kontraktoren  $C_2^r$  und  $C_3^r$  vorgestellt. Im Folgenden wird die Implementierung beschrieben. Für den Algorithmus für Kontraktor  $C_2^r$  sei auf [Vog20] verwiesen. Es ist Algorithmus 3 dargestellt, um die Berechnung durch Kontraktor  $C_3^r$  zu erläutern.

---

### Algorithm 3 Pseudo-Code: Kontraktor $C_3^r$

---

**Eingabe:**  $[\Delta t_{kl}], [a_{kl}], [m_k](\cdot), [m_l](\cdot), n_i$

**Ausgabe:**  $[\Delta t_{kl}], [a_{kl}], [m_k](\cdot), [m_l](\cdot)$

---

```

1: for  $i = 1$  to  $n_i$  do //Durchführung mehrerer Iterationen
2:   for  $j = 0$  to  $n_k$  do // Iteration über Tube  $[m_k](\cdot)$ 
3:      $[t_{k_j}] := [j \cdot \delta_k, (j + 1) \cdot \delta_k];$ 
4:      $[t_{s_j}] := [t_{k_j}] + [\Delta t_{kl}];$ 
5:     if  $[t_{s_j}] \subset [t_l]$  then //Ist in Zeitdomain von Tube  $[m_l](\cdot)$ 
6:        $[m_{k_j}] := [m_k](t_{k_j});$  //Auswertung von Tube
7:        $[t_{l_j}] := [m_l]^{-1}(\frac{1}{a} \cdot [m_{k_j}], t_{s_j});$ 
8:        $[\Delta t_{kl}] := [\Delta t_{kl}] \cap ([t_{l_j}] - [t_{k_j}]);$  //Kontraktion von  $[\Delta t_{kl}]$ 
9:        $[t_{s_j}] := [t_{k_j}] + \Delta t_{kl};$ 
10:       $[m_{l_j}] := [m_l]([t_{s_j}]);$  //Auswertung von Tube
11:      if  $[0, 0] \notin [m_{l_j}]$  then // Verhinderung Teilung durch 0
12:         $[a_j] := \frac{[m_{k_j}]}{[m_{l_j}]};$ 
13:         $[a_{kl}] := [a_{kl}] \cap ([a_j]);$  //Kontraktion von  $[a]$ 
14:      end if
15:       $[m_k]([t_{k_j}]) := [m_k]([t_{k_j}]) \cap ([a] \cdot [m_l]([t_{s_j}]));$  //Kontraktion von Slice von  $[m_k](\cdot)$ 
16:    end if
17:  end for
18:
19:  for  $j = 0$  to  $n_k$  do // Iteration über Tube  $[m_l](\cdot)$ 
20:     $[t_{l_j}] := [j \cdot \delta_l, (j + 1) \cdot \delta_l];$ 
21:     $[t_{s_j}] := [t_{l_j}] - [\Delta t_{kl}];$ 
22:    if  $[t_{s_j}] \subset [t_k]$  then //Ist in Zeitdomain von Tube  $[m_k](\cdot)$ 
23:       $[m_{l_j}] := [m_l](t_{l_j});$  //Auswertung von Tube
24:       $[t_{k_j}] := [m_k]^{-1}([a] \cdot [m_{l_j}], t_{s_j});$ 
25:       $[\Delta t_{kl}] := [\Delta t_{kl}] \cap ([t_{l_j}] - [t_{k_j}]);$  //Kontraktion von  $[\Delta t_{kl}]$ 
26:       $[t_{s_j}] := [t_{l_j}] - \Delta t_{kl};$ 
27:       $[m_{k_j}] := [m_k]([t_{s_j}]);$  //Auswertung von Tube
28:      if  $[0, 0] \notin [m_{k_j}]$  then // Verhinderung Teilung durch 0
29:         $[a_j] := \frac{[m_{l_j}]}{[m_{k_j}]};$ 
30:         $[a_{kl}] := [a_{kl}] \cap ([a_j]);$  //Kontraktion von  $[a]$ 
31:      end if
32:       $[m_l]([t_{l_j}]) := [m_l]([t_{l_j}]) \cap ([a] \cdot [m_k]([t_{s_j}]));$  //Kontraktion von Slice von  $[m_l](\cdot)$ 
33:    end if
34:  end for
35: end for

```

---

### 5.2.3 Erweiterungen

Um den vorgestellten Ansatz aus Abschnitt 5.2.2 für reale Signale einzusetzen, sind Konkretisierungen notwendig.

Falls die Signalenergie einen vorher festgelegten Schwellwert übersteigt, soll der Vergleich der Daten eines Mikrofonpaares durchgeführt werden. In Abhängigkeit von a-priori-Informationen wird der Schwellwert adaptiv berechnet. Berücksichtigt wird dabei die Rauschcharakteristik in der Umgebung der Messungen und der Charakteristik der zu erwarteten Signale.

Weiterhin hängt das Ergebnis stark von dem untersuchten Grenzen ab und die Berechnungszeit variiert stark von der Anzahl der Samples innerhalb eines Frames.

### 5.2.4 Ausblick - Übertragung auf mehrere Soundquellen

Mit dem entwickelten InTiD-Verfahren kann bisher nur die TDoA zu einer einzelnen Soundquelle geschätzt werden. In diesem Abschnitt wird der Ansatz auf mehrere Soundquellen erweitert.

Seien im Folgenden  $n_s$  Soundquellen gegeben, die die akustischen Signale  $s_1(t), \dots, s_{n_s}(t)$  ausstrahlen. Die Signale werden von den Mikrofonpaaren  $m_k(t)$  und  $m_l(t)$  wahrgenommen. Die Mikrofon-signale sind in Bezug zu den Signalen  $s_i(t)$  um  ${}^{(i)}t_k$  oder  ${}^{(i)}t_l$  zeitlich verschoben und die Amplitude ist um  ${}^{(i)}\tilde{a}_k$  oder  ${}^{(i)}\tilde{a}_l$  gedämpft. Für  $n_s = 2$  ist ein exemplarisches Beispiel in Abbildung 5.11 gegeben.

Die Mikrofon-signale ergeben sich aus der linearen Kombination aller ausgesendeten Sound-signalen wie folgt:

$$\begin{aligned} m_k(t) &= {}^{(1)}\tilde{a}_k \cdot s_1(t - {}^{(1)}t_k) + \dots + {}^{(n_s)}\tilde{a}_k \cdot s_{n_s}(t - {}^{(n_s)}t_k) \\ m_l(t) &= {}^{(1)}\tilde{a}_l \cdot s_1(t - {}^{(1)}t_l) + \dots + {}^{(n_s)}\tilde{a}_l \cdot s_{n_s}(t - {}^{(n_s)}t_l). \end{aligned} \quad (5.54)$$

Wenn die Signale  $s_i$  nicht bekannt sind, kann keine direkte Relation zwischen den verschiedenen Mikrofonen hergestellt werden. Im Gegensatz dazu ist es möglich bei der Anwesenheit einer einzelnen Soundquelle eine direkte Relation herzustellen.

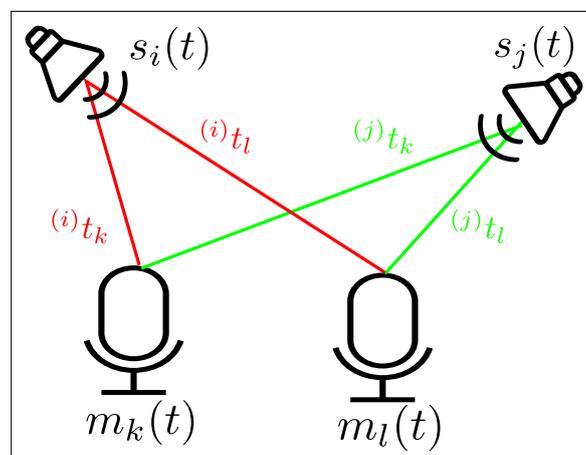


Abbildung 5.11: Die akustischen Signale von zwei Soundquellen werden durch zwei Mikrofone empfangen.

Sei angenommen, dass die Signale  $s_i$  jeweils Anteile in unterschiedlichen Frequenzbereichen besitzen, dann können die verschiedenen Signale voneinander getrennt werden:

$$\begin{aligned} {}^{(i)}m_k(t) &= {}^{(i)}\tilde{a}_i \cdot s_i(t - {}^{(i)}t_k) \\ {}^{(i)}m_l(t) &= {}^{(i)}\tilde{a}_l \cdot s_i(t - {}^{(l)}t_l). \end{aligned}$$

Hierbei bezeichnen  ${}^{(i)}m_k$  und  ${}^{(i)}m_l$  die Mikrofonsignale  $m_k(t)$  und  $m_l(t)$ , die Anteile von Signal  $s_i$  besitzen.

Im Folgenden wird das InTiD-Verfahren separat für jedes getrennte Signal durchgeführt:

$${}^{(i)}m_k(t) = {}^{(i)}a_{kl} \cdot {}^{(i)}m_l(t - {}^{(i)}\Delta t_{kl}) \quad (5.55)$$

Hierbei bezeichnet  ${}^{(i)}a_{kl}$  die Skalierung von dem Mikrofonsignal  $m_k$  und  $m_l$  für die Soundquelle  $s_i$  und  ${}^{(i)}\Delta t_{kl}$  die zeitliche Differenz  ${}^{(i)}t_k - {}^{(i)}t_l$ . Zur Visualisierung dient Abbildung 5.11.

Das Ergebnis besteht aus einer Menge von zeitlichen Differenzen zwischen den jeweiligen Mikrofonen für die getrennten Signale.

# Kapitel 6

## Experimentelle Untersuchungen und Ergebnisse

Im folgenden Kapitel wird das Soundquellenkartierungsverfahren sowohl in einer simulierten Umgebung als auch an einem mobilen Roboter evaluiert. Nachdem die einzelnen Komponenten des Kartierungsverfahrens separat evaluiert werden, erfolgt die Untersuchung des Gesamtsystems. Das entwickelte Verfahren wird mit zwei klassischen Verfahren verglichen, um Vor- und Nachteile zu identifizieren.

### 6.1 Randbedingungen der Evaluation

Zunächst werden die Randbedingungen der Evaluation genannt und darauffolgend die Ziele und der Ablauf der Evaluation beschrieben. Weiterhin werden die verwendete Simulationsumgebung und die eingesetzte Hardware näher beschrieben.

#### 6.1.1 Ziele und Ablauf der Evaluation

In der vorliegenden Dissertation liegt das Ziel darin, die dreidimensionalen Positionen von Soundquellen innerhalb einer metrischen Karte zu ermitteln. Für die Berechnung sind sowohl extrahierte akustische Merkmale der Mikrofondaten (vgl. Kapitel 5.1) als auch die Positionen der Mikrofone (vgl. Kapitel 4) notwendig. Die Evaluierung wird in den vier Stufen A, B, C und D durchgeführt (vgl. Abbildung 6.1).

#### **A: Soundquellenkartierung mit fehlerfreien Messwerten**

In Abschnitt 6.3 wird angenommen, dass fehlerfreie Messwerte für die akustischen Merkmale und Mikrofonpositionen vorliegen. Hierfür werden simulierte Daten verwendet. Dies hat den Vorteil, dass die tatsächlichen Werte (engl. *ground truth*) bekannt sind. Auf Basis dieser Werte erfolgt die Soundquellenkartierung. Es wird einerseits untersucht, ob eine korrekte Kartierung von mehreren Soundquellen ermittelt werden kann. Andererseits wird überprüft, ob das Verfahren geometrische Mehrdeutigkeiten berücksichtigt. Die Ergebnisse der intervall-basierten Kartierung werden mit den Resultaten von zwei Vergleichsverfahren gegenübergestellt.

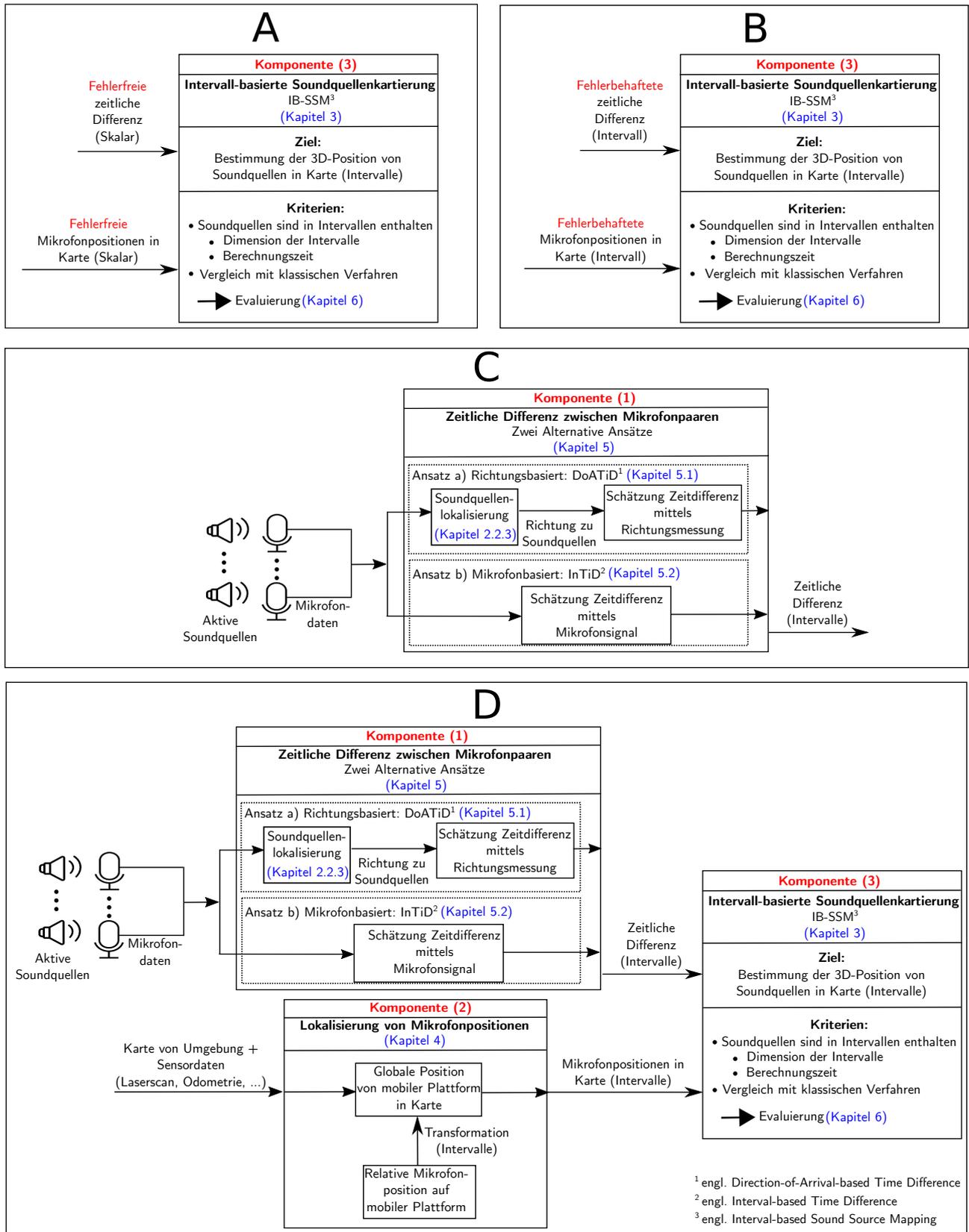


Abbildung 6.1: Aufbau der vierstufigen Evaluation.

**B: Soundquellenkartierung mit fehlerbehafteten Messwerten**

In Abschnitt 6.4 werden fehlerbehaftete Messwerte berücksichtigt, um den Einsatz in einer realen Umgebung zu ermöglichen. Es wird untersucht, wie die Ergebnisse des intervall-basierten Verfahrens und der beiden Vergleichsverfahren beeinflusst werden. Um die fehlerhaften Messwerte für das entwickelte Kartierungsverfahren anzuwenden, erfolgt die Modellierung der Unsicherheit der akustischen Merkmale (TDoA) und der Mikrofonposition durch ein intervall-basiertes Fehlermodell. Für die Position der Mikrofone wird ein systematischer Fehler angenommen. Dabei wird einerseits untersucht, wie die Ergebnisse der Kartierung beeinflusst werden, falls die relative Mikrofonposition inkorrekt angenommen wird (vgl. Abschnitt 4.1). Andererseits wird untersucht, wie sich die Annahme einer inkorrekten globalen Position von der mobilen Plattform innerhalb der metrischen Karte auf die Ergebnisse auswirkt (vgl. Abschnitt 4.2).

**C: Akustische Merkmale - Extraktion der TDoA und Modellierung der Unsicherheit**

Die beiden alternativen Verfahren DoATiD und InTiD zur Ermittlung der TDoA (vgl. Kapitel 5) werden in Abschnitt 6.5 untersucht.

DoATiD beruht auf Schätzungen der DoA (vgl. Abschnitt 5.1). Diese Schätzung hängt stark von u. a. dem verwendeten Soundquellenlokalisierungsverfahren, der Charakteristik der Soundquellen und der Raumcharakteristik ab. Um die Vielzahl von Abhängigkeiten für folgende Untersuchungen zu verringern, wird in einer simulierten Umgebung die geometrisch ermittelte DoA zu unterschiedlichen Soundquellen als Messung aufgefasst. Bevor die Schätzung der TDoA beurteilt werden kann, wird in Abschnitt 6.5.1.1 die Genauigkeit der DoA-Schätzung betrachtet. Hierfür wird in einer simulierten Umgebung für eine festgelegte Anzahl von Soundsignalen eine Evaluierung durchgeführt. Dabei wird exemplarisch das SRP-PHAT-Lokalisierungsverfahren ausgewählt, um den Einsatz in realen Experimenten zu ermöglichen. SRP-PHAT wird näher untersucht, da es in vorherigen Untersuchungen im Vergleich zu anderen Verfahren gute Ergebnisse im realen Umfeld lieferte. In Abschnitt 6.5.1.2 werden für die simulierten DoA-Messungen die Grenzen für die TDoAs durch ein universelles Verfahren innerhalb DoATiD berechnet (vgl. Abschnitt 5.1.1) und mit den tatsächlichen Werten (engl. *ground truth*) verglichen. In Abschnitt 6.5.1.3 wird die berechnete DoA vom SRP-PHAT-Verfahren betrachtet und die TDoAs durch das konkretisierte DoATiD-Verfahren (vgl. Abschnitt 5.1.5) berechnet.

Neben dem richtungsbasierten Verfahren DoATiD zur Schätzung der TDoA wird in Abschnitt 6.5.2 das mikrofonsignal-basierte Verfahren InTiD untersucht. InTiD führt intervall-basierte Berechnungen auf den Mikrofondaten aus. In einer simulierten Umgebung wird für eine festgelegte Anzahl von unterschiedlichen Soundsignalen eine Evaluierung durchgeführt. Hierbei werden unterschiedliche Signalmodelle untersucht. Die Evaluierung von InTiD bleibt beschränkt auf eine Soundquelle.

**D: Gesamtsystem Soundquellenkartierung mit Modellierung der Unsicherheiten**

In Abschnitt 6.6.1.2 wird die intervall-basierte Soundquellenkartierung unter Berücksichtigung der extrahierten akustischen Merkmale und der modellierten Unsicherheiten durchgeführt und evaluiert. Hierbei wird betrachtet, ob die Positionen der Soundquellen richtig erfasst werden und wie hoch die Genauigkeit ist. Das entwickelte intervall-basierte Verfahren wird mit einem probabilistischen Verfahren und der Lösung eines Optimierungsproblems verglichen. In Abschnitt 6.6.3 erfolgt eine Parameterstudie, die den Einfluss unterschiedlicher Parameter auf die Ergebnisse der Soundquellenkartierung hervorhebt. Zum Abschluss wird in Abschnitt 6.6.4 gezeigt, dass das intervall-basierte Verfahren in einer realen Umgebung erfolgreich eingesetzt werden kann.

### Übersicht der Abhängigkeiten:

Für den Vergleich des intervall-basierten Verfahren werden zwei klassische Verfahren verwendet. Das erste Verfahren ist ein probabilistischer Ansatz, der auf dem erweiterten Kalman-Filter (EKF) basiert. Das zweite Verfahren ist ein optimierungs-basierter Ansatz, der durch das Levenberg-Marquardt-Verfahren (LM) gelöst wird. Die drei Ansätze verwenden unterschiedliche Repräsentationen der Eingangsdaten (respektive Beobachtungen), um die Positionen der Soundquellen zu berechnen. In Abbildungen 6.2 sind die Eingangsdaten für die drei Verfahren dargestellt.

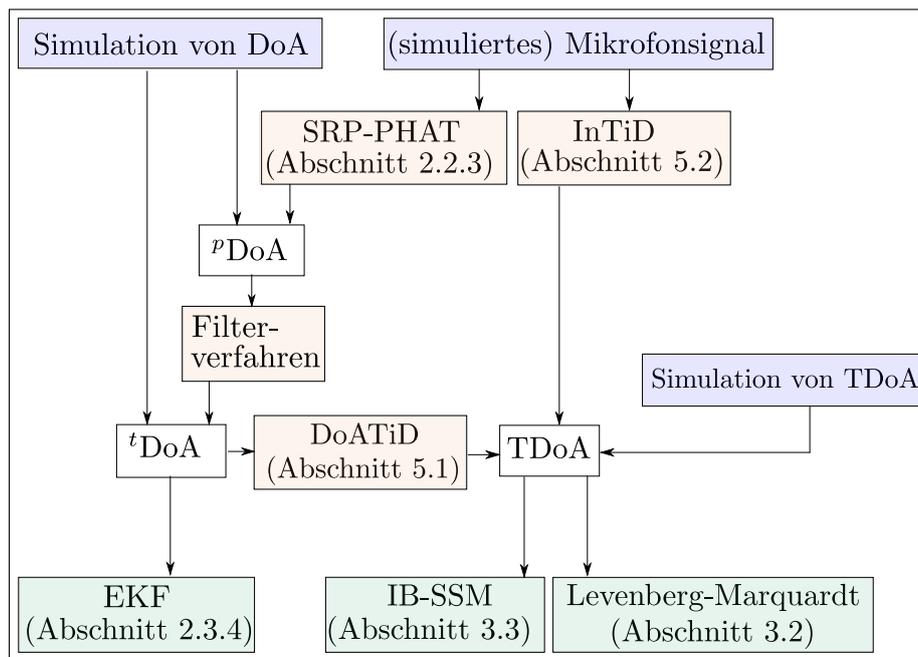


Abbildung 6.2: Übersicht der Eingangsdaten (blau) der eingesetzten Verfahren (grün).

Der EKF-Ansatz verwendet gefilterte DoA-Schätzung  ${}^t\text{DoA}$ , um die Kartierung durchzuführen. Dabei existieren drei Ansätze, um  ${}^t\text{DoA}$  zu ermitteln. Einerseits kann  ${}^t\text{DoA}$  fehlerfrei aus der Simulation vorgegeben werden. Andererseits können fehlerbehaftete Lösungskandidaten  ${}^p\text{DoA}$  und zusätzliche Fehlmessungen in der Simulation erzeugt oder durch das SRP-PHAT-Verfahren aus den Mikrofonsignalen berechnet werden. Im weiteren Schritt wird  ${}^t\text{DoA}$  durch ein Filterverfahren (z. B. M3K aus Abschnitt 5.1.2 oder GM-PHD aus Abschnitt 5.1.3) berechnet.

Für IB-SSM und das LM-Verfahren wird die TDoA benötigt. Für die Schätzung der TDoA existieren drei Ansätze. Die TDoA kann fehlerfrei aus der Simulation vorgegeben werden. Weiterhin kann die TDoA durch das InTiD-Verfahren aus den Mikrofonsignalen berechnet werden. Andererseits kann mithilfe der DoA-Schätzungen die TDoA durch das DoATiD-Verfahren berechnet werden.

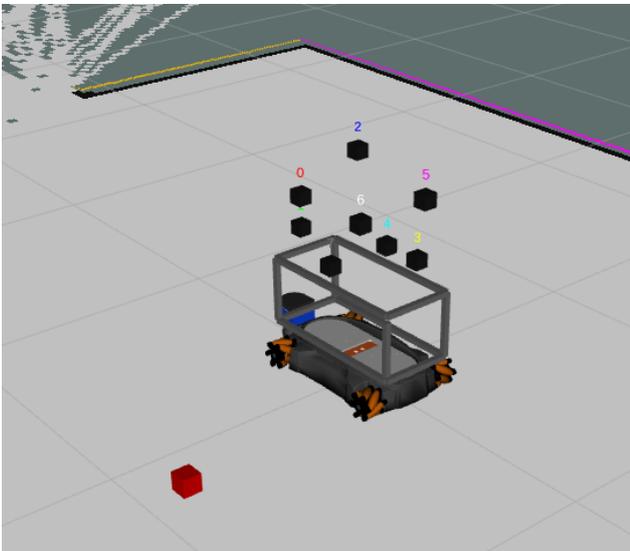
### Anmerkungen:

In dieser Dissertation wird die Komponente (2) nicht separat evaluiert (vgl. Abbildung 6.1 - Teil D). Die Folgerungen der angenommenen Unsicherheiten der Mikrofonpositionen auf die Ergebnisse der Kartierung sind in Evaluation B enthalten (vgl. Abschnitt 6.4). Es wird angenommen, dass die gewählten Grenzen für die Unsicherheit der Lokalisierung der Mikrofonpositionen korrekt gewählt wurden. In weiterführenden Arbeiten kann das Konzept auf eine garantierte Lokalisierung erweitert werden (vgl. Abschnitt 4.4). In diesem Fall ist es notwendig, die Komponente (2) separat zu evaluieren.

### 6.1.2 Experimentalumgebung

Zunächst wird das Verfahren in einer simulierten Umgebung evaluiert. Hierfür wurde in dieser Dissertation eine vereinfachte akustische Simulation implementiert. Die entwickelten Methoden werden zunächst in der Simulation unter Verwendung der Simulationssoftware Gazebo [KH04] evaluiert (vgl. Abbildung 6.3a). Um die Daten zwischen den entwickelten Modulen auszutauschen, wird das Robot Operating System (ROS) [Qui+09] verwendet. Das vorgestellte intervall-basierte Kartierungsverfahren IB-SSM wird mit der Bibliothek IBEX implementiert [Nin15].

Weiterhin werden reale Experimente unter Laborbedingungen in einem geschlossenen Raum durchgeführt. Ein Mikrofonarray, das auf einem mobilen Roboter montiert wurde, diente zur Aufzeichnung der Mikrofonsignale (vgl. Abbildung 6.3b). Im Raum werden an festen Positionen Lautsprecher montiert, die vordefinierte Signale abspielen. Neben dem Roboter und den Lautsprechern befinden sich keine weiteren Soundquellen im Raum.



(a) Simulationsumgebung



(b) Experimental-Plattform

Abbildung 6.3: Die Evaluation wurde in einer Simulationsumgebung (a) und an einem realen Versuchsaufbau (b) durchgeführt.

Alle Berechnungen werden auf einen handelsüblichen Computer (i7-6700@3,40 GHz x 8, 16 GB RAM) durchgeführt.

### 6.1.3 Roboterplattform und Sensorik

In dieser Dissertation wird das prototypische Mikrofonarray 8USBSound der kanadischen Universität Sherbrooke verwendet, das von der Firma IntRoLab vertrieben wird. Hierbei werden 8 Mikrofone eingesetzt. Dabei beträgt die Abtastfrequenz 44.100 Hz. Die Daten der verschiedenen Kanäle werden zeitlich synchronisiert. Die Anordnung der Mikrofone wird in den Experimenten dieser Arbeit variiert.

Das 8USBSound-Mikrofonarray wird auf einer mobilen Plattform, einem youBot der KUKA AG, angebracht. Die youBot-Plattform kann omnidirektional in allen Richtungen auf der Ebene bewegt werden und beinhaltet ausreichend Platz, um unterschiedliche Sensoren anzubringen.

Zur Lokalisierung der mobilen Plattform wird ein 2D-Laserscanner des Types SICK LMS200 eingesetzt. Als Soundquellen werden Lautsprecher von Mobiltelefonen verwendet.

## 6.2 Versuchsdurchführung

Im ersten Schritt wird die Soundkartierung in einer simulierten Umgebung betrachtet. Die Dimension beträgt  $(7,6\text{ m} \times 5,5\text{ m} \times 2\text{ m})$ . Zunächst wird eine Trajektorie durch die simulierte mobile Plattform abgefahren und Messungen an fünf unterschiedlichen Positionen durchgeführt. Der Versuchsaufbau ist in Abbildung 6.4 visualisiert.

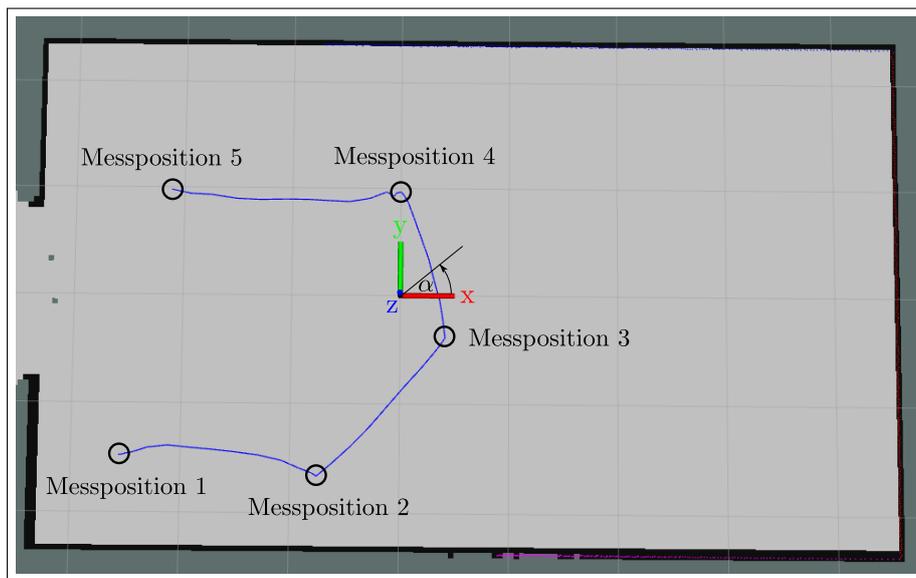


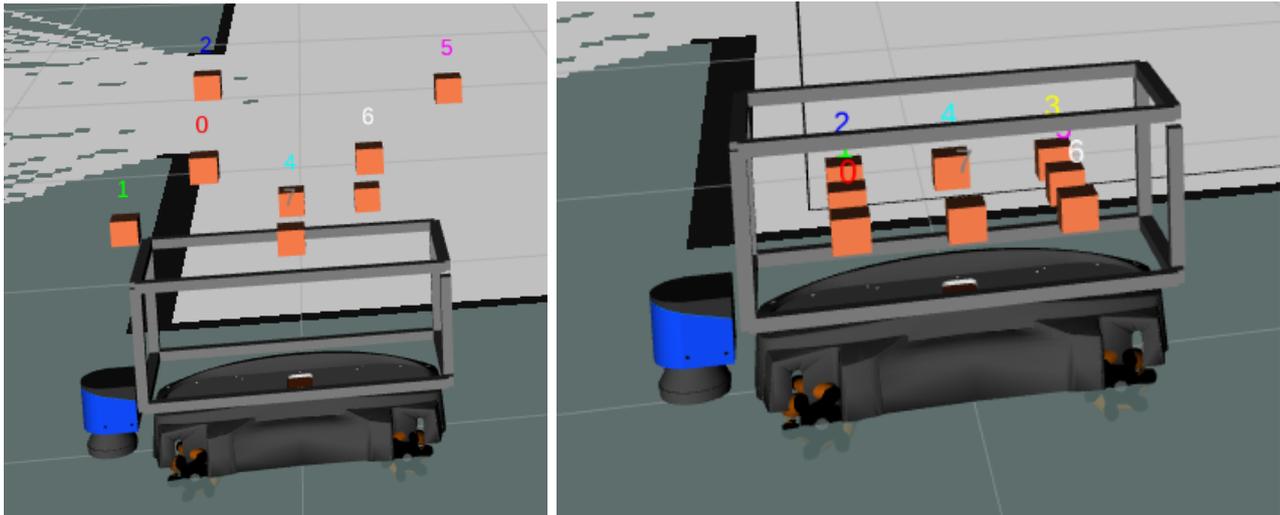
Abbildung 6.4: Der Verlauf der Trajektorie (blau) der mobilen Plattform und die fünf Messpositionen in der simulierten Umgebung.

Die Messposition und Orientierungen der mobilen Plattform sind in Tabelle 6.1 dargestellt.

Tabelle 6.1: Posen, an denen die Messungen durchgeführt wurden.  $\alpha$  kennzeichnet die Rotation um die z-Achse (vgl. Koordinatensystem in Abbildung 6.4).

Index	Position: (x;y;z) [m]	Orientierung: $\alpha$ [°]
1	(-2,57;-1,49;0,00)	345,37
2	(-0,78;-1,67;0,00)	186,31
3	(0,39;-0,39;0,00)	175,53
4	(-0,03; 0,96;0,00)	221,19
5	(-2,15; 0,98;0,00)	13,29

In den folgenden Experimenten werden die Anzahl und die Position der Soundquellen variiert. Insgesamt werden acht Mikrofone verwendet. Somit resultieren  $n_p = \binom{8}{2} = 28$  Mikrofonpaare. Für die Anordnung der Mikrofone auf der mobilen Plattform werden zwei unterschiedliche Konfigurationen betrachtet. Die Koordinaten der Mikrofone sind für die beiden Konfigurationen in Tabelle 6.2 und Tabelle 6.3 dargestellt und in Abbildung 6.5 visualisiert.



(a) Konfiguration 1

(b) Konfiguration 2

Abbildung 6.5: Dreidimensionale (a) und zweidimensionale (b) Konfiguration der Anordnung von acht Mikrofonen.

Konfiguration 1 (vgl. Abbildung 6.5a) besitzt eine dreidimensionale Struktur. Im Gegensatz dazu sind die Mikrofone von Konfiguration 2 (vgl. Abbildung 6.5b) auf einer Ebene angeordnet. Für Konfiguration 2 resultieren somit geometrische Mehrdeutigkeiten bei der Kartierung.

Tabelle 6.2: Koordinaten Konfiguration 1

Mikrofon	x[m]	y[m]	z[m]
$m_0$	-0,15	-0,1225	0,635
$m_1$	-0,31	0,000	0,485
$m_2$	-0,15	0,1225	0,735
$m_3$	0,15	0,1225	0,485
$m_4$	0,00	0,1225	0,485
$m_5$	0,31	0,0000	0,735
$m_6$	0,15	-0,1225	0,635
$m_7$	0,00	-0,1225	0,485

Tabelle 6.3: Koordinaten Konfiguration 2

Mikrofon	x[m]	x[m]	z[m]
$m_0$	-0,15	-0,1225	0,285
$m_1$	-0,15	0,0000	0,285
$m_2$	-0,15	0,1225	0,285
$m_3$	0,15	0,1225	0,285
$m_4$	0,00	0,1225	0,285
$m_5$	0,15	0,0000	0,285
$m_6$	0,15	-0,1225	0,285
$m_7$	0,00	-0,1225	0,285

## 6.3 Evaluation A: Fehlerfreie Messwerte

Die Kartierung wurde zunächst mit fehlerfreien Messwerten in der Simulation durchgeführt. Das Ziel bestand darin, zu überprüfen, ob eine korrekte Kartierung der Soundquellen durchgeführt wird. Weiterhin wurde die Genauigkeit der Lösung ermittelt und mit zwei klassischen Verfahren verglichen.

### 6.3.1 Parameter

Zunächst wurden fehlerfreie Messwerte für die TDoAs, Schallgeschwindigkeit und die Mikrofonpositionen angenommen. Die Schätzung der TDoA aus Kapitel 5 wurde daher nicht betrachtet. Im Folgenden werden die Parameter für die jeweiligen Verfahren beschrieben.

#### IB-SSM - Intervall-basierte Kartierung

Nachdem die gesamten Messungen durchgeführt wurden, wird die intervall-basierte Kartierung berechnet. Hierbei wurde der Kontraktor  $\mathcal{C}$  aus Gleichung (3.14) in Kombination mit dem SIVIA-Verfahren verwendet. Da zu Beginn keine Aussage über die Positionen der Soundquellen getroffen werden konnte, wurde die Startlösung mit der gesamten Dimension der virtuellen Umgebung ( $7,6 \text{ m} \times 5,5 \text{ m} \times 2 \text{ m}$ ) assoziiert. Weiterhin wurde die Genauigkeit  $\varepsilon$  des SIVIA-Verfahrens in den folgenden Experimenten variiert.

#### Levenberg-Marquard (LM) - Vergleichsverfahren 1

Der Ansatz aus Abschnitt 3.2 wurde als Vergleichsverfahren berücksichtigt. Dieser löst den Zusammenhang der TDoAs, der Mikrofonpositionen und der Position der Soundquellen aus Gleichung 3.1 durch Lösung eines Optimierungsproblems. Zur Lösung wurde die Methode der kleinsten Quadrate mittels Levenberg-Marquardt-Verfahren verwendet. Dieser Ansatz wird im Folgenden als LM-Verfahren bezeichnet. Für die Evaluierung wurde die Gesamtzahl der Soundquellen als bekannt vorausgesetzt. Die Datenzuordnung zwischen den TDoAs und der zugehörigen Quelle wurde als nicht bekannt vorausgesetzt. In dieser Arbeit wurde keine Strategie verwendet, um die Anzahl der Hypothesen zu verringern. Die Berechnung des LM-Verfahrens erfolgte, nachdem die gesamten Messungen durchgeführt wurden.

#### Erweiterter Kalman-Filter (EKF) - Vergleichsverfahren 2

Als zweites Vergleichsverfahren wurde der probabilistische Ansatz aus Abschnitt 2.3.4.2 verwendet. Dieser Ansatz wird im Folgenden als EKF-Ansatz referenziert. Im Gegensatz zu den anderen beiden Verfahren verwendet der EKF-Ansatz Richtungsmessungen anstatt der TDoA (vgl. Abbildung 6.2). Für die Evaluierung wurde angenommen, dass die (gefilterte) DoA-Schätzung fehlerfrei ist. Weiterhin wurde die Gesamtzahl der Soundquellen als bekannt vorausgesetzt. Insgesamt wurden 100 initiale Startlösungen ausgewählt, die gleichverteilt in der virtuellen Umgebung angeordnet wurden. Die maximale Hypothesenanzahl wurde auf 50 beschränkt. Nur für die Mikrofonkonfiguration 2 wurde die geometrische Mehrdeutigkeit im Sensormodell des EKF-Ansatzes berücksichtigt. Da der EKF-Ansatz die Schätzung rekursiv anpasst, erfolgte die Berechnung unmittelbar, nachdem eine neue Messung ausgeführt wurde.

### 6.3.2 Experiment 1 - Einzelne Soundquelle - ohne geometrische Mehrdeutigkeiten

Es wurde eine einzelne Soundquelle an der Position  $(-1,0 \text{ m}; 0,5 \text{ m}; 0,5 \text{ m})$  platziert und Konfiguration 1 verwendet.

Im Anschluss wurde IB-SSM durchgeführt. Die Genauigkeit der Lösung ist abhängig vom Parameter  $\varepsilon$  des SIVIA-Verfahrens. Eine Visualisierung der Lösung ist für drei Auflösungen in Abbildung 6.6 dargestellt.

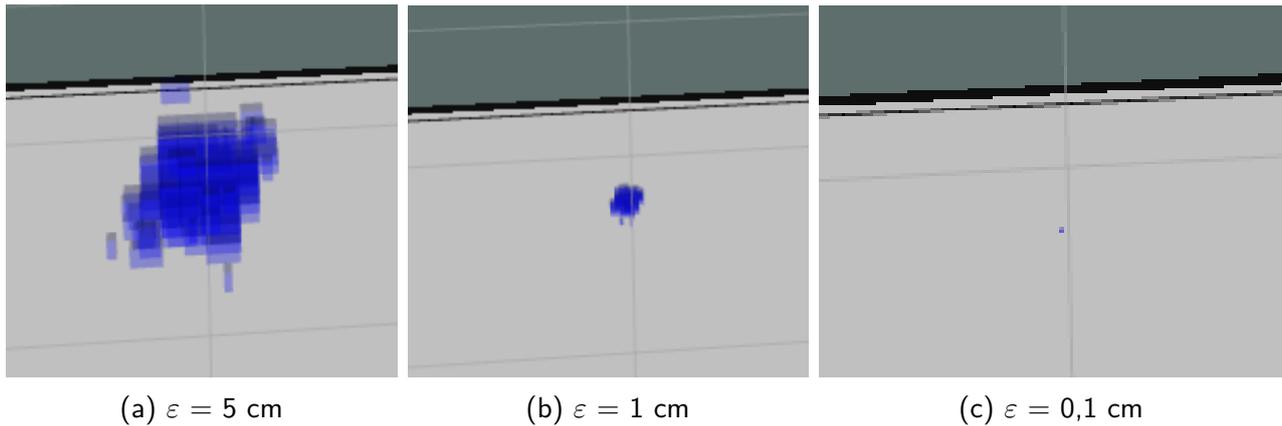


Abbildung 6.6: Ergebnis der intervall-basierten Kartierung für unterschiedliche Auflösung  $\varepsilon$ .

Für die resultierenden Intervall-Boxen wurde eine umhüllende Box ermittelt und die Dimensionen und der Schwerpunkt in Tabelle 6.4 darstellt.

Tabelle 6.4: Ergebnisse der intervall-basierten Soundquellenkartierung einer Quelle für unterschiedliche Auflösungen.

$\varepsilon$ [cm]	Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dimension: (dx;dy,dz) [cm]
5	23392,562211	$(-1,02;0,49;0,52)$	$(26,29;31,64;28,13)$
1	171,232224	$(-1,00;0,50;0,49)$	$(5,20;5,27;6,25)$
0,1	0,192990	$(-1,00;0,50;0,50)$	$(0,61;0,54;0,59)$
0,01	0,000125	$(-1,00;0,50;0,50)$	$(0,05;0,05;0,05)$

Es konnte gezeigt werden, dass die tatsächliche Soundposition in jedem Fall in der umhüllenden Box liegt.

Im nächsten Schritt wurden die beiden Vergleichsverfahren durchgeführt. Die Ergebnisse sind in Tabelle 6.5 dargestellt. Dabei wurde für den EKF-Ansatz die Hypothese dargestellt, die mit dem höchsten Gewicht assoziiert ist.

Tabelle 6.5: Ergebnisse der Positionsschätzung durch LM und EKF

Verfahren:	Schwerpunkt: (x;y;z) [m]
EKF	$(-1,01;0,50;0,50)$
LM	$(-1,00;0,50;0,50)$

Weiterhin ist in Tabelle 6.6 die Berechnungszeit des IB-SSM-Verfahrens für unterschiedliche Auflösungen und der beiden Vergleichsverfahren dargestellt.

Tabelle 6.6: Rechenzeit der Kartierung für unterschiedliche Verfahren.

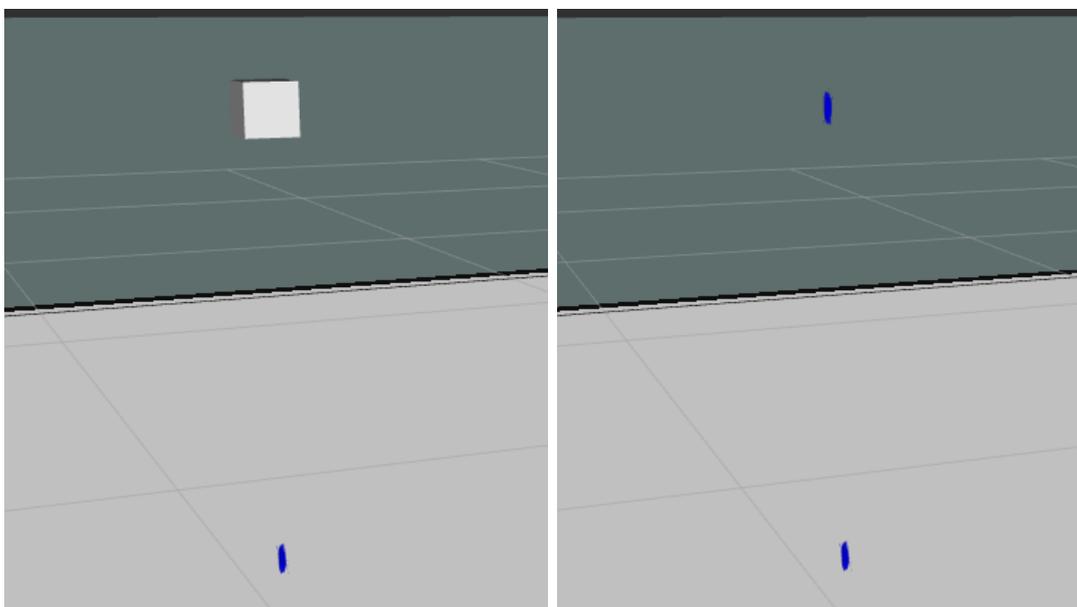
Verfahren:	Rechenzeit: [ms]
EKF	9
LM	12
IB-SSM $\varepsilon = 5$ cm	1083
IB-SSM $\varepsilon = 1$ cm	1279
IB-SSM $\varepsilon = 0,1$ cm	1800
IB-SSM $\varepsilon = 0,01$ cm	2317

Im Vergleich zu den beiden klassischen Verfahren wies das IB-SSM-Verfahren einen deutlich höheren Rechenaufwand auf. Für eine Auflösung von  $\varepsilon = 0,01$  cm betrug der zeitliche Faktor 257 höher. Die Rechenzeit der beiden Vergleichsverfahren war nahezu identisch.

### 6.3.3 Experiment 2 - Einzelne Soundquelle - geometrische Mehrdeutigkeiten

Im nächsten Experiment wurde betrachtet, ob das intervall-basierte Kartierungsverfahren geometrische Mehrdeutigkeiten berücksichtigt. Hierfür wurde Konfiguration 2 gewählt. Es wurde eine Soundquelle berücksichtigt an der gleichen Position wie im Experiment 1.

Die Ergebnisse von IB-SSM sind in Tabelle 6.7 dargestellt und in Abbildung 6.7 visualisiert.



(a) Wahre Soundquellenposition (weiß).

(b) Ergebnisse des IB-SSM-Verfahrens.

Abbildung 6.7: Geometrische Mehrdeutigkeit für eine Soundquelle. Dargestellt sind die zwei Lösungsmengen, die bei der intervall-basierten Kartierung resultieren.

Es kann erkannt werden, dass IB-SSM zwei räumlich getrennte Lösungsmengen berechnet. Die tatsächliche Lösung ist dabei enthalten.

Tabelle 6.7: Ergebnisse der Soundquellenkartierung unter geometrischen Mehrdeutigkeiten.

Cluster	Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dim: (dx;dy;dz) [cm]
1	2,50	(-1,0;0,50;0,50)	(1,03;0,73;3,32)
2	2,65	(-1,0;0,50;0,07)	(1,05;0,83;3,03)

Für das LM-Verfahren wurde je nach Wahl des Startwerts eine der beiden Lösungen aus Tabelle 6.8 angenommen. Es kann somit nicht garantiert werden, dass die wahre Lösung berechnet wird, falls die Berechnung für einen einzelnen Startwert ausgeführt wird.

Tabelle 6.8: Ergebnisse der Soundquellenkartierung durch LM-Verfahren unter geometrischen Mehrdeutigkeiten.

Index	Schwerpunkt: (x;y;z) [m]
1	(-1,00 m; 0,50 m; 0,07 m)
2	(-1,00 m; 0,50 m; 0,05 m)

Für den EKF-Ansatz wurde die Symmetrie-Ebene des Mikrofonarrays für die Berechnung berücksichtigt. Es ergaben sich für jede Messung zwei unterschiedliche Richtungen, die sich durch den Elevationswinkel  $\phi$  unterschieden. Zunächst wurde für die Standardabweichung des Messrauschens  $\sigma_Q$  aus Gleichung (2.80) der Wert 0,01 rad gewählt.

Das Ergebnis ist für die zwei Hypothesen mit dem höchsten Gewicht  $\omega$  in Tabelle 6.9 dargestellt.

Tabelle 6.9: Ergebnisse der Soundquellenkartierung durch EKF-Ansatz unter geometrischen Mehrdeutigkeiten.

Hypothese [Index ]	Schwerpunkt: (x;y;z) [m]	Gewicht
1	(-1,01;0,50;0,50)	0,86
2	(-1,01;0,50;0,08)	0,10

Für diesen Durchlauf enthielt der LM-Ansatz die wahre Position. Im Anschluss wurde  $\sigma_Q$  variiert und die Berechnung des EKF-Ansatzes wiederholt. Die Ergebnisse sind in Abbildung 6.8 visualisiert.

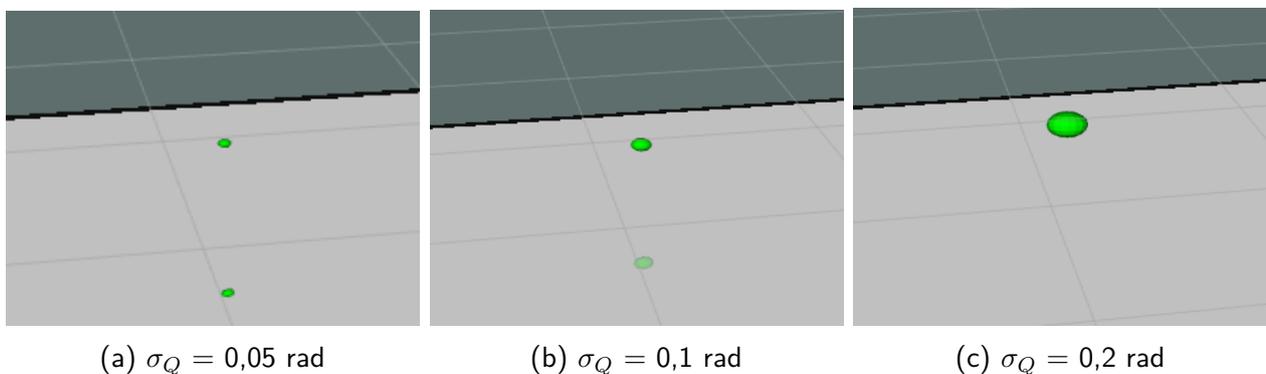


Abbildung 6.8: Ergebnisse der Soundquellenkartierung für den EKF-Ansatz (grün) für unterschiedlicher Modellierung des Messrauschens.

Falls  $\sigma_Q$  größer als 0,2 rad gewählt wurde, konnte gezeigt werden, dass durch die Hypothesenreduktion eine mögliche Lösung verworfen wurde (vgl. Abbildung 6.8c). Somit ist nicht garantiert, dass die tatsächliche Lösung durch den EKF-Ansatz berechnet wird.

### 6.3.4 Experiment 3 - Mehrere Soundquellen - ohne geometrische Mehrdeutigkeiten

Im nächsten Schritt wurden drei Soundquellen betrachtet und die Mikrofonkonfiguration 1 gewählt. Die Koordinaten der Soundquellen sind in Tabelle 6.10 dargestellt.

Tabelle 6.10: Positionen von drei Soundquellen in der simulierten Umgebung.

Index	Position: (x;y;z) [m]
1	(-1,0;0,5;0,5)
2	( 1,0;0,8;0,2)
3	(-1,0;-0,1;1,2)

Das Ergebnis des IB-SSM-Verfahrens ist durch mehrere Intervall-Boxen beschrieben. Räumlich zusammengehörige Boxen wurden durch das K-Means-Clustering ermittelt. Im Anschluss wurde für jede Gruppierung eine umhüllende Box ermittelt. Die Ergebnisse sind in Tabelle 6.11 dargestellt.

Tabelle 6.11: Die Ergebnisse der intervall-basierten Kartierung für drei Soundquellen bei einer Genauigkeit von  $\varepsilon = 0,01$  cm

Cluster	Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dimension: (dx;dy;dz) [cm]
1	2,10	(-1,00;-0,10;1,20)	(0,82;2,16;1,18)
2	0,33	( 1,00;0,80;0,20)	(0,82;0,50;0,80)
3	0,20	(-1,00;0,50;0,50)	(0,61;0,54;0,59)

Es konnte gezeigt werden, dass die tatsächlichen Positionen der Soundquellen in den umhüllenden Boxen enthalten sind.

Die Ergebnisse für den EKF-Ansatz sind in Tabelle 6.12 für die beste Hypothese dargestellt. Es kann erkannt werden, dass diese Lösung stark von der tatsächlichen Lösung abweicht.

Tabelle 6.12: Die Ergebnisse des EKF-Ansatzes für drei Soundquellen.

Index	Schwerpunkt: (x;y;z) [m]
1	(-1,01;-0,10;1,20)
2	(0,00;0,30;0,25)
3	(0,74;-0,10;0,00)

Für das LM-Verfahren ergeben sich für die fünf Messpositionen und die drei Soundquellen insgesamt  $(3!)^5 = 7776$  Hypothesen.

Die Ergebnisse der drei Verfahren sind in Abbildung 6.9 dargestellt.

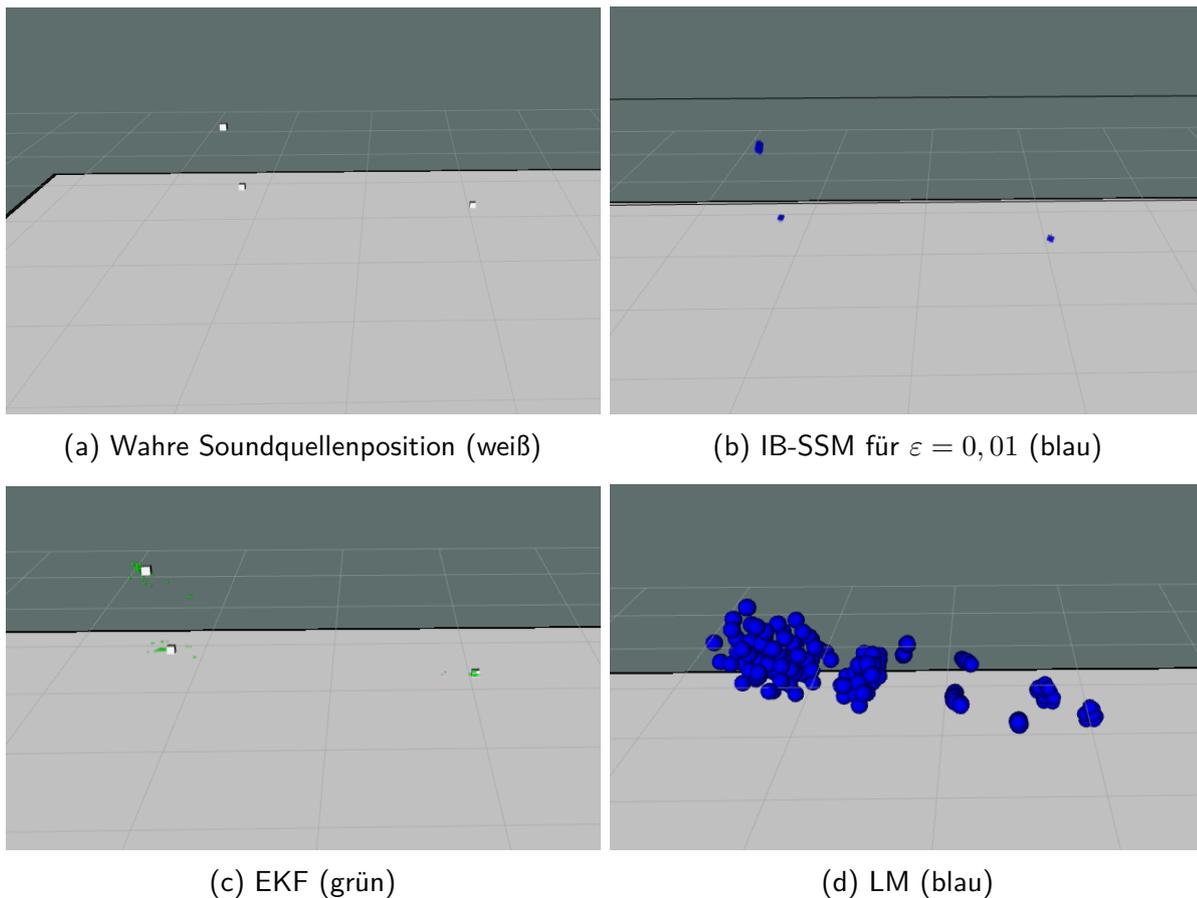


Abbildung 6.9: Ergebnisse der Soundquellenkartierung für drei Soundquellen.

Die berechneten Cluster durch das IB-SSM-Verfahren enthalten die tatsächlichen Lösungen. Für das LM-Verfahren ergeben sich viele Hypothesen. Eine genaue Schätzung der Positionen von den Soundquellen kann für das LM-Verfahren nicht berechnet werden.

Weiterhin ist in Tabelle 6.13 die Rechenzeit der unterschiedlichen Verfahren dargestellt. Der EKF-Ansatz benötigt die geringste Zeit. Die Rechenzeit vom IB-SSM-Verfahren wird stark beeinflusst von der Genauigkeit des Lösungsverfahrens. Die längste Rechenzeit benötigt das LM-Verfahren, da eine Vielzahl von Hypothesen berechnet werden müssen.

Tabelle 6.13: Rechenzeit der Kartierung von drei Soundquellen für unterschiedliche Verfahren.

Verfahren:	Rechenzeit: [s]
EKF	0,149
IB-SSM $\varepsilon = 5$ cm	6,7
IB-SSM $\varepsilon = 1$ cm	11,8
IB-SSM $\varepsilon = 0,1$ cm	15,9
LM	80

### 6.3.5 Experiment 4 - Mehrere Soundquellen - geometrische Mehrdeutigkeiten

In diesem Versuch wurde das vorherige Experiment 3 mit der Mikrofonkonfiguration 2 durchgeführt. Dabei wurde betrachtet, ob die Verfahren geometrische Mehrdeutigkeiten von mehreren Soundquellen adressieren können.

Die Ergebnisse der intervall-basierten Kartierung sind in Tabelle 6.14 dargestellt.

Tabelle 6.14: Die Ergebnisse der intervall-basierten Soundkartierung für drei Quellen und einer Genauigkeit von  $\varepsilon = 0,1$  cm

Cluster	Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dim: (dx;dy;dz) [cm]
1	2,57	(1,00;0,80;0,20)	(0,93;0,49;5,67)
2	2,50	(-1,00;0,50;0,50)	(1,03;0,73;3,32)
3	7,78	(-1,00;-0,01;1,20)	(1,39;1,62;3,45)
4	2,65	(-1,00;0,50;0,07)	(1,05;0,83;3,03)
5	3,44	(1,00;0,80;0,36)	(1,03;0,51;6,54)

Hierbei wurden insgesamt fünf Cluster ermittelt. Drei Cluster entsprechen den Positionen der drei Soundquellen und zwei Cluster sind mit geometrischen Mehrdeutigkeiten der Lösung assoziiert.

Weiterhin sind die Ergebnisse für den EKF-Ansatz für  $\sigma_Q = 0,1$  rad in Tabelle 6.15 für die beste Hypothese dargestellt. Hierbei weichen die Mittelwerte der berechneten Positionen stark von den tatsächlichen Positionen ab.

Tabelle 6.15: Die Ergebnisse des EKF-Ansatzes für drei Soundquellen.

Index	Schwerpunkt: (x;y;z) [m]
1	(-1,00;-0,10;1,19)
2	(0,00;0,07;0,16)
3	(-0,01;-0,01;0,01)

Die Ergebnisse der drei Verfahren sind in Abbildung 6.10 dargestellt. Die berechneten Cluster durch das IB-SSM-Verfahren enthalten die tatsächlichen Lösungen. Der EKF-Ansatz war in der Lage die geometrischen Mehrdeutigkeiten zu berücksichtigen. Falls jedoch, dass Messrauschen erhöht wird, ist nicht garantiert, dass die alternativen Lösungen angenommen werden. Für das LM-Verfahren ergeben sich aufgrund der geometrischen Mehrdeutigkeiten eine Vielzahl von Hypothesen. Eine genaue Schätzung der Positionen von den Soundquellen kann für das LM-Verfahren nicht berechnet werden.

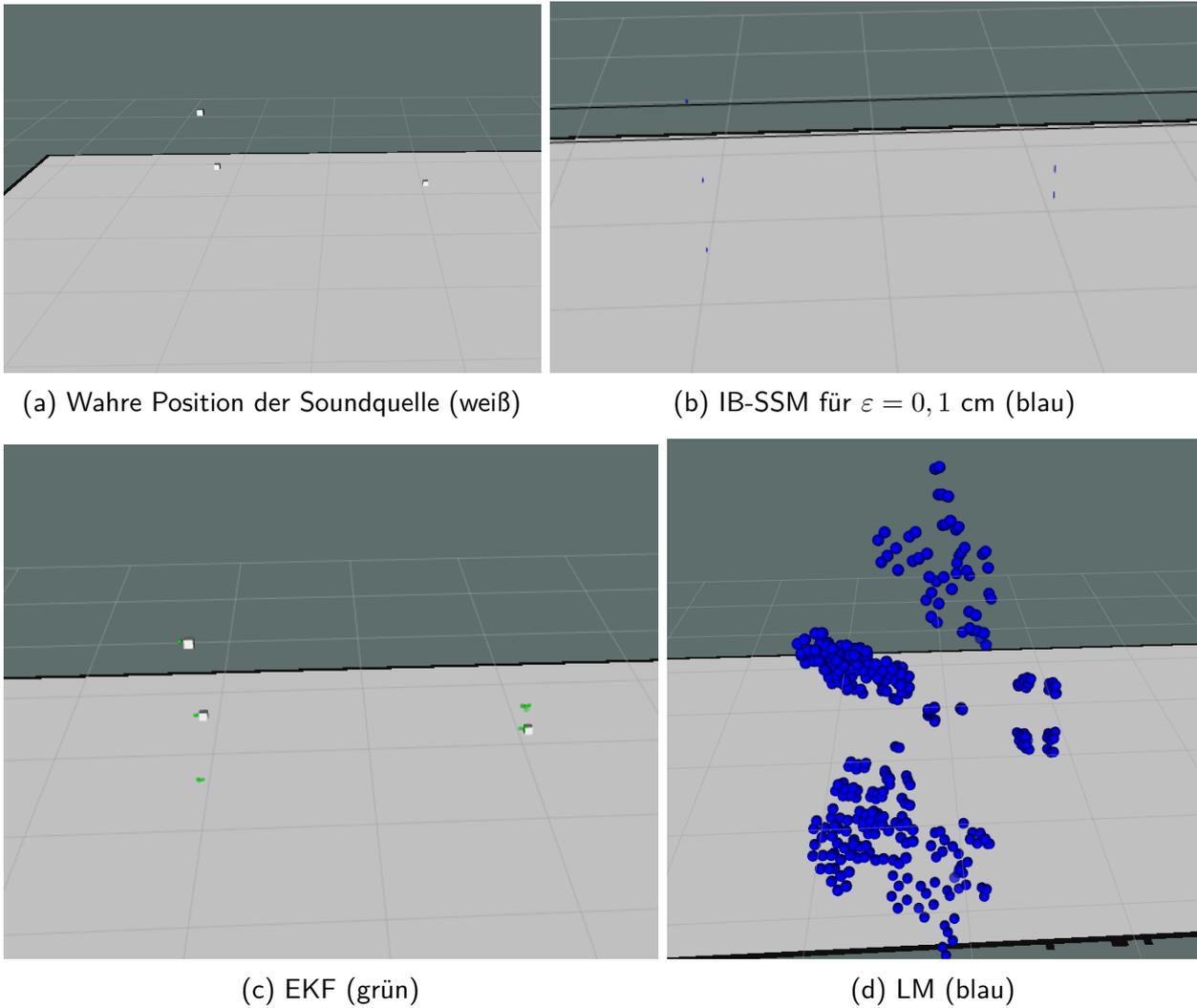


Abbildung 6.10: Ergebnisse der Soundquellenkartierung für drei Soundquellen unter Berücksichtigung geometrischer Mehrdeutigkeiten.

Weiterhin zeigt Tabelle 6.16 die Rechenzeit der unterschiedlichen Verfahren.

Tabelle 6.16: Rechenzeit der Kartierung von drei Soundquellen für unterschiedliche Verfahren.

Verfahren:	Rechenzeit: [s]
EKF	0,3
IB-SSM $\epsilon = 5$ cm	13,4
IB-SSM $\epsilon = 1$ cm	46,2
IB-SSM $\epsilon = 0,1$ cm	86,6
LM	340,9

Im Vergleich zu Experiment 3 benötigen alle Verfahren eine längere Rechenzeit, da geometrische Mehrdeutigkeiten auftreten. Die Rechenzeit vom EKF-Ansatz erhöht sich nur geringfügig. Die Rechenzeit des IB-SSM-Verfahrens für  $\epsilon = 0,1$  und das LM-Verfahren stiegen um den Faktor 4.

## 6.4 Evaluation B: Fehlerbehaftete Messwerte

Die Annahme von fehlerfreien Messwerten ist in der Realität nicht erfüllt. Im Weiteren wurde untersucht, wie die Lösung des IB-SSM-Verfahrens und der beiden Vergleichsverfahren von fehlerbehafteten Messwerten beeinflusst werden. Die Betrachtung beschränkte sich auf den Abtastfehler der Mikrofon-signale und einen systematischen Fehler der Mikrofonanordnung.

### 6.4.1 Experiment 5 - Einzelne Soundquelle - Abtastfehler

Wie in Kapitel 5 dargestellt, ist es nicht möglich die TDoA exakt zu erfassen. Die ermittelte TDoA kann aufgrund der Abtastung der Mikrofon-signale an diskreten Zeitpunkten um zwei Zeitschritte (Samples) vor oder hinter der tatsächlichen TDoA abweichen (vgl. Abbildung 5.5 in Abschnitt 5.1.1). Die Berücksichtigung dieser zeitlichen Abweichung im IB-SSM-Verfahren wird als Sampling-Erweiterung bezeichnet.

In diesem Experiment wurde untersucht, wie das Ergebnis der Soundkartierung durch die folgenden vier Fälle beeinflusst wird:

- (a) Abtastfehler tritt auf und IB-SSM berücksichtigt keine Sampling-Erweiterung,
- (b) Abtastfehler tritt auf und IB-SSM berücksichtigt Sampling-Erweiterung,
- (c) Abtastfehler tritt nicht auf und IB-SSM berücksichtigt keine Sampling-Erweiterung, sowie
- (d) Abtastfehler tritt nicht auf und IB-SSM berücksichtigt Sampling-Erweiterung.

In diesem Experiment wurde eine einzelne Soundquelle mit einer fehlerbehafteten TDoA-Messung betrachtet. Dabei wurde zufällig für jedes Mikrofonpaar die tatsächliche TDoA um -2 bis +2 Zeitschritte verschoben und als Eingang für die Soundquellenkartierung verwendet. Es wurde Mikrofon-konfiguration 1 und eine Samplingrate von 44.100 Hz verwendet.

Für Testfall (a) und (c) wurde die Unsicherheit der TDoA nicht im intervall-basierten Verfahren IB-SSM berücksichtigt. Im Anschluss wurde in Testfall (b) und (d) für IB-SSM eine Sampling-Erweiterung der TDoA von -2 bis 2 Samples berücksichtigt. Die Ergebnisse aller vier Testfälle sind in Tabelle 6.17 dargestellt.

Tabelle 6.17: Ergebnisse der Soundquellenkartierung unter Berücksichtigung von Abtastfehlern mit  $\varepsilon = 0,01$ .

Testfall	Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dim: (dx;dy;dz) [cm]
(a)	-	-	-
(b)	0,53	(-1,0;0,5;0,5)	(0,76;0,72;0,98)
(c)	0,000125	(-1,0;0,5;0,5)	(0,05;0,05;0,05)
(d)	2089,18	(-1,0;0,5;0,5)	(14,28;12,93;11,32)

Für Testfall (a) lieferte das Constraint-Satisfaction-Problem als Lösungsmenge die leere Menge. Als Vorteil kann direkt erkannt werden, dass das Fehlermodell für die Unsicherheit der TDoA nicht korrekt gewählt wurde. Eine vergleichbare Aussage kann nicht direkt für probabilistische Verfahren oder optimierungsbasierte Ansätze gefolgert werden.

Für Testfall (b) konnte gezeigt werden, dass es garantiert ist, dass die tatsächliche Lösung in den berechneten Intervallen enthalten ist. Das berechnete Volumen ist um den Faktor 4.240 größer als für Testfall (c), bei dem kein Abtastfehler auftritt und berücksichtigt wird.

Weiterhin kann in Testfall (d) erkannt werden, dass das Volumen der umhüllenden Intervallboxen sehr groß ist, falls im IB-SSM-Verfahren eine Sampling-Erweiterung berücksichtigt wird, jedoch nicht auftritt. Im Vergleich zu Testfall (b) ist das Volumen um den Faktor 3.942 größer.

Für die Testfälle (b) bis (d) entsprach der Schwerpunkt der berechneten Intervallboxen exakt der tatsächlichen Position.

Im Weiteren wurden die Vergleichsverfahren untersucht. Für das LM-Verfahren wurde keine Modellierung des Fehlers der TDoA vorgenommen. Es wurde der gleiche Samplingfehler wie bei Testfall (a) und (b) berücksichtigt. Insgesamt wurde die folgende Position berechnet: (-1,00 m; 0,51 m; 0,48 m). Dabei ergab sich eine Abweichung von 2,3 cm von der tatsächlichen Position.

Der EKF-Ansatz konnte nicht zum Vergleich herangezogen werden, da die Formulierung des mathematischen Modells durch die Richtungsschätzung gegeben ist (vgl. Abbildung 6.2). Wie in Teil C der Evaluation noch gezeigt wird, kann jedoch die Richtungsschätzung für reale Messwerte stark von der tatsächlichen Richtung abweichen.

### 6.4.2 Experiment 6 - Einzelne Soundquelle - Orientierungsfehler

Im Weiteren wurde eine Soundquelle betrachtet und ein systematischer Orientierungsfehler der Mikrofone auf der mobilen Plattform modelliert. Der Orientierungsfehler bezieht sich auf den Azimut-Winkel  $\theta$ . Es wurde die Mikrofonkonfiguration 1 gewählt, und somit keine geometrischen Mehrdeutigkeiten in diese Untersuchung einbezogen.

In folgender Untersuchung wurde der Orientierungsfehler zwischen 1 und 20° variiert und die gleichen Grenzen für die Modellierung des Fehlers beim IB-SSM-Verfahren verwendet. Es konnte gezeigt werden, dass IB-SSM keine Lösung liefert, falls der Orientierungsfehler nicht in der Modellierung berücksichtigt wird. Die Ergebnisse unter Berücksichtigung des Modellierungsfehlers sind in Tabelle 6.18 dargestellt.

Tabelle 6.18: Ergebnisse der Soundquellenkartierung unter Berücksichtigung von Orientierungsfehler mit  $\varepsilon = 0,1$  cm.

Orientierungsfehler [°]	Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dim: (dx;dy;dz) [cm]
1	2,99	(-1,00;0,50;0,50)	(1,36;1,08;2,05)
2	6,73	(-1,00;0,50;0,49)	(1,75;1,31;2,93)
5	13,11	(-0,99;0,50;0,49)	(2,28;1,51;3,81)
10	39,81	(-0,99;0,50;0,49)	(3,32;2,28;5,27)
20	130,58	(-0,99;0,50;0,48)	(5,36;3,08;7,91)

Es konnte gezeigt werden, dass die tatsächliche Lösung in jedem Fall in den berechneten Intervallen enthalten ist. Bei einem Orientierungsfehler von 10° beträgt die Entfernung zwischen dem Schwerpunkt der umhüllenden Intervallbox und der wahren Soundposition weniger als 1 cm.

Zum Vergleich wurden die Ergebnisse des EKF-Ansatzes und des LM-Verfahrens in Abbildung 6.11 dargestellt.

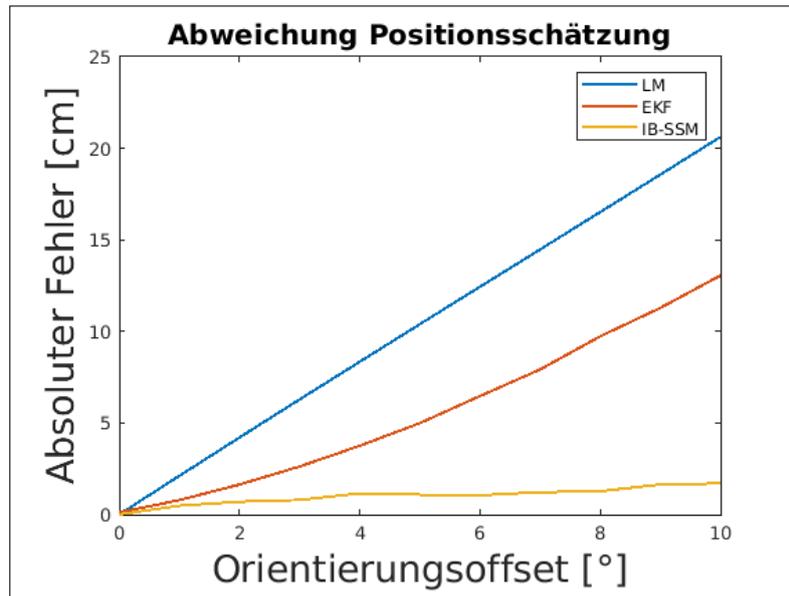


Abbildung 6.11: Absoluter Fehler der Abweichung zwischen berechneter Soundposition und wahrer Position. Es wurde das Levenberg-Marquardt-Verfahren (LM), ein Kalman-Filter-Ansatz (EKF) und das intervall-basierte Kartierungsverfahren (IB-SSM) verwendet.

Bei steigendem Orientierungsfehler steigt der absolute Fehler der beiden Vergleichsverfahren sehr stark an. Bei einem Orientierungsfehler von  $10^\circ$  ergaben sich für den LM-Ansatz eine Abweichung von 20,6 cm und für den EKF-Ansatz eine Abweichung von 13,1 cm von der tatsächlichen Position. Für IB-SSM lag die Abweichung bei 1,7 cm.

### 6.4.3 Experiment 7 - Einzelne Soundquelle - Mikrofonpositionsfehler

In diesem Experiment wurde die Unsicherheit der Position der einzelnen Mikrofone betrachtet. Dabei wurde die Konfiguration 1 verwendet und die tatsächliche Mikrofonposition in der xy-Ebene variiert. Es wurde für die Mikrofonposition jedes einzelnen Mikrofons zufällig von der ursprünglichen Position verschoben. Mit der neuen Position wurden fehlerfreie TDoAs für die jeweiligen Mikrofone ermittelt. Für den Kartierungsprozess wurde die Mikrofonposition von Konfiguration 1 betrachtet.

Für den intervall-basierten Ansatz konnte gezeigt werden, dass keine Lösung ermittelt werden kann, falls die Unsicherheit der Mikrofonpositionen nicht modelliert wird.

Im Weiteren wurde für das IB-SSM-Verfahren angenommen, dass die tatsächliche Mikrofonposition innerhalb einer Intervallbox liegt. Diese hat ihren Ursprung auf der angenommenen Mikrofonposition und eine Ausdehnung in die x- und y-Richtung.

Die Ergebnisse für das IB-SSM-Verfahren sind in Tabelle 6.19 dargestellt.

Tabelle 6.19: Ergebnisse der Soundquellenkartierung unter Berücksichtigung von Positionsfehler der einzelnen Mikrofone mit  $\varepsilon = 0,01$ .

Positionsfehler [cm]	Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dim: (dx;dy;dz) [cm]
0,5	5.256	(-1,00;0,50;0,50)	(16,6;18,5;17,2)
1	20.462	(-1,00;0,50;0,50)	(26,4;29,0;26,7)
2	207.328	(-1,01;0,42;0,41)	(46,7;65,7;67,5)
5	1.700.474	(-1,00;0,42;0,65)	(115,1;113,9;129,8)

Es kann erkannt werden, dass sich die tatsächliche Lösung in den berechneten Intervallen befindet. Jedoch ist die Schätzung sehr konservativ. Für einen Positionsfehler von 5 cm in x- und y-Richtung beträgt die Länge in jeder Dimension der berechneten Intervallboxen über einen Meter.

Die Ergebnisse für das LM-Verfahren sind in Tabelle 6.20 dargestellt.

Tabelle 6.20: Ergebnisse der Soundquellenkartierung durch das LM-Verfahren unter Berücksichtigung von Positionsfehler der einzelnen Mikrofone.

Positionsfehler [cm]	Position: (x;y;z) [m]	Abweichung [cm]:
0,5	(-1,00;0,51;0,50)	0,68
1	(-1,00;0,51;0,50)	1,35
2	(-0,98;0,52;0,50)	2,73
5	(-1,00;0,42;0,65)	6,95

Hierbei zeigt sich, dass die Abweichung vergleichbar mit den Ergebnissen des IB-SSM-Verfahrens sind.

Der EKF-Ansatz wurde für den Vergleich nicht herangezogen, da die Formulierung durch die DoA-Schätzungen gegeben ist.

## 6.5 Evaluation C: Zeitdifferenzschätzung

In den weiteren Experimenten dieses Kapitels wurde die Berechnung der zeitlichen Differenzen auf Basis von den zwei alternativen Ansätzen betrachtet. Zunächst wurde DoATiD aus Kapitel 6.5.1 und im Anschluss InTiD aus Kapitel 6.5.2 untersucht.

### 6.5.1 Richtungs-basierte Schätzung (DoATiD)

In Kapitel 5.1 wurde beschrieben, wie die zeitlichen Intervalle berechnet werden, wenn Richtungs-schätzungen von aktiven Soundquellen durch klassische Verfahren verwendet werden. Abbildung 6.12 visualisiert die Abhängigkeiten bei der Bestimmung der TDoA-Intervalle, falls DoA-Schätzungen verwendet werden.

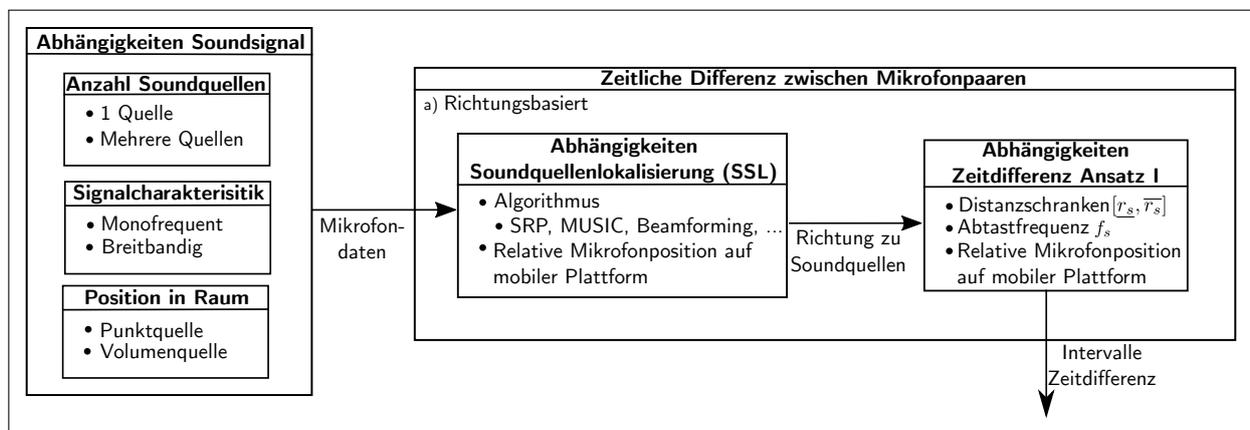


Abbildung 6.12: Abhängigkeiten bei der Bestimmung der TDoA-Intervalle unter Verwendung eines richtungsbasierten Ansatzes.

So ist das Resultat der TDoA-Intervalle von den Soundsignalen, den Soundquellenlokalisierungs-Verfahrens und den internen Parametern abhängig.

Im Abschnitt 6.5.1.1 wird in Abhängigkeit eines Soundquellenlokalisierungsverfahrens zunächst die Richtungsschätzung (DoA) unter der Berücksichtigung verschiedener Soundsignale evaluiert. In Vorarbeiten für diese Dissertation wurden unterschiedliche Schätzverfahren implementiert und untersucht. Im Folgenden wird ausschließlich auf das SRP-PHAT-Verfahren weiter eingegangen. Nachdem die DoA-Schätzung beurteilt wurde, wird in Abschnitt 6.5.1.2 auf die Ermittlung der TDoA eingegangen.

### 6.5.1.1 DoA-Schätzung durch SRP-PHAT

In diesem Abschnitt wird das SRP-Verfahren für die Richtungsschätzung kompakt untersucht. Da das Ergebnis durch unterschiedliche Einflussfaktoren beeinflussen wird, erfolgt die Evaluation des Verfahrens für eine vorgegebene Menge von Signalen. Die Evaluation beschränkt sich auf die Mikrofonkonfiguration 1 (vgl. Abbildung 6.5a).

Zur Beurteilung des Winkelfehlers der DoA-Schätzungen wird sowohl der absolute Fehler des Azimutwinkels  $e_a$  und des Elevationswinkel  $e_e$  als auch der totale Fehler  $e_t$  berücksichtigt:

$$\begin{aligned} \text{Absoluter Fehler Azimut: } e_a &= |\theta - \theta^G| \\ \text{Absoluter Fehler Elevation: } e_e &= |\phi - \phi^G| \\ \text{Absoluter Fehler Total: } e_t &= |\theta - \theta^G| + |\phi - \phi^G|. \end{aligned} \quad (6.1)$$

Hierbei bezeichnen  $\theta$  und  $\phi$  den geschätzten Azimut- und Elevationswinkel und  $\theta^G$  und  $\phi^G$  den tatsächlichen Winkeln.

#### Experiment 8 - Einzelne Soundquelle - Maximaler Fehler der DoA für SRP-PHAT

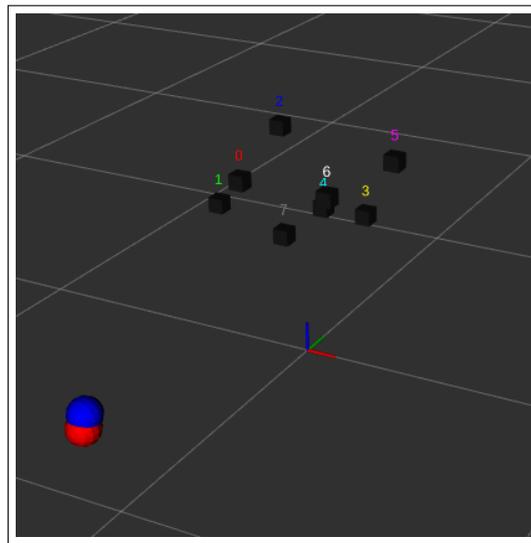


Abbildung 6.13: Eine Soundquelle wird in der simulierten Umgebung angeordnet und die Richtung wird mittels SRP-PHAT-Verfahren geschätzt. Hierfür wird Mikrofonkonfiguration 1 verwendet. blau: wahre DoA, rot: geschätzte DoA durch SRP-PHAT-Verfahren.

Zunächst wurde in einer simulierten Umgebung eine einzelne Soundquelle angeordnet. Als Signalquelle wurde weißes Rauschen verwendet. Die Interpolationsrate betrug eins und die Abtastfrequenz 44.100 Hz. Die Anzahl der Samples betrug 1024. Es wurden 163.842 Testkandidaten auf einer Kugel erzeugt und im Anschluss die obere Hälfte der Einheitskugel berücksichtigt.

In dem folgenden Experiment wurde der Elevations-Winkel fixiert und der Azimut-Winkel der Soundquelle variiert. Der absolute Fehler bezogen auf den Azimut-Winkel ist in Abbildung 6.14 dargestellt.

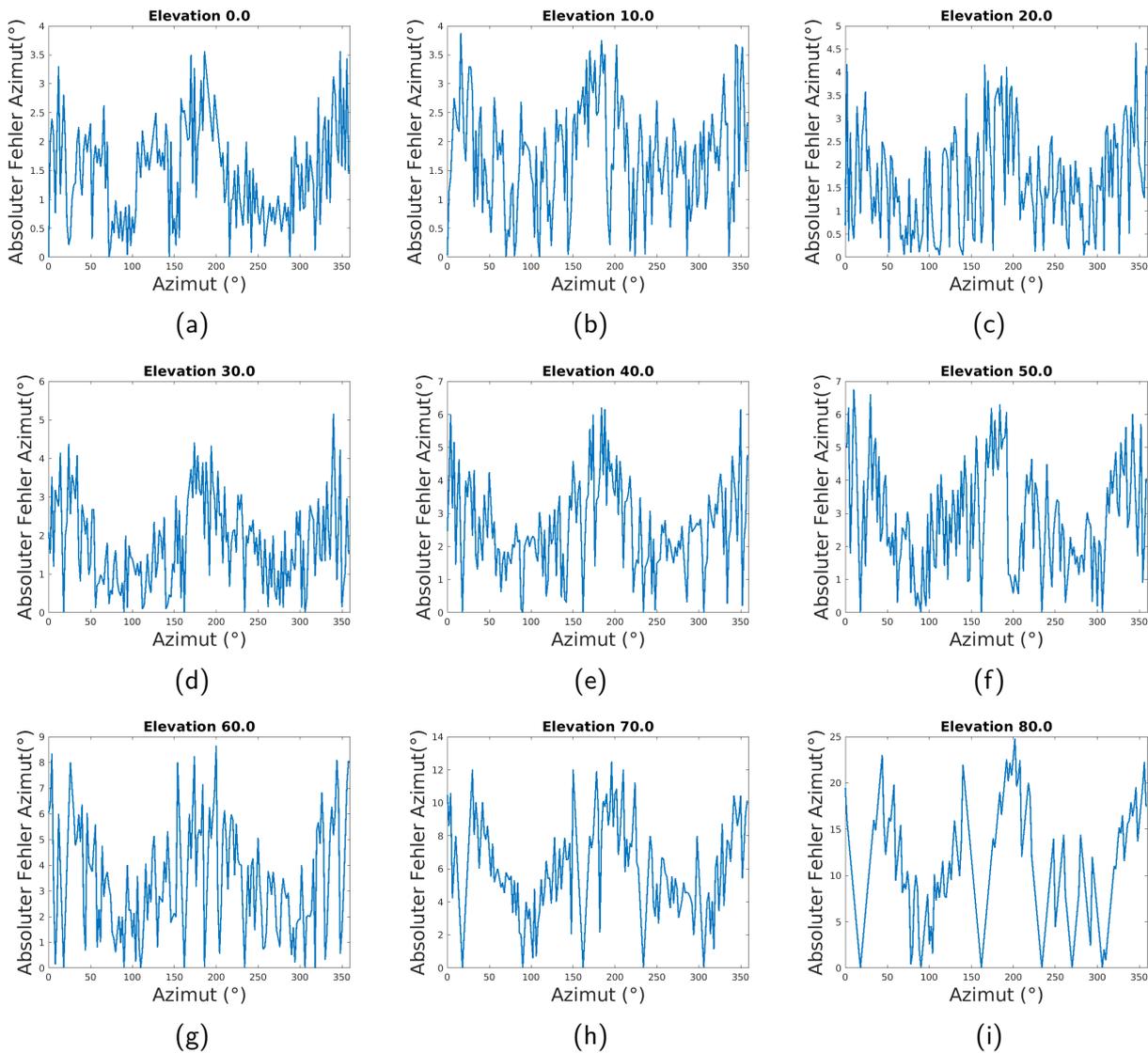


Abbildung 6.14: Absoluter Fehler der DoA-Schätzung beim SRP-PHAT-Verfahren für weißes Rauschen und Konfiguration 1.

In Tabelle 6.21 sind die Fehler der DoA-Schätzungen für unterschiedliche Elevationswinkel  $\theta$  dargestellt.

Es ist erkennbar, dass der Fehler mit steigender Elevation zunimmt. Weiterhin besitzt der Azimut-Winkel einen hohen Einfluss auf den absoluten Fehler.

Tabelle 6.21: Fehler der DoA-Messung für eine Soundquelle

$\theta$	Mean $e_t$	Min $e_t$	Max $e_t$	Mean $e_a$	Min $e_a$	Max $e_a$	Mean $e_e$	Min $e_e$	Max $e_e$
0	1,8	0	4,1	1,5	0	3,6	0,4	0	2,5
10	4,1	1	7,3	1,8	0	3,9	2,3	0	6,8
20	4,3	0,4	7,5	1,6	0	4,6	2,7	0	6,7
30	3,1	0	7,8	1,8	0	5,2	1,3	0	6,3
40	4,1	0,4	9,2	2,6	0	6,2	1,5	0	4,5
50	4,8	0,4	9,6	2,8	0	6,8	1,9	0	5,1
60	5,3	0,8	10,4	3,4	0	8,6	1,9	0,1	5,2
70	7,2	1,2	13,9	5,9	0	12,5	1,3	0	3,3
80	12,3	1,1	25,7	10,8	0	24,8	1,5	0	3,6

### Experiment 9 - Mehrere Soundquellen - SRP-PHAT

Im Folgenden wurde untersucht, ob die Richtung von mehreren aktiven Soundquellen durch das SRP-PHAT-Verfahren ermittelt werden kann. Hierfür wurden vier Soundquellen mit einem Abstand von 20 m zum Ursprung des Mikrofonarrays angeordnet. Für 100 Messpositionen wurden die Positionen der Soundquellen kontinuierlich auf einer sinusförmigen Trajektorie variiert. Es wurden durch das SRP-PHAT-Verfahren die zugehörigen DoA-Schätzungen mit den maximalen Werten verwendet. In diesem Experiment wurde pro Soundposition jeweils eine Messung durchgeführt. Die Signale der vier Soundquellen entsprachen weißem Rauschen. In Abbildung 6.15 ist der Versuchsaufbau dargestellt.

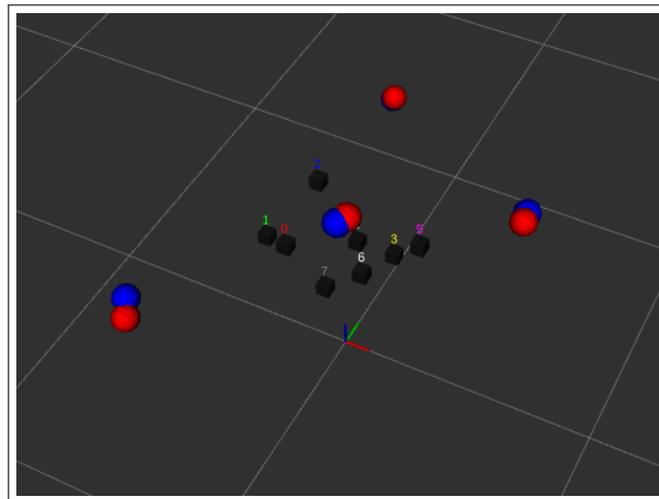


Abbildung 6.15: DoA-Schätzung zu vier gleichzeitig aktiven Soundquellen.  
blau: wahre DoA, rot: geschätzte DoA durch SRP-PHAT-Verfahren.

Der Verlauf der Winkel der jeweiligen Soundquellen und die zugehörige Schätzung sind in Abbildung 6.16 dargestellt. Die jeweiligen DoA-Schätzungen wurden für die Darstellung in Abbildung 6.16 mit der nächsten tatsächlichen Lösung (*ground truth*) assoziiert. Hierbei wurde die DoA-Schätzung betrachtet, die die minimale euklidische Distanz der projizierten Richtungsschätzungen auf die Einheitskugel besitzt. In Tabelle 6.22 sind die absoluten Fehler  $e_a$  und  $e_e$  für die vier Soundquellen dargestellt. Hierbei wurden für die 100 Messpositionen der minimale, maximale und mittlere Fehler sowie die resultierende Standardabweichung ermittelt.

Es konnte gezeigt werden, dass es mit dem verwendeten Algorithmus möglich ist die DoA von vier gleichzeitig aktiven Soundquellen zu ermitteln.

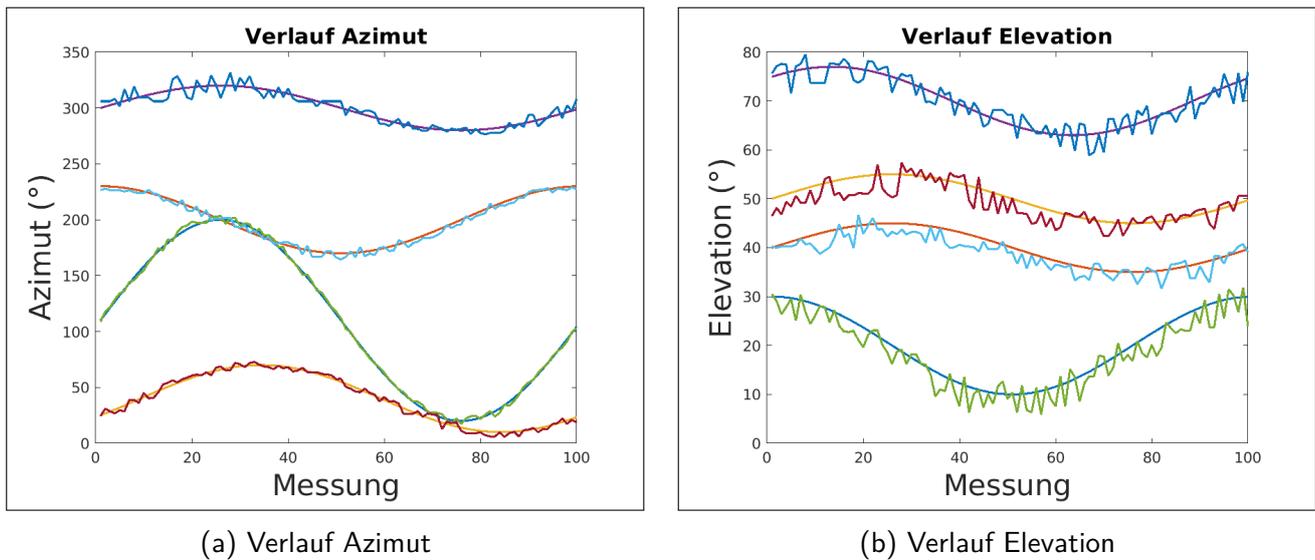


Abbildung 6.16: Ergebnisse der Richtungsschätzung des SRP-PHAT-Verfahrens für vier Soundquellen und der tatsächliche sinusförmige Verlauf für Azimut (a) und Elevation (b).

Tabelle 6.22: Fehler der DoA-Schätzungen für zwei Soundquellen.

	Mittelwert [°]	Minimum [°]	Maximum [°]	Standardabweichung [°]
Azimut 1	1,9	0	4,7	1,2
Azimut 2	2,7	0,2	5,7	1,4
Azimut 3	2,4	0	5,9	1,4
Azimut 4	4,2	0,1	11,7	2,9
Elevation 1	2,2	0	6,1	1,5
Elevation 2	1,6	0	4,5	1,1
Elevation 3	1,9	0	4,7	1,3
Elevation 4	2,2	0	6,1	1,5

### 6.5.1.2 Zeitliche Differenz durch simulierte DoA-Schätzung

Im letzten Experiment wurde dargestellt, dass die DoA-Schätzung stark von der Position der Soundquellen und der verwendeten Mikrofonanordnung abhängig ist. Ein weiterer Faktor ist der verwendete Soundquellenlokalisierungsalgorithmus. In den vorherigen Experimenten blieb die Auswertung auf den SRP-PHAT-Algorithmus beschränkt.

Um die genannten Abhängigkeiten bei der Ermittlung der Intervalle der TDoA zu verringern, wird im Folgenden auf eine simulierte DoA-Schätzung zurückgegriffen. Dabei ist in Abbildung 6.17 dargestellt, wie das folgende Vorgehen vereinfacht wird.

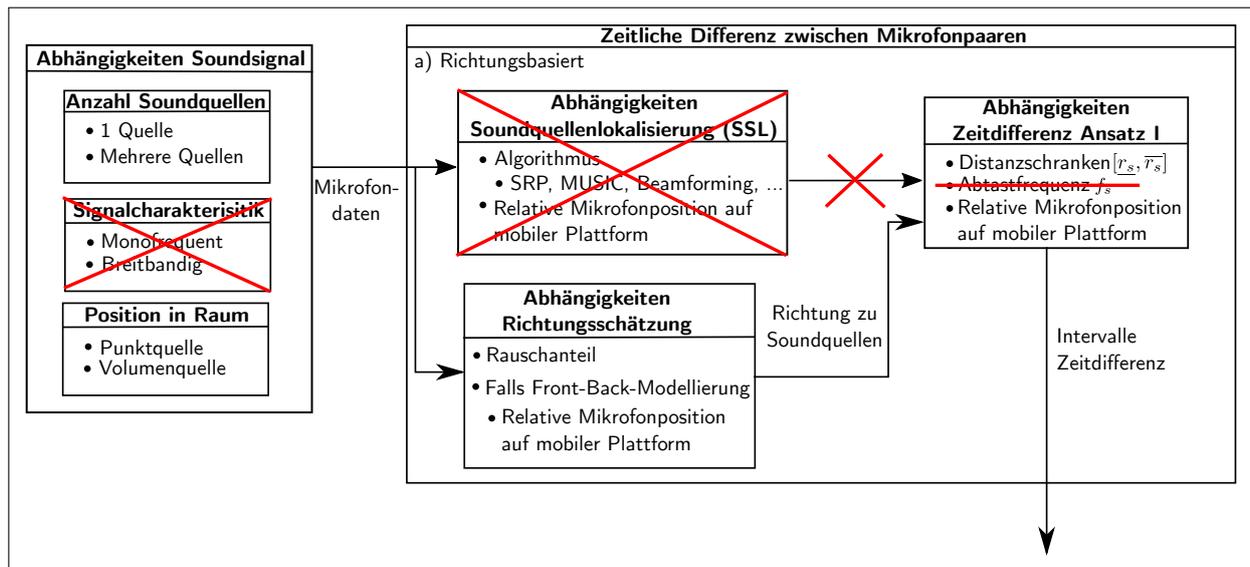


Abbildung 6.17: Zeitliche Differenz - Richtungsbasiert mit Vereinfachung.

Um einen Bezug zu den ermittelten Zeitintervallen zu erhalten, sind in Tabelle 6.23 die physikalischen maximalen Zeitdifferenzen zwischen den jeweiligen Mikrofonpaaren der beiden Mikrofonarrays aufgelistet. Dabei wurde die Zeitdauer ermittelt, die eine akustische Welle benötigt, um von dem ersten zum zweiten Mikrofon eines Paares zu gelangen.

Tabelle 6.23: Ermittelte TDoA-Intervalle für eine einzelne Soundquelle für Konfiguration 1.

Mikrofonpaar	Maximale TDoA [ms]	Mikrofonpaar	Maximale TDoA [ms]
0	0,734	14	0,852
1	0,773	15	1,391
2	1,213	16	1,169
3	0,947	17	1,113
4	1,421	18	0,438
5	0,876	19	0,938
6	0,620	20	0,839
7	0,938	21	0,839
8	1,391	22	1,217
9	0,974	23	0,947
10	1,953	24	0,716
11	1,458	25	0,657
12	0,974	26	1,217
13	1,141	27	0,620

### Experiment 10 - Einzelne Soundquelle - DoATiD - universaler Ansatz

Zunächst wurde eine einzelne Soundquelle mit einer Distanz von 4 m zu dem Ursprung des Mikrofonarrays angeordnet. Die Richtung betrug  $45^\circ$  in Azimut und  $30^\circ$  in Elevation. Im ersten Schritt wurde der Bereich  $[r_i, \bar{r}_i]$  für die Berechnung des Kegelsegmentes (vgl. Abschnitt 5.1.1) auf  $[1,5 \text{ m}; 7,0 \text{ m}]$  gesetzt. Zunächst wurde keine Sampling-Erweiterung betrachtet.

Tabelle 6.24: Ermittelte TDoA-Intervalle für eine einzelne Soundquelle

Paar	$\Delta t_{kl}^G$ [ms]	$\Delta t_{kl}$ [ms]	$\Delta \bar{t}_{kl}$ [ms]	$w([\Delta t_{kl}])$ [ms]
0	-0,25	-0,266	-0,177	0,09
1	0,561	0,478	0,574	0,095
2	0,871	0,822	1,047	0,226
3	0,586	0,544	0,737	0,193
4	1,208	1,152	1,209	0,057
5	0,564	0,554	0,578	0,024
6	0,122	0,091	0,242	0,151
7	0,811	0,655	0,84	0,185
8	1,12	1,088	1,224	0,136
9	0,836	0,81	0,913	0,103
10	1,458	1,328	1,47	0,141
11	0,814	0,753	0,82	0,067
12	0,372	0,357	0,419	0,062
13	0,309	0,248	0,569	0,321
14	0,025	-0,03	0,259	0,288
15	0,647	0,63	0,675	0,046
16	0,003	-0,02	0,098	0,118
17	-0,439	-0,483	-0,236	0,247
18	-0,285	-0,31	-0,278	0,033
19	0,338	0,105	0,382	0,277
20	-0,307	-0,471	-0,268	0,203
21	-0,748	-0,805	-0,731	0,074
22	0,622	0,415	0,66	0,244
23	-0,022	-0,161	0,01	0,171
24	-0,464	-0,495	-0,453	0,042
25	-0,644	-0,65	-0,576	0,074
26	-1,086	-1,113	-0,91	0,203
27	-0,442	-0,463	-0,334	0,129

Die Ergebnisse sind in Tabelle 6.24 dargestellt. Hierbei sind für jedes Mikrofon die tatsächliche TDoA, das geschätzte Intervall der TDoA und die Breite  $w[\Delta t]$  des Intervalls angegeben. Es ist zu erkennen, dass sämtliche Intervalle die wahre Zeitdauer beinhaltet. Je genauer die Informationen über die Grenzen der Distanz  $[\underline{r}_i; \bar{r}_i]$  sind, desto kleiner werden die resultierenden TDoA-Intervalle. Falls eine obere Grenze von 100 m gewählt wird, erhöht sich im Schnitt die Intervallbreite um den Faktor 1,7.

Im weiteren Schritt wurden 100.000 zufällig gewählte Soundpositionen erzeugt, und das vorherige Experiment wiederholt. Bei der Erzeugung wurde berücksichtigt, dass die Soundquellenpositionen jeweils im Bereich  $[\underline{r}_i; \bar{r}_i]$  liegen. Als Ergebnis konnte gezeigt werden, dass die ermittelten Intervalle in jedem Fall die wahre TDoA enthielten.

Im Folgenden wurde keine weitere Untersuchung durchgeführt, wie sich die TDoA-Schätzung verändert, falls ein systematischer Fehler auftritt (z. B. durch ungenaue Information der Mikrofonpositionen), da dieser Fehler in der Soundquellenkartierung (vgl. Kapitel 3) berücksichtigt wird.

### 6.5.1.3 Zeitliche Differenz durch SRP-PHAT-Verfahren

In vorherigem Abschnitt konnte gezeigt werden, dass die TDoA-Intervalle auf Basis von fehlerfreien DoA-Messungen berechnet werden können. Im Folgenden wird das Verfahren für ein reales Signal untersucht und das SRP-PHAT-Verfahren eingesetzt.

#### Experiment 11 - Einzelne Soundquelle - DoATiD mit SRP-PHAT

Die Grenzen für  $[\underline{r}_i; \bar{r}_i]$  wurden wie im vorherigen Experiment auf den Bereich [1,5 m;7,0 m] gesetzt. Für das Signal der Soundquelle wurde weißes Rauschen verwendet. Es wurden fünf Messwerte gesammelt und hieraus die Winkelgrenzen für das Kegelsegment ermittelt (vgl. Abschnitt 5.1.4). Da ein reales Signal betrachtet wurde, das zu diskreten Zeitpunkten abgetastet wird, wurde das TDoA-Intervall um eine Sampling-Erweiterung  $\Delta t_e$  von 0,045 ms ergänzt (vgl. Gleichung (5.6)).

Tabelle 6.25: Ermittelte TDoA-Intervalle für eine einzelne Soundquelle - durch SRP-Verfahren

Paar	$\Delta t_{kl}^G$ [ms]	$\Delta t_{kl}$ [ms]	$\Delta \bar{t}_{kl}$ [ms]	$w([\Delta t_{kl}])$ [ms]	Enthalten ohne $\Delta t_e$
0	-0,25	-0,258	-0,054	0,204	nein
1	0,561	0,4	0,631	0,23	ja
2	0,871	0,858	1,177	0,32	nein
3	0,586	0,573	0,867	0,294	nein
4	1,208	1,04	1,261	0,22	ja
5	0,564	0,521	0,626	0,105	nein
6	0,122	0,088	0,351	0,263	nein
7	0,811	0,5	0,843	0,343	nein
8	1,12	1,07	1,277	0,207	ja
9	0,836	0,786	0,967	0,181	ja
10	1,458	1,14	1,473	0,333	nein
11	0,814	0,62	0,826	0,206	nein
12	0,372	0,3	0,451	0,151	ja
13	0,309	0,272	0,732	0,459	nein
14	0,025	-0,012	0,422	0,434	nein
15	0,647	0,584	0,704	0,119	ja
16	0,003	-0,062	0,166	0,228	ja
17	-0,439	-0,497	-0,094	0,403	ja
18	-0,285	-0,356	-0,239	0,116	ja
19	0,338	-0,092	0,357	0,449	nein
20	-0,307	-0,611	-0,289	0,322	nein
21	-0,748	-0,871	-0,724	0,147	nein
22	0,622	0,218	0,642	0,424	nein
23	-0,022	-0,301	-0,005	0,297	nein
24	-0,464	-0,561	-0,44	0,121	nein
25	-0,644	-0,692	-0,474	0,218	ja
26	-1,086	-1,127	-0,734	0,393	nein
27	-0,442	-0,48	-0,215	0,266	nein

Die Ergebnisse sind in Tabelle 6.25 dargestellt. Zu erkennen ist, dass die wahre TDoA in jedem Fall in den berechneten Intervallen enthalten ist, falls die Sampling-Erweiterung betrachtet wird.

Im Vergleich zum Experiment im vorherigen Abschnitt ist jedoch die Breite der Intervalle deutlich größer. Interessant ist, dass ohne Berücksichtigung der Sampling-Erweiterung  $\Delta t_e$  die wahre TDoA für 18 von 28 Mikrofonpaaren nicht enthalten ist. Weiterhin kann die Erweiterung der Grenzen  $[\underline{r}_i; \bar{r}_i]$  diesen Effekt nicht ändern. Die Ursache der falschen Grenzen der TDoA ist durch die fehlerhafte DoA-Schätzung begründet, dargestellt in Abbildung 6.18.

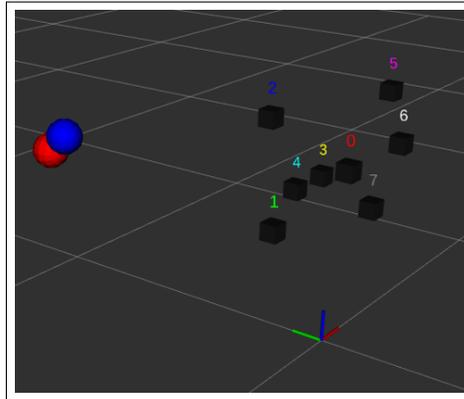


Abbildung 6.18: Fehler der DoA-Schätzung. blau: wahre DoA, rot: geschätzte DoA durch SRP-PHAT-Verfahren.

### 6.5.2 Mikrofon-basierte Schätzung (InTiD)

Nachdem im vorherigen Abschnitt dargestellt wurde, wie die TDoA-Intervalle auf Basis von DoA-Messungen berechnet wurden, erfolgte in diesem Abschnitt die Evaluierung des intervall-basierten Verfahrens InTiD. Das Verfahren führt die Berechnung auf den Mikrofondaten durch. In Kapitel 5.2.1 wurden mehrere Kontraktoren vorgestellt, um die TDoA zwischen Mikrofonpaaren aus den Mikrofonsignalen zu ermitteln. In dem vorliegenden Kapitel werden diese Kontraktoren evaluiert. Es wird ausschließlich die Anwesenheit einer einzelnen Soundquelle vorausgesetzt. In Abbildung 6.19 sind die Abhängigkeiten des Verfahrens visualisiert.

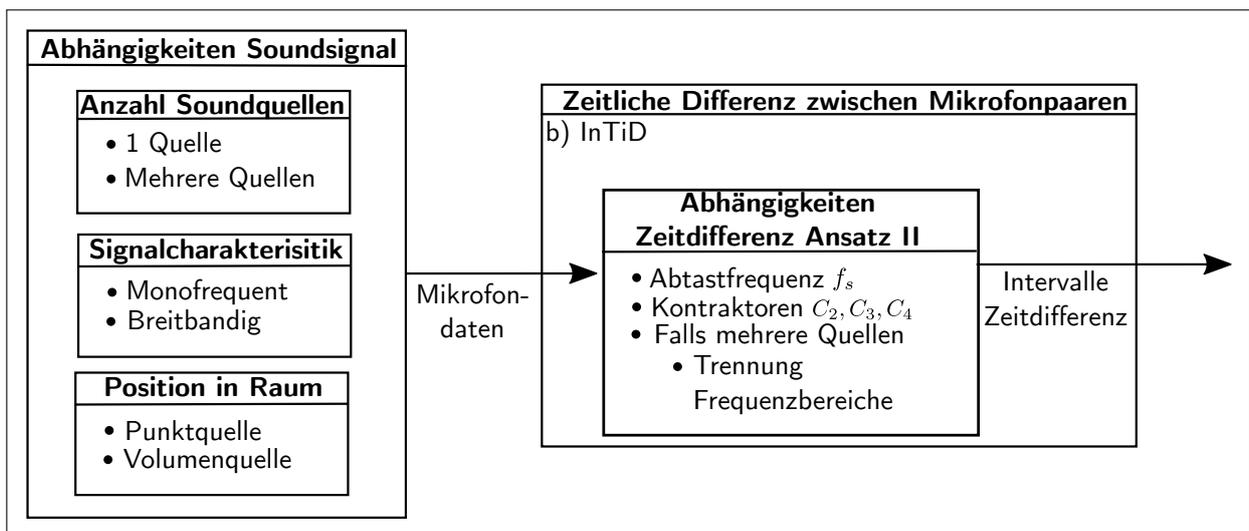


Abbildung 6.19: Abhängigkeiten für InTiD.

### 6.5.2.1 InTiD - Zeitliche Differenzen - Einzelne Soundquelle

Wie in vorherigen Abschnitt wurde eine einzelne Soundquelle mit Abstand 4 m zu dem Ursprung der Mikrofonkonfiguration 1 angeordnet. Die Richtung betrug  $45^\circ$  in Azimut und  $30^\circ$  in Elevation. Als Signal der Soundquellen wurde weißes Rauschen betrachtet.

#### Experiment 12 - InTiD ohne Modellierung der Dämpfung

Es wurde zunächst die akustische Dämpfung in der Simulation vernachlässigt. Die Länge der Datenblöcke betrug 180 Elemente. Es wurde fünf Messungen nacheinander durchgeführt und im Anschluss das Gesamtergebnis ermittelt. Im ersten Schritt wurden die Intervalle der zeitlichen Differenzen  $[\Delta t_{kl}]$  für alle Mikrofonpaare mittels Kontraktor  $C_2$  ermittelt. Der initiale Suchbereich von  $[\Delta t_{kl}]$  wurde mit der minimalen und maximalen zeitlichen Differenz belegt, die abhängig von der euklidischen Distanz der jeweiligen Mikrofonpaare und der Schallgeschwindigkeit  $c$  ist (vgl. Tabelle 6.23). Diese Grenzen wurden um die Sampling-Erweiterung ergänzt. Die Ergebnisse mit dem tatsächlichen Werten (ground truth)  $\Delta t_{kl}^G$  sind in Tabelle 6.26 angegeben.

Tabelle 6.26: Ermittelte TDoA-Intervalle durch InTiD-Ansatz

Paar $kl$	$\Delta t_{kl}^G$ [ms]	$[\Delta t_{kl}]$ [ms]	$w([\Delta t_{kl}])$ [ms]
0	-0,25	[-0,27;-0,20]	0,068
1	0,561	[0,52;0,57]	0,045
2	0,871	[0,84;0,88]	0,045
3	0,586	[0,54;0,59]	0,045
4	1,208	[1,18;1,22]	0,045
5	0,564	[0,52;0,57]	0,045
6	0,122	[0,09;0,14]	0,045
7	0,811	[0,75;0,82]	0,068
8	1,12	[1,07;1,13]	0,068
9	0,836	[0,77;0,84]	0,068
10	1,458	[-2,00;1,93]	3,926
11	0,814	[0,75;0,82]	0,068
12	0,372	[0,32;0,39]	0,068
13	0,309	[0,29;0,34]	0,045
14	0,025	[-0,02;0,05]	0,068
15	0,647	[0,63;0,68]	0,045
16	0,003	[-0,05;0,05]	0,091
17	-0,439	[-0,45;-0,39]	0,068
18	-0,285	[-0,32;-0,25]	0,068
19	0,338	[0,29;0,36]	0,068
20	-0,307	[-0,34;-0,29]	0,045
21	-0,748	[-0,77;-0,70]	0,068
22	0,622	[0,59;0,66]	0,068
23	-0,022	[-0,05;0,00]	0,045
24	-0,464	[-0,48;-0,41]	0,068
25	-0,644	[-0,68;-0,63]	0,045
26	-1,086	[-1,11;-1,04]	0,068
27	-0,442	[-0,45;-0,39]	0,068

Es ist zu erkennen, dass die wahren Werte in jedem Fall in den Intervallen enthalten sind. Weiterhin kann der Tabelle entnommen werden, dass die Breite der Intervalle deutlich kleiner ist, als bei den Experimenten, in denen DoATiD eingesetzt wurde.

### Experiment 13 - InTiD mit Modellierung der Dämpfung - SIVIA

Im nächsten Experiment wurde die akustische Dämpfung in der Simulation berücksichtigt. Zur Vereinfachung wurde jedes Mikrofonsignal mit einer positiven Zahl skaliert. Um  $[\Delta t_{kl}]$  und die Amplituden-Skalierung  $[a_{kl}]$  zwischen den jeweiligen Mikrofonen zu berechnen, wurde im ersten Experiment der SIVIA-Algorithmus mit dem Kontraktor  $C_2$  verwendet. Es wurde eine Genauigkeit von  $\varepsilon = 0,1$  für die Breite  $w([a_{kl}])$  gewählt, die für die Aufnahme in die Lösungsmenge entscheidend ist. Der initiale Suchbereich von  $[a_{kl}]$  wurde auf  $[0,1;10]$  gesetzt. Die Länge der Blöcke betrug 180 Elemente und als Signal wurde weißes Rauschen gewählt. Es wurde eine Messung durchgeführt. Die Ergebnisse und der Ground-Truth  $\Delta t_{kl}^G$  und  $a_{kl}^G$  sind in Tabelle 6.27 dargestellt.

Tabelle 6.27: Ermittelte TDoA-Intervalle unter Berücksichtigung von Dämpfung durch InTiD-Ansatz mit SIVIA und Kontraktor  $C_2$

Paar $kl$	$\Delta t_{kl}^G [ms]$	$[\Delta t_{kl}] [ms]$	$w([\Delta t_{kl}]) [ms]$	$a_{kl}^G$	$[a_{kl}]$
0	-0,25	[-0,29; -0,20]	0,091	2	[1,96; 2,03]
1	0,561	[0,50; 0,59]	0,091	1	[0,95; 1,03]
2	0,871	[0,82; 0,91]	0,091	4	[3,74; 4,04]
3	0,586	[0,52; 0,61]	0,091	1	[0,95; 1,03]
4	1,208	[1,16; 1,25]	0,091	5	[4,66; 5,05]
5	0,564	[0,50; 0,59]	0,091	3	[2,88; 3,04]
6	0,122	[0,07; 0,16]	0,091	1	[0,95; 1,03]
7	0,811	[0,75; 0,84]	0,091	0,5	[0,49; 0,56]
8	1,12	[1,07; 1,16]	0,091	2	[1,80; 2,03]
9	0,836	[0,77; 0,86]	0,091	0,5	[0,49; 0,56]
10	1,458	[1,41; 1,50]	0,091	2,5	[2,27; 2,58]
11	0,814	[0,75; 0,84]	0,091	1,5	[1,34; 1,57]
12	0,372	[0,32; 0,41]	0,091	0,5	[0,49; 0,56]
13	0,309	[0,27; 0,36]	0,091	4	[3,74; 4,04]
14	0,025	[-0,02; 0,07]	0,091	1	[0,87; 1,03]
15	0,647	[0,61; 0,70]	0,091	5	[4,66; 5,05]
16	0,003	[-0,05; 0,05]	0,091	3	[2,73; 3,04]
17	-0,439	[-0,48; -0,39]	0,091	1	[0,87; 1,03]
18	-0,285	[-0,34; -0,25]	0,091	0,25	[0,18; 0,25]
19	0,338	[0,29; 0,39]	0,091	1,25	[1,11; 1,34]
20	-0,307	[-0,36; -0,27]	0,091	0,75	[0,64; 0,80]
21	-0,748	[-0,79; -0,70]	0,091	0,25	[0,18; 0,25]
22	0,622	[0,59; 0,68]	0,091	5	[4,66; 5,05]
23	-0,022	[-0,07; 0,02]	0,091	3	[2,73; 3,04]
24	-0,464	[-0,50; -0,41]	0,091	1	[0,95; 1,03]
25	-0,644	[-0,70; -0,61]	0,09	0,6	[0,56; 0,64]
26	-1,086	[-1,13; -1,04]	0,091	0,2	[0,18; 0,25]
27	-0,442	[-0,48; -0,39]	0,091	0,333	[0,25; 0,41]

Es konnte gezeigt werden, dass sowohl  $\Delta t_{kl}^G$  als auch  $a_{kl}^G$  in den ermittelten Intervallen enthalten sind. Die Breite der Intervalle  $w([\Delta t_{kl}])$  der jeweiligen TDoAs ist größer als im vorherigen Experiment, in dem keine Dämpfungseffekte berücksichtigt wurden.

#### Experiment 14 - InTiD mit Modellierung der Dämpfung - Kontraktor

Im nächsten Schritt wurde anstelle von SIVIA der Kontraktor  $C_3$  verwendet. Hierbei wurde eine Messung durchgeführt und für den Kontraktor zwei Iterationsschritte ermittelt (vgl. Algorithmus 3). Die Ergebnisse sind in Tabelle 6.28 dargestellt.

Tabelle 6.28: Ermittelte TDoA-Intervalle unter Berücksichtigung von Dämpfung durch den InTiD-Ansatz mit Kontraktor  $C_3$

Paar $kl$	$\Delta t_{kl}^G [ms]$	$[\Delta t_{kl}] [ms]$	$w([\Delta t_{kl}]) [ms]$	$a_{kl}^G$	$[a_{kl}]$
0	-0,25	[-0,29 ; -0,20]	0,091	2	[1,90 ; 2,04]
1	0,561	[0,50 ; 0,59]	0,091	1	[0,95 ; 1,02]
2	0,871	[0,82 ; 0,91]	0,091	4	[3,79 ; 4,09]
3	0,586	[0,52 ; 0,61]	0,091	1	[0,95 ; 1,02]
4	1,208	[1,16 ; 1,25]	0,091	5	[4,74 ; 5,11]
5	0,564	[0,50 ; 0,59]	0,091	3	[2,84 ; 3,06]
6	0,122	[0,07 ; 0,16]	0,091	1	[0,95 ; 1,02]
7	0,811	[0,75 ; 0,84]	0,091	0,5	[0,47 ; 0,51]
8	1,12	[1,07 ; 1,16]	0,091	2	[1,90 ; 2,04]
9	0,836	[0,77 ; 0,86]	0,091	0,5	[0,47 ; 0,51]
10	1,458	[-1,16 ; 1,50]	2,653	2,5	[0,10 ; 3,16]
11	0,814	[0,75 ; 0,84]	0,091	1,5	[1,42 ; 1,53]
12	0,372	[0,32 ; 0,41]	0,091	0,5	[0,47 ; 0,51]
13	0,309	[0,27 ; 0,36]	0,091	4	[3,79 ; 4,09]
14	0,025	[-0,02 ; 0,07]	0,091	1	[0,95 ; 1,02]
15	0,647	[0,61 ; 0,70]	0,091	5	[4,74 ; 5,11]
16	0,003	[-0,05 ; 0,05]	0,091	3	[2,84 ; 3,06]
17	-0,439	[-0,48 ; -0,39]	0,091	1	[0,95 ; 1,02]
18	-0,285	[-0,34 ; -0,25]	0,091	0,25	[0,24 ; 0,26]
19	0,338	[0,29 ; 0,39]	0,091	1,25	[1,18 ; 1,28]
20	-0,307	[-0,36 ; -0,27]	0,091	0,75	[0,71 ; 0,77]
21	-0,748	[-0,79 ; -0,70]	0,091	0,25	[0,24 ; 0,26]
22	0,622	[0,59 ; 0,68]	0,091	5	[4,74 ; 5,11]
23	-0,022	[-0,07 ; 0,02]	0,091	3	[2,84 ; 3,06]
24	-0,464	[-0,50 ; -0,41]	0,091	1	[0,95 ; 1,02]
25	-0,644	[-0,70 ; -0,61]	0,09	0,6	[0,57 ; 0,61]
26	-1,086	[-1,13 ; -1,04]	0,091	0,2	[0,19 ; 0,20]
27	-0,442	[-0,48 ; -0,39]	0,091	0,33	[0,32 ; 0,34]

Es konnte gezeigt werden, dass sowohl  $\Delta t_{kl}^G$  als auch  $a_{kl}^G$  in den ermittelten Intervallen enthalten sind. Im Vergleich zum SIVIA-Verfahren mit  $\varepsilon = 0,1$  sind die resultierenden Intervallbreiten ein wenig größer für  $a$ , jedoch ist die Berechnungszeit deutlich geringer. Eine Übersicht der Gesamtzeit der beiden Verfahren und der Berechnung für ein Mikrofonpaar ist in Tabelle 6.29 dargestellt.

So unterscheidet sich die Gesamtzeit der beiden Verfahren um den Faktor 6. Jedoch ist die Berechnungszeit beider intervall-basierten Verfahren deutlicher höher im Vergleich zu den richtungs-basierten Ansatz DoATiD. Um die Berechnungszeit weiter zu reduzieren, könnten weniger Samples (somit Slides der Tubes) oder weniger Mikrofonpaare berücksichtigt werden.

Tabelle 6.29: Vergleich der Rechenzeit zwischen SIVIA und Kontraktor  $C_3$  unter Berücksichtigung von 28 Mikrofonpaaren.

Verfahren	Gesamtzeit [ms]	Median [ms]	Minimum [ms]	Maximum [ms]
SIVIA mit $\varepsilon = 0,1$	59758,3	2134,3	1005,0	4154,72
Kontraktor $C_3$ : 2 Iterationen	10084,0	360,2	188,6	685,5

## 6.6 Evaluation D: Gesamtsystem - Soundquellenkartierung

In diesem Abschnitt wird die Kartierung von Soundquellen im Gesamtsystem untersucht. Die Ergebnisse der Soundquellenkartierung wird in Abschnitt 6.6.1 unter Verwendung von DoATiD und in Abschnitt 6.6.2 unter Verwendung von InTiD untersucht. Im Anschluss wird in Abschnitt 6.6.3 der Einfluss unterschiedlicher Parameter für die Kartierung evaluiert. Zum Abschluss wird in Abschnitt 6.6.4 gezeigt, dass die entwickelte Soundquellenkartierung in realen Umgebungen erfolgreich eingesetzt werden kann.

### 6.6.1 Richtungs-basierte Schätzung (DoATiD)

Um die Abhängigkeit des Soundquellenlokalisierungs-Verfahrens und der Charakteristik der Signalquellen zu vernachlässigen, wurden zunächst in Abschnitt 6.6.1.1 fehlerfreie Richtungsmessungen betrachtet. Im Anschluss wurden in Abschnitt 6.6.1.2 simulierte Mikrofon-signale verwendet. Die potenziellen Richtungsschätzungen wurden durch das SRP-PHAT-Verfahren durchgeführt und die intervall-basierte Kartierung untersucht.

Es wurden die fünf Messpositionen von Abschnitt 6.2 betrachtet. Weiterhin wurde eine einzelne Soundquelle betrachtet und an der Position (-1,0 m; 0,5 m; 0,9 m) platziert. Als Mikrofonanordnung wurde die Konfiguration 1 verwendet.

#### 6.6.1.1 Kartierung mit simulierten DoA

Es wurden fehlerfreie Richtungsmessungen angenommen. Für das richtungs-basierte Verfahren DoATiD wurden die Grenzen  $[r_s, \bar{r}_s]$  auf den Bereich  $[0,5 \text{ m}; 9,6 \text{ m}]$  gesetzt. Weiterhin wurde die Genauigkeit  $\varepsilon$  des SIVIA-Verfahrens auf 1 cm gesetzt.

Das Ergebnis des intervall-basierten Verfahrens ist in Tabelle 6.30 dargestellt.

Es konnte gezeigt werden, dass die wahre Lösung enthalten ist.

Tabelle 6.30: Ergebnisse der intervall-basierten Soundquellenkartierung unter Verwendung der richtungsbasierten Methode.

Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dimension: (dx;dy;dz) [cm]
51.417	(-0,95;0,49;0,94)	(33,19;25,49;32,03)

Für das LM-Verfahren wurde der Mittelwert der resultierenden Intervalle für die TDoA verwendet. Die Schätzung der Position betrug (-1,00; 0,52; 0,78).

Der EKF-Ansatz verwendet Richtungsmessungen anstelle der TDoAs. Daher unterscheidet sich das Ergebnis nicht von denen fehlerfreien Annahmen in Evaluation A und wurde hier nicht untersucht.

### 6.6.1.2 Kartierung mittels SRP-PHAT

Statt fehlerfreie Richtungsschätzungen zu verwenden, wurde im folgenden Experiment eine Richtungsschätzung mit dem SRP-PHAT-Verfahren durchgeführt. Dabei wurde eine potenzielle Richtung betrachtet. Das Ergebnis des intervall-basierten Verfahrens ist in Tabelle 6.31 dargestellt.

Tabelle 6.31: Ergebnisse der intervall-basierten Soundquellenkartierung unter Verwendung der richtungsbasierten Methode.

Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dimension: (dx;dy;dz) [cm]
33.767	(-1,04;0,49;0,75)	(30,71;28,13;39,10)

Es konnte gezeigt werden, dass die wahre Lösung in den berechneten Intervallen enthalten ist.

Die Ergebnisse des LM-Verfahrens und der besten Hypothese des EKF-Ansatzes sind in Tabelle 6.32 dargestellt.

Tabelle 6.32: Ergebnisse der Vergleichsverfahren unter Verwendung von DoATiD.

Verfahren	Schwerpunkt: (x;y;z) [m]	Absoluter Fehler [cm]
LM	(-1,04;0,52;0,65)	25,8
EKF ( $\sigma_Q = 0,3$ rad)	(-1,04;0,50;0,76)	14,5

### 6.6.2 Mikrofon-basierte Schätzung (InTiD)

In diesem Experiment wurde der mikrofon-basierte Ansatz InTiD für die Ermittlung der TDoA untersucht. Akustische Dämpfung wurde nicht modelliert und somit der Kontraktor  $\mathcal{C}_2$  für die Berechnung verwendet. Es wurden alle 28 Mikrofonpaare verwendet, 150 Samples des Signals betrachtet und eine Iteration der Berechnung durchgeführt.

Das Ergebnis des intervall-basierten Verfahrens ist in Tabelle 6.33 dargestellt. Die wahre Lösung ist in den berechneten Intervallen enthalten.

Für das LM-Verfahren wurde der Mittelwert der resultierenden Intervalle für die TDoA verwendet. Die Schätzung der Position betrug (-1,13; 0,56; 0,71) und die Abweichung 23,8 cm.

Da der EKF-Ansatz Richtungsmessungen verwendet, kann die ermittelte TDoA nicht berücksichtigt werden.

Tabelle 6.33: Ergebnisse der intervall-basierten Soundquellenkartierung unter Verwendung von In-TiD.

Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dimension: (dx;dy;dz) [cm]
20.698	(-1,01;0,48;0,90)	(30,09;27,25;25,24)

### 6.6.3 Parameterstudie - Einfluss unterschiedlicher Parameter

In diesem Experiment wurde der Einfluss unterschiedlicher Parameter auf die Genauigkeit der Soundquellenkartierung evaluiert.

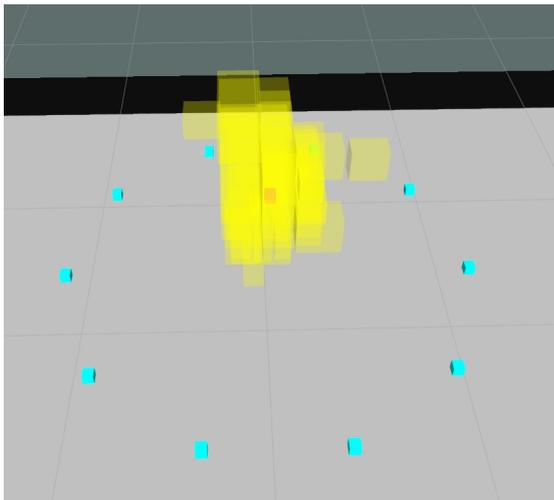
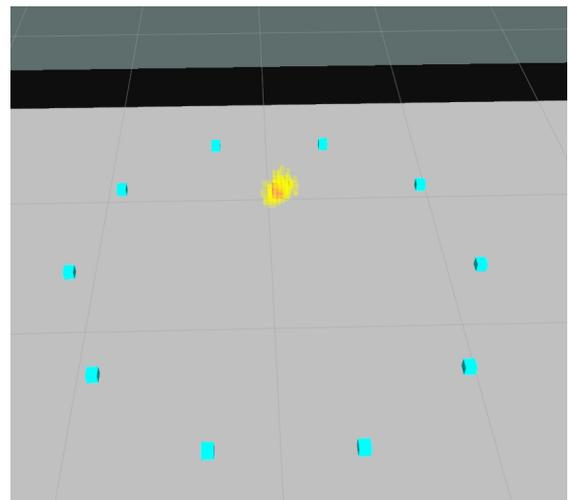
(a) Genauigkeit  $\varepsilon = 30$  cm(b) Genauigkeit  $\varepsilon = 5$  cm

Abbildung 6.20: Einfluss der Genauigkeit von SIVIA. Messpositionen (blau), Soundquellenpositionen (rot), Lösung von IB-SSM (gelb)

Der Fokus der Parameterstudie bilden die folgenden Parameter:

- $n_l$  : Anzahl von Messpositionen
- $d$  : Distanz zwischen Messposition und Soundquelle
- $\varepsilon$  : Genauigkeit des Branch-and-Bound-Algorithmus *SIVIA* zur Lösung des CSPs
- $\Delta t_e$  : Sampling-Erweiterung der TDoA.

Um die Evaluation in diesem Abschnitt zu vereinfachen, wurde nicht auf unterschiedliche Mikrofonkonfiguration eingegangen. Für das folgende Experiment wurden die Positionen der Messungen auf einem Kreis angeordnet und eine einzelne Soundquelle im Zentrum mit einer Höhe von einem Meter platziert. Die Anzahl der Messpositionen  $n_l$  wurde zwischen 1 bis 20, der Radius des Kreises zwischen 0,5 m bis 2 m und die Genauigkeit  $\varepsilon$  von *SIVIA* zwischen 0,1 und 0,3 m variiert. Hierbei spezifiziert  $\varepsilon$  die Dimension der Intervallboxen, dargestellt in Abbildung 6.20. Die ermittelte TDoA aus Kapitel 5 wurde um die Sampling-Erweiterung  $\Delta t_e$  in Gleichung (5.6) erweitert. In diesem Experiment wurde eine minimale Abtastfrequenz  $f_s$  von 8.000 Hz gewählt. Dies entspricht einer Abtastzeit  $t_s$  von 0,125 ms. Daher wurde die TDoA Sampling-Erweiterung  $\Delta t_e$  zwischen 0 ms und  $2/f_s = 0,25$  ms gesetzt. Für alle Kombinationen dieser genannten Parameter wurde das gesamte Volumen der Intervall-Boxen, sowie die Berechnungszeit des IB-SSM-Algorithmus bei Verwendung des globalen Kontraktors aus Gleichung (3.14) berücksichtigt. Die Ergebnisse sind in Abbildung 6.21 visualisiert.

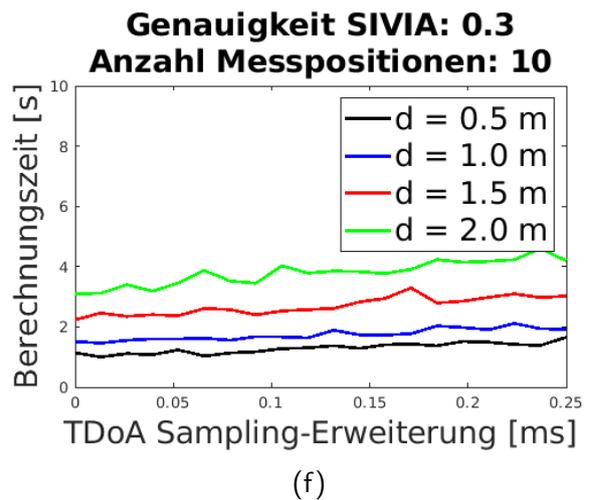
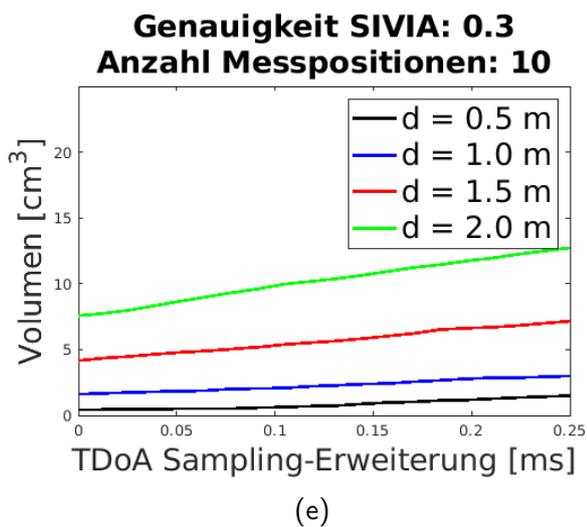
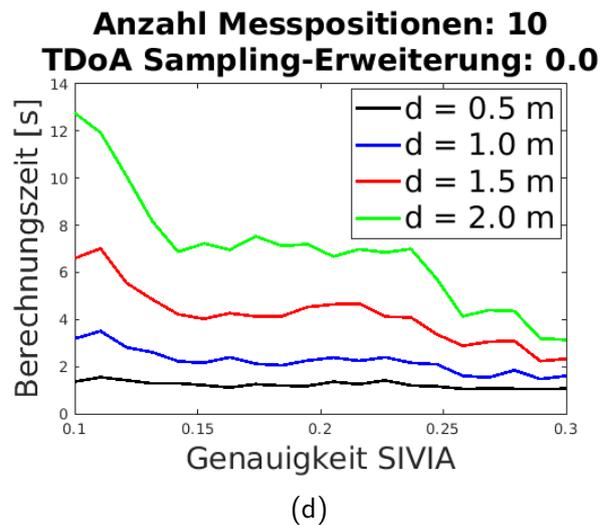
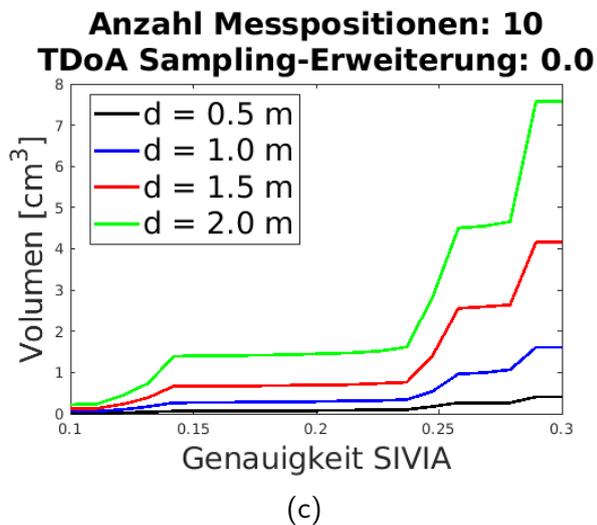
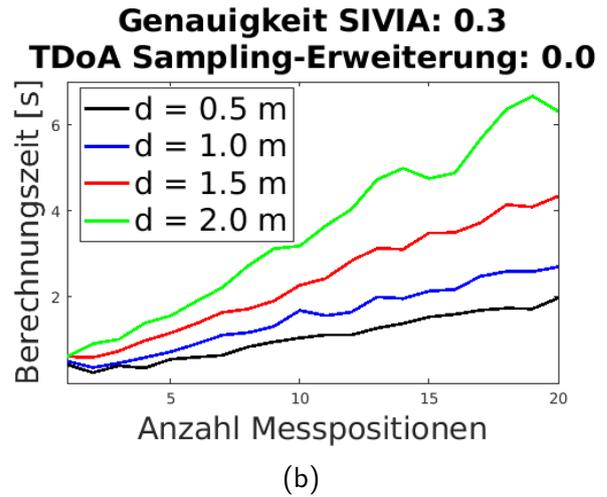
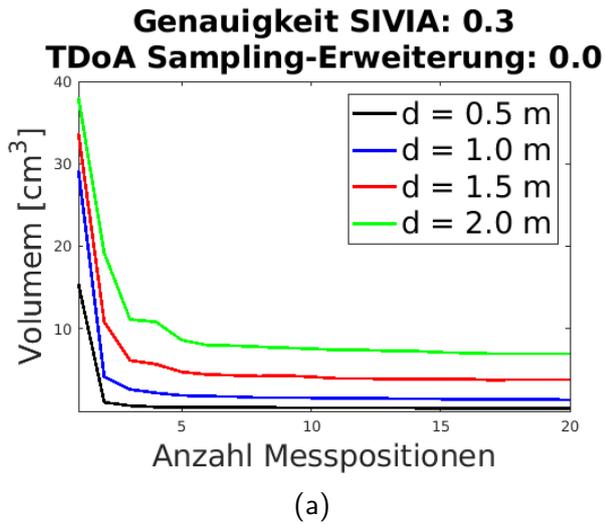


Abbildung 6.21: Einfluss des gesamten Volumens aller Intervallboxen und der Berechnungszeit von: (a) + (b) Anzahl der Messpositionen, (c) + (d) Genauigkeit von SIVIA, (e) + (f) TDoA-Sampling-Erweiterung.

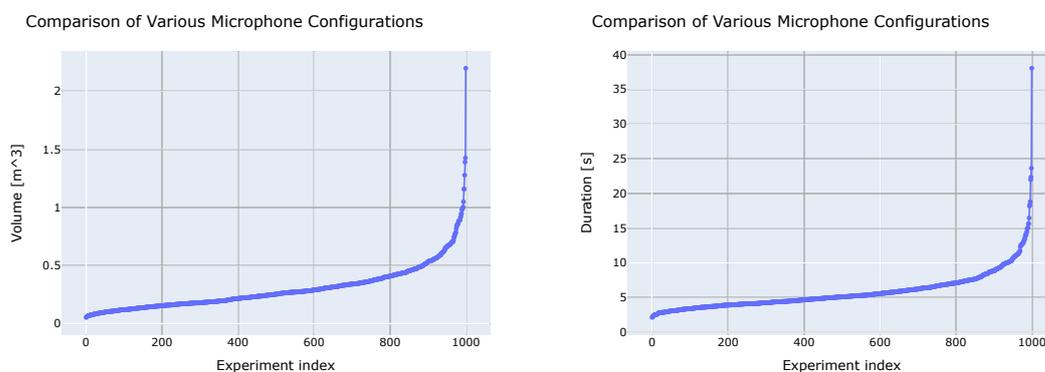
Abbildungen 6.21a und 6.21b zeigen, dass eine höhere Anzahl von Messpositionen zu einer verbesserten Lösung führen, jedoch der Rechenaufwand steigt. Ab einer maximalen Anzahl von Messpositionen ist nur eine kleine Verbesserung möglich. Es kann erkannt werden, dass die Distanz zwischen dem Mikrofonarray und der Soundquelle wichtig ist. Messungen an zwei unterschiedlichen Positionen mit einer Distanz von 0,5 m zu einer Soundquelle führen zu einer genaueren Lösung als 20 Messungen mit einer Distanz von 1 m.

Weiterhin zeigen Abbildungen 6.21c und 6.21d den Einfluss des Volumens und der Rechenzeit in Abhängigkeit von der Genauigkeit  $\varepsilon$ . Für  $d = 2$  m ergibt die Verringerung von  $\varepsilon$  (höhere Genauigkeit) eine Reduktion des Volumens. Das Volumen verringert sich von 7,6 auf  $0,2 \text{ cm}^3$ . Jedoch steigt die Berechnungszeit von 3,1 auf 12,7 s. Für  $d = 0,5$  m ist die Reduktion des Volumens hoch (von 0,4 auf  $0,0129 \text{ cm}^3$ ) und die Berechnungszeit ist näherungsweise konstant (von 1,06 auf 1,36 Sekunden). Als Resultat muss ein Kompromiss zwischen der Genauigkeit und der Berechnungszeit gefunden werden.

Die Abbildungen 6.21e und 6.21f zeigen den Einfluss des Volumens und der Berechnungszeit in Abhängigkeit von der Sampling-Erweiterung  $\Delta t_e$ . Für eine Abtastfrequenz von 8.000 Hz (respektive  $\Delta t_e = 0,25$  ms) und einer Abtastfrequenz von 96.000 Hz (respektive  $\Delta t_e = 0,02$  ms) unterscheidet sich das Volumen der korrespondierenden Intervallboxen bei  $d = 2$  m zwischen  $7,9 \text{ cm}^3$  und  $12,7 \text{ cm}^3$ , bei einem Abstand von  $d = 0,5$  m zwischen  $1,49 \text{ cm}^3$  und  $0,40 \text{ cm}^3$ . Die Berechnungszeit erhöht sich sehr gering für steigende Sampling-Erweiterung  $\Delta t_e$ .

### 6.6.3.1 Einfluss der Mikrofonanordnung

In diesem Experiment wurde die Abhängigkeit der Mikrofonanordnung auf die Ergebnisse der Soundkartierung betrachtet. Wie im vorherigen Experiment wurde zunächst eine Soundquelle im Zentrum eines Kreises angeordnet. Es wurden  $n$  Messungen äquidistant auf einem Kreis mit Durchmesser  $d$  durchgeführt. Während des Experimentes wurde die Anordnung der  $m_n$  Mikrofone zufällig variiert. Die Position der Mikrofone wurde eingeschränkt auf einen räumlichen Bereich, der durch eine Box repräsentiert wurde. Die Dimension der Box wurde so gewählt, dass sie dem verfügbaren Platz auf einer realen mobilen Robotikplattform entspricht.



(a) Volumen in Abhängigkeit der Mikrofonanordnung (b) Zeitdauer in Abhängigkeit der Mikrofonanordnung

Abbildung 6.22: Variation der Mikrofonkonfiguration bei fixierter Position der Soundquelle und Messposition

Es wurden 1.000 Experimente durchgeführt bei denen zufällige Mikrofonpositionen erzeugt wurden. Im Anschluss wurde die intervall-basierte Soundquellenkartierung durchgeführt. In Abbildung 6.22 sind die Ergebnisse dargestellt.

Als Beurteilungskriterium wurde das Volumen der geschätzten Position der Soundquelle als auch die Berechnungszeit verwendet.

Die Mikrofonanordnungen mit dem zugehörigen geringsten und höchsten Gesamtvolumen der Soundkartierung sind in Abbildung 6.23 dargestellt.

Comparison of microphone configurations

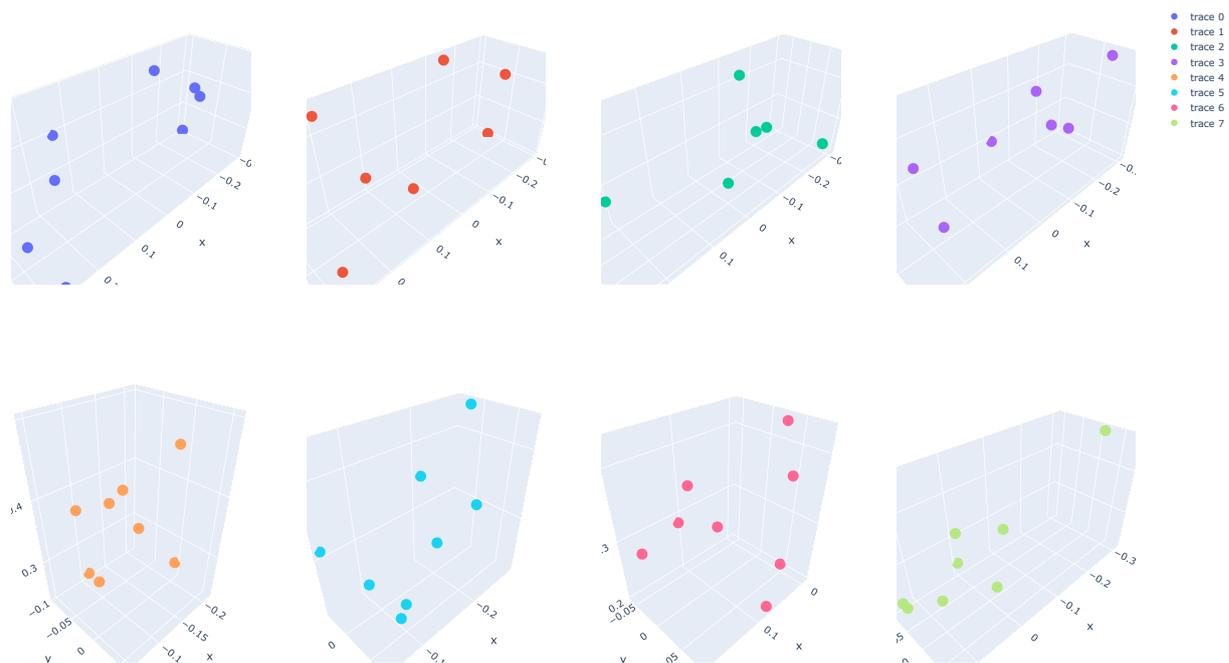
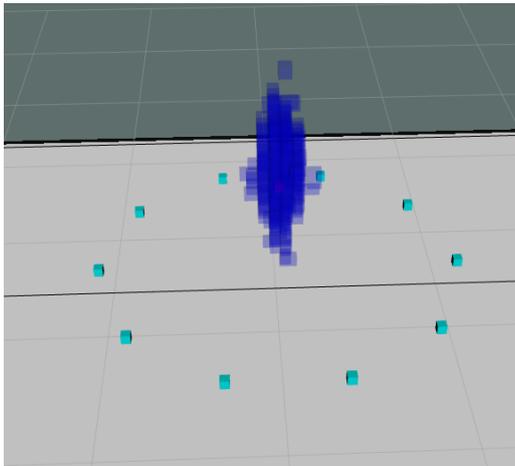


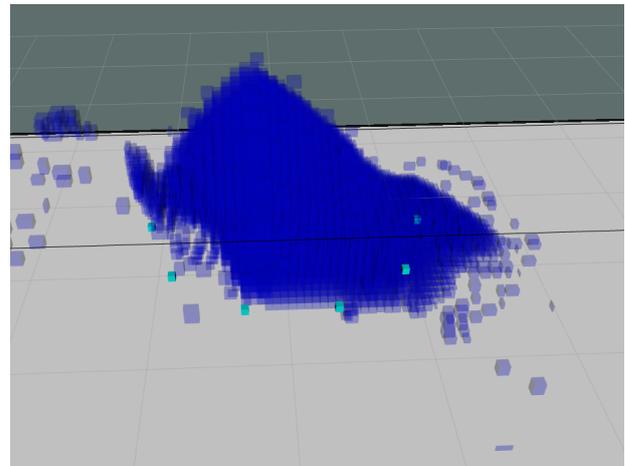
Abbildung 6.23: Darstellung der Mikrofonkonfigurationen. Erste Zeile: beste Konfigurationen von links nach rechts. Zweite Zeile: schlechteste Konfigurationen von links nach rechts

Schließlich sind in Abbildung 6.24 die besten und schlechtesten Soundquellenkartierungsergebnisse in Bezug auf das Gesamtvolumen dargestellt. Die zugehörigen Mikrofonanordnungen in Bezug zur mobilen Plattform sind in Abbildung 6.25 dargestellt.

Es kann erkannt werden, dass die Anordnung der Mikrofone einen starken Einfluss auf die Ergebnisse der Soundquellenkartierung aufweisen. Bereits in Abschnitt 6.5.1.1 wurde dargestellt, dass die Genauigkeit der Richtungsschätzung von der Richtung der aktiven Soundquelle abhängig ist. Die Richtungsschätzung beeinflusst das Ergebnis der Soundquellenkartierung.

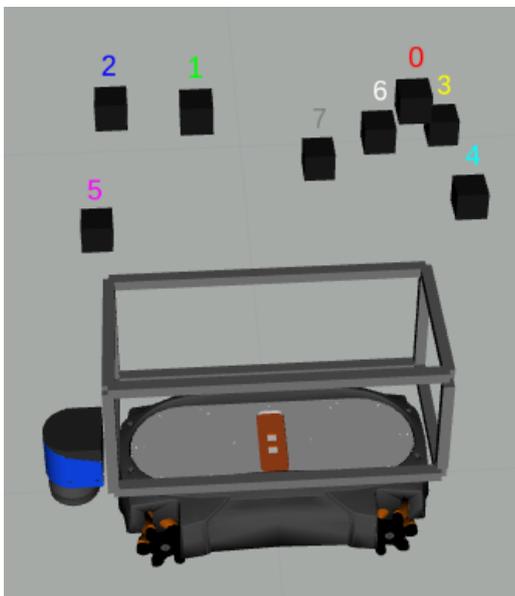


(a) Bestes Ergebnis

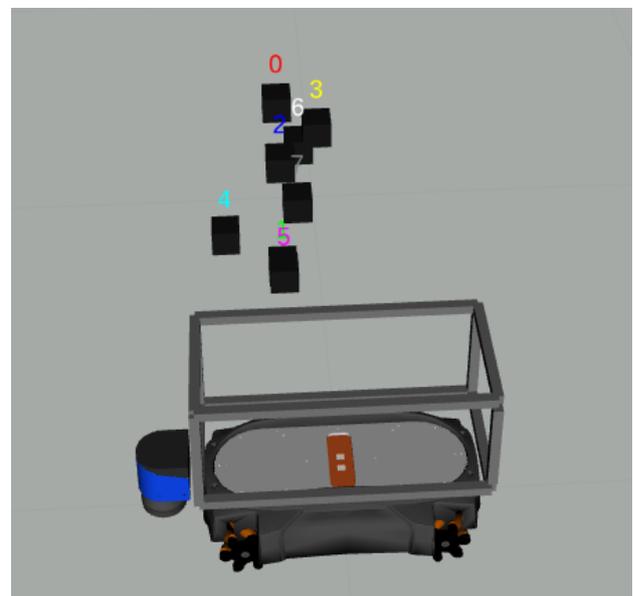


(b) Schlechtestes Ergebnis

Abbildung 6.24: Vergleich der Ergebnisse der Soundquellenkartierung für die beste und schlechteste Mikrofonkonfiguration.



(a) Beste Mikrofonkonfiguration



(b) Schlechteste Mikrofonkonfiguration

Abbildung 6.25: Darstellung der besten und schlechtesten Mikrofonkonfiguration für das durchgeführte Experiment.

### 6.6.4 Proof-of-Concept in realer Umgebung

In diesem Experiment wurde gezeigt, dass das entwickelte intervall-basierte Kartierungsverfahren in realer Umgebung erfolgreich anwendbar ist.

Das Experiment wurde in einem geschlossenen Raum (11,7 m × 6 m × 3,3 m) durchgeführt mit einem KUKA youBot, der mit einem Mikrofonarray (IntRoLab 8SoundUSB) mit acht Kanälen ausgerüstet war. Die Abtastfrequenz betrug 44.100 Hz. Es wurde die Mikrofonkonfiguration 1 verwendet (vgl. Abbildung 6.3b). Die Evaluation wurde auf eine einzelne aktive Soundquelle beschränkt. Diese wurde an Position (1,09 m; 5,92 m; 0,7 m) platziert. Dabei wurde ein Lautsprecher eines Mobiltelefons verwendet, der Sprache aussendete.

Das ODAS<sup>15</sup> framework (Open embeddeD Audition System) wurde verwendet, um die potenziellen Richtung zu den Soundquellen zu ermitteln. Dabei wurde der *SRP-PHAT-HSDA* Algorithmus mit der *M3K-Methode* verwendet [GM19]. Weiterhin wurde *gmapping*, eine openSLAM Implementierung [GSB07] und *amcl* (engl. *Adaptive Monte-Carlo localization*) zur Lokalisierung des Roboters innerhalb der Umgebung verwendet. Beides sind Pakete innerhalb des Robot Operating System (ROS) [Qui+09]. Alle Algorithmen wurden auf der Hardware des Roboters ausgeführt.

Zunächst wurde der Suchbereich restringiert auf (5,6 m × 4,3 m × 2,0 m) und in Abbildung 6.26 dargestellt. Ein Lokalisierungsfehler von 0,03 m in x- und y-Richtung wurde angenommen. Für eine Dauer von 5 s wurden die Richtungsschätzungen des SRP-PHAT-Verfahrens durch den M3K-Ansatz (vgl. Abschnitt 5.1.2) gefiltert. Nachdem die getracketen Richtungsschätzungen an einer Messposition gesammelt wurden, erfolgt die Soundquellenkartierung. Der richtungsbasierte Ansatz DoATiD aus Abschnitt 5.1.1 wurde verwendet, um die TDoA zu ermitteln. Dafür wurden die Grenzen  $[\underline{r}_i; \bar{r}_i]$  auf den Bereich [1,5 m; 7,0 m] gesetzt.

Die Genauigkeit  $\varepsilon$  des SIVIA-Verfahrens wurde schrittweise erhöht, um die Rechenzeit zu reduzieren (vgl. Tabelle 6.34).

Tabelle 6.34: Genauigkeit  $\varepsilon$  an den unterschiedlichen Messpositionen.

Messposition:	1	2	3	4
$\varepsilon$ [cm]	12,0	7,2	4,3	2,5

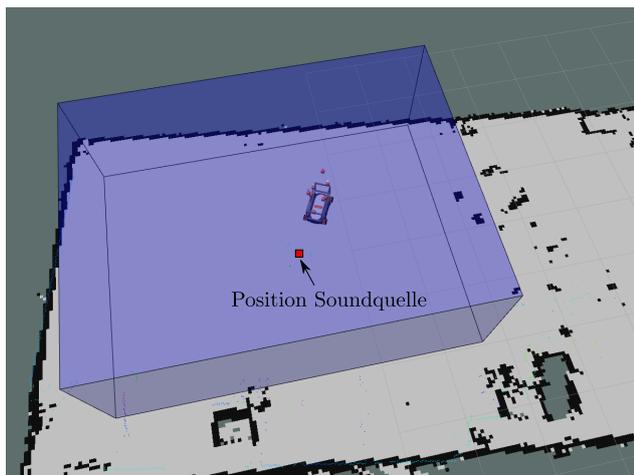
Die Ergebnisse der Kartierung sind in Abbildung 6.26 dargestellt, nachdem die jeweiligen Messungen durchgeführt wurden. Nach der vierten Messungen wurde die Genauigkeit auf  $\varepsilon = 1$  cm gesetzt und das Ergebnis in Tabelle 6.35 dargestellt.

Tabelle 6.35: Ergebnisse der intervall-basierten Soundquellenkartierung nach der vierten Messung im Realexperiment.

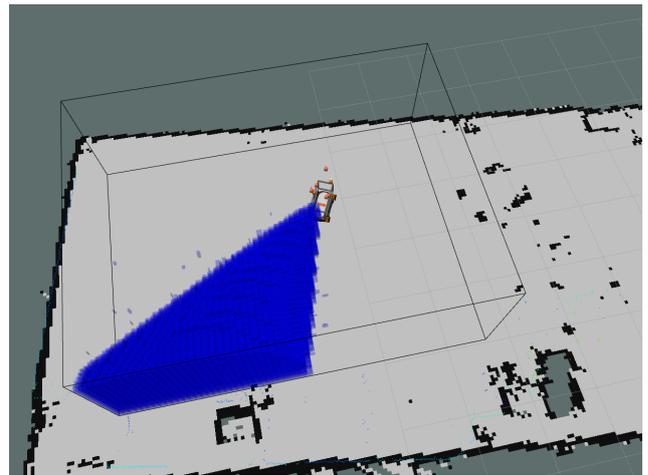
Volumen: [cm <sup>3</sup> ]	Schwerpunkt: (x;y;z) [m]	Dim: (dx;dy;dz) [cm]
23.376	(1,01;5,98;0,61)	(27,47;27,23;31,25)

Es konnte gezeigt werden, dass die wahre Soundposition in den Intervallen der berechneten Lösung enthalten ist.

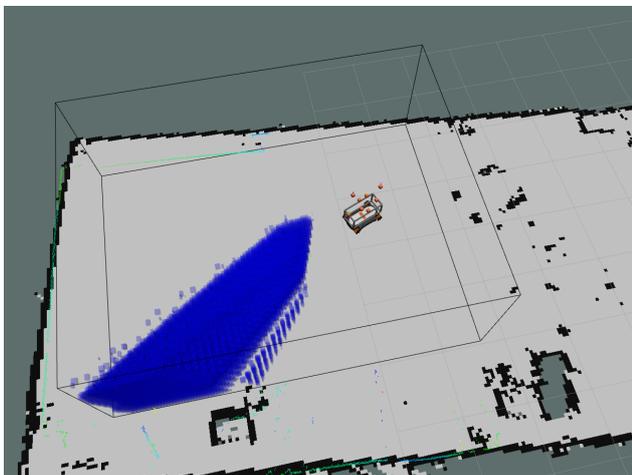
<sup>15</sup><http://odas.io>



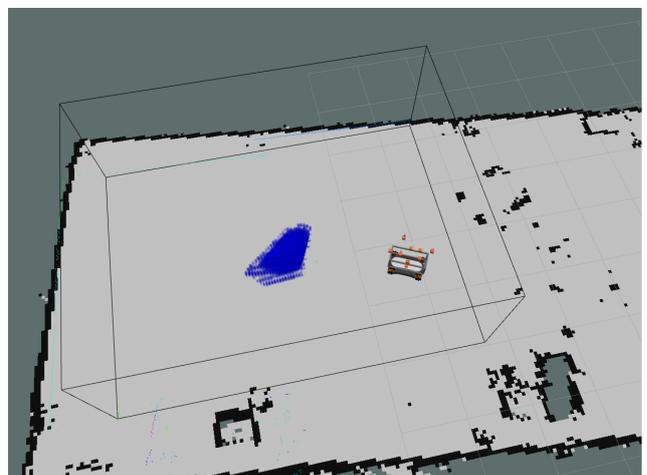
(a) Start



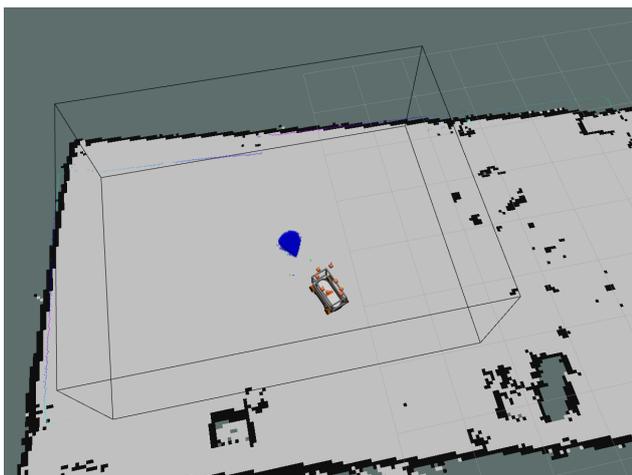
(b) Nach Messung 1



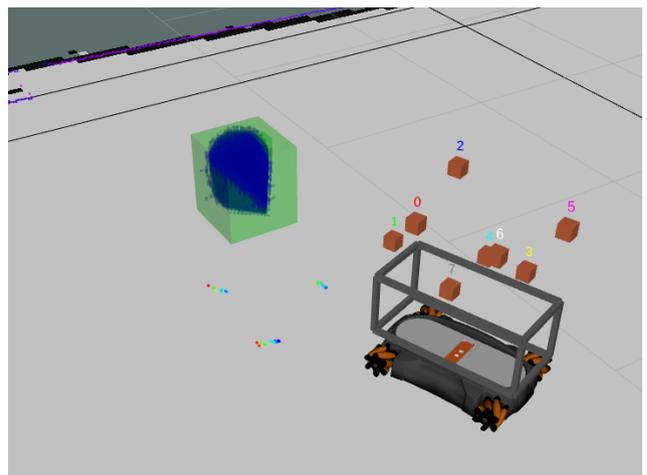
(c) Nach Messung 2



(d) Nach Messung 3



(e) Nach Messung 4



(f) Nachberechnung 4

Abbildung 6.26: Resultat von IB-SSM (potenzielle Soundquellen Positionen als Boxen in blau) in einer realen Umgebung. Nach Durchführung der vierten Messung wurde eine Intervallbox ermittelt, die die gesamten Ergebnisse umschließt (grün).

# Kapitel 7

## Bewertung und Diskussion

Das Ziel der Evaluierung in Kapitel 6 war es, die entwickelte intervall-basierte Soundquellenkartierung und die Schätzverfahren der *TDoA* zu untersuchen. Das entwickelte Kartierungsverfahren wurde verglichen mit einem probabilistischen Ansatz (EKF-Ansatz) und einem Verfahren, das auf der Lösung eines Optimierungsproblems (LM-Verfahren) basiert.

Es konnte gezeigt werden, dass die berechneten Grenzen des intervall-basierten Kartierungsverfahrens die korrekten Positionen der Soundquellen umschließen. Weiterhin wurde gezeigt, dass geometrische Mehrdeutigkeiten und systematische Fehler in dem entwickelten Verfahren berücksichtigt werden.

### 7.1 Soundquellenkartierung mit fehlerfreien Messwerten

Um zu zeigen, dass das entwickelte Kartierungsverfahren *IB-SSM* korrekte Lösungen liefert, wurden in Abschnitt 6.3 zunächst fehlerfreie Messwerte innerhalb einer simulierten Umgebung berücksichtigt. Dabei wurde angenommen, dass sowohl die *TDoA* als auch die Positionen der Mikrofone (innerhalb einer metrischen Karte) bekannt sind. Da die tatsächlichen Werte (engl. *ground truth*) innerhalb der Simulation vorlagen, konnte eine Aussage über Korrektheit und die Genauigkeit der Lösung getroffen werden. In vier aufeinanderfolgenden Experimenten (a), (b), (c) und (d) wurde die Komplexität der Versuchsdurchführung schrittweise erhöht:

- (a) Einzelne Soundquelle ohne geometrischen Mehrdeutigkeiten (Abschnitt 6.3.2),
- (b) Einzelne Soundquelle mit geometrischen Mehrdeutigkeiten (Abschnitt 6.3.3),
- (c) Mehrere Soundquellen ohne geometrischen Mehrdeutigkeiten (Abschnitt 6.3.4),
- (d) Mehrere Soundquellen mit geometrischen Mehrdeutigkeiten (Abschnitt 6.3.5).

Einerseits wurde die Anzahl der aktiven Soundquellen variiert. Andererseits wurde durch die Wahl der Mikrofonanordnung beeinflusst, ob für die Lösung der Kartierung geometrische Mehrdeutigkeiten resultierten. Für Experimente (a) bis (d) konnte gezeigt werden, dass die berechneten Grenzen von *IB-SSM* in allen Fällen die tatsächlichen Soundpositionen enthielten. Im Gegensatz dazu, konnten die beiden Vergleichsverfahren nicht garantieren, dass die korrekte Lösung für die Experimente (b) bis (d) ermittelt wurde.

Für das Experiment (a) haben alle Verfahren die korrekte Lösung ermittelt. Als wesentliche Limitierung erwies sich die benötigte Rechenzeit von *IB-SSM*. Während die Vergleichsverfahren die

Lösung nach wenigen Millisekunden lieferten, betrug die Rechenzeit für *IB-SSM* einige Sekunden. Es konnte gezeigt werden, dass die Rechenzeit von der gewählten Genauigkeit des verwendeten *Branch-and-Bound-Verfahrens SIVIA* abhängig ist (vgl. Abschnitt 2.4.5). Diese Abhängigkeit erweist sich im Umkehrschluss als Stärke. Denn die Genauigkeit der Lösung der Kartierung kann direkt vorgegeben werden. Auch wenn die Genauigkeit des Verfahrens verringert wird, ist weiterhin garantiert, dass die tatsächliche Lösung in den berechneten Grenzen liegt. Als Folgerung sind diese Grenzen jedoch konservativer. Für die meisten Einsatzgebiete stellt dies keine Einschränkung dar, da keine hochpräzise Positionsschätzung notwendig ist.

Für das Experiment (b) wurde untersucht, ob die Verfahren geometrische Mehrdeutigkeiten adressieren können. Hierbei wurde eine symmetrische Mikrofonkonfiguration verwendet, die auf einer Ebene angeordnet war. Aufgrund der symmetrischen Anordnung ergaben sich zwei mögliche Lösungen für die Soundposition - entweder oberhalb oder unterhalb der Ebene. Ohne zusätzliche Information kann nicht entschieden werden, welche der beiden Lösungen der tatsächlichen Position entspricht. Daher ist es notwendig, dass die jeweiligen Verfahren beide Lösungsmengen berücksichtigen. Es konnte gezeigt werden, dass das *IB-SSM-Verfahren* beide Lösungsmengen ermittelte. Dabei war garantiert, dass die tatsächliche Position der Soundquelle in den berechneten Grenzen enthalten war.

Bei dem LM-Verfahren beeinflusste die Wahl des Startwertes, welche der beiden Lösungen ermittelt wurden. Um zu garantieren, dass die korrekte Lösung berechnet wird, muss die Optimierung des LM-Verfahrens mit unterschiedlichen Startwerten durchgeführt werden.

Für den EKF-Ansatz wurden beide Lösungen ermittelt, falls die Symmetrie-Ebene der Mikrofonanordnung direkt im Ansatz berücksichtigt wurde. Jedoch musste aufgrund der steigenden Anzahl von Hypothesen eine Reduktion durchgeführt werden. Dabei konnte nicht garantiert werden, dass Hypothesen von den beiden alternativen Lösungen nach der Reduktion erhalten blieben. Nachdem die Messunsicherheit des EKF-Ansatzes in dem Experiment erhöht wurde, konnte gezeigt werden, dass eine der beiden möglichen Lösung verworfen wurde.

In Experiment (c) wurde das Verfahren für die Anwesenheit von mehreren aktiven Soundquellen untersucht und zunächst keine geometrischen Mehrdeutigkeiten berücksichtigt. Für das *IB-SSM-Verfahren* konnte gezeigt werden, dass alle Soundquellen in den berechneten Intervallboxen enthalten waren. Weiterhin zeigte sich der Vorteil des *IB-SSM-Verfahrens*. Im Gegensatz zu den beiden Vergleichsverfahren musste weder die Anzahl der Soundquellen bekannt sein, noch musste eine Datenzuordnung zwischen den Messungen und den geschätzten Soundpositionen durchgeführt werden. Insbesondere resultierten für das LM-Verfahren eine hohe Anzahl von möglichen Datenzuordnungen. Bereits bei drei Soundquellen und fünf Messpositionen resultierten 7.776 Hypothesen.

In Experiment (d) wurde die Komplexität erhöht, indem einerseits mehrere Soundquellen betrachtet und andererseits geometrische Mehrdeutigkeiten berücksichtigt wurden. Für das *IB-SSM-Verfahren* konnte gezeigt werden, dass alle Soundquellen in den berechneten Intervallboxen enthalten waren. Bedingt durch die geometrischen Mehrdeutigkeiten wurden weitere Lösungsbereiche berechnet. Die Einschränkungen der beiden Vergleichsverfahren, die in Experiment (b) und (c) beobachtet wurden, traten kombiniert auf. So war der Startwert für das LM-Verfahren entscheidend, ob die korrekte Lösung berechnet wurde. Da jedoch aufgrund der unbekanntenen Datenzuordnung nicht bekannt ist, welche Hypothese der tatsächlichen Lösung entspricht, kann keine Aussage über die Position getroffen werden. Weiterhin konnte der EKF-Ansatz aufgrund der Hypothesenreduktion nicht garantieren, dass die korrekten Soundpositionen berechnet wurden.

Es sei darauf hingewiesen, dass keine akustischen Reflexionen in den Experimenten berücksichtigt wurden. Einerseits berücksichtigt *IB-SSM* direkt in der Modellierung akustische Reflexionen (vgl. Abschnitt 3.4). Hierbei werden Richtungsmessungen, die durch Reflexionen hervorgerufen wurden, werden verworfen, wenn sie nicht konsistent mit der Lösung sind. Andererseits berücksichtigen beide Vergleichsverfahren keine akustischen Reflexionen. Dies verstärkt die Problematik der Datenzuordnung.

Zusammengefasst konnte für die vier Experimente gezeigt werden, dass *IB-SSM* den beiden Vergleichsverfahren deutlich überlegen ist, falls mehrere aktive Soundquellen in der Umgebung vorhanden sind und/oder eine Mikrofonanordnungen betrachtet wird, die geometrische Mehrdeutigkeiten hervorruft. Falls jedoch nur eine einzelne Soundquelle betrachtet und eine Mikrofonanordnung verwendet wird, die keine geometrischen Mehrdeutigkeiten hervorruft, dann sind die Vergleichsverfahren besser geeignet.

## 7.2 Soundquellenkartierung mit fehlerbehafteten Messwerten

In realen Umgebungen können die Messwerte nicht fehlerfrei beobachtet werden. Daher wurde in Abschnitt 6.4 untersucht, wie die Verfahren durch fehlerbehaftete Messwerte beeinflusst werden. Es wurde berücksichtigt, dass akustische Signale nur zu diskreten Zeitpunkten wahrgenommen werden. Als Folgerung kann die *TDoA* nicht exakt erfasst werden. Weiterhin wurde berücksichtigt, dass die Positionen der Mikrofone innerhalb der gegebenen metrischen Karte einen systematischen Fehler aufweisen können.

Bereits im vorherigen Abschnitt wurde auf die Einschränkungen der beiden Vergleichsverfahren aufgrund der Datenzuordnung bei der Anwesenheit mehrerer Soundquellen und der geometrischen Mehrdeutigkeiten eingegangen. Daher beschränkte sich die Evaluation in Abschnitt 6.4 auf eine einzelne Soundquelle ohne Betrachtung von geometrischen Mehrdeutigkeiten.

Falls im *IB-SSM-Verfahren* keine korrekten Unsicherheiten für die Messwerte berücksichtigt wurden, zeigten alle Experimente in Abschnitt 6.4, dass keine konsistente Lösung berechnet werden konnte. Als Ergebnis resultierte eine leere Lösungsmenge. Die Vergleichsverfahren hingegen konnten auch mit fehlerhaften Messwerten eine Lösung ermitteln. Je nach Experiment wich diese Lösung jedoch stark von der tatsächlichen Lösung ab. Zusammengefasst kann durch das Ergebnis (leere Lösungsmenge) des *IB-SSM-Verfahrens* erkannt werden, ob die Unsicherheiten der Messwerte inkorrekt modelliert wurden. Für die Vergleichsverfahren ist diese Aussage nicht zutreffend.

Falls im *IB-SSM-Verfahren* korrekte Grenzen für die Unsicherheit der Messwerte berücksichtigt wurden, konnten in allen Experimenten gezeigt werden, dass die tatsächliche Lösung in den berechneten Grenzen enthalten war.

In Abschnitt 6.4.1 wurde die Unsicherheit der *TDoA* betrachtet. Die Ergebnisse der beiden Vergleichsverfahren wichen nur geringfügig von der tatsächlichen Position ab. Im Vergleich zur Lösung von *IB-SSM*, bei dem fehlerfreie Messwerte angenommen und keine Unsicherheiten modelliert wurden, vergrößerte sich das berechnete Volumen der Lösungsmenge von  $0,000125 \text{ cm}^3$  auf  $0,53 \text{ cm}^3$ . Es konnte gezeigt werden, dass das Volumen der Lösungsmenge deutlich größer ist, falls eine zu kon-

servative Schätzung der Unsicherheit der *TDoA* im Modell berücksichtigt wird, diese Unsicherheit in den Experimenten jedoch geringer ist.

In den Abschnitten 6.4.2 und 6.4.3 wurde ein systematischer Fehler der Mikrofonposition berücksichtigt. Dabei wurde in Abschnitt 6.4.2 angenommen, dass die Orientierung des Mikrofonarrays in der globalen Karte eine konstante Abweichung besaß. Es konnte belegt werden, dass die wahre Position in den berechneten Grenzen des *IB-SSM*-Verfahrens enthalten war. Der Schwerpunkt der umhüllenden Intervallbox war nur geringfügig von der wahren Position entfernt. Im Gegensatz dazu wiesen die Lösungen des *LM*-Verfahrens und des *EKF*-Ansatzes eine große Abweichung zur wahren Position auf. Bei einem Orientierungsfehler von  $10^\circ$  ergaben sich 21 cm Abweichung für das *LM*-Verfahren, 13 cm für den *EKF*-Ansatz, jedoch nur etwa 2 cm für das *IB-SSM-Verfahren*. In Abschnitt 6.4.3 wurde angenommen, dass die einzelnen Mikrofonpositionen eine Unsicherheit aufwiesen. Es konnte gezeigt werden, dass *IB-SSM* Grenzen berechnete, die die tatsächlichen Positionen enthielten. Deutlicher Nachteil von *IB-SSM* ist, dass diese Grenzen sehr konservativ waren. Bei einem Positionsfehler von 5 cm betrug die Abmessungen der berechneten Intervallboxen über einen Meter.

### 7.3 Akustische Merkmale - Extraktion der *TDoA* und Modellierung der Unsicherheit

In Abschnitt 6.5 wurde untersucht, wie die *TDoA* berechnet und deren Unsicherheit modelliert werden kann. Hierfür wurden zwei alternative Ansätze untersucht. Im ersten Schritt wurde das richtungs-basierte Verfahren *DoATiD* aus Abschnitt 5.1 evaluiert. Für die Schätzung der Richtung wurde *SRP-PHAT* angewendet. Da für die Berechnung der *TDoA* die Genauigkeit der *DoA* wichtig ist, wurde die Genauigkeit in Abschnitt 6.5.1.1 in einer simulierten Umgebung betrachtet. Das Ziel bestand darin, die Abweichung zwischen der geschätzten und der tatsächlichen *DoA* zu ermitteln. Hierfür wurde systematisch eine einzelne Soundquelle in unterschiedlichen Richtungen platziert und die Abweichung ermittelt. Es konnte belegt werden, dass die Genauigkeit maßgeblich von der Richtung abhängig ist, in der sich die Soundquelle befand. Für die untersuchte Mikrofonanordnung betrug der maximale Winkelfehler  $25^\circ$ . Falls die Position der Soundquelle ausschließlich anhand der *DoA* ermittelt wurde, resultierte für das Resultat eine große Abweichung. Diese Erkenntnis wurde bereits in Abschnitt 7.2 diskutiert. Dort führte die inkorrekte Annahme der Orientierung von der mobilen Plattform zu einer großen Abweichung für das *EKF*-Verfahren. Eine große Stärke des *IB-SSM*-Verfahrens ist, dass auch aus fehlerhaften Richtungsschätzungen eine korrekte *TDoA* berechnet werden kann. Die Ursache liegt darin, dass die akustischen Signale nur zu diskreten Zeitpunkten wahrgenommen werden. Somit können unterschiedliche Richtungen mit der gleichen (diskrete) *TDoA* assoziiert sind.

Im darauffolgenden Experiment wurde untersucht, ob *SRP-PHAT* mehrere Richtungsmessungen zur gleichen Zeit detektieren ermitteln kann. Hierfür wurden vier gleichzeitig aktive Soundquellen betrachtet, deren Positionen auf einer festgelegten Trajektorie variiert wurden. Es konnte gezeigt werden, dass für alle Soundquellen die Richtung ermittelt werden konnte. Die maximale Abweichung betrug für die vorgegebenen Trajektorien  $12^\circ$ .

Nachdem die Richtungsschätzung ermittelt wurde, folgte in Abschnitt 6.5.1.2 eine Untersuchung, wie aus dieser Schätzung die *TDoA* ermittelt werden konnte. Hierfür wurden zunächst fehlerfreie Richtungsmessungen betrachtet. Im Anschluss wurden Richtungsmessungen verwendet, die durch das *SRP-PHAT*-Verfahren aus den Mikrofonsignalen berechnet wurden. Durch das *DoATiD*-Verfahren

wurden die resultierenden Intervalle der  $TDoA$ s ermittelt. Für beide Experimente konnte gezeigt werden, dass alle wahren  $TDoA$ s in den berechneten Intervallen enthalten waren, falls der Abtastfehler berücksichtigt wurde.

In Abschnitt 6.5.2 wurde das *InTiD-Verfahren* aus Abschnitt 5.2 zur Schätzung der  $TDoA$  untersucht. Da innerhalb dieser Dissertation keine Kalibrierung der Mikrofone bezüglich der Amplitude vorausgesetzt wird (vgl. Abschnitt 1.3) musste die akustische Dämpfung innerhalb des *InTiD-Verfahrens* berücksichtigt werden. Für die Berechnung der  $TDoA$  als auch der Dämpfung wurden zwei Ansätze untersucht. Dabei wurden das *SIVIA-Verfahren* und im Anschluss der Kontraktor  $C_3$  verwendet. Es konnte belegt werden, dass die beiden Ansätze korrekte Grenzen für die  $TDoA$ s berechneten, sich jedoch im Rechenaufwand unterschieden. So besaß der *SIVIA-Ansatz* eine um den Faktor 6 höhere Rechenzeit. Außerdem konnte gezeigt werden, dass die Breite der ermittelten Grenzen von *InTiD* deutlich kleiner war als die Ergebnisse von *DoATiD*. Die Ursache liegt darin, dass im *DoATiD-Verfahren* nur die Richtungen zu den Soundquellen berücksichtigt wird. Jedoch steht keine Information über die Entfernung zu den Soundquellen zur Verfügung. Somit wird eine sehr konservative Schätzung durchgeführt.

Insgesamt konnte gezeigt werden, dass sowohl *DoATiD* als auch *InTiD* verwendet werden können, um die  $TDoA$  zu schätzen. Der Vorteil von *DoATiD* liegt darin, dass das Verfahren zuverlässig für die Schätzung der  $TDoA$  für mehrere aktive Soundquellen eingesetzt werden kann. Weiterhin ist die Rechenzeit von *DoATiD* deutlich geringer. Als Nachteil erwies sich, dass für *DoATiD* nicht mathematisch garantiert werden kann, dass die korrekte Lösung in den berechneten Grenzen liegt. Diese Garantie kann nur gegeben werden, falls ausschließlich Berechnungen mithilfe der Intervallarithmetik durchgeführt werden. Die Limitation von *InTiD* liegt darin, dass der Ansatz derzeit nicht auf mehrere Soundquellen übertragen werden kann, falls keine Information über die Frequenzanteile der unterschiedlichen Soundquellen vorliegt.

## 7.4 Gesamtsystem Soundquellenkartierung mit Modellierung der Unsicherheit

In Abschnitt 6.6 wurde das Gesamtsystem evaluiert. Hierfür wurden die geschätzten Intervalle der  $TDoA$  durch *DoATiD* und *InTiD* ermittelt. Diese Grenzen wurden im Anschluss für die Kartierung verwendet. Die Evaluation beschränkte sich auf eine Soundquelle, um einen Vergleich zwischen *InTiD* und *DoATiD* zu ermöglichen.

In allen Experimenten konnte belegt werden, dass *IB-SSM* korrekte Grenzen für die wahre Position der Soundquelle ermittelte. *IB-SSM* kombiniert mit *InTiD* lieferte eine Lösung, die ein geringeres Volumen aufwies, als die Kombination von *IB-SSM* mit *DoATiD*. Dies steht im Einklang mit den Ergebnissen in Abschnitt 6.5.2. Dort konnte belegt werden, dass die ermittelten Grenzen von *InTiD* für die Intervalle der  $TDoA$  eine geringere Breite aufwiesen als für *DoATiD*.

Im Anschluss wurde eine Parameterstudie für *IB-SSM* durchgeführt. Es konnte gezeigt werden, dass die Entfernung zwischen der Soundquelle und der Messposition des Mikrofonarrays einen großen Einfluss besaß. Je geringer diese Entfernung war, desto genauer konnte die Lösung ermittelt werden.

Im Anschluss wurde die Anordnung der Mikrofone variiert. Es konnte gezeigt werden, dass die Mikrofonanordnung einen großen Einfluss auf die Ergebnisse der Soundquellenkartierung besaß. Die Auswahl einer optimalen Mikrofonanordnung ist nicht Bestandteil dieser Dissertation und wurde in den Experimenten nicht untersucht.

Zum Abschluss wurde in Abschnitt 6.6.4 gezeigt, dass das intervall-basierte Verfahren *IB-SSM* erfolgreich in einem realen Umfeld durchgeführt werden kann, wenn eine einzelne Soundquelle betrachtet wird.

## 7.5 Bezug zur Zielsetzung

Hervorzuheben ist, dass in dieser Arbeit alle gesetzten Ziele ( $Z_1$ ) bis ( $Z_5$ ) aus Abschnitt 1.2.2 erfüllt wurden. Es wurden bestehende Kartierungsverfahren von Soundquellen analysiert und deren Einschränkungen beim Einsatz in der mobilen Robotik in den Abschnitten 7.1 und 7.2 identifiziert ( $Z_1$ ). Um die Einschränkungen zu adressieren, wurde das intervall-basierte Kartierungsverfahren *IB-SSM* entwickelt. Es konnte belegt werden, dass *IB-SSM* geometrische Mehrdeutigkeiten und akustische Reflexionen berücksichtigt ( $Z_2$ ). Um die Unsicherheit der Position von Mikrofonen innerhalb einer metrischen Karte zu beschreiben, wurde ein Fehlermodell entwickelt ( $Z_3$ ). Darauffolgend wurden die zwei neuen Verfahren *DoATiD* und *InTiD* vorgestellt, um die *TDoA* und deren Unsicherheit zu berechnen ( $Z_4$ ). Zum Abschluss wurde eine experimentelle Evaluierung der entwickelten Ansätze sowohl anhand simulierter Daten als auch anhand realer Messwerte durchgeführt ( $Z_5$ ).

## 7.6 Zukünftige Forschungsfragen

In der vorliegenden Dissertation wurden für die Kartierung ausschließlich Mikrofonensignale verwendet. In weiterführenden Arbeiten sollte untersucht werden, wie zusätzliche Sensormodalitäten eingesetzt werden können. Zum Beispiel könnten visuelle Sensoren (wie Farbkameras) eingesetzt werden. Diese zusätzliche Information könnte genutzt werden, um inkonsistente Lösungen aus dem Lösungsraum zu entfernen.

In der Parameterstudie in Abschnitt 6.6.3 wurde gezeigt, dass die Mikrofonanordnung und die Anfahrtpunkte einen großen Einfluss auf die Genauigkeit der Ergebnisse einnehmen. Gegenstand zukünftiger Arbeiten sollte die Entwicklung einer Exploration sein, die optimale Anfahrtpunkte berechnet. Hierbei sollte die Geometrie des Raumes, die Kartierungsergebnisse aus dem vorherigen Rechenschritt und die Mikrofonengeometrie berücksichtigt werden.

In Abschnitt 6.5.1.2 konnte belegt werden, dass die dreidimensionale Richtungsschätzung von der Orientierung der Mikrofonanordnung abhängig ist. Um eine verbesserte Schätzung zu erhalten, sollte in weiterführenden Arbeiten untersucht werden, wie die Orientierung und Position der mobilen Mikrofonanordnung angepasst werden kann.

# Kapitel 8

## Zusammenfassung

In der vorliegenden Arbeit wurde untersucht, wie die dreidimensionale Position von statischen Soundquellen innerhalb einer metrischen Karte bestimmt werden kann. Hierfür wurden klassische Kartierungsverfahren und auftretende Probleme beim Einsatz in der mobilen Robotik analysiert. Bestehende akustische Verfahren können nicht direkt ohne Anpassung in der mobilen Robotik eingesetzt werden. Neben Fehldetektion durch Umgebungsrauschen und akustischen Reflexionen resultieren technische Einschränkungen bezüglich der verwendeten Mikrofone.

In dieser Dissertation wurde ein neues Verfahren (*IB-SSM*) entwickelt, um durch ein mobiles Mikrofonarray die 3D-Kartierung von mehreren Soundquellen durchzuführen. Dabei wurde angenommen, dass die Soundquellen ausschließlich durch die Mikrofon-signale erfasst werden. Andere Sensoren wurden nur eingesetzt, um das mobile Mikrofonarray in einer metrischen Karte zu lokalisieren. Die Übersicht des in dieser Dissertation entwickelten Lösungsansatzes ist in Abbildung 8.1 dargestellt.

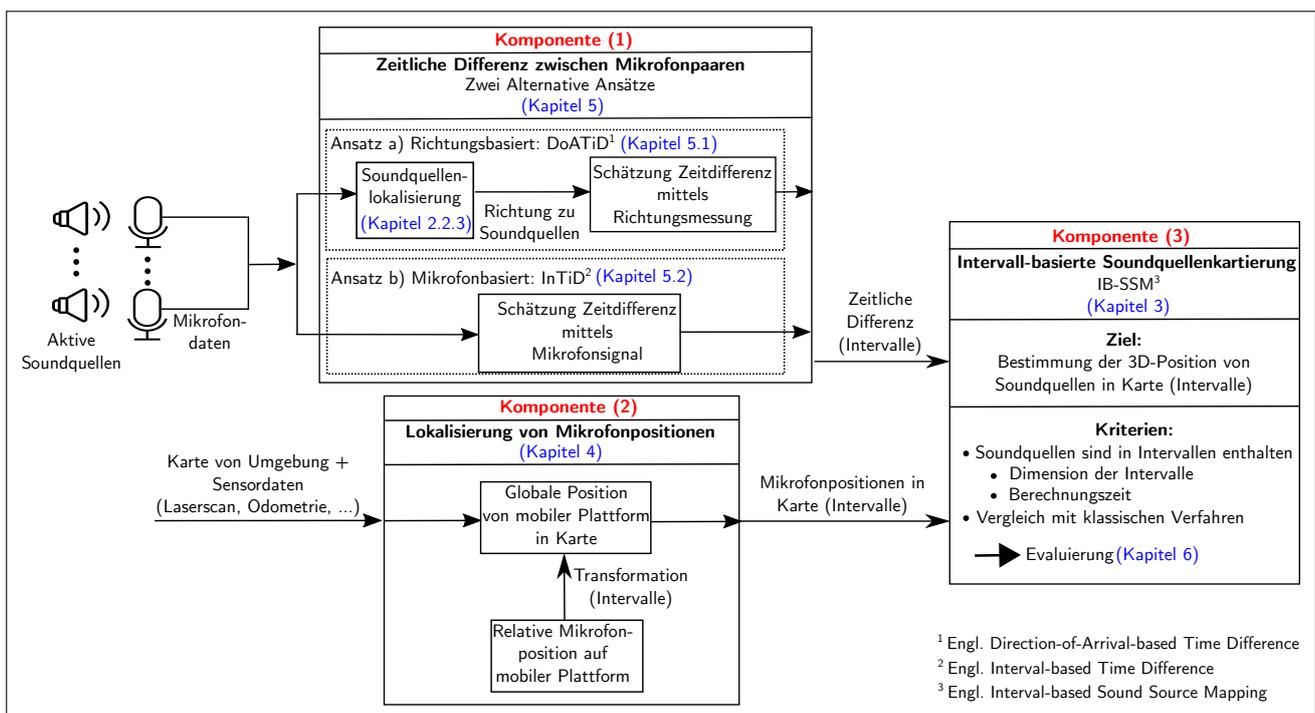


Abbildung 8.1: Systematischer Aufbau des wissenschaftlichen Beitrages dieser Dissertation.

In dem entwickelten Verfahren zur Soundquellenkartierung wurden im Vergleich zu klassischen Ansätzen einige Einschränkungen aufgehoben. Einerseits wird die Anzahl der aktiven Soundquellen nicht als bekannt vorausgesetzt. Insbesondere muss bei der Anwesenheit mehrerer Soundquellen keine Datenzuordnung zwischen einer Messung und einer möglichen Soundposition durchgeführt werden. Andererseits wurde statt einer zweidimensionalen Formulierung, der Ansatz für den dreidimensionalen Raum modelliert. Geometrische Mehrdeutigkeiten werden bei klassischen Verfahren durch die Randbedingungen in der Versuchsdurchführung aufgehoben. Im Gegensatz dazu wurden in dieser Dissertation die Mehrdeutigkeiten explizit mit modelliert. Für ein einzelnes Mikrofonpaar und einer akustischen Messung ergibt sich ein Kegel, auf dem sich die Soundquelle befindet. Für die Berechnung des Kegels werden die Positionen eines Mikrofonpaares und akustische Merkmale der Mikrofonsignale verwendet. Dabei wird in dieser Arbeit die *TDoA* (vgl. Abschnitt 3.1) genutzt.

Weil Messungen an unterschiedlichen Positionen durchgeführt und mehrere Mikrofonpaare berücksichtigt werden, können die Positionen der Soundquellen näher eingeschränkt werden. Um die resultierenden Abhängigkeiten zu lösen, wurde ein *Constraint Satisfaction Problem (CSP)* formuliert. Für die Lösung wurde in dieser Arbeit ein *Branch-and-Bound-Algorithmus* eingesetzt, der auf der Intervall-Arithmetik basiert. Hierbei werden mathematische Berechnung auf begrenzten Mengen durchgeführt. Es konnte in dieser Arbeit gezeigt werden, dass für fehlerfreie Messwerte der intervall-basierte Algorithmus präzise Lösungen ermittelt. Falls jedoch die Messwerte von den korrekten Werten abweichen, resultiert als Ergebnis der Kartierung eine leere Lösungsmenge. Dies fundiert auf der Eigenschaft, dass durch den Einsatz der Intervall-Arithmetik inkonsistente Lösungen verworfen werden. Sofern jedoch die Grenzen der Unsicherheit der Messwerte korrekt gewählt werden, sind die Ergebnisse der intervall-basierten Kartierung mathematisch garantiert und somit ist die korrekte Lösung in der Lösungsmenge enthalten. Daher wurde in dieser Dissertation untersucht, wie die Grenzen für die Unsicherheiten der Messwerte geschätzt werden können. Hierbei wurden die Unsicherheiten der *TDoA* und der Mikrofonpositionen berücksichtigt.

Für die Mikrofone wurde die Ausdehnung der Membran ermittelt und die Ungenauigkeit der Mikrofonposition innerhalb der metrischen Karte geschätzt. Als Resultat wird die Position eines einzelnen Mikrofons durch eine dreidimensionale Box repräsentiert.

Um die *TDoA* und die korrespondierende Unsicherheit zu schätzen, wurden in dieser Dissertation zwei alternative Ansätze *DoATiD* und *InTiD* vorgestellt. Diese Verfahren können kombiniert werden, und somit kann die Robustheit gegenüber Messfehlern erhöht werden.

*DoATiD* basiert auf den Richtungsschätzungen von bekannten Ansätzen der Soundquellenlokalisierung. Um eine robuste Schätzung innerhalb realer Experimente zu erhalten, wurde ein Filterverfahren verwendet. Hierbei wurde ein Kriterium definiert, um zu folgern, ob eine Richtungsschätzung von einer Soundquelle stammt oder durch Rauschen verursacht wurde. Da die Lautstärke der Soundquelle nicht bekannt ist, kann nicht allein auf Basis der Richtungsschätzung die *TDoA* ermittelt werden. Falls die Soundquelle weit entfernt von dem Mikrofonarray angeordnet ist, kann die akustische Ausbreitung durch eine ebene Welle approximiert und die *TDoA* berechnet werden. In dieser Arbeit wurden Grenzen für die Distanz zu der Soundquelle in Abhängigkeit der Raumdimension gesetzt und die *TDoA* geschätzt. Weiterhin wurden Abtastfehler des Mikrofonsignals berücksichtigt.

*InTiD* beruht auf einer mikrofon-basierten Methode, die direkt mittels der Intervallarithmetik die Mikrofondaten verarbeitet. Hierfür wurden unterschiedliche Kontraktoren entwickelt, die für jedes Mikrofonpaar die *TDoA* ermitteln. Es konnte in der Evaluation gezeigt werden, dass die ermittelten

---

Grenzen für die *TDoA* deutlich kleiner waren im Vergleich zu *DoATiD*. Falls mehrere Soundquellen betrachtet werden, kann jedoch das *InTiD*-Verfahren nicht angewendet werden, ohne Kenntnis über die Frequenzanteile der jeweiligen Signale zu besitzen.

Es erfolgte in Kapitel 6 eine mehrstufige Evaluation mit simulierten als auch realen Messwerten. Als Vergleichsverfahren dienten ein probabilistischer Ansatz und ein Optimierungsverfahren. Um eine vergleichbare Evaluation durchzuführen, wurde in dieser Arbeit ein bestehender probabilistischer Ansatz von der zweidimensionalen auf die dreidimensionalen Beschreibung erweitert. Für die beiden Vergleichsverfahren musste im Gegensatz zum entwickelten intervall-basierten Verfahren die Anzahl der Soundquellen bekannt sein. Falls geometrische Mehrdeutigkeiten auftreten, konnten die Vergleichsverfahren nicht garantieren, dass die korrekte Lösung berechnet wird. Das intervall-basierte Kartierungsverfahren hat in allen Fällen korrekte Grenzen berechnet, die die wahre Position enthielten. Es konnte gezeigt werden, dass die Genauigkeit der Lösung des intervall-basierten Verfahrens vorgegeben und somit die Rechenzeit beeinflusst werden kann. Dabei wird die Garantie nicht aufgehoben, dass die korrekte Lösung in der Lösungsmenge enthalten ist. Im Gegensatz dazu kann bei einem Optimierungsansatz auch die Genauigkeit (beschrieben durch das Abbruchkriterium) vorgegeben werden, jedoch ist nicht garantiert, dass die korrekte Lösung gefunden wird.

Schließlich wurde gezeigt, dass das intervall-basierte Verfahren in realen Umgebungen anwendbar ist.



# Literaturverzeichnis

- [AAM99] F. Asono, H. Asoh und T. Matsui. „Sound source localization and signal separation for office robot “Jijo-2”“. In: *Proceedings. 1999 IEEE/SICE/RSJ. International Conference on Multisensor Fusion and Integration for Intelligent Systems. MFI'99 (Cat. No. 99TH8480)*. IEEE. 1999, S. 243–248.
- [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 (1979), S. 943–950.
- [Aco21] Acoustical Society. „Saving Lives During Disasters by Using Drones by Macarena Varela“. <https://acoustics.org/1aspa5-saving-lives-during-disasters-by-using-drones-macarena-varela/>. Zugriff: 08.12.21. 2021.
- [AD15] S. Argentieri und P. Danes P.and Soueres. „A survey on sound source localization in robotics: From binaural to array processing methods“. In: *Computer Speech & Language* 34.1 (2015), S. 87–112.
- [AH14] X. Alameda-Pineda und R. Horaud. „A geometric approach to sound source localization from time-delay estimates“. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.6 (2014), S. 1082–1095.
- [Alg+01] V. R. Algazi *et al.* „The cipic hrtf database“. In: *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. IEEE. 2001, S. 99–102.
- [Apt03] K. Apt. *Principles of constraint programming*. Cambridge university press, 2003.
- [AS72] D. Alspach und H. Sorenson. „Nonlinear Bayesian estimation using Gaussian sum approximations“. In: *IEEE transactions on automatic control* 17.4 (1972), S. 439–448.
- [Bad+09] A. Badali *et al.* „Evaluating real-time audio localization algorithms for artificial audition in robotics“. In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2009, S. 2033–2038.
- [Ban+13] Y. Bando *et al.* „Posture estimation of hose-shaped robot using microphone array localization“. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2013, S. 3446–3451.

- [Bar75] E. Bar-Shalom Y. and Tse. „Tracking in a cluttered environment with probabilistic data association“. In: *Automatica* 11.5 (1975), S. 451–460.
- [Bar80] P. Barton. „Digital beam forming for radar“. In: *IEEE Proceedings F (Communications, Radar and Signal Processing)*. Bd. 127. 4. IET. 1980, S. 266–277.
- [Bas+95] H. E. Bass *et al.* „Atmospheric absorption of sound: Further developments“. In: *The Journal of the Acoustical Society of America* 97.1 (1995), S. 680–683.
- [BCH08] J. Benesty, J. Chen und Y. Huang. *Microphone array signal processing*. Bd. 1. Springer Science & Business Media, 2008.
- [Ber19] H. Bernstein. *Elektroakustik: Mikrofone, Klangstufen, Verstärker, Filterschaltungen und Lautsprecher*. Springer-Verlag, 2019.
- [Bey21] Beyerdynamic. „Tauchspulmikrofon - Beyerdynamic TG V70 (s)“. <https://www.beyerdynamic.de/tg-v70s.html>. Zugriff: 19.05.21. 2021.
- [BFC90] Y. Bar-Shalom, T. E. Fortmann und P. G. Cable. „Tracking and data association“. 1990.
- [BFF16] R. Bomhardt, M. de la Fuente Klein und J. Fels. „A high-resolution head-related transfer function and three-dimensional ear model database“. In: *Proceedings of Meetings on Acoustics 172ASA*. Bd. 29. 1. Acoustical Society of America. 2016, S. 050002.
- [BJ14] A. Bethencourt und L. Jaulin. „Solving non-linear constraint satisfaction problems involving time-dependant functions“. In: *Mathematics in Computer Science* 8.3-4 (2014), S. 503–523.
- [BL95] Y. Bar-Shalom und X. Li. *Multitarget-multisensor tracking: principles and techniques*. Bd. 19. YBs Storrs, CT, 1995.
- [Bla01] D. T. Blackstock. „Fundamentals of physical acoustics“. 1274-1276. 2001.
- [BM09] J. Bourgeois und W. Minker. „Time-domain beamforming and blind source separation“. In: *Lecture Notes in Electrical Engineering*. Springer-Verlag (2009).
- [Boh87] D. A. Bohn. „Environmental effects on the speed of sound“. In: *Audio Engineering Society Convention 83*. Audio Engineering Society. 1987.
- [Bra09] K. F. Braun. „Electrical oscillations and wireless telegraphy“. In: *Nobel Lecture, December 11.1909* (1909), S. 226–245.
- [Bre94] A. S. Bregman. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

- [Cal09] L. Calmes. „Biologically inspired binaural sound source localization and tracking for mobile robots.“ Dissertation. RWTH Aachen University, 2009.
- [CBG09] I. Cohen, J. Benesty und S. Gannot. *Speech processing in modern communication: challenges and perspectives*. Bd. 3. Springer Science & Business Media, 2009.
- [CH99] S. Cavaco und J. Hallam. „A biologically plausible acoustic azimuth estimation system“. In: *Proceedings of IJCAI-99 Workshop on Computational Auditory Scene Analysis (CASA'99)*. 1999, S. 78–87.
- [CUI21] CUI. „CMC-4012-40P ECM-Mikrofon“. <https://www.cuidevices.com/product/audio/microphones/electret-condenser-microphones/cmc-4012-40p>. Zugriff: 24.06.21. 2021.
- [CW99] C. Cheng und G. H. Wakefield. „Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space“. In: *Audio Engineering Society Convention 107*. Audio Engineering Society. 1999.
- [Dor21] Brian Dorey. „Amazon Echo“. <https://www.briandorey.com/post/echo-dot-4th-gen-smart-speaker-teardown>. Zugriff: 19.05.21. 2021.
- [Dud77] D. E. Dudgeon. „Fundamentals of digital array processing“. In: *Proceedings of the IEEE* 65.6 (1977), S. 898–904.
- [Dun+15] F. Dunn *et al.* *Springer handbook of acoustics*. Springer, 2015.
- [Ent21] Binaural Enthusiast. „B1-E Dummy Head“. <https://binauralenthusiast.com/product/etiam-ullamcorper-dolor-5/>. Zugriff: 15.04.21. 2021.
- [Eul06] S. Euler. *Spracherkennung*. Springer, 2006, S. 15–27.
- [Eve+15] J. Even *et al.* „Audio augmented point clouds for applications in robotics“. In: *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE. 2015, S. 4846–4851.
- [Eve+17a] J. Even *et al.* „Probabilistic 3-D Mapping of Sound-Emitting Structures Based on Acoustic Ray Casting“. In: *IEEE Transactions on Robotics* 33.2 (2017), S. 333–345.
- [Eve+17b] C. Evers *et al.* „Source tracking using moving microphone arrays for robot audition“. In: *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE. 2017, S. 6145–6149.
- [Foy76] W. H. Foy. „Position-location solutions by Taylor-series estimation“. In: *IEEE Transactions on Aerospace and Electronic Systems* 2 (1976), S. 187–194.

- [Fré+12] M. Fréchet *et al.* „Integration of sound source localization and separation to improve dialogue management on a robot“. In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2012, S. 2358–2363.
- [GM19] F. Grondin und F. Michaud. „Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations“. In: *Robotics and Autonomous Systems* 113 (2019), S. 63–80.
- [GM95] W. G. Gardner und K. D. Martin. „HRTF measurements of a KEMAR“. In: *The Journal of the Acoustical Society of America* 97.6 (1995), S. 3907–3908.
- [Gro+21] F. Grondin *et al.* „ODAS: Open embedded audition system“. In: *arXiv preprint arXiv:2103.03954* (2021).
- [GSB07] G. Grisetti, C. Stachniss und W. Burgard. „Improved techniques for grid mapping with rao-blackwellized particle filters“. In: *IEEE transactions on Robotics* 23.1 (2007), S. 34–46.
- [Hu+11] J. Hu *et al.* „Simultaneous localization of a mobile robot and multiple sound sources using a microphone array“. In: *Advanced Robotics* 25.1-2 (2011), S. 135–152.
- [Inf21] Infineon. „IM67D130A MEMS-Mikrofon“. <https://www.infineon.com/cms/de/product/sensor/mems-microphones/mems-microphones-for-automotive/im67d130a>. Zugriff: 24.06.21. 2021.
- [Ish+09] C. T. Ishi *et al.* „Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments“. In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2009, S. 2027–2032.
- [Jau+01] L. Jaulin *et al.* „Interval analysis“. In: *Applied interval analysis*. Springer, 2001, S. 11–43.
- [Jau09] L. Jaulin. „Robust set-membership state estimation; application to underwater robotics“. In: *Automatica* 45.1 (2009), S. 202–206.
- [Jov+21] S. Jovanoska *et al.* „Passive Sensor Processing and Data Fusion for Drone Detection“. In: *NATO STO Meeting Proceedings: MSG-SET-183 Specialists' Meeting on Drone Detectability: Modelling the Relevant Signature*. 2021.
- [JUD95] S. J. Julier, J. K. Uhlmann und H. F. Durrant-Whyte. „A new approach for filtering nonlinear systems“. In: *Proceedings of 1995 American Control Conference-ACC'95*. Bd. 3. IEEE. 1995, S. 1628–1632.
- [Kal+13] N. Kallakuri *et al.* „Probabilistic approach for building auditory maps with a mobile microphone array“. In: *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE. 2013, S. 2270–2275.

- [Kal60] R. E. Kalman. „A new approach to linear filtering and prediction problems“. In: *Transaction of the ASME—Journal of Basic Engineering* (1960), S. 35–45.
- [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 (1976), S. 320–327.
- [KH04] N. Koenig und A. Howard. „Design and use paradigms for gazebo, an open-source multi-robot simulator“. In: *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Bd. 3. IEEE. 2004, S. 2149–2154.
- [Kid15] K. Kido. „Digital Fourier Analysis: Fundamentals“. In: *Springer-verlag New York Incorporated* (2015).
- [Kim+08] C. Kim *et al.* „Robust estimation of sound direction for robot interface“. In: *IEEE International Conference on Robotics and Automation, (ICRA 2008)*. IEEE. 2008, S. 3475–3480.
- [KMJ19] S. Karthick, K. Madhav und K. Jayavidhi. „A comparative study of different eye tracking system algorithms“. In: *AIP Conference Proceedings*. Bd. 2112. 1. AIP Publishing LLC. 2019, S. 020171.
- [KND06] F Keyrouz, Y Naous und K Diepold. „A new method for binaural 3-D localization based on HRTFs“. In: *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*. Bd. 5. IEEE. 2006, S. V–V.
- [KNO15] U. Kim, K. Nakadai und H. G. Okuno. „Improved sound source localization in horizontal plane for binaural robot audition“. In: *Applied Intelligence* 42.1 (2015), S. 63–74.
- [Lev44] K. Levenberg. „A method for the solution of certain non-linear problems in least squares“. In: *Quarterly of applied mathematics* 2.2 (1944), S. 164–168.
- [LFL13] Hong Liu, Zhuo Fu und Xiaofei Li. „A two-layer probabilistic model based on time-delay compensation for binaural sound localization“. In: *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE. 2013, S. 2705–2712. DOI: 10.1109/ICRA.2013.6630949.
- [Lid04] D. R. Lide. *CRC handbook of chemistry and physics*. Bd. 85. CRC press, 2004.
- [Lin73] R. Bruce Lindsay. „Acoustics: Historical and Philosophical Development, Dowden, Hutchinson & Ross“. In: *Inc., PA* (1973), S. 194–201.
- [LW12] M. Langerwisch und B. Wagner. „Guaranteed mobile robot tracking using robust interval constraint propagation“. In: *International Conference on Intelligent Robotics and Applications*. Springer. 2012, S. 354–365.

- [LZG94] J. Lu, H. Zou und J. F. Greenleaf. „Biomedical ultrasound beam forming“. In: *Ultrasound in medicine & biology* 20.5 (1994), S. 403–428.
- [Ma+15] F. Ma *et al.* „Finite element determination of the head-related transfer function“. In: *Journal of Mechanics in Medicine and Biology* 15.05 (2015), S. 1550066.
- [Mac+67] J. MacQueen *et al.* „Some methods for classification and analysis of multivariate observations“. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Bd. 1. 14. Oakland, CA, USA. 1967, S. 281–297.
- [Mah03] R. Mahler. „Multitarget Bayes filtering via first-order multitarget moments“. In: *IEEE Transactions on Aerospace and Electronic systems* 39.4 (2003), S. 1152–1178.
- [ME85] H. Moravec und A. Elfes. „High resolution maps from wide angle sonar“. In: *Proceedings. 1985 IEEE international conference on robotics and automation*. Bd. 2. IEEE. 1985, S. 116–121.
- [Mer13] A. Mertins. *Signaltheorie: Grundlagen der Signalbeschreibung, Filterbänke, Wavelets, Zeit-Frequenz-Analyse, Parameter- und Signalschätzung*. 3. Aufl. Springer-Verlag, 2013.
- [MEW09] John C Murray, Harry R Erwin und Stefan Wermter. „Robotic sound-source localisation architecture using cross-correlation and recurrent neural networks“. In: *Neural Networks* 22.2 (2009), S. 173–189. DOI: 10.1016/j.neunet.2009.01.013.
- [MG91] J. C. Middlebrooks und D. M. Green. „Sound localization by human listeners“. In: *Annual review of psychology* 42.1 (1991), S. 135–159. DOI: 10.1146/annurev.ps.42.020191.001031.
- [Mic+20] S. Michaud *et al.* „3D Localization of a Sound Source Using Mobile Microphone Arrays Referenced by SLAM“. In: *Intelligent Robots and Systems (IROS), 2020 IEEE/RSS International Conference on*. 2020.
- [Miz+11] T. Mizumoto *et al.* „Design and implementation of selectable sound separation on the Texai telepresence system using HARK“. In: *2011 IEEE International Conference on Robotics and Automation*. IEEE. 2011, S. 2130–2137.
- [Mös05] M. Möser. *Technische Akustik*. Bd. 8. Springer Berlin, 2005.
- [Mul+11a] J. Mullane *et al.* „A random-finite-set approach to Bayesian SLAM“. In: *IEEE Transactions on Robotics* 27.2 (2011), S. 268–282.
- [Mul+11b] J. Mullane *et al.* *Random Finite Sets for Robot Mapping & SLAM: New Concepts in Autonomous Robotic Map Representations*. Bd. 72. Springer Science & Business Media, 2011.

- [MW01] A. Mahajan und M. Walworth. „3D position sensing using the differences in the time-of-flights from a wave source to various receivers“. In: *IEEE Transactions on Robotics and Automation* 17.1 (2001), S. 91–94.
- [Nak+00] K. Nakadai *et al.* „Active audition for humanoid“. In: *AAAI/IAAI*. 2000, S. 832–839.
- [Nak+03] K. Nakadai *et al.* „Applying scattering theory to robot audition system: Robust sound source localization and extraction“. In: *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*. Bd. 2. IEEE. 2003, S. 1147–1152.
- [Nak+05] K. Nakadai *et al.* „Sound source tracking with directivity pattern estimation using a 64 ch microphone array“. In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2005, S. 1690–1696.
- [Nak+09] H. Nakajima *et al.* „Real-time sound source orientation estimation using a 96 channel microphone array“. In: *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2009, S. 676–683.
- [Nak+11] K. Nakamura *et al.* „Intelligent sound source localization and its application to multimodal human tracking“. In: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2011, S. 143–148.
- [Neu12] A. Neubauer. *DFT-Diskrete Fourier-Transformation*. Springer-Verlag, 2012.
- [Neu21] Neumann. „U 87 Ai Studio Set Kondensatormikrofon“. <https://en-de.neumann.com/u-87-ai>. Zugriff: 24.06.21. 2021.
- [New21] Newegg. „ClearOne 910 - Beamforming Microphone Array 2“. <https://www.newegg.com/clearone-910-3200-201/p/0ZU-0007-00062>. Zugriff: 19.05.21. 2021.
- [Ngu+16] V. Q. Nguyen *et al.* „Localizing an Intermittent and Moving Sound Source Using a Mobile Robot“. In: *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE. 2016.
- [Nin15] J. Ninin. „Global optimization based on contractor programming: An overview of the IBEX library“. In: *International Conference on Mathematical Aspects of Computer and Information Sciences*. Springer. 2015, S. 555–559.
- [NNI12] K. Nakamura, K. Nakadai und G. Ince. „Real-time super-resolution sound source localization for robots“. In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2012, S. 694–699.
- [NOK01] K. Nakadai, H. G. Okuno und H. Kitano. „Epipolar geometry based sound localization and extraction for humanoid audition“. In: *Intelligent Robots and Systems,*

2001. *Proceedings. 2001 IEEE/RSJ International Conference on*. Bd. 3. IEEE. 2001, S. 1395–1401.
- [Oku+12] K. Okutani *et al.* „Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter“. In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2012, S. 3288–3293.
- [ON15] Hiroshi G Okuno und Kazuhiro Nakadai. „Robot audition: its rise and perspectives“. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2015, S. 5610–5614. DOI: 10.1109/ICASSP.2015.7179045.
- [Pra14] K. M. Prabhu. *Window functions and their applications in signal processing*. Taylor & Francis, 2014.
- [Qui+09] M. Quigley *et al.* „ROS: an open-source Robot Operating System“. In: *ICRA workshop on open source software*. Bd. 3. 3.2. Kobe, Japan. 2009, S. 5.
- [Ray07] L. Rayleigh. „XII. On our perception of sound direction“. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13.74 (1907), S. 214–232.
- [Rey09] D. A. Reynolds. „Gaussian mixture models.“ In: *Encyclopedia of biometrics* 741 (2009), S. 659–663.
- [RJC09] O. Reynet, L. Jaulin und G. Chabert. „Robust TDOA passive location using interval analysis and contractor programming“. In: *2009 International Radar Conference Surveillance for a Safer World (RADAR 2009)*. IEEE. 2009, S. 1–6.
- [RM17] C. Rascon und I. Meza. „Localization of sound sources in robotics: A review“. In: *Robotics and Autonomous Systems* 96 (2017), S. 184–210.
- [Sas+10] Y. Sasaki *et al.* „Map-generation and identification of multiple sound sources from robot in motion“. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2010, S. 437–443.
- [SBT20] M. Sewtz, T. Bodenmüller und R. Triebel. „Robust MUSIC-Based Sound Source Localization in Reverberant and Echoic Environments“. In: *Intelligent Robots and Systems (IROS), 2020 IEEE/RSJ International Conference on*. 2020.
- [Sch21] Schoeps. „Kleinmembran-Mikrofon - Schoeps Colette CMC-64“. <https://schoeps.de/produkte/colette.html>. Zugriff: 19.05.21. 2021.
- [Sch86] R. Schmidt. „Multiple emitter location and signal parameter estimation“. In: *IEEE transactions on antennas and propagation* 34.3 (1986), S. 276–280.

- [Sen21] Sennheiser. „Deckenmikrofon TeamConnect Ceiling 2“. <https://de-de.sennheiser.com/tcc2>. Zugriff: 19.05.21. 2021.
- [SF14] G. R. Sinambari und I. M. Fallen. *Ingenieurakustik*. 5. Aufl. Springer, 2014.
- [SRT16] Y. Sasaki, T. Ryo und H. Takemura. „Probabilistic 3D Sound Sources Mapping Using Moving Microphone Array“. In: *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE. 2016.
- [Su+16] D. Su *et al.* „Robust Sound Source Mapping Using Three-Layered Selective Audio Rays for Mobile Robots“. In: *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE. 2016.
- [Tam+05] Yuki Tamai *et al.* „Three ring microphone array for 3d sound localization and separation for mobile robot audition“. In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2005, S. 4172–4177. DOI: 10.1109/IROS.2005.1545095.
- [Tho21] Thomann. „Bändchenmikrofon - Coles 4038 Studio Ribbon“. [https://www.thomann.de/de/coles\\_4038\\_studio\\_ribbon.htm](https://www.thomann.de/de/coles_4038_studio_ribbon.htm). Zugriff: 19.05.21. 2021.
- [Tho90] J. E. Thorner. „Approaches to sonar beamforming“. In: *IEEE Technical Conference on Southern Tier*. IEEE. 1990, S. 69–78.
- [Thr+02] S. Thrun *et al.* „Robotic mapping: A survey“. In: *Exploring artificial intelligence in the new millennium* 1.1-35 (2002), S. 1.
- [Thr02] S. Thrun. „Probabilistic robotics“. In: *Communications of the ACM* 45.3 (2002), S. 52–57.
- [TK15] R. Takeda und K. Komatani. „Performance comparison of MUSIC-based sound localization methods on small humanoid under low SNR conditions“. In: *Humanoid Robots (Humanoids), 2015 IEEE-RAS 15th International Conference on*. IEEE. 2015, S. 859–865.
- [VM06] B. Vo und W. Ma. „The Gaussian mixture probability hypothesis density filter“. In: *IEEE Transactions on signal processing* 54.11 (2006), S. 4091–4104.
- [Vog20] R. Voges. „Bounded-error visual-LiDAR odometry on mobile robots under consideration of spatiotemporal uncertainties“. Dissertation. Hannover: Institutionelles Repositorium der Leibniz Universität Hannover, 2020.
- [VSD05] B. Vo, S. Singh und A. Doucet. „Sequential Monte Carlo methods for multitarget filtering with random finite sets“. In: *IEEE Transactions on Aerospace and electronic systems* 41.4 (2005), S. 1224–1245.

- [WB06] D. Wang und G. J. Brown. *Computational auditory scene analysis: Principles, algorithms, and applications*. Wiley-IEEE Press, 2006.
- [WBS54] R. S. Woodworth, B. Barber und H. Schlosberg. *Experimental psychology*. Oxford und IBH Publishing, 1954.
- [Wor20] World Robotics. „Executive Summary World Robotics 2020 Service Robots“. [https://ifr.org/img/worldrobotics/Executive\\_Summary\\_WR\\_2020\\_Service\\_Robots.pdf](https://ifr.org/img/worldrobotics/Executive_Summary_WR_2020_Service_Robots.pdf). Zugriff: 22.08.21. 2020. URL: [https://ifr.org/img/worldrobotics/Executive\\_Summary\\_WR\\_2020\\_Service\\_Robots.pdf](https://ifr.org/img/worldrobotics/Executive_Summary_WR_2020_Service_Robots.pdf).
- [WWS20] W. Willems, A. Wagner und D. Stricker. *Schallschutz: Bauakustik*. Bd. 2. Auflage. Springer, 2020.
- [YAZ12] K. Youssef, S. Argentieri und J. Zarader. „Towards a systematic study of binaural cues“. In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2012, S. 1004–1009.
- [Zah02] P. Zahorik. „Auditory display of sound source distance“. In: *Proc. Int. Conf. on Auditory Display*. 2002, S. 326–332.

# Liste der Veröffentlichungen

## **Autor**

Rauschenberger, A. und Wagner B. (2020): Interval-Based Sound Source Mapping for Mobile Robots, Proceedings International Conference on Informatics in Control, Automation and Robotics (ICINCO), S. 335-345

Rauschenberger, A. und Wagner B. (2022): 3D Sound Source Mapping using Interval-based Methods, Informatics in Control, Automation and Robotics, ICINCO 2020, Lecture Notes in Electrical Engineering, vol 793. Springer, Cham, S. 432-455



# Liste der betreuten Arbeiten

Erweiterung eines Versuches zur Demonstration von steuerungsinterpretierten Petri-Netzen, Bachelorarbeit, 2021

Kartierung von statischen Soundquellen mittels Extended Mixture Kalman Filtering an einem mobilen Roboter, Bachelorarbeit, 2021

Entwicklung einer Evaluierungssoftware für die intervall-basierte Soundquellenkartierung, Studienarbeit, 2020

Lokalisierung von statischen Soundquellen mittels GEVD-MUSIC-Verfahren an einem mobilen Roboter, Masterarbeit, 2020

Erweiterung einer Soundsimulation um akustische Reflexions- und Dämpfungseffekte in Gazebo, Masterarbeit, 2020

Lokalisierung von statischen Soundquellen mittels MUSIC-Verfahren an einem mobilen Roboter, Masterarbeit, 2020

Intervall-basierte Positionsbestimmung von mehreren Soundquellen an einem mobilen Roboter, Masterarbeit, 2020

Visuelles Tracking der Position von Mikrofonen und Soundquellen, Studienarbeit, 2019

Lokalisierung von statischen Soundquellen mittels SRP-Verfahren an einem mobilen Roboter, 2019

Entwicklung einer sprachlichen Benutzerschnittstelle und einer dynamischen Aufgabenplanung für einen mobilen Serviceroboter, Masterarbeit 2019

Erweiterung einer mobilen Plattform zur Erstellung von Zeichnungen, Bachelorarbeit, 2018

Lokalisierung von Soundquellen mittels Deep Learning, Masterarbeit 2018

Soundquellenlokalisierung mittels Beamformer-basierten Ansätzen an einem mobilen Roboter, Bachelorarbeit 2018

Development of a Voice Control System for a Mobile Service Robot, Masterarbeit, 2018

Entwicklung eines Verfahrens zum optimalen Scheduling von Fahrzeugen am Beispiel des Finishbereiches einer Automobilfertigung, Masterarbeit, 2016



# Über den Autor

Name: Axel Rauschenberger  
Geburtsdatum: 28. Oktober 1988  
Geburtsort: Burgwedel

## Ausbildung

2008 Abitur am Gymnasium Isernhagen in Altwarmbüchen  
2008 - 2009 Zivildienst, Pestalozzi-Stiftung, Burgwedel  
2009 - 2015 Studium der Computergestützten Ingenieurwissenschaften  
an der Leibniz Universität Hannover  
mit den Abschlüssen B. Sc. und M. Sc.

## Wissenschaftliche Arbeit

2015 - 2022 Wissenschaftlicher Mitarbeiter  
am Institut für Systems Engineering, Fachgebiet Echtzeitsysteme  
an der Leibniz Universität Hannover

## Berufliche Tätigkeit

seit März 2022 Entwicklungsingenieur  
bei der Robert Bosch GmbH