

Object Detection and Classification for Mobile Platforms Using 3D Acoustic Imaging

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
Dipl.-Wirtsch.-Ing. Marco Moebus
geboren am 19.09.1979 in Kronberg

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof. Dr. Salim Bouzerdoum
Tag der Einreichung:	15.10.2010
Tag der mündlichen Prüfung:	13.12.2010

D 17
Darmstädter Dissertation
Darmstadt, 2011

Acknowledgments

First and foremost, I would like to thank Prof. Dr.-Ing. Abdelhak Zoubir for his support and encouragement during the last years. It was a pleasure to work for him and I appreciate his ability to create an atmosphere of cooperation and openness in his group. I also thank Prof. Dr. Salim Bouzerdoum and Prof. Dr. Mats Viberg for their interest in my work and the fruitful discussions we had.

My thanks go to the many colleagues at the institute who are responsible for making my time there an enjoyable and formative experience. I especially thank Weaam Alkhaldi, Saïd Aouada, Ramon Brcic, Chris Brown, Luke Cirillo, Christian Debes, Raquel Fandos, Ulrich Hammes, Philipp Heidenreich, Stefan Leier, Michael Muma, and Michael Rübsamen.

I am also grateful to the colleagues from my industry partner and want to thank them for their support. I appreciate their willingness to not only cooperate on this topic, but also to give me access to hardware and measurement equipment which were substantial for obtaining real data measurements.

Most importantly, I would like to thank my wife and my family: Thank you for your love and support!

Zusammenfassung

Diese Doktorarbeit thematisiert das Problem der Objekt-Entdeckung und -Klassifizierung für mobile Plattformen wie Autos oder Roboter mit Hilfe von Ultraschall-Sensorgruppen. Hierbei wird ein räumlich breites Anregungssignal ausgesendet und von einer zwei-dimensionalen Sensorgruppe empfangen. Mit Hilfe von adaptiven Beamforming-Verfahren werden die zurück reflektierten Echos dann verarbeitet und können in einem drei-dimensionalen Bild der Szene dargestellt werden. Zunächst beschreiben wir einige Entwicklungsprinzipien, die sich aus dem Betrieb in von Menschenhand geschaffener Umgebung sowie physikalischen Bedingungen ergeben, so z.B. durch die geringe Schallgeschwindigkeit in Luft. Darüber hinaus stellen wir ein Kalibrierungsverfahren vor, welches insbesondere für Ultraschall-Sensorgruppen eine parametrische Korrektur von Positionsfehlern der Sensoren ermöglicht. Bisherige Verfahren benötigen hierzu entweder viele Kalibrierungsquellen oder können Positionsfehler nur nicht-parametrisch und damit manchmal unzureichend korrigieren. Aufgrund des hohen bzw. steigenden Kostendrucks in der Automobil- und Robotikbranche ist hierbei besonders wichtig, ein solches System mit hoher Leistungsfähigkeit, aber nur einer geringen Anzahl von Sensoren betreiben zu können. Daher werden in dieser Arbeit Verfahren dargestellt, die es erlauben, schwach besetzte Sensorgruppen zu entwerfen, die eine hohe räumliche Auflösung ermöglichen, dabei jedoch auch gute Rauschunterdrückung aufweisen. Dadurch können Objekte verlässlich und ausreichend genau dargestellt werden und heben sich gut vom Hintergrund ab. Die entwickelten Methoden basieren auf der Theorie minimaler Redundanz, die in dieser Arbeit für den zwei-dimensionalen Fall angewandt und erweitert wird. In einem weiteren Teil der Arbeit beschäftigen wir uns mit der Detektion von Personen mit Hilfe dieser Ultraschall-Sensorgruppen. Wir zeigen, dass sich aus den verarbeiteten Echos einfache geometrische Merkmalen extrahieren lassen, anhand derer Personen mit einer Genauigkeit von knapp 97 Prozent von anderen Objekten unterschieden werden können. Die dazu notwendigen Klassifikatoren basieren auf linearer und quadratischer Diskriminantenanalyse und besitzen daher geringe Komplexität. Darüber hinaus ist es auch anhand eines einzelnen Bildes möglich, die Haltung einer Person genauer zu klassifizieren. So entwickeln wir Klassifikatoren, die anhand einer Kombination von geometrischen und statistischen Merkmalen erkennen, ob eine Person läuft oder steht. Die hierbei erreichte Genauigkeit beträgt mehr als 87 Prozent. Die entwickelten Methode werden nicht nur in Simulationen angewendet und bewertet, sondern auch auf reale Messdaten angewandt. Die Daten wurden mit Hilfe mehrerer Prototypen von akustischen Sensorgruppen aufgenommen. Dabei wurden sowohl innerhalb als auch außerhalb von Gebäuden verschiedene Szenarien aufgenommen.

Abstract

This thesis addresses the problem of obstacle detection and classification for mobile platforms such as robots using acoustic imaging. To obtain an acoustic image of a scene, a spatially broad signal is transmitted and the object's reflections are received by a 2D array of acoustic receivers. The resulting data is processed using adaptive beamforming and can be translated into a three-dimensional image. Since the targeted platforms operate in man-made environments, we first develop design principles derived from physical constraints such as the slow speed of sound in air. Furthermore, we present a calibration method which is specifically well suited for acoustic arrays and parametrically corrects for position errors of the sensors. Other methods are either limited, e.g., by the need of a high number of calibration sources, or can correct such errors non-parametrically and therefore sometimes insufficiently. The increasing cost pressure for domestic robots demands to operate using cheap hardware, which favors the use of acoustic imaging. Such cost constraints require to use highly sparse 2D array designs which still allow to resolve objects clearly and result in acoustic images with a distinct discrimination between object echoes and background. Thus, we develop methods to design non-uniform, sparse arrays which possess reasonable spatial resolution together with good noise suppression. The presented methods apply minimum-redundancy theory in the two-dimensional case and extend it in order to control the redundancy. We also address the problem of human detection and develop feature sets for a corresponding binary classifier. As a result, humans can be discriminated from other objects using only a three-dimensional feature space and simple classifiers such as Linear Discriminant Analysis or Quadratic Discriminant Analysis with an accuracy of almost 97 percent. We also present geometrical and statistical features which allow the classification of humans with respect to their pose, meaning that we can distinguish whether a person is walking or standing with a classification accuracy of more than 87 percent. All developed methods are applied not only to simulation data, but also to real data measurements. The data was obtained using several prototypes of real acoustic array systems in indoor and outdoor environments.

Contents

Zusammenfassung	III
Abstract	V
List of Acronyms	XI
List of Symbols	XIII
1 Introduction	1
1.1 Motivation	1
1.2 Overview	2
2 Fundamentals	5
2.1 Fundamentals of Array Signal Processing	5
2.1.1 Signal Model	5
2.1.2 Beamforming	8
2.1.3 Subspace Methods	11
2.1.4 Coherent Sources	13
2.1.5 Robust Beamforming	14
2.2 Fundamentals of Classification	15
2.2.1 Discriminant Functions	16
2.2.1.1 Fisher's Linear Discriminant	16
2.2.1.2 Generalized Linear Discriminants	17
2.2.2 Support Vector Machines	17
2.2.2.1 Maximum Margin Classifier	18
2.2.2.2 Soft Margin Classifier	19
2.2.2.3 Choice of the Kernel Function	21
3 Design of Acoustic Imaging Systems	22
3.1 Design Principles for Acoustic Imaging Systems	22
3.1.1 Data Processing	22
3.1.2 Assumptions and Basic Characteristics	24
3.1.3 System Setup	26
3.1.4 Real Data Examples	27
3.2 Calibration Techniques	32
3.2.1 Fundamentals	32
3.2.2 Global Calibration Techniques	34
3.2.3 Local Calibration Techniques	35

3.2.4	Parametric Maximum-Likelihood Estimation of Position Errors	36
3.2.5	Proposed Low-complexity Estimation Procedure	37
3.2.6	Results and discussion	38
4	Sparse Array Design	45
4.1	Introduction	45
4.2	Problem Formulation	46
4.3	Minimum Redundancy Theory	47
4.3.1	Fundamental Concept	47
4.3.2	Two-Dimensional Difference Sets	49
4.4	Forward Inclusion Approach	49
4.5	Lattice Structure	54
4.5.1	Randomization of Lattices	55
4.6	Experimental Results	57
4.6.1	Simulations	57
4.6.2	Kernel Bandwidth	59
4.6.3	Acoustic Imaging	62
5	Human Detection and Classification	67
5.1	Introduction	67
5.2	Segmentation	72
5.3	Feature Extraction	75
5.3.1	Modeling the Acoustic Signature	75
5.3.2	Geometric Features	78
5.3.2.1	Elliptic Torso Fitting	78
5.3.2.2	Generic Shape Parameters	79
5.3.3	Statistical Features	79
5.3.3.1	Hill Estimator	80
5.3.3.2	Power-related Tail Parameters	80
5.3.3.3	Depth-related Parameters	81
5.4	Feature Selection	81
5.5	Results	83
5.5.1	Experimental Setup	83
5.5.2	Modeling Results	84
5.5.3	Human Detection Performance	86
5.5.4	Pose Classification Performance	88
6	Conclusions and Outlook	92
6.1	Conclusions	92
6.2	Outlook	93

Appendix	95
Bibliography	95
Curriculum Vitae	105
Publications	107
A.1 Internationally Refereed Publications	107
A.2 Filed Patent Applications	107

List of Acronyms

AIC	Akaike Information Criterion
AWGN	Additive White Gaussian Noise
DOA	Direction-Of-Arrival
DS	Difference Set
FFT	Fast Fourier Transform
GMM	Gaussian-Mixture-Model
HPBW	Half-Power-Beam-Width
iid	independent and identically distributed
ISLR	Integrated-SideLobe-Ratio
LDA	Linear Discriminant Analysis
MALSO	Maneuvering Aids for Low-Speed Operation
MDL	Minimum Description Length
ML	maximum likelihood
MLE	maximum likelihood estimator
MRA	Minimum-Redundancy Array
mRMR	minimal-Redundancy-Maximal-Relevance
MSE	mean square error
MUSIC	MUltiple-SIgnal-Classification
QDA	Quadratic Discriminant Analysis
RMSE	root mean square error
SLL	Side-Lobe-Level
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
TDS	Two-Dimensional Difference Set

TOF	Time-Of-Flight
ULA	Uniform Linear Array
URA	Uniform Rectangular Array
UCA	Uniform Circular Array
US	Ultrasound

List of Symbols

$(\cdot)^T$	transpose of a vector or matrix
$(\cdot)^H$	Hermitian of a vector or matrix
a_i	magnitude of the array response of the i th array element
\mathbf{a}	array manifold vector
A	area of an image segment
A_c	convex area of an image segment
\mathbf{A}	array steering matrix
$B(\theta, \phi), B(\mathbf{k})$	beam-pattern
c	speed of propagation
c_k	class label of the k th observation
(c_θ, c_ϕ)	contour pixel of an image segment
\mathcal{C}	contour of a foreground region in an image
\mathcal{C}_e	contour of an ellipsoid region in an image
$\tilde{\mathcal{C}}_e$	contour of a standardized ellipse
\mathcal{C}_n	class label of the n th class
cv	convexity
D	number of (calibration) sources, dimensionality of feature space
$D(\mathcal{S}, \mathcal{C}_n)$	averaged mutual information between a feature set and a class label
f_c	center-frequency
f_s	sampling frequency
fps	frames per second
\mathbf{f}	feature vector
F	volume fraction of a Gaussian that is part of the foreground region
h	bandwidth parameter

\mathbf{I}	identity matrix
$J(\mathbf{w})$	class separability
\mathbf{J}	selection matrix
k	wave number
\mathbf{k}	wave vector
K	number of observations, number of Gaussians in a Gaussian-Mixture-Model
L	aperture length
m	number of order statistics used for tail estimation
\mathbf{m}	mean vector
$\mathbf{M}(x, z)$	array position function describing an array geometry
$\mathbf{n}(t)$	noise vector
N	number of array elements in a dense array, number of classes in a classification problem
N_x, N_z	aperture length along x, z -axes (normalized to $\frac{1}{2}\lambda$)
\mathbf{p}	position vector
$P(\theta, \phi), P(\mathbf{k})$	spatial power spectrum
\mathbf{P}_A^\perp	orthogonal projection matrix of \mathbf{A}
q	complex calibration coefficient
\mathbf{Q}	calibration matrix
r	range
\mathbf{r}	range vector between two points
R	redundancy of an array geometry
$R(\mathcal{S})$	redundancy of a feature set
\mathbf{R}_{XX}	covariance matrix of process X
$\mathbf{s}(t)$	signal data vector

$s_i(t)$	signal data of i th source
$\mathbf{s}(x, z)$	local sensor density
\mathcal{S}	feature set
\mathcal{S}_i	ellipsoid image region
$\mathbf{S}(x_i, z_i)$	conditional global sensor density
\mathbf{S}	diagonal matrix with the eigenvalues of an ellipse
\mathbf{u}	unit vector
\mathbf{U}	unitary matrix
\mathbf{w}	weighting vector
$\mathbf{W}(x, z)$	Kernel function
$\mathbf{W}(\theta, \phi)$	real-valued weighting matrix
$\mathbf{x}(t)$	array data vector
x	cross-range coordinate
X	random process
$X^{(i)}$	i th order statistic of process X
$\tilde{x}_{0.8}^{(i)}$	i th ordered data sample above the 80-percentile
$y(t)$	output of a beamformer
$y(\mathbf{f})$	output of a classifier
y	range coordinate
z	height coordinate
$\boldsymbol{\beta}$	parameter vector in a Gaussian-Mixture-Model
Γ	transformation function
ϵ	residual error
ζ	magnitude error of an array element
η	phase error of an array element

θ	elevation angle
$(\bar{\theta}, \bar{\phi})$	center of gravity in the foreground region of an image
κ	number of array elements in a sparse array ($\kappa < N$)
λ	wavelength
$\boldsymbol{\mu}$	mean vector of a Gaussian
ν	an integer represented by a Difference Set
Ξ	complex weighting matrix
ρ	correlation coefficient
$\boldsymbol{\rho}$	position error for an array element
σ	standard deviation of a random process
Σ	covariance matrix
τ	delay
ϕ	azimuth angle
$\Phi(\mathbf{x}, \mathbf{y})$	kernel function of a classifier
Ψ	stacked position vectors of all array elements

Chapter 1

Introduction

This thesis addresses the problem of object detection and classification in the close surroundings of mobile platforms such as robots by means of acoustic imaging. The imaging systems operate using an array of acoustic receivers and a single transmitter. Due to the cost-sensitive nature of the application-related markets, it is important to obtain good imaging performance using a large array aperture, but only a relatively low number of sensors. Thus, this thesis addresses the problem of sparse array design using minimum-redundancy theory and demonstrates how array geometries can be designed which allow high-resolution images and good noise suppression with only a limited number of array elements. This allows precise object detection with minimal resources for the applications of interest. Additionally, we develop statistical and geometrical features which allow to reliably detect and distinguish humans from other objects. Furthermore, we demonstrate how to classify whether a human is standing or walking based on a single acoustic image and a nine-dimensional feature set.

1.1 Motivation

In the following we motivate the use of acoustic imaging in robotic applications more closely. We discuss its advantages and disadvantages and also how acoustic imaging can mitigate limitations of other sensor entities in this context.

In the field of robotics, the demand for higher autonomy increases the requirements for reliable sensing of the environment. Additionally, the fast growing market and ambitious goals also increase cost pressure for the sensor systems, mainly because service robots can only be sold for significantly less than industrial robots [Lie09, Lit09]. One of the biggest markets is Japan, where the government has identified robotics as a core technology in the future assistance of elderly people (e.g. [Cab08]). A quickly developing market for such robots is also seen in other parts of the world, e.g., in Europe, where demographic trends similar to Japan can be observed. Here, a growth rate of 4 percent or more is expected for the next years [Myo09, Lit09]. Many projects have been set up which aim to achieve higher level of autonomy of robots, e.g. projects such as "Humanoids with auditory and visual abilities in populated spaces" (HUMAVIPS), "Interactive Urban Robot" (IURO), "European robotic pedestrian assistant" (EUROPA)

or "Knowledgeable Service Robots for Aging" (KSERA). They all share the need of reliable and precise sensing capabilities, such that the robots can interact more efficiently and in a broad variety of human environments. They are designed to assist humans and operate not only in households and clinical institutions, but also generally in urban, populated environments. Thus, it is crucial for them to become aware of the presence of humans in order to fulfill their tasks. Only when a robot detects the presence of humans, it can respond meaningfully, e.g., it can step out of the way of the human, address the person and offer help, etc.

In this context, acoustic imaging systems can help to detect obstacles in the surroundings of robots and to improve the robot's understanding of the surrounding scene by object classification. Acoustic imaging can enhance the robot's capabilities especially in situation where lighting is insufficient, which is often the case for a robot's operation in urban scenarios or in indoor operations. Additionally, range information is directly obtained for each object in the scene, which can be difficult and expensive to obtain from optical sensors. Obtaining reliable range information was also the reason why originally single ultrasound sensors were employed in robotics [AW89].

From the above described application, the objectives of this work are derived. Our goal is to create acoustic imaging systems which can be used for object detection and classification in the surroundings of a mobile platform such as a robot. As mentioned before, such systems operate mostly at low platform speed and can reliably detect objects in the surroundings, which is crucial especially in severe lighting conditions. The environment in which such systems are most valuable are indoor scenarios for robotic applications and generally urban traffic scenarios. Due to the cost-sensitive applications, the acoustic imaging system is required to use highly sparse sensor arrays.

1.2 Overview

In this section, we give an overview on the structure of the thesis and shortly present the content of the chapters. The thesis is structured as follows: In Chapter 2, we introduce fundamentals which are necessary to understand the work in the following chapters. We discuss the basic signal model used in array processing and commonly used direction-finding algorithms as well as the basic concepts of pattern recognition and classification together with a definition of the classifiers used in this work. In Chapter 3, we present the design principles for the problem of acoustic imaging together with some basic assumptions about the signals and the propagation medium. We also give a short description of the real array systems which were built during the course of this work

and have been used for the real data measurements. This is followed by a discussion of the array calibration problem. Due to inevitable errors in real systems, calibration is required to compensate errors which occur in any real array imaging system due to manufacturing tolerances, etc. Moreover, we demonstrate in this chapter how position errors in an acoustic array affect the performance of the system and how they can be corrected by a low-complexity calibration procedure. In Chapter 4, we discuss the problem of sparse array design. Here, we describe design approaches which allow highly sparse sensor arrays which exhibit low sidelobes. The approaches are based on the theory of minimum-redundancy. The results of this chapter do not only apply to applications in robotics, but are valid for any array system which employs two-dimensional (2D) arrays, e.g., arrays for ultrasonography and other medical imaging systems. After this emphasis on the *design* of acoustic imaging systems, we focus on the functional level of such systems and address the problem of human presence detection in Chapter 5. Here, we present a parametric and a non-parametric method to distinguish between humans and other objects present in a scene. We present a low-dimensional feature set which allows to achieve a correct classification rate of almost 97 percent using simple classifiers. Additionally, we show that it is possible to even further classify the pose of a human, more particularly whether the person is walking or standing. The obtained correct classification rate for this problem is higher than 87 percent. Finally, we conclude the findings of this thesis and give an outlook on future work in Chapter 6.

Chapter 2

Fundamentals

2.1 Fundamentals of Array Signal Processing

In this section, we describe the fundamental signal models and estimation methods that are applied when a sensor array is employed to spatial problems such as spatial spectrum estimation, be it imaging or Direction-Of-Arrival (DOA) estimation, waveform estimation and spatial filtering.

2.1.1 Signal Model

To introduce the standard signal model, we consider a narrow-band signal from unknown direction (θ, ϕ) and wavelength λ impinging on an array with N elements, with ϕ being the azimuth and θ the elevation angle (see Fig. 2.1). The position of the i th element in the array is denoted by position vector $\mathbf{p}_i, i = 1, \dots, N$. The output of the array at time t is denoted by

$$\mathbf{x}(t) = \mathbf{a}(\mathbf{k})s(t) + \mathbf{n}(t) \quad (2.1)$$

where

$$\mathbf{a}(\mathbf{k}) = \begin{pmatrix} a_1(\mathbf{k}, \mathbf{p}_1) \\ \vdots \\ a_N(\mathbf{k}, \mathbf{p}_N) \end{pmatrix}$$

is the *array manifold vector* which models the spatial characteristics such as phase delays and attenuation of the signal impinging on the array's sensors, $s(t)$ is the complex baseband signal and $\mathbf{n}(t)$ is assumed to be spatially white noise with variance σ_n^2 .¹ If $D > 1$ signals impinge on the array, the output will be the superposition of the single received waveforms, e.g.,

$$\mathbf{x}(t) = \sum_d^D \mathbf{a}(\mathbf{k}_d)s_d(t) + \mathbf{n}(t), \quad (2.2)$$

¹Since the sensor positions are normally fixed, we drop the dependence of \mathbf{a} on the sensor positions in the notation for most of this work except in Chapter 3.2, where the positions are assumed to be not perfectly known.

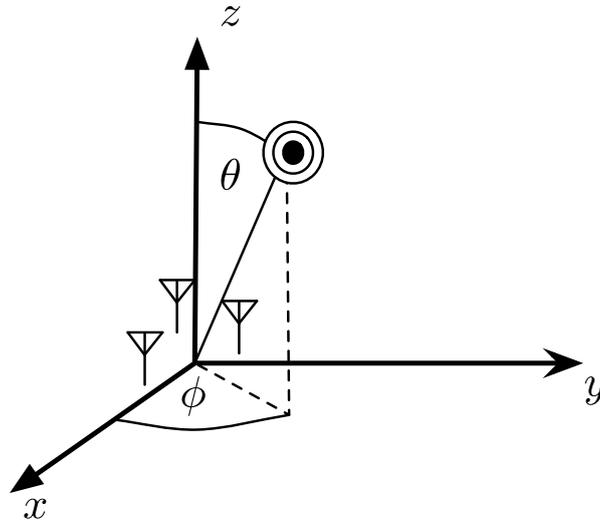


Figure 2.1: Example of a signal impinging on a sensor array. Then angle θ denotes elevation, ϕ denotes azimuth.

or, in matrix form

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad (2.3)$$

where

$$\mathbf{A} = [\mathbf{a}(\mathbf{k}_1) \quad \mathbf{a}(\mathbf{k}_2) \quad \dots \quad \mathbf{a}(\mathbf{k}_D)] \quad (2.4)$$

and

$$\mathbf{s}(t) = \begin{bmatrix} s_1(t) \\ \vdots \\ s_D(t) \end{bmatrix}. \quad (2.5)$$

As $\mathbf{a}(\mathbf{k})$ models the response of the array to a signal with wavelength λ , its form depends on the type of signal source and its distance to the array. Most commonly, it is assumed that signals are excited from point sources emitting spherical wavefronts. Thus, the phase delay at the array's sensors depends not only on the direction, but also on the curvature of the wavefront impinging on the array. Therefore, the distance \mathbf{r}_i between the i th array element and the source can also be taken into account using the *near-field model*, where \mathbf{a} is a function of both direction and distance \mathbf{r}_i between array and source, such that (see [HB91])

$$a_i = \frac{1}{\|\mathbf{p} - \mathbf{r}_i\|} \exp(j\frac{2\pi}{\lambda}\|\mathbf{p} - \mathbf{r}_i\|) \quad .$$

However, if the distance between array and source is sufficiently large, one can choose to neglect the effect of the wavefront curvature and model the incoming wave as a plane

wave, resulting in the *far-field model*. This model is only applicable if the approximation of the wavefront as a plane wave can be justified. The critical distance after which this can be safely assumed is decided on by some rules of thumb, e.g., the *Fraunhofer distance* which states that

$$\|\mathbf{r}\| \gg \frac{2L^2}{\lambda},$$

with L being the largest dimension of the antenna array. The resulting form for the array manifold vector is then simply

$$\mathbf{a}(\mathbf{k}) = \begin{pmatrix} e^{j\mathbf{k}^T \mathbf{p}_1} \\ \vdots \\ e^{j\mathbf{k}^T \mathbf{p}_N} \end{pmatrix}, \quad (2.6)$$

where

$$\mathbf{k} = -\frac{2\pi}{\lambda} \begin{pmatrix} \sin(\theta) \cos(\phi) \\ \sin(\theta) \sin(\phi) \\ \cos(\theta) \end{pmatrix} \quad (2.7)$$

is the wave vector expressed in Cartesian coordinates. Since $\|\mathbf{k}\| = k = \frac{2\pi}{\lambda}$ is the wave number and the vector in Eq. (2.7) denotes a vector in unit space, \mathbf{k} simply refers to an impinging wave with wavelength λ and points into the direction of its arrival. As it is only the phase *differences* which contain the information about the direction of the signals, $\mathbf{a}(\mathbf{k})$ can be normalized such that $a_1 = 1$ without loss of information. If the signals stem not from point sources, but are spatially extended in their dimensions, they can be modeled as the superposition of point sources. Alternatively, one can model the signal by a spatial basis function: A point source would correspond to a spatial dirac delta function, but a spatially extended signal, e.g. due to fading or local scattering at the source is modeled by a physically justifiable basis function such as a Gaussian [Tap02, BV98]. Often, the sensor arrays in an application are of a regular geometry, e.g., a Uniform Linear Array (ULA) or a Uniform Rectangular Array (URA) where the distances between sensor's position are uniform. This regularity is then also present in the array response vector which then shows a Vandermonde structure. This can be exploited for an efficient implementation of the array signal processing methods. For example, using conventional beamforming as explained below will result in the possibility to perform DOA estimation using a spatial Fourier transform, for which the Vandermonde structure in a ULA or a URA leads to a spatial Fast Fourier Transform (FFT). Also the symmetry in other geometries such as Uniform Circular Arrays (UCAs) can be exploited by an FFT by transforming the array into a domain where the array is then a *virtual ULA* (see e.g. [DD94a, DD94b, DD94c]).

2.1.2 Beamforming

The term *beamforming* denotes a technique where the array elements are weighted such that its spatial characteristics can be manipulated. This allows to control the directivity of the array in order to spatially filter the received data, i.e. suppress noise and interference from undesired directions. The two-dimensional spatial power spectrum $P(\mathbf{k})$ of a signal scenario can be estimated by applying a weighting vector $\mathbf{w}(\mathbf{k})$ to the array data using an estimate $\hat{\mathbf{R}}_{XX}$ of the spatial covariance matrix \mathbf{R}_{XX} of the received data in $\mathbf{x}(t)$. By doing so, we obtain the filter output

$$y(t) = \mathbf{w}(\mathbf{k})^H \mathbf{x}(t) .$$

The resulting spatial spectrum estimate is then

$$\hat{P}(\mathbf{k}) = \mathbf{w}(\mathbf{k})^H \hat{\mathbf{R}}_{XX} \mathbf{w}(\mathbf{k}) . \quad (2.8)$$

While imaging applications demand for a high accuracy of $\hat{P}(\mathbf{k})$ in the region of interest, the only figure of merit for DOA estimation is the accuracy of the estimator of (θ, ϕ) . In beamforming, the estimator is typically

$$(\hat{\theta}, \hat{\phi}) = \arg \max_{(\theta, \phi)} \hat{P}(\theta, \phi) . \quad (2.9)$$

A natural choice for $\hat{\mathbf{R}}_{XX}$ is the sample covariance matrix, as it is the maximum likelihood estimator (MLE) of \mathbf{R}_{XX} in white Gaussian noise [VVB88]. Using K data samples, it is defined as

$$\hat{\mathbf{R}}_{XX} = \frac{1}{K} \sum_{t=1}^K \mathbf{x}(t) \mathbf{x}(t)^H .$$

The most intuitive choice for $\mathbf{w}(\mathbf{k})$ results in the *delay-and-sum* beamformer, which is also called *Bartlett beamformer* [HJOK85]. Here, the elements of $\mathbf{w}(\mathbf{k})$ are simply chosen according to the array manifold vector such that the occurring phase differences are compensated by delaying all array channels such that their output is coherent again. As the signal from direction \mathbf{k}_l is recorded with phase shifts in all data channels, choosing

$$\mathbf{w}(\mathbf{k}_l) = \frac{1}{N} \mathbf{a}(\mathbf{k}_l)$$

will weigh all channels differently such that a signal impinging with \mathbf{k}_l is summed up coherently. Since other signals and the spatial noise impinge from other directions than the look direction, this results in a gain in Signal-to-Noise Ratio (SNR) because they add up non-coherently, effectively reducing their power in $y(t)$. The resulting power spectrum estimate is

$$P_{\text{Bartlett}}(\mathbf{k}) = \frac{\mathbf{a}^H(\mathbf{k}_l) \hat{\mathbf{R}}_{XX} \mathbf{a}(\mathbf{k}_l)}{\mathbf{a}^H(\mathbf{k}_l) \mathbf{a}(\mathbf{k}_l)} .$$

In Figure 2.2, we show a 1D-example of a signal scenario with two uncorrelated sources of equal power and an SNR of 20dB impinging on an ULA with 8 elements. The sources come from directions $\theta_1 = 90^\circ$ and $\theta_2 = 80^\circ$ and the noise was assumed to be spatially white. Figure 2.2 (a) depicts the beam-pattern of the array when it is steered to $\theta_l = \theta_1$. While we see that the resulting beam-pattern is maximally sensitive in

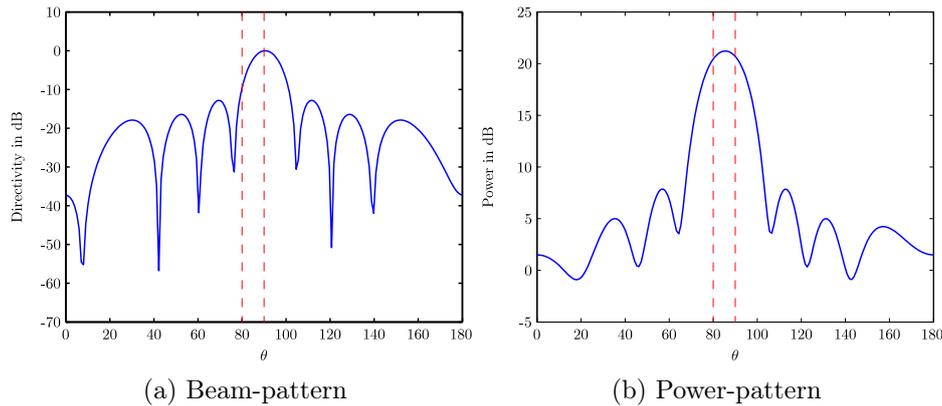


Figure 2.2: (a) Example of the Bartlett beam-pattern for a Uniform Linear Array with 8 elements steered at the signal impinging from $\theta_l = 90^\circ$. A second signal impinges from $\theta_0 = 80^\circ$ and will be received with an attenuation of 10dB. (b) Resulting power-pattern $P(\theta)$ when scanning along the θ -axis. The two signals can not be resolved.

direction θ_l , the array is less sensitive to other directions. The signal coming from θ_2 will only be attenuated by approximately 10dB. Thus, it will be received and interferes with the desired signal although it stems from an undesired direction. In fact, since Bartlett beamforming is a spatial form of a Fourier transform, the corresponding beam-pattern of a ULA will result in a spatial sinc-function due to the regular sampling of space. Thus, the resulting power-pattern is the estimator spatially equivalent to the periodogram. Like the periodogram, it has limited resolution capabilities, thus, it can not resolve the two sources in this example (see Figure 2.2 (b)). However, it is the MLE to find the direction of single sources in white Gaussian noise.

To increase the resolution of the beamformer, one can determine the weight vector adaptively based on the data samples at hand. A popular example of such a beamformer with higher resolution is the so-called *Capon beamformer* [Cap69] which selects the weight vector $\mathbf{w}(\mathbf{k}_l)$ based on the recorded signal data in order to suppress power from undesired directions. To achieve this, the beamforming problem is formulated as an optimization problem where

$$\begin{aligned} \mathbf{w}(\mathbf{k}_l) &= \arg \min_{\mathbf{w}} (P(\mathbf{w})) \\ \text{subject to } \mathbf{w}(\mathbf{k}_l)^H \mathbf{a}(\mathbf{k}_l) &= 1 . \end{aligned} \quad (2.10)$$

The solution of this problem yields

$$\mathbf{w}_c(\mathbf{k}_l) = \frac{\hat{\mathbf{R}}_{XX}^{-1} \mathbf{a}(\mathbf{k}_l)}{\mathbf{a}^H(\mathbf{k}_l) \hat{\mathbf{R}}_{XX}^{-1} \mathbf{a}(\mathbf{k}_l)} \quad (2.11)$$

and the resulting power spectrum is

$$P_{\text{Capon}}(\mathbf{k}) = \frac{1}{\mathbf{a}^H(\mathbf{k}_l) \hat{\mathbf{R}}_{XX}^{-1} \mathbf{a}(\mathbf{k}_l)} . \quad (2.12)$$

We can interpret this result as follows: The information in $\hat{\mathbf{R}}_{XX}$ is exploited such that the beam-pattern is adapted to the present signals in space. In the beam-pattern, nulls will be placed in the direction of any interference and maximal directivity remains only in the direction of interest (see Figure 2.3 (a)). This is achieved at the cost of higher directivity in directions where only noise is present. This adaptive directivity results in a much higher SNR gain and improved resolution capabilities. In Figure 2.3 (b), the resulting power pattern shows that, in contrast to Bartlett's beamformer, Capon's beamformer can clearly resolve the two signals in the signal scenario of the previous example.

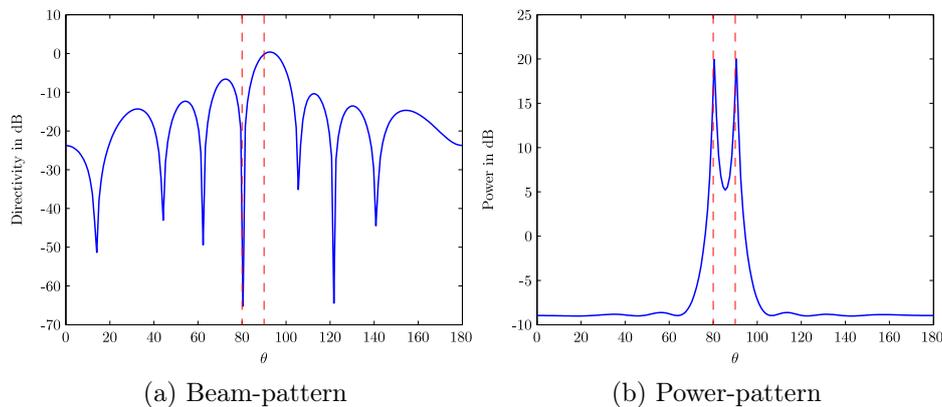


Figure 2.3: (a) Example of the Capon beam-pattern for the same Uniform Linear Array steered at $\theta_l = 90^\circ$. The second signal from $\theta_0 = 80^\circ$ is taken into account by the beamformer and will be almost completely suppressed. (b) Resulting power-pattern $P(\theta)$ when scanning along the θ -axis. The two signals are clearly resolved.

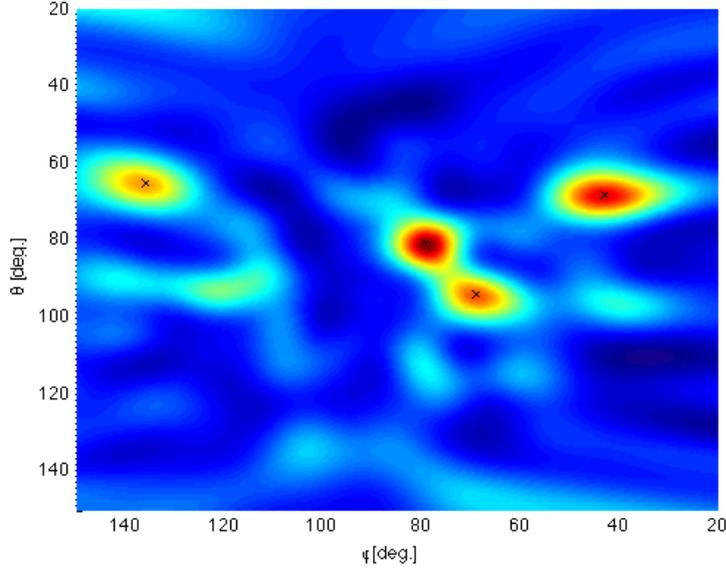


Figure 2.4: Example of the 2D power spectrum of four acoustic point sources using Capon beamforming. The true position of the sources is marked by the black crosses.

2.1.3 Subspace Methods

In addition to the beamforming approach to DOA estimation, there exists another category of methods which try to separate the data into signal and noise subspaces [VO91]. This is based on the fact that, following the standard signal model introduced in Section 2.1.1, the covariance matrix can be decomposed into

$$\mathbf{R}_{XX} = \mathbf{A}\mathbf{R}_{SS}\mathbf{A}^H + \sigma_n^2\mathbf{I} = \mathbf{U}_S\mathbf{\Lambda}_S\mathbf{U}_S^H + \mathbf{U}_N\mathbf{\Lambda}_N\mathbf{U}_N^H,$$

where $\mathbf{U}_S, \mathbf{U}_N$ are unitary matrices containing the signal and noise eigenvectors, spanning the corresponding subspaces [Van02, KV96]. $\mathbf{\Lambda}_S, \mathbf{\Lambda}_N$ are diagonal matrices containing the signal and noise eigenvalues, respectively. If D signals impinge on an array with N elements, $\mathbf{\Lambda}_S$ contains the D largest eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_D > \sigma_n^2$. The smallest $N - D$ eigenvalues all correspond to the noise subspace and are equal under the common assumption of spatially uniform noise. Thus, we have $\mathbf{\Lambda}_N = \sigma_n^2\mathbf{I}$.

Based on this structure of the true covariance matrix, subspace methods seek to separate the eigenvectors in the estimated matrix $\hat{\mathbf{R}}_{XX}$. The most popular representative is the Multiple-Signal-Classification (MUSIC) algorithm which defines a spatial spectrum by

$$P_{\text{MUSIC}}(\mathbf{k}) = \frac{\mathbf{a}^H(\mathbf{k})\mathbf{a}(\mathbf{k})}{\mathbf{a}^H(\mathbf{k})\hat{\mathbf{\Pi}}^\perp\mathbf{a}(\mathbf{k})},$$

where $\hat{\Pi}^\perp = \hat{\mathbf{U}}_N \hat{\mathbf{U}}_N^H$ [Sch86]. This is often referred to as a *pseudo-spectrum*, because due to the subspace projection it has no physical meaning anymore. It is simply the result of minimizing the power in the estimated noise subspace by the MUSIC algorithm. Alternatively, one could also change the numerator to $\mathbf{a}^H(\mathbf{k}) \hat{\mathbf{U}}_S \hat{\mathbf{U}}_S^H \mathbf{a}(\mathbf{k})$ and maximize the power in the estimated signal subspace. Since the DOAs of the signals are found by projections of the data on subspaces, there is no scanning of the environment involved and it is not possible to establish direct measures of directivity such as a beam-pattern for such methods.

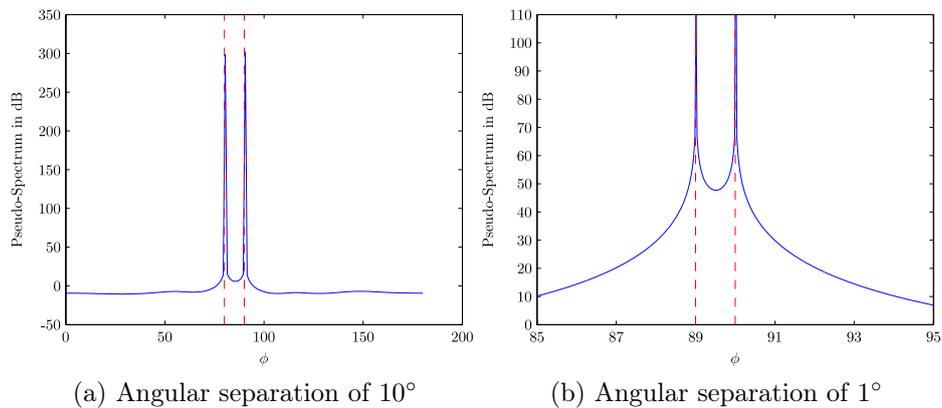


Figure 2.5: Power-pattern obtained using Multiple-Signal-Classification for different angular separations. Clearly, Multiple-Signal-Classification can resolve the two sources well, even when they stem from directions extremely close to each other.

Subspace methods are often called *high-resolution* direction-finding methods because they have a spatial resolution capability that is by far better than that of beamforming methods. However, this is only true when the signals stem from point sources and the number of signals impinging on the array is distinct because the subspace decomposition is based on discrete eigenvectors corresponding to single points in the array manifold space. While this allows to obtain good DOA estimation performance, the result of the methods is only the estimation of a set of distinct DOAs. Apart from that, the methods do not allow to directly measure parameters such as signal strength in a sense that is physically meaningful. Thus, subspace-based methods are not suitable to create acoustic images in the case of spatially extended source signals, only if the application requires simple parameters such as the DOAs of the sources. Furthermore, the methods require an eigen-decomposition of $\hat{\mathbf{R}}_{XX}$, which is computationally complex and can be a limiting factor in the applicability in a real-time system. If the source signals are spatially spread, there exist parametric methods that model the spatial signature based on *a priori* knowledge [MWW96, WWMR94]. However, this knowledge is only rarely available and is also only applicable to DOA estimation problems.

Thus, when the spatial spread of the sources is not negligible, but unknown, the use of beamforming methods is preferable. Generally, the algorithms require knowledge about the number of sources present in the data. This knowledge is usually unavailable and the number of signals has to be estimated using information criteria such as the Akaike Information Criterion (AIC) or Minimum Description Length (MDL) (see [SS04, Aka74, WK85, WZ89, WZ88] for an overview of information criteria).

2.1.4 Coherent Sources

While it is generally fruitful to exploit information from the data, the performance of adaptive methods is degraded when the signal scenarios are complicated. While all methods suffer from an inaccurate estimation of the covariance matrix due to a small sample size or low SNR, specific problems can arise when the signals impinging on the array are correlated or even coherent [SK93, SS98, LVL05, KV96]. While coherent or strongly correlated signals may generally occur in communications, radar and sonar due to jamming or multi-path propagation, they also occur in acoustic imaging when similar objects reflect the signal back to the array. This effect can reduce the resolution capabilities of spatial methods or even lead to their failure. For example, when signal sources are correlated, it is harder for Capon's beamformer to suppress signals from directions other than the look direction when they are correlated to a signal from that direction. More severely, *coherent* signals even lead to a rank-deficiency in \mathbf{R}_{XX} , meaning that signal eigenvectors will deviate into the noise subspace and the signal subspace will be reduced in dimension [KV96]. Thus, adaptive methods, which rely on the estimated covariance matrix, may fail because they normally assume a full-rank covariance matrix in the underlying signal model. To overcome these problems, there exist decorrelation methods that allow to improve the performance in the situation of coherent signals [SWK85, EJS82]. If only two sources are coherent, one can apply forward-backward (FB) averaging to the array. This means that in addition to $\hat{\mathbf{R}}_{XX}$ a *backward* matrix is constructed using $\hat{\mathbf{R}}_{XX}^*$ and a selection matrix \mathbf{J} with non-zeros only on the anti-diagonal, which reverses the order of the array's elements. The two covariance matrices are averaged such that

$$\hat{\mathbf{R}}_{\text{FB}} = \frac{1}{2}(\hat{\mathbf{R}}_{XX} + \mathbf{J}\hat{\mathbf{R}}_{XX}^*\mathbf{J}) .$$

which is then used in lieu of $\hat{\mathbf{R}}$. This effectively decorrelates the two coherent sources. If $D > 2$ sources are coherent, one can use *spatial smoothing* which generalizes the idea of FB averaging by using at least D sub-arrays and the corresponding covariance matrices $\mathbf{R}_d, d = 1, \dots, D$ [SK85]. If the sub-arrays are identical in shape, the corresponding

covariance matrices are assumed to be identical up to a scaling factor which depends only on the geometric distance between the sub-arrays. Thus, averaging the smaller covariance matrices leads to a *smoothed* covariance matrix

$$\mathbf{R}_D = \frac{1}{D} \sum_{d=1}^D \mathbf{R}_d$$

In [SK85], it is shown that any additional sub-array increases the rank of \mathbf{R}_D . Thus, we can always restore the full rank if the array is sufficiently large, i.e., the number of array elements has to be $N \geq 2D$. Although the original idea of spatial smoothing was developed for ULAs, it can be applied to arbitrary array geometries using the array interpolation technique [FW92, WF93]. Note that Bartlett's beamformer, although worse in performance than other techniques previously discussed, does not suffer from any correlation between sources because it is signal-independent.

2.1.5 Robust Beamforming

The beamforming methods described above make several assumptions which might not be fulfilled in practice. They assume perfect knowledge about the array manifold vector and assume sufficient estimation accuracy of the covariance matrix, meaning that SNR or sample size are sufficiently high. However, due to many reasons such as production or operational errors, the array manifold might differ from the used model or signal conditions may become challenging [CZO87]. In such situations, adaptive beamforming can easily perform worse than beamformers with fixed weight vectors when they make wrong assumptions. For example, it may happen due to model errors that there is a mismatch between the desired direction θ_0 and the true steering direction $\theta_0 + \Delta$. Capon's beamformer, for example, would then try to suppress the actual signal from direction θ_0 in order to minimize the overall power under the constraint that maximal directivity in direction $\theta_0 + \Delta$ is preserved. Clearly, the best way to solve this type of problems is to reduce the uncertainty in the model, which can be achieved by testing the array by a calibration procedure (see Section 3.2) and use a more realistic model of the array manifold vector. However, this is not always possible due to time or cost constraints. Thus, alternatively, one can modify the beamforming methods directly such that they rely on milder assumptions about the array manifold and the signal scenario. This approach, known as *robust beamforming*, increases robustness with respect to model mismatch at the cost of reduced spatial resolution. For example, when the sample support is small, the inverse of the covariance matrix estimate $\hat{\mathbf{R}}_{XX}$ typically becomes numerically unstable. In order to allow a more stable matrix inversion, an adaptive beamformer can be robustified by adding a constant to the main diagonal

of $\hat{\mathbf{R}}_{XX}$. This is equivalent to adding artificial noise and allows to stabilize the main diagonal of $\hat{\mathbf{R}}_{XX}$ [LSW03]. This technique, which is called *diagonal loading*, is a general means of robustification of adaptive beamforming techniques. The resulting covariance matrix $\tilde{\mathbf{R}}_{XX}$ can be described as

$$\tilde{\mathbf{R}}_{XX} = \hat{\mathbf{R}}_{XX} + \sigma_{DL}^2 \mathbf{I}. \quad (2.13)$$

The choice of the loading value σ_{DL}^2 can only be determined optimally if a measure of uncertainty on the existing errors is available, otherwise, it has to be set empirically (e.g. [LSW03]). Often, one finds that $\frac{\sigma_{DL}^2}{\hat{\sigma}_n^2} = 10$ is used, where $\hat{\sigma}_n^2$ is an estimate of the average power on the main diagonal of $\tilde{\mathbf{R}}_{XX}$ [Van02]. The concept of diagonal loading provides an easy way to balance the degree of adaptivity of a beamformer. For example, if the artificial noise power is increased, the main diagonal of $\tilde{\mathbf{R}}_{XX}$ becomes more dominant and the beamformer performs more like Bartlett's beamformer. In [Ric07, Ric10], the correlation between the two beamformers is studied and it is shown how the correct degree of adaptivity can improve the spatial resolution for DOA estimation problems.

2.2 Fundamentals of Classification

Classification is the task to assign a class label \mathcal{C}_n , where $n = 1, \dots, N$, to an input vector \mathbf{f} in order to *classify* it as belonging to one of N discrete classes. Clearly, the definition of classes is problem-dependent and while it is most commonly assumed that the data classes are disjoint, this assumption may not hold for specific problems, e.g., when a person walks, it will resemble a standing person during certain parts of the movement (see Section 5.5). The input vector consists of *features* that are extracted in prior stages, they describe specific characteristics from data derived from a lower level, e.g., an image that is obtained from raw sensor data or specific descriptors of that image. A classifier divides the feature space into *decision regions* and the different classification methods differ in the way they formulate and obtain the *decision boundaries* between the classes. In the context of classification, the array signal processing pursued to obtain 3D images can be interpreted simply as a way to create data from which features are extracted. In this section, we introduce some of the approaches to machine learning that we use for the classification of acoustic imaging data. However, we do restrict the discussion here to the applied classifiers and do not cover other powerful approaches, e.g., neural networks or Markov random fields. Moreover, we present the theory in the context of two-class problems for simplicity and because we do not apply the methods to multi-class problems in this thesis. For a more complete and in-depth introduction to the general field of classification, we refer the reader to [Bis07, DHS01].

2.2.1 Discriminant Functions

A *discriminant function* takes a D -dimensional input vector \mathbf{f} and maps it directly to a class label \mathcal{C}_n by some transformation of the data. In its simplest form, the function is linear such that

$$y(\mathbf{f}) = \mathbf{w}^T \mathbf{f} + w_0, \quad (2.14)$$

where \mathbf{w} is a weighting vector and w_0 is a threshold sometimes called *bias*. If $y(\mathbf{f}) \geq 0$, the input is assigned to \mathcal{C}_1 and to \mathcal{C}_2 otherwise. If $K > 2$, one would construct K discriminant functions and obtain decisions by linearly combining them. Therefore, the decision boundary is defined by $y(\mathbf{f}) = 0$ and is a hyperplane with dimension $(D - 1)$.

2.2.1.1 Fisher's Linear Discriminant

In its most simple form, the input \mathbf{f} is weighted linearly as shown above. The task of obtaining a good classifier can then be interpreted as the task to geometrically find the projection direction, represented by \mathbf{w} , that separates the data well into the two classes. Fisher's Linear Discriminant defines a criterion for this class-separability by taking both the distance *between* classes as well as the distribution *within* the classes into account. The inter-class distance, also called between-class scatter, is measured by

$$\mathbf{m}_1 - \mathbf{m}_2,$$

where \mathbf{m}_i is the mean vector of the i th class. We use it in its quadratic form, denoted by

$$\Sigma_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \quad (2.15)$$

as the inter-class scatter matrix. To measure the total intra-class distance of the data, we can simply use the sum of the covariance matrices inside the classes, each given by

$$\Sigma_i = \sum_{\mathbf{f} \in \mathcal{C}_i} (\mathbf{f} - \mathbf{m}_i)(\mathbf{f} - \mathbf{m}_i)^T. \quad (2.16)$$

This results in the intra-class scatter matrix

$$\Sigma_W = \Sigma_1 + \Sigma_2. \quad (2.17)$$

When the data is weighted by \mathbf{w} , these distances can be expressed as $\mathbf{w}^T \Sigma_B \mathbf{w}$ and $\mathbf{w}^T \Sigma_W \mathbf{w}$, respectively. To find the weighting vector that maximizes the class separability, the ratio of inter-and intra-class distances

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \Sigma_B \mathbf{w}}{\mathbf{w}^T \Sigma_W \mathbf{w}} \quad (2.18)$$

is maximized with respect to \mathbf{w} . Assuming Σ_W to be non-singular, the optimal \mathbf{w} is then

$$\mathbf{w}_{\text{opt}} = \Sigma_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) .$$

When this approach is used for classification, it is denoted Linear Discriminant Analysis (LDA). It implicitly assumes that the classes share common covariance matrix, which can be unrealistic. However, the input has limited sample size and mostly no knowledge about the class distribution or its parameters \mathbf{m}_i, Σ_i is available. Thus, these parameters have to be estimated using training data.

2.2.1.2 Generalized Linear Discriminants

The principle of LDA can be generalized to a weighting function that consists of a linear combination of more complex function, e.g. by

$$y(\mathbf{f}) = \sum_{i=0}^d a_i g(\mathbf{f}) , \quad (2.19)$$

where $g(\cdot)$ is some arbitrary function of the input. For example, a Quadratic Discriminant Analysis (QDA) uses a quadratic function of the input vector \mathbf{f} and finds weightings based on that. As a result, the decision boundaries in the feature space will also be of quadratic form. Clearly, this allows better class separation in cases where the classes are not well separated. At the same time, a more complex function $g(\cdot)$ can be the result of more general assumptions on the data, e.g., that the classes have different covariance matrices.

2.2.2 Support Vector Machines

A powerful method to build a classifier is to form the decision boundaries not from functions of a specific form, but from the data directly. In contrast to parametric models which find a weighting vector \mathbf{w} and then project any test data, the idea of Support Vector Machines (SVMs) is to determine the decision boundary based on single input vectors from the training data, the *support vectors*, which are close to the boundary. The boundary is chosen such that the maximum of data points in the training data is lying on the correct side of the boundary. It can be shown that finding the parameters of the boundary is always a convex optimization problem. However, training of a SVM can be computationally complex and in practice, some parameters have to be chosen manually which can highly influence the performance. We will show

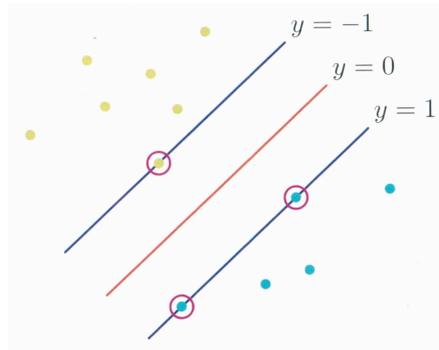


Figure 2.6: The margin of a decision boundary is defined by the smallest orthogonal distance to any of training samples. A Support Vector Machine will choose a boundary such that this margin is maximal. The circles show the support vectors as the samples which determine the boundary. Source: [Bis07]

the general idea behind that concept by introducing the *maximum margin* SVMs, followed by an extension of the discussion to *soft margin* SVMs which are mostly used in practice.

2.2.2.1 Maximum Margin Classifier

To understand the concept of SVMs, let us first define the *margin* of a classifier as the smallest distance between the decision boundary and any of the training samples. The decision boundary is defined in a more general form than in Section 2.2.1 as

$$y(\mathbf{f}) = \mathbf{w}^T v(\mathbf{f}) + w_0, \quad (2.20)$$

where $v(\cdot)$ denotes a function that transforms the feature space into some higher-dimensional space in which it might be easier to separate the classes from each other. For later reference, we also introduce the *kernel function*, denoted by

$$\Phi(\mathbf{x}, \mathbf{y}) = v(\mathbf{x})^T v(\mathbf{y}), \quad (2.21)$$

at this point. Figure 2.6 visualizes that for all points in the feature space which lie on the decision boundary, we have $y(\mathbf{f}) = 0$. Under the assumption of linearly separable classes $\mathcal{C}_1, \mathcal{C}_2$, all samples of \mathcal{C}_1 lie on one side of the boundary and all samples of \mathcal{C}_2 lie on the other side, resulting in

$$y(\mathbf{f}_i) > 0 \quad \mathbf{f}_i \in \mathcal{C}_1 \quad (2.22)$$

$$y(\mathbf{f}_j) < 0 \quad \mathbf{f}_j \in \mathcal{C}_2. \quad (2.23)$$

Given K training samples $\mathbf{f}_1, \dots, \mathbf{f}_K$ and the corresponding binary class labels $c_k \in \{-1, 1\}$, we can write this in a general form as

$$c_k y_k(\mathbf{f}_k) = c_k(\mathbf{w}^T v(\mathbf{f}_k) + w_o) \geq 0 \quad k = 1, \dots, K. \quad (2.24)$$

To find a decision boundary that separates the classes, we are interested to find the parameters \mathbf{w} , w_o which satisfy this condition. Since the distance of any point \mathbf{f} to the boundary is generally given by $|y(\mathbf{f})|/||\mathbf{w}||$, a maximum-margin SVM can be created by choosing the solution which maximizes the distance between the decision boundary and all training data points, which can be expressed as

$$\arg \max_{\mathbf{w}, w_o} \frac{1}{||\mathbf{w}||} \min_k (c_k(\mathbf{w}^T v(\mathbf{f}_k) + w_o)) . \quad (2.25)$$

To solve this problem more efficiently, it is convenient to convert into the equivalent quadratic programming problem [Bis07, SC08]

$$\arg \min_{\mathbf{w}, w_o} ||\mathbf{w}|| \quad (2.26)$$

subject to

$$c_k(\mathbf{w}^T v(\mathbf{f}_k) + w_o) \geq 0 \quad k = 1, \dots, K. \quad (2.27)$$

Thus, from all possible solutions for Eq. (2.20), the one with the minimal norm $||\mathbf{w}||$ will be selected. Since the constraints represent the oriented distance from each sample of the training data to the possible decision boundary, at least one constraint will be active while the majority of constraints are inactive in any data clusters. Only the samples with active constraints determine the margin and are therefore called support vectors. The reformulated quadratic programming problem can be solved using Lagrange multipliers for which various techniques exist [NW99, BB99].

2.2.2.2 Soft Margin Classifier

Using the maximum-margin approach, a misclassified training sample leads to violation of one constraints, meaning that the problem is not solvable anymore. Thus, the classifier will always try to obtain a decision boundary that correctly classifies *all* training data correctly. As a result, this approach only guarantees good performance if the class conditional probabilities do not overlap, i.e., if the data is linearly separable in the feature space. Otherwise, the found solution separates the training data perfectly, but might not perform well on new data, because the decision boundary established from the training data was not general. While this phenomenon, also known as *overfitting*, can occur with any classifier if the training data is not representative, hard-margin

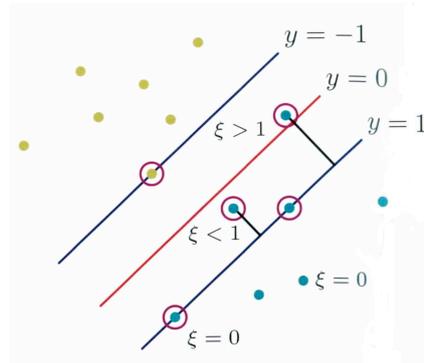


Figure 2.7: Example of a soft margin of a decision boundary. Training samples close to the boundary result in a penalty term $\epsilon < 1$. Misclassified samples are allowed, but lead to a larger penalty with $\epsilon > 1$. Source: [Bis07]

SVMs are specifically sensitive to it because they do not rely on specific forms of the decision boundary. If they establish the decision boundary based on too many support vectors, the found classifier will not perform well in the general case. More severely, depending on the kernel function used, the data might be transformed to a space with much higher dimensionality, which increases the possibility of a decision boundary that is too specific. The *soft-margin* SVMs overcome this limitation by introducing slack variables into the constraints, such that misclassified samples are allowed at the cost of a penalty term that depends on the distance of the sample to the decision boundary [CST00]. Denoting the slack variables ϵ_k for each sample, and $\epsilon_n \geq 0$ for all k , Eq. (2.24) is reformulated to

$$c_k y_k (\mathbf{f}_k) = c_k (\mathbf{w}^T v(\mathbf{f}_k) + w_o) \geq 1 - \epsilon_k \quad k = 1, \dots, K. \quad (2.28)$$

Thus, correctly classified training samples close to the boundary are penalized with $\epsilon_k < 1$. Misclassified samples are allowed, but weighted with a penalty $\epsilon_k > 1$ that scales with the distance to the decision boundary, which is illustrated in Figure 2.7. The objective function then includes the overall penalty by adding a weighted sum of the slack variables as

$$\arg \min_{\mathbf{w}, w_o} \|\mathbf{w}\| + C \sum_{k=1}^K \epsilon_k, \quad (2.29)$$

where C is a constant. This allows to penalize misclassification, and, if C is sufficiently large, the found solution will minimize the number of errors in the training set when the class distributions overlap.

2.2.2.3 Choice of the Kernel Function

While the choice of the kernel function directly affects the performance of the classifier, unfortunately there are no theoretical guidelines which help to choose which kernel function is appropriate for a given problem. Thus, the choice of a kernel is made based on experience and often data analysis. Popular kernels include

- the linear kernel $\Phi(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \cdot \mathbf{y}$,
- radial basis functions, e.g., Gaussian functionals $\Phi(\mathbf{x}, \mathbf{y}) = \exp\left[-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right]$, with σ being determined problem-specific by validation procedures,
- and polynomial kernels, e.g. $\Phi(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d$ with polynomial order d .

Generally, the more complex the kernel function is, the higher the dimensionality of the transformation domain. SVMs solve the classification problem by linear separation of the data in this higher-dimensional space. However, the high dimensionality can also be problematic, especially if the available training data is limited. In fact, solving the problem can become infeasible because the data might be transformed such that it does not form any reasonable clusters in the transformation space, which makes it impossible to recognize patterns and separate them.

Chapter 3

Design of Acoustic Imaging Systems

In this chapter, we present general design principles for acoustic imaging systems for the applications of interest. After some basic definitions and an overview of the signal processing chain, we present some useful assumptions about the transmission medium and the signals, as well as some basic characteristics of an acoustic imaging system which operates in air. After we show system-specific properties of the prototypes used in this thesis, we also show several real-data examples. Additionally, we present calibration techniques which are needed as a first step after the production of an acoustic array in order to compensate for inevitable production errors and tolerances. After a short review of some general techniques, we present a parametric approach which is specifically suited for acoustic arrays and how this approach can be combined with traditional calibration methods.

3.1 Design Principles for Acoustic Imaging Systems

The term *acoustic imaging* denotes techniques which use acoustic signals in order to create images of an object or a scene of interest. In general, an acoustic signal is sent out and the reflections are recorded and processed in order to form an image, although there also exist passive approaches that strive to simply visualize the originating location of sounds. In this thesis, we focus on the use of acoustic arrays which allow to process the recorded reflections in order to estimate the spatial location and shape of reflectors (see also [Ste00, MT00, MT94, PH96]). In the following, we will introduce the assumptions we make with respect to signal model and propagation, followed by a description of the general steps needed to create a three-dimensional, acoustic image from the recorded reflection data. We will then briefly describe the acoustic imaging system that has been developed and used throughout this work.

3.1.1 Data Processing

To generate 3D images of a scene in air, the main limitation one has to deal with is the slow speed of propagation. In contrast to other typical imaging applications, we therefore do not perform beamforming to *transmit* the signal. A better strategy is to

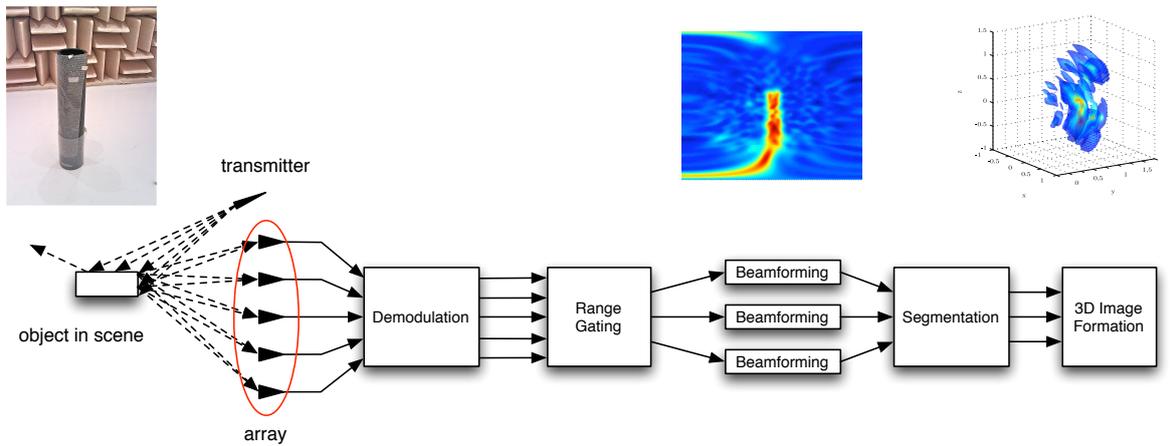


Figure 3.1: Flowchart of an active acoustic imaging system using an array. A signal is transmitted into the scene and reflected by objects. A part of those reflections is received by the receiver array. After demodulation, the signal can be divided into range gates based on the time-of-flight of the echoes. For each range gate, an estimate of the spatial spectrum is obtained using beamforming. To obtain the final image of the scene, reflections sources are determined for each range gate and merged into a 3D image.

illuminate the scene to be analyzed by a single acoustic source which can be part of the array or be placed close to the array. Thus, we are able to image the scene by processing the back-scattered reflections from only one transmit pulse and maintain a scan period suitable for a real-time application. In Figure 3.1, it can be seen how a three-dimensional (3D) image can be created using a receiver array of acoustic sensors: After a single transmitter sends a spatially broad pulse into the scene, the reflections are partly reflected back to the sensor array from objects in the image. Those reflections are then recorded by the array's sensors and demodulated into the complex base-band under use of the known excitation signal. The time series data is then divided into range gates. From each range gate, a 2D spatial power spectrum is estimated using adaptive beamforming, resulting in a 2D image. Based on these 2D images, regions of interest can be identified using some segmentation technique, e.g. global or adaptive thresholding. The found regions correspond to areas from which high power intensity reflections were recorded, i.e., they correspond to reflecting parts of objects in the scene. Finally, the 2D images are transformed from polar to cartesian coordinates and are merged to form a 3D image. Alternatively, one can avoid continuous processing of all channels by detecting echoes and run the beamforming only for the segments in which echoes occur. One can then estimate the range from the Time-Of-Flight (TOF) for each and process the corresponding segments in the 2D (θ, ϕ) -space. One can apply matched filtering and obtain a noise estimate by analyzing the signal of one reference sensor up to a time τ_{min} , which corresponds to the minimal distance r_{min} of an object.

Since no echoes are assumed to be present in this interval, an estimate $\hat{\sigma}_n^2$ of the noise floor is calculated. Note that since the noise is assumed to be Additive White Gaussian Noise (AWGN), it is sufficient to set r_{min} to a small value (e.g. $r_{min} = 20$ cm).

The estimated noise floor $\hat{\sigma}_n^2$ is then used to determine a threshold to detect echo segments, where echoes have to occur with a minimal duration of 1 ms. These segments are then processed individually by the beamforming algorithm, assuming a range calculated from the start of the echo segment, resulting in a dynamic focusing system. Note that the translation of the TOF into range assumes a direct path echo. Additionally, due to the possible overlap of different reflections, the length of the echo segments might vary. In that case, the later echo is assigned the same τ as the first one, possibly introducing a small range error for some parts of the analyzed scene.

Many of the existing adaptive approaches in array signal processing have been developed for far-field conditions and a finite number of point sources. The imaging system must also operate on objects that are close and have a non-negligible spatial spread, such that these algorithms can not always be applied in this problem. We therefore restrict the system to using beamforming algorithms which do not rely on the assumption of point-sources, such as the Capon beamformer (see Section 2.1.2). To obtain an image from a processed echo segment, we scan the environment on a hemisphere with a fine, 2D grid in the θ, ϕ -space and calculate the received power from each point in a specific range gate. To construct the 3D images, we generally need to decide which areas of the 2D images contain reflections. Generally, this involves searching for local peaks at different ranges and deciding based on some segmentation criterion how large the areas should be and whether the overall peak region is large enough to be considered significant and likely to be related to an object. While the images are obtained using beamforming which inevitable has a finite resolution, also the reflections themselves result in areas of monotonically decreasing power intensity. Thus, compared to traditional image processing problems derived for optical images, the regions of interest in the images can be found using relatively low-complexity segmentation methods based on adaptive thresholding and similar concepts. In this work we have used the EM algorithm for segmentation which is described in more detail in Section 5.2.

3.1.2 Assumptions and Basic Characteristics

In the following, we make some assumptions about the signal excitation and the physical conditions which are reasonable for acoustic imaging systems and the applications of interest:

1. The scene is illuminated by a narrow-band acoustic signal with center-frequency f_c and wavelength λ , emitted from a single acoustic sensor at a fixed position.
2. Echoes are recorded by an N -element dense array of isotropic acoustic sensors with uniform noise σ_n at each element.
3. The array operates in air, i.e. signals propagate in a homogeneous linear medium with constant propagation speed (as opposed to human tissue or water) [Boh88].
4. Objects closer than 1m are processed using a near-field signal model, such that the propagation of the sound echoes can be modeled using Fresnel's approximation. This follows directly from the wavelength of an acoustic signal in a range of 40 – 60kHz as well as the array aperture suitable for the application in a robot.
5. Additionally, the objects are assumed to have a solid surface, resulting in large acoustic impedance differences between air and the materials. This results in hard echoes from the objects. Almost no signal energy is lost due to diffusion into the object, i.e. the signal does not penetrate the object's surface.
6. Further, without loss of generality, we assume that the center of the array lies in the origin of the coordinate system.

When sound is reflected from massive, solid objects, there are three sources of reflected echoes that are visible to the array: As is illustrated in Figure 3.2, there is a direct reflection that occurs from power reflected orthogonally from a planar surface of the object in a specular way. Additionally, there can be ground reflections where the signal is reflected from the object to ground and vice versa. Clearly, the time-of-flight of such a reflection is longer due to the indirect path it takes. Moreover, such echoes will impinge on the array from a lower angle and, thus, appear to stem from a reflection below ground. However, to correct this, we simply have to change the sign of the height coordinate of those reflections. Clearly, this type of reflection does only occur when the object has reflecting areas close to the ground. Most importantly, all solid objects reflect sounds from their edges, where power is reflected as a superposition of spherical waves. In Figure 3.3 (b), we give a real data example where those three distinct regions are clearly visible. If the object surface is not smooth, the reflection process becomes more complicated. The echoes return not only from the three sources discussed above. Additionally, the sound wave will be reflected from many parts of the surface which are orthogonal to the direct path between the array and the surface. This leads to reflections which are spatially more diffuse (see Figure 3.3 (c)). According to acoustic theory, this can be modeled as the superposition of reflections from point sources which form the surface of the reflecting areas. This effect enables us to obtain information

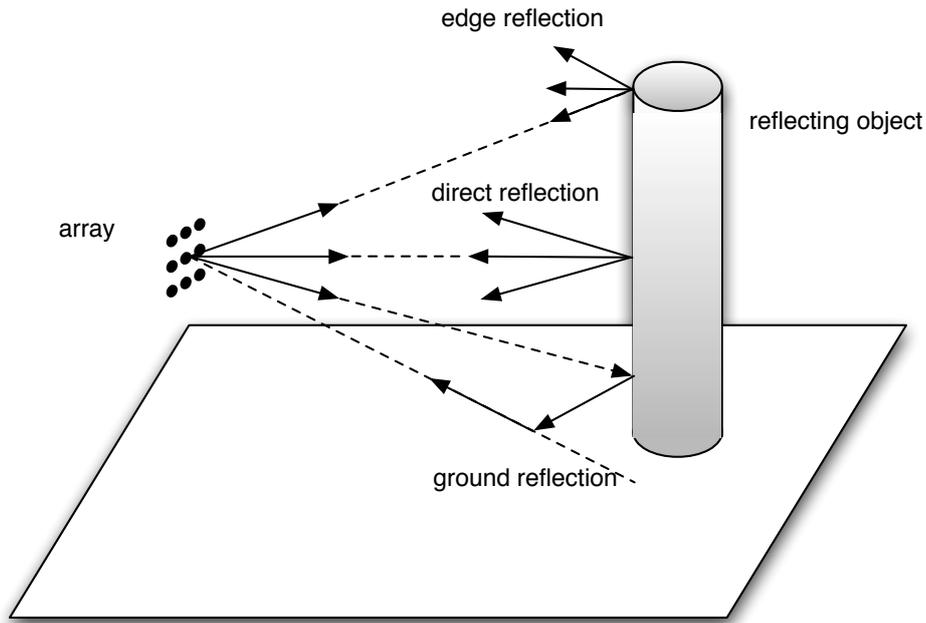


Figure 3.2: There are three sources of captured reflections from solid surfaces. Besides the direct orthogonal reflection from the object surface, the signal is reflected spherically at all edges of the object. Thus, the array will always receive edge reflections from objects. The third source of reflection are ground reflections. Due to their indirect path, they have a longer time-of-flight and the position of reflection has to be corrected.

about the texture of the object's surface as well as the discrimination between many artificial, man-made objects and natural objects, which typically do not have a smooth surface (see Chapter 5).

3.1.3 System Setup

To obtain real acoustic data, we have two systems available: First, a synthetic aperture approach which synthesizes any array geometry by a single receiver mounted on a high-precision 2D positioning system in the xz -plane. This implies that the environmental parameters such as object's position, temperature, etc. are stationary during the measurements to guarantee reproducibility of the experiments. This can safely be assumed to be true since the synthesis of an array does not exceed a time interval of a few minutes, even for large arrays. Both the fixed transmitter and the receiver are piezo-electric devices with a membrane of diameter 6.9 mm and a resonance frequency of $f_c = 48$ kHz. The transmitter is specified to have a beam pattern such that the 3dB-cutoff area is approximately 60° in azimuth and elevation. Its membrane is excited by a sinusoidal signal at frequency f_c with a duration of $100 \mu\text{s}$, resulting in a

narrow-band excitation signal of that frequency and a duration of 1 ms. The received analog signals at the array channels are band-limited before they are sampled at a rate of $f_s = 200$ kHz. The data is then demodulated to obtain the complex base-band signals. The described system is mainly useful for the validation of simulation results in the array geometry design. In addition to this synthetic aperture system, several array geometries were produced by our industry partner, both uniform dense arrays as well as nonuniform sparse arrays. They allow to capture a scene using a single excitation signal with a frame rate that is limited only by propagation time and hardware. For example, due to the propagation time necessary to record reflections from objects 10m away, the frame rate of a narrow-band system has an upper bound of

$$\text{fps}_{\max} = \frac{1}{2 \frac{10m}{c}} = \frac{343}{20s} \approx 17 \frac{1}{s}.$$

While this would be sufficient for real-time scene analysis, it requires fast hardware which can be a limiting factor in the design of imaging systems for cost-sensitive applications, e.g., for domestic robots.

3.1.4 Real Data Examples

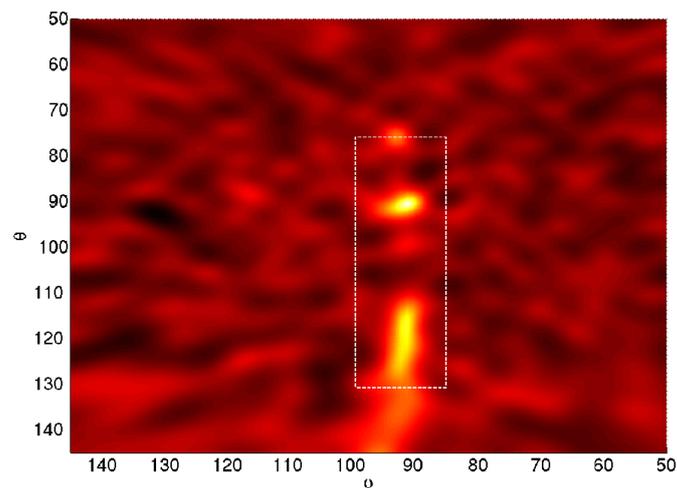
To demonstrate the nature of the obtained acoustic images, we show some real data examples of different objects. The images have been obtained using one of the real array prototypes. In Figure 3.3, we show a pole object and its corresponding acoustic images for different surfaces. The pole object is one of the reference objects of the Maneuvering Aids for Low-Speed Operation (MALSO) (ISO 17386:2010, [ISO10]). We can clearly see the three types of echoes distinctively in the image. To demonstrate the effect of different surfaces on the scattering process, we compare this to the same pole covered with bubble wrap, which results in a surface structured in a dimension comparable to λ . Only part of the smooth surface fully reflects the transmitted signal back to the array due to the specular nature of the scattering. While the general level of power is lower, reflections are observed from the whole object, since more regions reflect power back to the array. One can also observe that, in contrast to the smooth surface, the width of the pole is visible in ϕ -dimension, since reflections do not only occur on a small fraction of the curved surface but on the whole front.

To demonstrate the scattering behavior of a square-edged, artificial object, we show in Figure 3.4 a cuboid cardboard box placed in front of the array at a distance of 1.35 m. It was mounted on top of a pedestal which was covered with acoustic damping material. The front side of the box had dimensions that translate into an angular

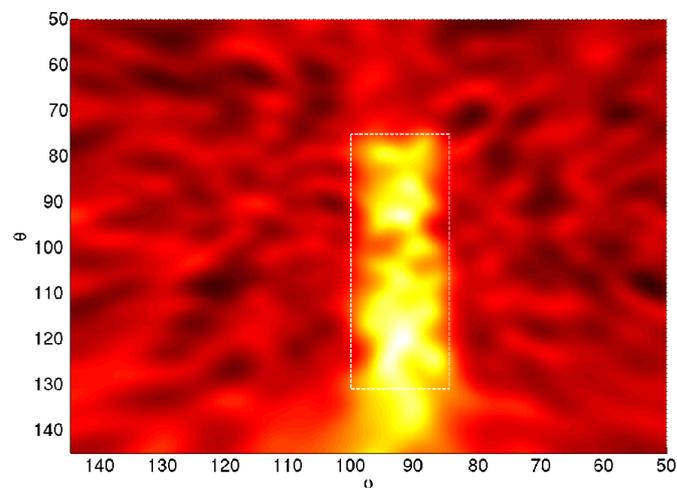
spread of $(\Delta\theta; \Delta\phi) = (13; 32)^\circ$ from the array's perspective. While the main peak is clearly the direct reflection from the front side of the box, one can also see echoes from the lateral edges as well as the bottom edge. The echo from the upper edge overlaps with the direct reflection. Echoes from the region $\theta > 115^\circ$ do not belong to the object, but are attenuated echoes from the pedestal. In Figure 3.5, we give an example of an acoustic image of a human standing in front of the array. It can be seen that the main reflective areas from the person are the head and torso. Depending on the pose and orientation, the shape of the torso echo will change while the head echo is less dependent on the exact pose. Additionally, arm and legs reflect echoes back to the array if the person moves them in a way that they have reflective areas orthogonal the direct line towards the array.



(a) Optical image of the reference pole.



(b) Acoustic image of the pole with a smooth surface.



(c) Acoustic image of the pole with a rough surface.

Figure 3.3: Example images of a reference pole object according to [ISO10]. The same object is shown with different surfaces. The acoustic images show the significant changes between specular scattering on a smooth surface and diffuse scattering from the rough surface.

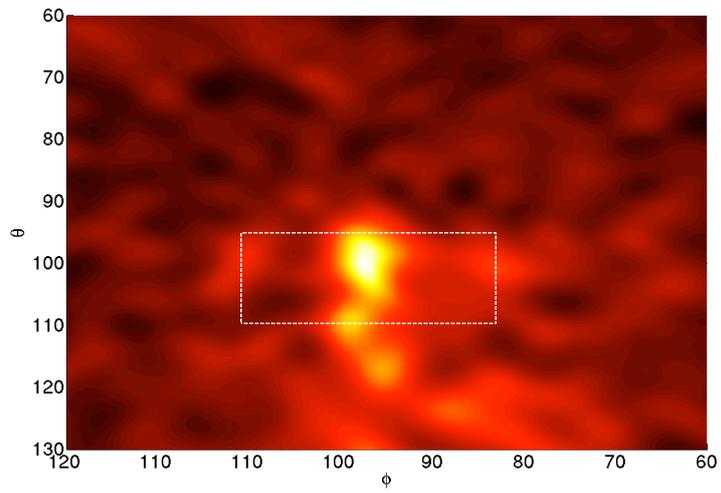
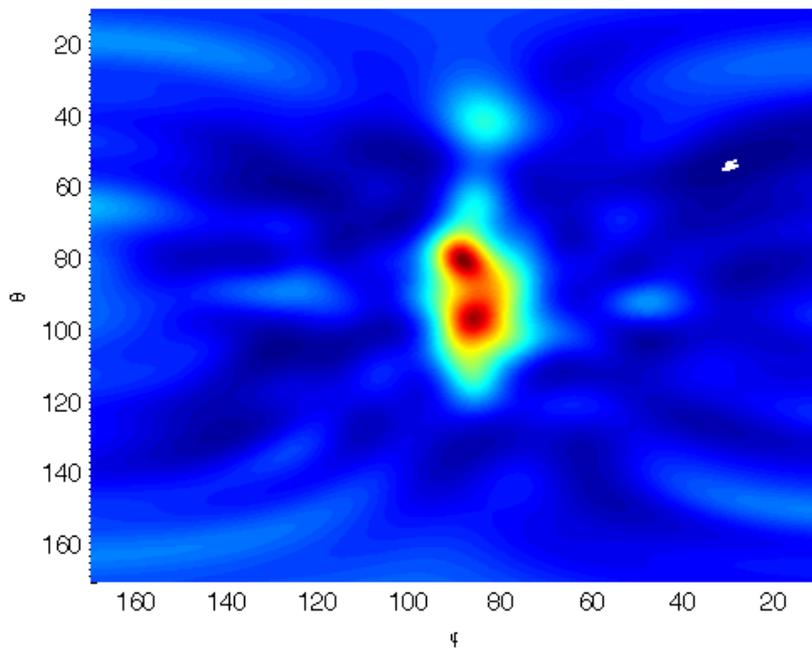


Figure 3.4: Image of a cuboid on a pedestal.



(a) Optical image



(b) Acoustic image

Figure 3.5: Example images of a human standing in front of the array. The person faces the array and stands still in relaxed pose.

3.2 Calibration Techniques

The performance of sensor arrays is well-known to be sensitive to errors or uncertainties in the model of the array manifold (see e.g. [VS94b, VS94a, SK92]). In practice, many of the array's characteristics can differ from their nominal values and are not precisely known *a priori*, e.g., due to imperfections in the hardware, manufacturing tolerances, mounting errors, etc. Thus, the sensor array is affected by gain and phase differences of the sensors, position errors and imbalances in receiver electronics that alter the array manifold. In order to account for these inevitable errors and production tolerances, every practical sensor array has to be calibrated such that the mismatch between assumptions and reality can be reduced. On the other hand, if the desired accuracy of the calibrated system is known, it may be possible to relax the tolerances required in production and compensate for that with the calibration procedure, potentially resulting in a manufacturing cost reduction. In this chapter, we will formulate the calibration problem and illustrate the common approaches in the literature. We will then describe a calibration method which is specifically suited for acoustic arrays and allows to compensate position errors with low complexity. We will present results obtained using simulations and real data measurements and discuss the performance in relation to other calibration methods.

3.2.1 Fundamentals

Following the standard signal model from Section 2.1.1, the output of the array can be described as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{S}(t) + \sigma_n \mathbf{I},$$

where $\mathbf{s}(t)$ describes the signal vector, \mathbf{A} is the array response matrix and $\sigma_n \mathbf{I}$ describes the uniform noise with power σ_n^2 at each sensor. Since \mathbf{A} models the spatial sensitivity of the overall array, the accuracy of the assumed model is crucial for the performance of any array signal processing techniques. The theoretical model displayed here differs from reality in that it assumes ideal isotropic, homogeneous sensor elements, perfect synchronization between the elements and no coupling effects. These assumptions are rarely true in practice due to manufacturing tolerances and imperfections. While it is possible to adjust the model in some aspects based on some nominal knowledge about the hardware, the specific properties of the single sensors vary and cannot be modeled accurately *a priori*. As a consequence, different sources of error arise when the nominal array response matrix is used. Some of these errors result in direction-independent offset errors, some in direction-dependent changes of the model. The true

array manifold $\mathbf{a}_t(\mathbf{k})$ is then unknown and has to be estimated. This is typically done using offline measurements with calibration sources. These sources send known signals from different directions. Based on the widely applied approach in [PK91], we model the deviations of the true array manifold from the nominal model by the use of a calibration matrix \mathbf{Q} such that

$$\mathbf{a}_t(\mathbf{k}) = \mathbf{Q}\mathbf{a}(\mathbf{k}) \quad . \quad (3.1)$$

Thus, finding the true array manifold can be reduced to finding a good estimate of \mathbf{Q} , based on measurements from D calibration sources from known directions (θ_d, ϕ_d) with $d = 1, \dots, D$. We denote the whole set of calibration angles by $(\theta_{\text{cal}}, \phi_{\text{cal}})$. This approach is termed *offline calibration* and serves as a tool to correct static errors in the array before its operation. Additionally, *online calibration* is based on the incoming signals during operation and can correct dynamic errors during operation, e.g., due to temperature changes. However, due to the lower amount of information typically available during online operation, these methods can not compensate larger errors. Depending on the types of present errors, \mathbf{Q} has to be modeled as a function of direction or can be constrained in its form, e.g., if no coupling is present, \mathbf{Q} is a diagonal matrix. If position errors are present, its coefficients are functions of the direction angles:

$$\mathbf{Q}(\mathbf{k}) = \mathbf{Q}(\theta, \phi) = \text{diag} \{q_1(\theta, \phi), \dots, q_N(\theta, \phi)\} \quad . \quad (3.2)$$

Under those assumptions, we seek the optimal $\mathbf{Q}(\theta, \phi)$ in order to solve the least-squares problem

$$\min_{\mathbf{Q}(\theta, \phi)} \epsilon = \|\mathbf{A}_t(\theta, \phi) - \mathbf{Q}(\theta, \phi)\mathbf{A}(\theta, \phi)\|_F \quad (3.3)$$

for all directions, where $\|\cdot\|_F$ stands for the Frobenius norm. Let us repeat here that the d th column of \mathbf{A} describes the array response vector for the d th source and is of the form

$$\mathbf{a}(\mathbf{k}_d, \boldsymbol{\psi}) = \begin{pmatrix} e^{j\mathbf{k}_d^T \mathbf{p}_1} \\ \vdots \\ e^{j\mathbf{k}_d^T \mathbf{p}_N} \end{pmatrix} \quad , \quad (3.4)$$

with $\boldsymbol{\psi} = \text{vec}\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$ denoting a vector with the stacked position vectors \mathbf{p}_n of all N sensors. The true array response for the calibration directions $\hat{\mathbf{A}}_t(\theta_{\text{cal}}, \phi_{\text{cal}})$ can be estimated by determining the principal eigenvector of the empirical covariance matrix $\hat{\mathbf{R}}$ of the data received from that single source [PK91].

In this work, we focus on offline calibration because the cheap production of the sensors introduces several errors which require a precise calibration. On the other hand, dynamic errors in acoustic imaging are not severely affecting the performance of the array and are neglected here. There exists a wealth of approaches to the problem of offline

calibration of a sensor array. They can be discriminated with respect to their assumptions about both the types of errors as well as their assumed effects on the array model. Methods which compensate only for direction-independent errors are *global* techniques, as they estimate \mathbf{Q} as single matrix valid for all directions. *Local* calibration methods, on the other hand, can take also direction-dependent errors into account, which results in a correction matrix that is also a function of direction. While both global and local methods directly estimate the calibration matrix, there are also approaches which try to estimate some parameters of the array, e.g., the sensor positions, parametrically based on some model [Vib10]. Those techniques can often be combined with local or global techniques.

3.2.2 Global Calibration Techniques

As just described above, global calibration approaches assume that all errors present in the array can be corrected by a *global* calibration matrix which is not a function of the angles θ and ϕ . Such techniques can account for gain and phase offset errors of the single elements, but, e.g., not for deviations of the single sensor beam-patterns. Additionally, it is mostly assumed in the literature that inter-element coupling is also direction-independent and, thus, can also be corrected (e.g. [PK91, DLX00]). Global methods estimate the calibration matrix by solving an optimization problem under the above assumptions as

$$\hat{\mathbf{Q}} = \arg \min_{\mathbf{Q}} \|\mathbf{Q}\mathbf{A}(\theta_{\text{cal}}, \phi_{\text{cal}}) - \hat{\mathbf{A}}_t(\theta_{\text{cal}}, \phi_{\text{cal}})\mathbf{\Xi}\|_F, \quad (3.5)$$

with $\mathbf{\Xi}$ being a complex weighting matrix. For $\mathbf{\Xi} = \mathbf{I}$, this leads to the solution [PK91]:

$$\hat{\mathbf{Q}} = \hat{\mathbf{A}}_t(\theta_{\text{cal}}, \phi_{\text{cal}})\mathbf{A}^\dagger(\theta_{\text{cal}}, \phi_{\text{cal}}), \quad (3.6)$$

where $\mathbf{A}^\dagger = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}$ is the Moore-Penrose pseudo-inverse of the ideal array response matrix \mathbf{A} . If no mutual coupling is present, this matrix will be diagonal and the solution is always unique. If coupling is affecting the sensors, the number of available calibration sources D must be larger than the number of elements N in order to obtain a unique solution for \mathbf{Q} . By choosing the weighting matrix $\mathbf{\Xi} \neq \mathbf{I}$, one can emphasize and possibly improve the classification for specific regions, see e.g. [LVL07]. An iterative procedure for the calculation of $\hat{\mathbf{Q}}$, that does not require matrix inversion has been given in [Hun00]. One general advantage of global calibration methods is that one can preprocess the array data with the inverse global calibration matrix before applying the beamforming algorithm. Moreover, only a single matrix is stored.

3.2.3 Local Calibration Techniques

Although global methods can compensate offset errors and compensate coupling between the single array elements, the techniques fail completely when direction-dependent errors are affecting the array. Thus, they are computationally attractive, but local methods are needed as soon as there is uncertainty about position errors or other direction dependent. The aim of local calibration is to find a correction matrix $\mathbf{Q}(\theta, \phi)$ which is a function of direction, such that

$$\mathbf{a}_t(\theta, \phi) = \mathbf{Q}(\theta, \phi)\mathbf{a}(\theta, \phi). \quad (3.7)$$

Typically, the calibration data is weighted and the weights depend on the direction of interest. For example, the algorithm proposed in [LLV06] uses the classical ideas of Pierre and Kaveh [PK91], but a direction dependent weight matrix is added which makes it possible to take direction dependent errors into account. The method does not assume prior knowledge about the errors, which makes it interesting in cases where the presence of errors is unknown and one has to assume that several types of direction dependent errors could be present. $\hat{\mathbf{Q}}(\theta, \phi)$ is calculated for every direction of interest, which means that the correction matrix is a function of the angle. The optimal $\hat{\mathbf{Q}}$ for direction (θ, ϕ) in a weighted least squares sense is given by:

$$\hat{\mathbf{Q}}(\theta, \phi) = \arg \min_{\mathbf{Q}} \|\hat{\mathbf{A}}_t((\theta, \phi)_{\text{cal}}) - \mathbf{Q}\mathbf{A}_i((\theta, \phi)_{\text{cal}})\mathbf{W}^{1/2}(\theta, \phi)\|_F,$$

where (θ, ϕ) is the current angle to be calibrated and $\mathbf{W}(\theta, \phi)$ is a diagonal weighting matrix for direction (θ, ϕ) with D non-zero elements. Since each correction matrix $\hat{\mathbf{Q}}(\theta, \phi)$ is calculated for a single direction, it is sufficient to model \mathbf{Q} as a diagonal matrix in order to capture any deviation from the nominal model.

The diagonal elements of $\hat{\mathbf{Q}}(\theta, \phi)$ can be calculated using the following equation:

$$\hat{q}_n(\theta, \phi) = \frac{\sum_{d=1}^D a_{nd}^* w_d(\theta, \phi) \hat{a}_{t,nd}}{\sum_{d=1}^D a_{nd}^* w_d(\theta, \phi) a_{nd}}, n = 1, \dots, N.$$

The weights $w_d(\theta, \phi)$ in \mathbf{W} should have the property that the calibration data from angles in the set $(\theta, \phi)_{\text{cal},j}$ which are close to (θ, ϕ) have strong weight and directions further away are less significant. Therefore, the chosen weighting function in [LLV06] is $w_j(\theta, \phi) = \exp(-h \cdot \|(\theta, \phi)_{\text{cal},j} - (\theta, \phi)\|^2)$. The bandwidth parameter h determines the amount of smoothing. A large bandwidth results in a narrow weighting function, giving low weight to more distant data. A too narrow weight function will give a worse interpolation between the calibration angles and less reduction of noise in the calibration data. On the other hand, choosing a small bandwidth will result in a DOA estimate

which is heavily influenced by the direction-dependent model errors. To choose a good value of h for a specific array, one can use various validation procedures, e.g. [LLV06] suggest to use a leave-one-out approach. The specific form of \mathbf{W} is not crucial for the performance of the calibration procedure and other forms of monotonically decreasing functions work well, too. Note that, while possible coupling effects present in the array should theoretically lead to direction-independent magnitude and phase errors, local methods implicitly assume diagonal calibration matrices and can not directly estimate coupling effects. As a consequence, local methods do inherently capture the effects of coupling in finding the transformation between theoretical and true array response matrices, but can not help to gain insight in the presence and structure of inter-element coupling, if needed. If such insight is needed while direction-dependent errors are present, one has to explicitly model an additional coupling matrix, estimated by global methods, together with the direction-dependent calibration matrix.

3.2.4 Parametric Maximum-Likelihood Estimation of Position Errors

In this section, we briefly describe an approach by [NS96] which uses a maximum likelihood estimation procedure for the true sensor positions stacked in $\boldsymbol{\psi}$ of an array with position errors and which motivated the development of our calibration method presented in the next section. The approach separates the problem into finding the true sensor positions and does not model position error effects in the calibration matrix, but directly as parameters of the array manifold. The calibration matrix \mathbf{Q} simply models the effects of other present errors identical to the general calibration problem formulation. The approach then alternates between the estimation of the true positions $\hat{\boldsymbol{\psi}}$ and estimation of the effects of all remaining errors in $\hat{\mathbf{Q}}$ until a convergence is reached. Thus, the approach couples the parametric approach with a global calibration. The MLE for the true position can be derived to be

$$\hat{\boldsymbol{\psi}} = \arg \min_{\boldsymbol{\psi}} \text{Tr}(\mathbf{P}_{\mathbf{A}^H(\boldsymbol{\psi})}^{\perp} \hat{\mathbf{A}}_t^H \hat{\mathbf{A}}_t), \quad (3.8)$$

with $\mathbf{P}_{\mathbf{A}^H(\boldsymbol{\psi})}^{\perp} = \mathbf{I} - \mathbf{A}^{\dagger} \mathbf{A}(\boldsymbol{\psi})$. The other errors present in the array are estimated by the least-squares expression in Eq. (3.6) based on the current estimate of the sensor positions. The final estimate is found by running a search procedure such as a damped Newton search while alternating between these two estimation steps until a convergence is reached.. This approach can also be combined with the previously mentioned local estimation, allowing an even more flexible correction of errors. However, to obtain a solution, a minimal number of mN calibration sources is required, with m being

the number of dimensions in which the positions are disturbed, e.g., if the positions are perturbed on a plane, $m = 2$. This constraint is of importance specifically for the calibration of 2D arrays where significantly more array elements are employed. Furthermore, the approach does not model the *effect* of position errors explicitly, it simply corrects the errors.

3.2.5 Proposed Low-complexity Estimation Procedure

The requirement of a minimal number of calibration sources is not of significance for small one-dimensional (1D) arrays, but can easily become perturbing for larger 2D arrays. To overcome these limitations, we present in this section an approach that explicitly models the phase effect of position errors on the array and does not require a minimal number of sources that depends on the array size. Assuming that each position vector \mathbf{p}_n is affected by an additive random vector $\boldsymbol{\rho}_n$, we can derive that the resulting additive phase error is proportional to the components of $\boldsymbol{\rho}$ because for any direction (θ, ϕ) we have

$$\mathbf{a}_t(\mathbf{k}) = \mathbf{a}_t(\theta, \phi) = \begin{pmatrix} e^{j\mathbf{k}_d^T(\mathbf{p}_1 + \boldsymbol{\rho}_1)} \\ \vdots \\ e^{j\mathbf{k}_d^T(\mathbf{p}_N + \boldsymbol{\rho}_N)} \end{pmatrix} = \mathbf{Q}(\theta, \phi)\mathbf{a}(\theta, \phi) \quad . \quad (3.9)$$

Thus, the position errors are included in the main diagonal of $\mathbf{Q}(\theta, \phi)$ and result in a phase shift

$$\angle q_n(\theta, \phi) = \mathbf{k}^T \boldsymbol{\rho}_n \quad (3.10)$$

for a signal with wave vector \mathbf{k} . Thus, using calibration sources from known directions (ϕ_d) , we can estimate the position errors under use of the coefficients q_n . The estimator for each sensor can then simply be obtained by a least-squares formulation as

$$\hat{\boldsymbol{\rho}}_n = \arg \min_{\boldsymbol{\rho}_n} \|\angle q_n - \mathbf{k}^T \boldsymbol{\rho}_n\|_F, \quad n = 1, \dots, N \quad . \quad (3.11)$$

Generally, the coefficients q_n can be estimated as (weighted) sum over all calibration sources (see [PK91, LLV06]):

$$\hat{q}_n = \frac{1}{D} \sum_{d=1}^D \frac{\hat{a}_t(\theta_d, \phi_d)}{a(\theta_d, \phi_d)}, \quad n = 1, \dots, N \quad . \quad (3.12)$$

In contrast, using eq. (3.12) in Eq. (3.11) removes the need for a minimal number of calibration sources and principally allows a position error estimate $\hat{\boldsymbol{\rho}}_n$ based already

on a few sources. For the simple case of a one-dimensional position error, e.g., if the sensors are only affected by position errors along the x -axis, Eq.(3.11) simplifies to

$$\rho_{x,n} = -\frac{1}{D} \sum_{d=1}^D \frac{\lambda}{2\pi u_x} \sphericalangle q_n(\theta_d, \phi_d), \quad n = 1, \dots, N. \quad (3.13)$$

This has the advantage that it is not only computationally much less complex than the method by [NS96], but also does not require any minimal number of calibration sources. If the true array manifold vectors $\mathbf{a}_t(\theta_d, \phi_d)$ can be estimated consistently, e.g., by the principal eigenvector of the calibration data covariance matrix, $\boldsymbol{\rho}_n$ will also be consistently estimated, assuming that no other direction dependent phase errors are present. The relation between $\boldsymbol{\rho}_n$ and $\sphericalangle q_n$ motivates also to estimate the phase errors nonparametrically using a polynomial fitting approach. Similar to [OV91, LLV06], we choose to model the coefficients $\hat{q}_n(\theta, \phi)$ over the whole domain of (θ, ϕ) , but fit $\sphericalangle q_n$ using a simplified, but still general approach in contrast to the exponential function proposed by [LLV06]. Motivated from the trigonometric functions that result from the position errors, we estimate the phase and magnitude of the coefficients by using a simple cubic spline interpolation. We found that using this approach, the calibration is more accurate and robust to a sparse calibration grid than using a specific functional form (see Section 3.2.6). This allows fitting \mathbf{Q} to the measured data from the calibration sources without the need of any smoothing parameter or the choice of specific basis functions. If the position errors are the only direction-dependent errors affecting the phase of the array manifold, the spline interpolation will correctly result in the estimation of a trigonometric function of the domain (θ, ϕ) or, equivalently, a linear function in the \mathbf{u} -space. This can also be combined with other parametric approaches to calibration, e.g., in the presence of coupling errors. Please note that the model-based approach assumes that there is no coupling present in the acoustic array. This is a reasonable assumption for acoustic imaging, since acoustic sensors receive sound signals without interference between each other. This stems from the fact that sound is a pressure wave and there is no electro-magnetic interaction between adjacent sensors by induction.

3.2.6 Results and discussion

In this section, we compare the proposed low-complexity parametric approach to other calibration methods in terms of its performance as a function of the number of calibration sources and as a function of the position errors. We ran 300 Monte Carlo simulations of a scenario where gain, phase and positions of the sensors were perturbed by random errors. The studied array is a 5×5 uniform rectangular array with position

errors uniformly distributed in a circle of radius $\|\boldsymbol{\rho}\| = 0.2\lambda$ on the aperture around the nominal positions. Additionally, the gain and phase of the sensors are affected by uniformly distributed random variables such that the true array response of the n th sensor $a_{t,n}$ can be expressed by the nominal response a_n , a random variable ζ representing magnitude errors, and another random variable η representing phase errors:

$$\begin{aligned} a_{t,n} &= (|a_n|(1 + \zeta))e^{j\angle(a_n)+\eta}, \\ \zeta &\sim \mathcal{U}(-0.15, 0.15), \\ \eta &\sim \mathcal{U}(0, \pi/6) \quad . \end{aligned} \tag{3.14}$$

Because we are interested only in acoustic arrays in this work, we assume that there is no coupling present. The calibration sources are placed uniformly across the grid, although this is not required. In Figure 3.6, the approach is compared to other methods for a varying number of available calibration sources. To vary the number of calibration sources, we uniformly increase the density of the calibration grid. Performance is measured by the resulting calibration error ϵ in decibel. It is illustrated that the global method by [Hun00] fails to model the present errors independent of the number of available sources and is only slightly correcting the nominal model as expected. The local methods perform better with an increasing number of calibration sources because they can model the array manifold more accurately in all directions the more dense the calibration grid becomes. While the approach by [LLV06] can correct the error partly, it performs worse than all other local methods. The computationally more complex method by [OV91] uses a rather generic polynomial functional to model the array manifold which allows a better calibration. Furthermore, we can see that fitting \mathbf{Q} using spline interpolation alone (denoted by "Q-space") already outperforms the other local calibration methods, but is worse than the method of [NS96] which is able to almost perfectly model the true array manifold if the condition on the minimal number of sources is fulfilled. When Q-space fitting is combined with parametric estimation as proposed (denoted by ParamLocal), we see that the performance is equal to the one of [NS96]. Additionally, it is capable to model the errors accurately even for only ten calibration sources because it does not require a minimal number of calibration sources. We want to emphasize that ϵ should not be interpreted as a measure for DOA estimation accuracy in array problems. However, it is meaningful to compare the methods in terms of this correction error as it is derived from the initial problem formulation. Similar results are obtained when the methods are compared with respect to the co-linearity between the true and calibrated array response. The order in performance is equal, however, the gap between the parametric and nonparametric approaches is less visible. In Figure 3.7, the performance of the different algorithms is evaluated when the effect of the position errors is varied from 0.01λ to 0.2λ . We see that both the proposed method and the method of [NS96] are able to correct the positions in every

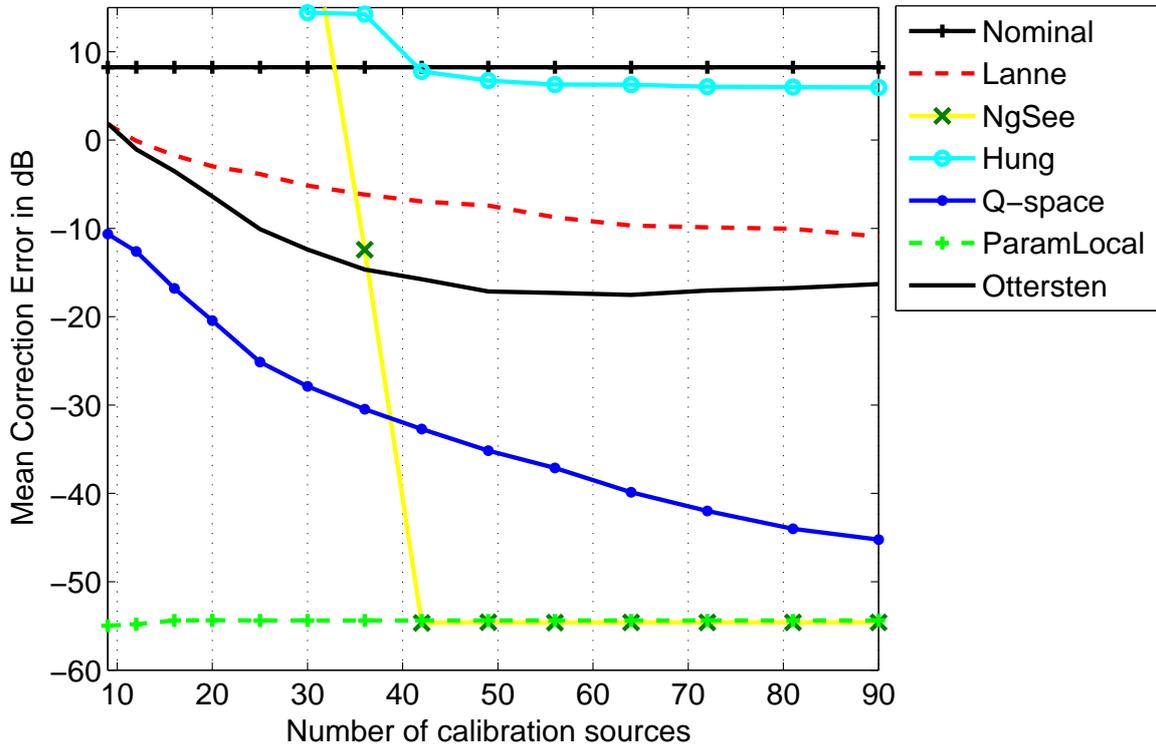


Figure 3.6: The calibration error is depicted vs. the number of calibration sources used. It can be seen that the global method fails to correct the errors due to wrong assumptions about the errors. The local methods all perform better with an increasing number of sources. The parametric method by [NS96] fails completely if the number of available sources is smaller than the minimal one. Our proposed method does not need this restriction and still compensates for the direction-dependent errors.

case and perform constantly well, while the performance of the other methods decreases with increasing position errors. This demonstrates that a calibration procedure that explicitly models position errors allows to decouple their effect on the array manifold completely from other errors. Nonparametric local methods can not compensate for the effect once the error is too large and the global method clearly fails, which is not surprising because it assumes that only direction-independent errors are present and is applied in an error scenario it is not designed for. Figure 3.8 shows the acoustic images of an acoustic array using a Capon beamformer with and without application of the calibration procedure. The array uses 30 omnidirectional acoustic receivers which were affected by gain, phase and position errors. The recorded scene consisted of four point sources located at $\theta = 136^\circ, 79^\circ, 69^\circ, 43^\circ$ and $\phi = 65^\circ, 81^\circ, 94^\circ, 68^\circ$, respectively. It is clearly visible that the calibration procedure improves the resolution of the array. Using the uncalibrated array, the present errors result in errors of the location estimation of the point sources, especially for the two sources which are located far off-broadside.

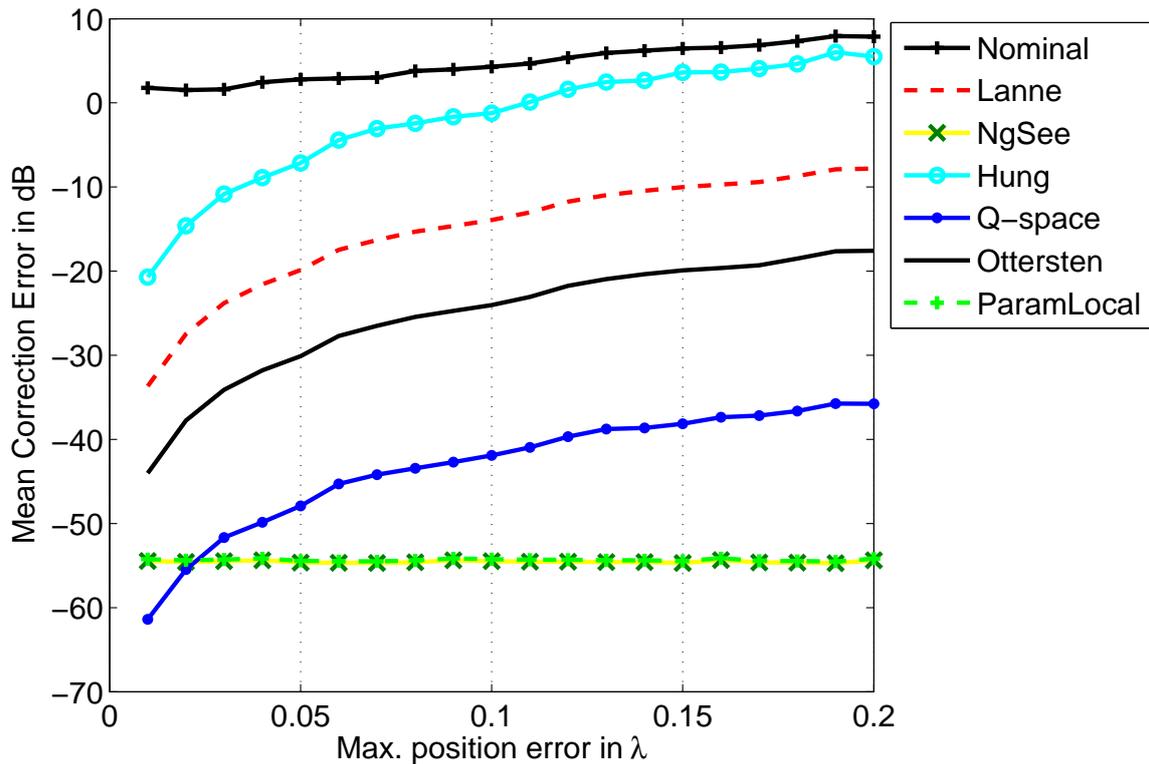
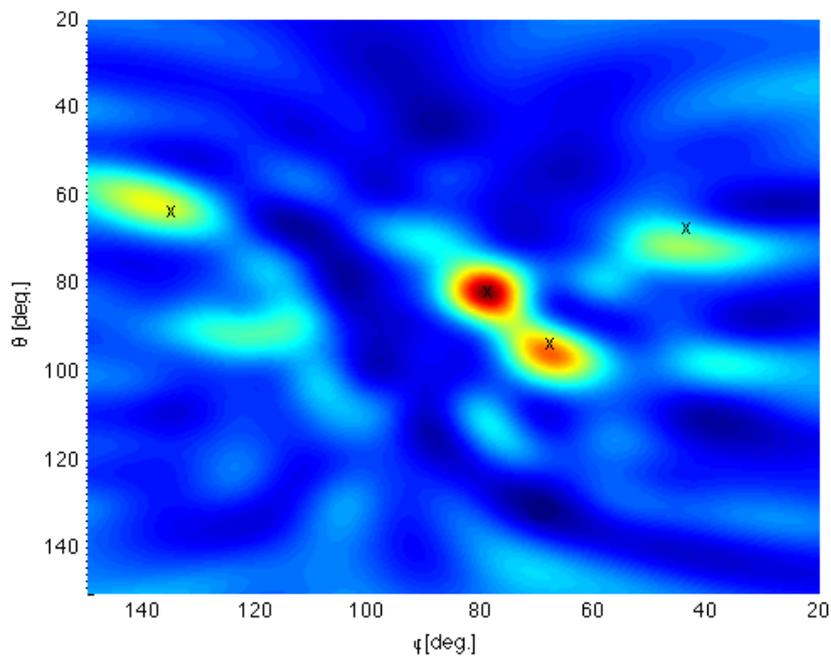


Figure 3.7: The mean correction error is depicted as a function of the scale of position errors. All methods except the parametric ones suffer from an increase of the error.

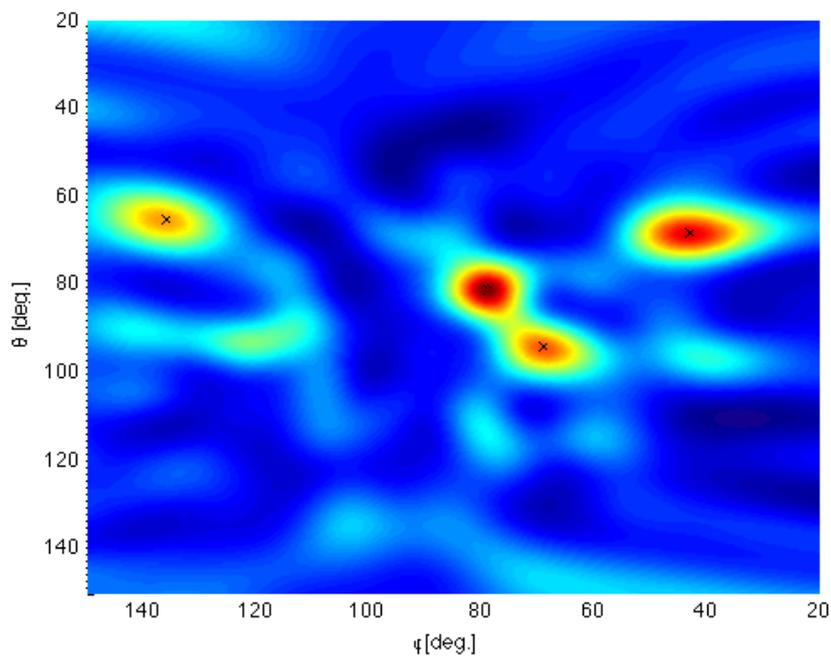
This can be compensated by the calibration, resulting in an improved location estimate of the point sources as well as more pronounced peaks in the acoustic image due to increased resolution. In Figure 3.9, we see another example of a human standing in broadside and facing of the array. It is visible that in the uncalibrated case, for a large number of reflectors, the reflected echoes are not clearly focused and it is hard to distinguish between main echoes and small scatterers. Again, also in this example, the calibration has the effect of increasing the focus and peak power of the reflectors. Additionally, clutter reflections in the background are reduced and the image allows a better interpretation of the depicted scene.

A drawback of combining a parametric estimator with Q-space fitting is its failure to accurately model the array manifold if there are other direction-dependent phase errors that are more dominant than the position errors. However, this is also valid for the approach of [NS96], that does not model other direction-dependent errors than position errors. In such a situation, the motivation for a parametric model for the position errors is reduced and a nonparametric local methods will perform better, unless one can model the other sources of direction-dependent errors. However, one can also use the model and the calibration results to analyze the impact of position errors, because

of its previously derived known effect on the array response. Due to the trigonometric relation wave vector and phase effect of the position errors, the interpolation between measured calibration sources leads to trigonometric functions with respect to direction for all array elements affected by a position error. As described above, the amplitude of these trigonometric functions is proportional to the magnitude of the error.

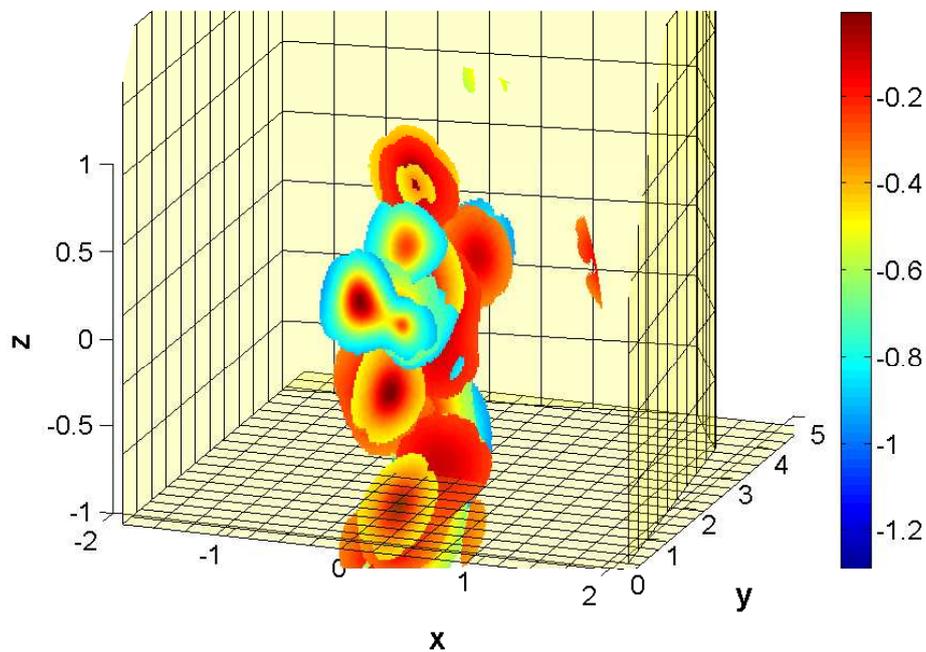


(a) Uncalibrated array with gain, phase and position errors

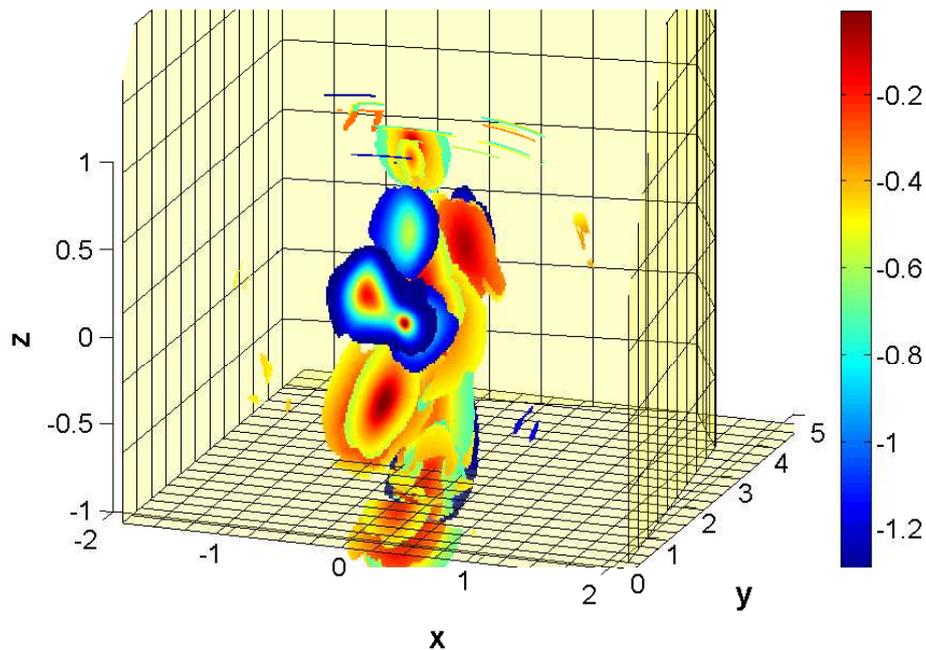


(b) Calibrated array

Figure 3.8: Comparison of the acoustic images in a scenario with four point sources. The echo peaks are more concentrated after calibration, thus objects can be better resolved. Also the position of the peaks is closer to the true positions.



(a) Uncalibrated array with gain, phase and position errors



(b) Calibrated array

Figure 3.9: A human standing in front of the array, facing the array. After calibration, the echoes from the object are more clearly visible and are more concentrated. Additionally, the noise suppression has been improved.

Chapter 4

Sparse Array Design

4.1 Introduction

As discussed in Chapter 1, acoustic imaging for robotic applications requires cheap hardware, but still demands for good spatial resolution and noise suppression. Especially the use of 2D arrays for 3D imaging is challenging, e.g. the high number of sensors and therefore active channels employed in such systems, leads to increased hardware and computational costs. This motivates the use of 2D *sparse* arrays which approximate the beam-pattern of a fully populated array using a lower number of sensors. However, the geometry of the array has to be optimized, as it fundamentally influences all performance parameters. There have been numerous approaches to design sparse 1D and 2D arrays using single- and multi-objective optimization techniques where Half-Power-Beam-Width (HPBW) and Side-Lobe-Level (SLL) are optimized. Common approaches design sparse arrays by either *thinning* a full array [AH02, AH02, HAIH01, Tru99, AHW⁺97, KS91, Kop92] or randomly placing a fixed number of sensors, sometimes under some geometrical constraints [JRS07, OM05, Ruf93, RL75]. These approaches allow to obtain good results, but rely on the evaluation of a cost function over a large search space and involve subjective weighting of the individual goals. To find good solutions, these methods require the creation of a large number of possible candidates from which one has to be selected. This allows little or no control over the properties of a found solution. In this chapter, we discuss design methods for sparse 2D arrays that are suitable for acoustic imaging applications, e.g., obstacle detection in robotics, or medical imaging diagnosis. Due to the slow wave propagation speed as well as hardware constraints, systems operating in air transmit in broadcast mode and therefore require a receiver beam-pattern without grating lobes.¹ The approach presented in this chapter enables the design of highly sparse 2D arrays that exhibit low SLLs. Additionally, direct control over the trade-off between sparsity and SLL is achieved. The approach is based on the theory of minimum redundancy and employs an iterative procedure that populates the array aperture with additional sensor elements. The criterion for the placement of additional sensor is formulated such that the co-array is manipulated to allow a good approximation of

¹In active systems that employ both transmit and receive beamforming, e.g. in medical imaging, it is common to optimize the two-way-beam-pattern of the array, i.e. grating lobes are allowed for the transmission beam-pattern and are suppressed in receiving mode by controlled nulling (see e.g. [LTB98]).

the performance of a full array. This allows to choose a suitable combination of SLL and sparsity while exhibiting a low computational complexity. These advantages come at the cost of an altered mainlobe shape which can become quasi-convex.

4.2 Problem Formulation

Based on the signal model presented in Section 2.1.1, we formulate the sparse array design problem in the context of acoustic imaging as the problem of finding a sparse array with κ elements that performs well in terms of object resolution and noise suppression. The reference of performance is the beam-pattern of a full Uniform Rectangular Array with $N = N_x \times N_z$ elements with $\lambda/2$ -spacing. Our goal is to reduce the number of sensor elements in this 2D planar array while retaining the resolution and suppression capabilities of the corresponding URA with respect to λ , such that the spatial power spectrum estimate is still consistent². In other terms, we want to approximate the beam-pattern of an URA by the use of a nonuniform sparse array with only $\kappa < N$ elements on the same aperture for all possible steering angles \mathbf{k}_l . For the imaging applications of interest, we measure performance in terms of the SLL and mainlobe width of the beam-pattern of a Bartlett beamformer as well as the sparsity of the used array, measured by the thinning factor κ/N . The SLL is defined as the distance between the maximum value in the mainlobe of the beam-pattern and the second-largest maximum value. Additionally, grating lobes have to be avoided by all cost as they lead to spurious peaks in the resulting images. Note that while the Integrated-SideLobe-Ratio (ISLR) is used for evaluation in some applications, we do not use it here because it is less important in the application of interest. This is due to the fact that we deal with situations where there are only relatively few scatterers simultaneously present in the image. Reducing κ can be interpreted as a measure of compressive spatial 2D sampling. In this context, we can say that we are interested to find the transformation function $\mathbf{\Gamma}(\cdot)$ which transforms the beam-pattern of the URA into the beam-pattern of the sparse array such that certain properties of the image obtained by the full, dense array are the desired parameters to be reliably estimated using the sparse sampling. For example, using Bartlett's beamformer, the beam-pattern of the sparse array looking at direction \mathbf{k}_l is

$$B_S(\mathbf{k})|_{\mathbf{k}_l} = \mathbf{w}_S^H(\mathbf{k}_l) \cdot \mathbf{a}_S(\mathbf{k}) \quad (4.1)$$

$$= \frac{1}{\kappa} \sum_{i=1}^{\kappa} a_i(\mathbf{k}_l)^H a_{S,i}(\mathbf{k}), \quad (4.2)$$

²The consistency of power spectra based on nonuniform sampling is proved in [Mar86, MC90].

where $\mathbf{w}_s, \mathbf{a}_s$ are the κ -element weighting vector and array manifold vector, respectively. The power-pattern of the sparse array is equivalent to the spatial nonuniform Fourier Transform of the data. Then, we can also express the beam-pattern in look direction \mathbf{k}_l of the sparse array $B_S(\mathbf{k})|_{\mathbf{k}_l}$ as a transformation of the beam-pattern of the corresponding full array $B_U(\mathbf{k})|_{\mathbf{k}_l}$

$$B_S(\mathbf{k})|_{\mathbf{k}_l} = \Gamma(B_U(\mathbf{k})|_{\mathbf{k}_l}) \quad (4.3)$$

$$= \Gamma(\mathbf{w}_U^H(\mathbf{k}_l)\mathbf{a}_U(\mathbf{k})), \quad (4.4)$$

where $\mathbf{a}_U(\mathbf{k})$ is the array manifold vector of the uniform array. Note that in the sequel, and without loss of generality of the following results, we assume a broad-side look direction, i.e. $(\theta, \phi) = (90^\circ, 90^\circ)$. Finding the transformation function is equivalent to finding a binary selection matrix \mathbf{J} with dimensions $\kappa \times N$ and $\text{rank}(\mathbf{J}) = \kappa$. It reduces the URA to a sparse array such that the array manifold vector of the sparse array is

$$\mathbf{a}_S(\mathbf{k}) = \mathbf{J}\mathbf{a}_U(\mathbf{k}). \quad (4.5)$$

4.3 Minimum Redundancy Theory

4.3.1 Fundamental Concept

To introduce the fundamental concept of minimal-redundancy theory, we will first consider a one-dimensional ULA with $\kappa = N_x$ elements at $x_1 = 0, x_2 = 1, \dots, x_N = N_x - 1$. Every integer up to $N_x - 1$ can be represented as a difference of two sensor elements $\nu = x_i - x_j$. The concept of minimum redundancy affords to analyze an array not in terms of its actual sensor positions, but their *differences* in position, referred to as the coarray [BV98]. If some differences occur multiple times, this will be reflected in redundancy in \mathbf{R}_{XX} and therefore determines important characteristics of the array, e.g., resolution and directivity.

Due to the Toeplitz structure of the covariance matrix of the ULA, the estimation of specific correlation lags is based on several phase differences in the array [Van02, ASG99a]. The estimation accuracy depends on the number of samples of this correlation lag and is denoted as spatial sensitivity (see Figure 4.1) [Mof68, LST93, Lin92]. In a *Minimum-Redundancy-Array* (MRA), the array is spanned on the same aperture using only the minimum $\kappa < N_x$ elements such that each integer difference ν occurs at least once, but as few times as possible. If we set

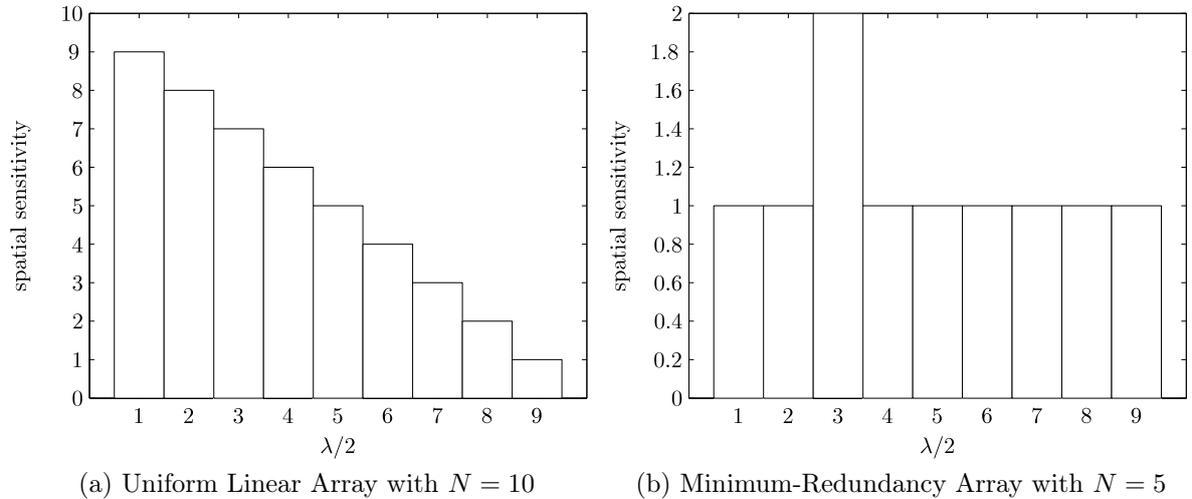


Figure 4.1: Example of the spatial sensitivities of two arrays with $L = 9$. (a) shows the high redundancy in a Uniform Linear Array, (b) the Minimum-Redundancy Array where redundancy is present except for the double occurrence of the distance of $3\lambda/2$.

$x_1 = 0, x_\kappa = N_x - 1$, the set of integers $x_1, x_2, \dots, x_\kappa$ forms a *restricted difference basis* [Lee56] with respect to $N_x - 1$, if

$$\nu = x_i - x_j \quad \forall 1 \leq i, j \leq \kappa \quad (4.6)$$

can be satisfied for all $0 < \nu \leq N_x - 1$. Such an array exhibits a low number of elements to span the aperture with the guarantee that no grating lobes are introduced in the beam pattern for any look direction. Additionally, it shows the same HPBW as the corresponding ULA on the same aperture. Due to the introduced gaps in the geometry, the SLL is considerably higher than for an ULA. The redundancy is measured as [Mof68]

$$R = \frac{\kappa(\kappa - 1)}{2(N_x - 1)}, \quad (4.7)$$

where $\frac{\kappa(\kappa-1)}{2}$ is the number of possible pairs between κ elements and $(N_x - 1)$ is the length of the ULA. Upper and lower bounds for R can be found in [Lee56, Bed86] for 1D arrays. It is also interesting to note from those difference sets, one can create cyclic difference sets on a larger scale [Kop92] and that the average beam-pattern of all cyclic difference set arrays is equivalent to the average of all $\binom{N}{k}$ possible array geometries [Lee99]. For DOA estimation problems, there exist approaches which augment the resulting covariance matrices such that they exhibit Toeplitz structure corresponding to a virtual array with a steering vector in Vandermonde structure. Using such virtual arrays, one can obtain a gain in DOA estimation accuracy compared to the smaller uniform arrays with the same number of elements [PBNH85, AGGS98, ASG99b, ASG99a].

4.3.2 Two-Dimensional Difference Sets

MRAs have been studied extensively in the 1D-case and optimal configurations have been found also for large κ [PPL90]. However, no optimal solution for 2D arrays is known yet. In fact, not even a criterion of optimality has been found so far [Kop92]. Efforts to create 2D sparse arrays have been made based on the concept of Two-Dimensional Difference Sets (TDSs) (e.g. [KS91, Lee99]). They can be obtained in several ways, e.g., by the multiplication of two orthogonal 1D restricted difference sets (see Figure 4.2 and refer to [Kop92] for other possibilities). Such a TDS with parameters $(N_x, N_z, \kappa, \Lambda)$ is defined as a set $\text{TDS} = \{(x_1, z_1), \dots, (x_\kappa, z_\kappa)\}$ of κ elements on a (N_x, N_z) integer grid. Position coordinates (x, z) are represented exactly Λ times as [Kop92]

$$\begin{aligned} x &= x_i - x_j \pmod{N_x}, 1 \leq i \leq N_x \\ z &= z_i - z_j \pmod{N_z}, 1 \leq j \leq N_z \end{aligned} \quad (4.8)$$

Similarly to the 1D case, a κ -element, *low*-redundancy array is created, if the array positions are uniquely represented by a TDS with parameters $(N_x, N_z, \kappa, 1)$. In [MD01], bounds for sparse 2D arrays are derived based on the 1D redundancy measure R and it is proven that this approach is asymptotically efficient, meaning that the lower redundancy bound is reached for $N_x, N_z \rightarrow \infty$. Note that while the first sidelobe of such low-redundancy arrays is high, the second sidelobe is mostly at approximately the same level than the sidelobes of the corresponding uniform array. Additionally, the spatial sensitivity is distributed on the same multiples of $\frac{\lambda}{2}$.

4.4 Forward Inclusion Approach

In this section, we address the problem of sparse array design using a forward inclusion approach based on minimum-redundancy theory in contrast to traditional thinning approaches. We propose to separate the design process of the sparse 2D array into two steps: First, we span an initial 2D array on an $N_x \times N_z$ lattice in a low-redundancy fashion using a $\text{TDS}(N_x, N_z, \kappa, 1)$. This is followed by an iterative procedure that includes additional sensor elements at specific points of the underlying lattice. The initial array can be represented by a position function

$$\mathbf{M}(x, z) = \sum_{i=1}^{\kappa} \delta(x - x_i, z - z_i) \quad (4.9)$$

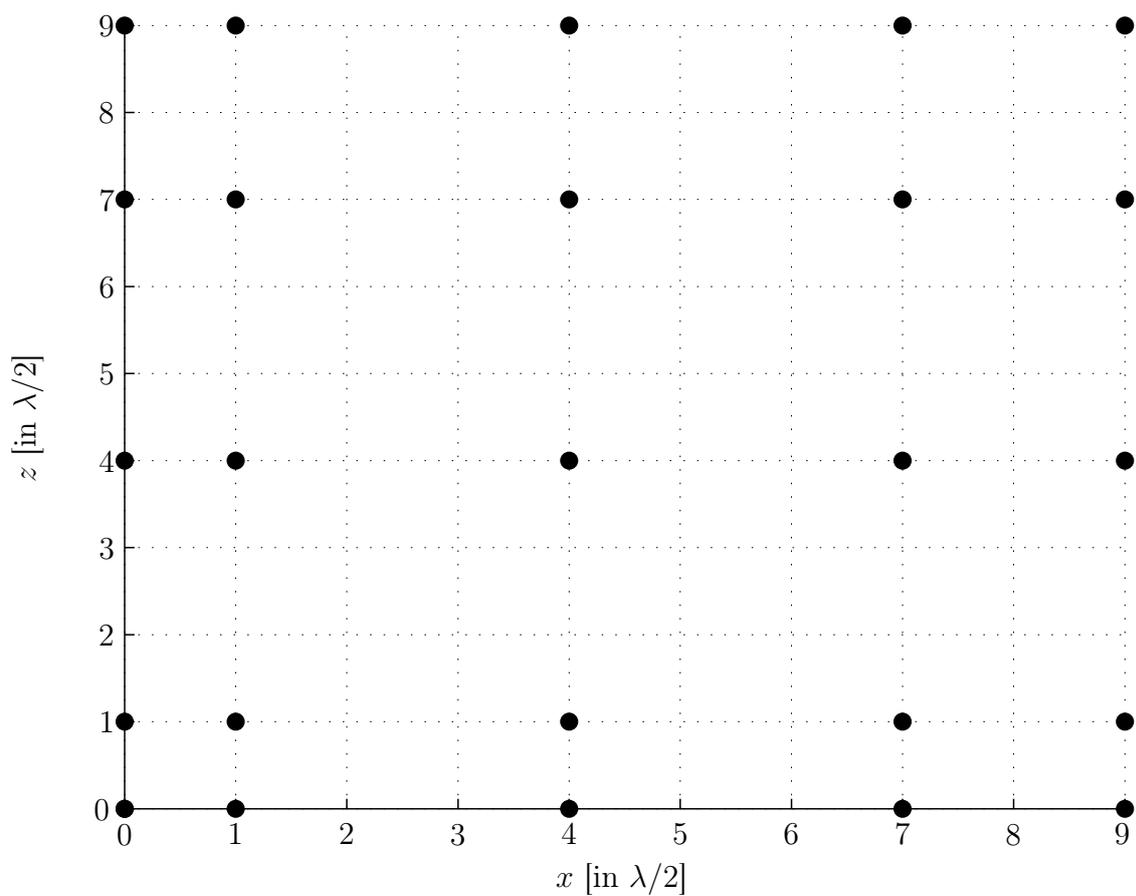


Figure 4.2: An initial array spanned by a Two-Dimensional Difference Set for a (10×10) -aperture.

which represents the distribution of elements on the aperture of the initial array in the (x, z) -plane (see Figure 4.2). It exhibits approximately the same mainlobe-width as the corresponding URA on the same aperture, using only a low number κ of elements.

After this initial step, the array is extended iteratively in order to fulfill a combination of the remaining design goals, namely high sparsity and a low SLL of the array. While it is clear that additional sensors will improve the overall sidelobe behavior of the array, we want to maximize the impact on the SLL for each additional sensor in order to obtain a good trade-off between the design goals. In [AGGS96], it is shown that increasing simply the number of continuous correlation lags does not necessarily increase DOA estimation accuracy. Thus, we choose a different approach which increases the spatial sensitivity for the smallest lag in the co-array. To achieve this, we have to increase the occurrence of the *smallest* sensor distances in the distribution of sensor elements. It is assumed and verified by simulation that this leads to low sidelobes. This assumption is derived from the thought that if we aim to increase the number of occurrences of the smallest distances in an array maximally, the effect on the co-array is most effective and adds maximally to approximate the co-array of the dense array. We can measure this effect by the local sensor density $\mathbf{s}(x, z)$, which can be easily obtained by

$$\mathbf{s}(x, z) = \mathbf{M}(x, z) * \mathbf{W}(x, z) \quad (4.10)$$

where $\mathbf{W}(x, z)$ is a 2D cuboid kernel function.³ Its bandwidth has to be chosen according to the overall aperture dimensions (see Section 4.6.2). By adding elements iteratively such that they increase uniformity maximally, we create a uniform subarray which grows with every iteration by a single element and, thus, varies in shape. However, this subarray exhibits a significantly lower SLL than the overall array. Due to the nature of TDSs, $\mathbf{s}(x, z)$ will exhibit a global maximum which serves as a reference point

$$(x_r, z_r) = \arg \max_{x, z} (\mathbf{s}(x, z)) \quad (4.11)$$

for the element inclusion. To determine the position of the additional element, we calculate the potential *global* sensor density which results if the element is placed at (x_i, z_i) . It is defined as

$$\mathbf{S}(x_i, z_i) = \sum_{n=1}^{\kappa} \mathbf{s}(x_n, z_n)|_{x_i, z_i}, \quad (4.12)$$

where $\mathbf{s}(x, z)|_{x_i, z_i}$ denotes the local density assuming an element in position (x_i, z_i) . Depending on the aperture and the initial array geometry, there might be several

³Although other kernel shapes are possible, we found that the resulting array geometries show poor sparsity for all aperture sizes. A uni-modal function with tapering edges seems to be inappropriate to model uniformity in $\mathbf{s}(x, z)$ and $\mathbf{S}(x_i, z_i)$, respectively.

candidate positions

$$(x_i, z_i)_c = \arg \max_i \mathbf{S}(x_i, z_i) \quad (4.13)$$

leading to a maximal $\mathbf{S}(x_i, z_i)$. In such circumstances, we choose the one $(x_i, z_i)_c$ which is closest to (x_r, z_r) , i.e.,

$$\min(\|(x_i, z_i)_c - (x_r, z_r)\|) \quad (4.14)$$

such that the array exhibits only one uniform subarray. This maximizes the contribution of the i th element in each iteration to maximally increase the spatial sensitivity of the overall array, independently of the aperture shape or the used initial TDS. We evaluate the array with respect to its SLL in the broad-side Bartlett beam-pattern and κ in each iteration. The desirable trade-off between these design goals can be determined by the formulation of a suitable stop criterion, e.g., a targeted SLL, a maximal number of elements or a combination thereof. In Table 1, we summarize the algorithm.

- | |
|---|
| Step 0: Create an initial array based on a TDS($N_x, N_z, \kappa, 1$). |
| Step 1: Determine the reference position (x_r, z_r) . |
| Step 2: Determine candidate positions $(x_i, z_i)_c$ as the maxima of $\mathbf{S}(x_i, z_i)$. |
| Step 3: Insert additional array element at :
$\min(\ (x_i, z_i)_c - (x_r, z_r)\)$. |
| Step 4: Evaluate the broad-side Bartlett beam-pattern of the array. |
| Step 5: Repeat steps 2-4 until the stop criterion is fulfilled. |

Table 4.1: Summary of the proposed algorithm.

In Figure 4.3, it is illustrated how the SLL is reduced by iteratively adding sensor elements to a TDS(18, 18, 50, 1) on a rectangular lattice. The increased aperture of the subarray leads to suppression of the first sidelobe and shapes the mainlobe towards the shape of the URA counterpart. For $\kappa = 95$, the suppression of the first four sidelobes closest to the mainlobe is sufficiently high such that they are transformed into a part of the mainlobe, which becomes non-convex, but monotonic at this point. The depicted sudden drop in SLL is therefore not due to a than a sudden change of a single sidelobe, but due to the change of the peak sidelobe location, because the previous sidelobes merge with the mainlobe, meaning that the ISLR changes gradually at this iteration. The mainlobe is convex in the upper region and now shows a widening pedestal where it merged with the previously present sidelobes. Note that the second sidelobe of a MRA is mostly close to the SLL of the corresponding uniform array.

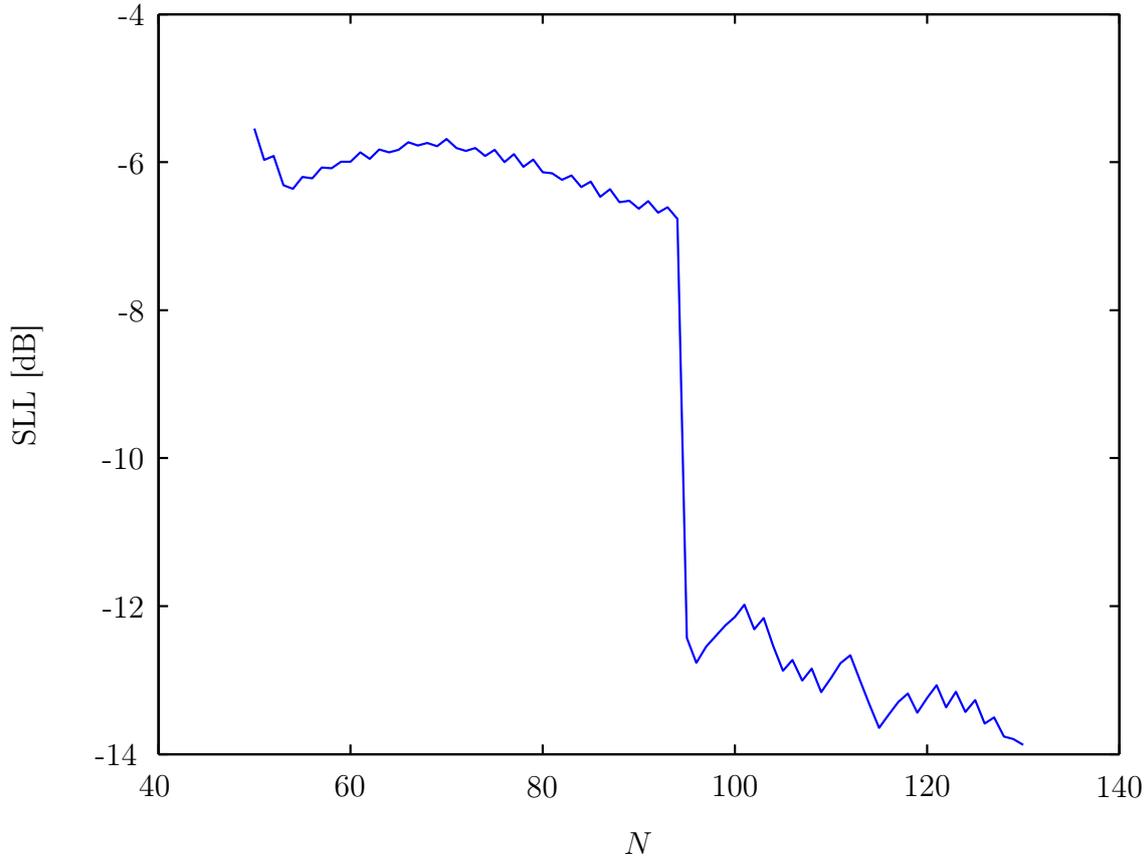


Figure 4.3: Development of the SLL of an array designed according to the proposed algorithm. The number of elements κ is iteratively increased and new elements are placed based on a rectangular lattice.

Thus, by suppressing the first set of sidelobes, the improved SLL will also be close to the SLL of the corresponding URA and a distinct change in SLL is created. The resulting two-dimensional beam-pattern is depicted in Figure 4.4. This enables us to identify the threshold where the first sidelobe merges into the main lobe which is critical with respect to source detection (and therefore also object detection), as the beam-pattern now is uni-modal in that region. We therefore can formulate a stop criterion with respect to the occurrence of this transformation. However, since the mainlobe is not *strictly* monotonic anymore, we can alternatively formulate a stop criterion that allows for an additional amount of iterations to compensate for further smoothing of the main lobe shape. Using this approach, we can design highly sparse arrays with the same HPBW as the corresponding URAs and a SLL which is close to the results obtained in [Kop08]. This comes at the cost of a quasi-convex mainlobe which is widened below the HPBW. The degree of widening depends on the number of additional elements one is willing to place in the aperture.

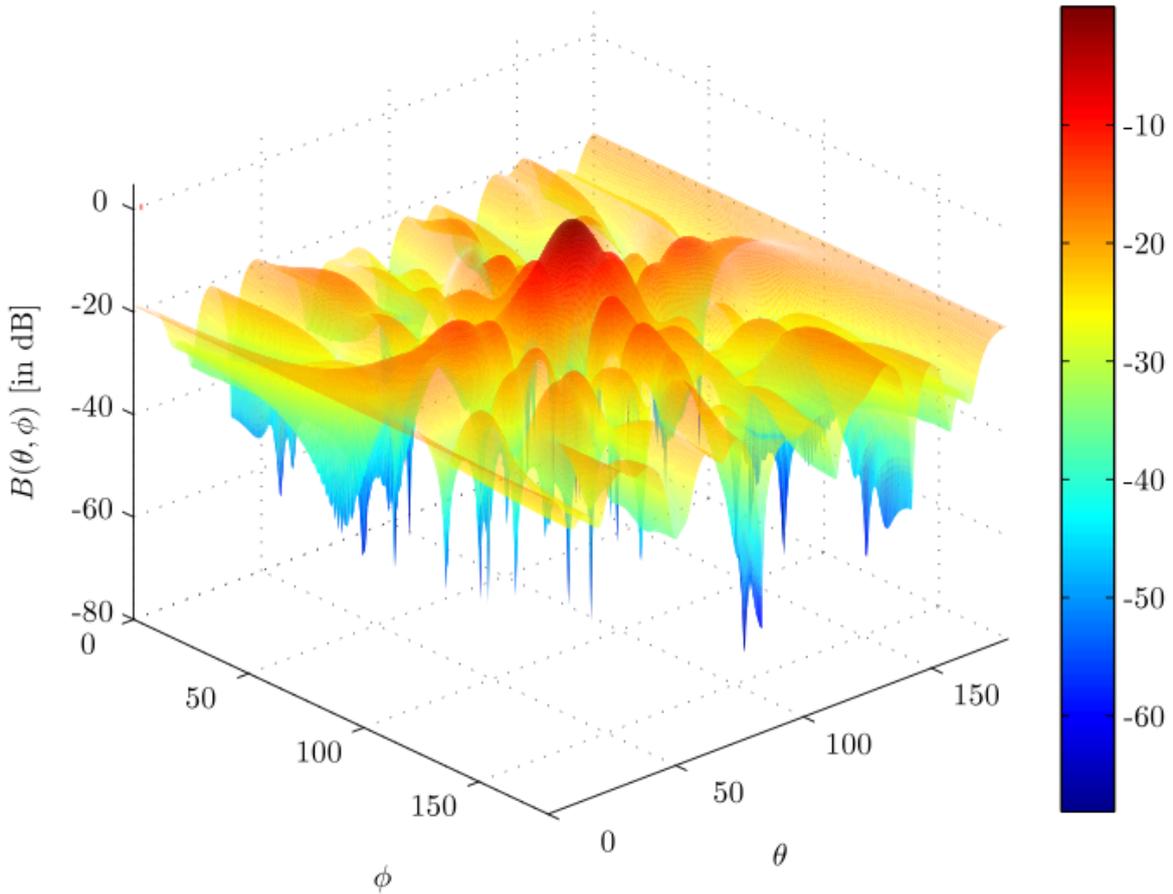


Figure 4.4: 2D-beam-pattern of a designed nonuniform array after suppression of the first sidelobes.

4.5 Lattice Structure

To control how elements are placed on the aperture, one can use different lattice structures. Besides a rectangular structure, which allows the potential use of any element of the corresponding URA, structures such as hexagonal or quincunx lattices allow the design of arrays based on other grid symmetries, potentially with even higher sparsity. In particular, using a quincunx lattice has the advantage that impinging signals are sampled across the whole aperture with the same correlation lags in both directions than with the URA, but require only $\kappa = \frac{N_x N_z}{2} \pm \frac{N_x N_z \pmod{2}}{2}$ sensor elements. Alternatively, one can interpret a quincunx pattern as a rectangular grid which is scaled by a factor of $\sqrt{2}$ and rotated by 45 degrees. For a broadside look-direction, the beam-pattern of a quincunx array is equivalent to the one of an URA. However, when looking off-broadside, the effective distance between adjacent sensor elements is larger than $\frac{\lambda}{2}$ and grating lobes in the beam-pattern occur. We will not go into details of hexagonal

sampling because we found similar performance behavior of the designed arrays for hexagonal lattice structures.

4.5.1 Randomization of Lattices

The main reason for high sidelobes in the beam-patterns of the nonuniform arrays stems from the fact that they are placed on regular grid structures. It is a well-known fact that random placement of sensor elements on the array aperture allows to obtain low sidelobe levels (e.g. [Lo64, Lo63, Ste72]). The price for this behavior is that sidelobes are positioned in an irregular fashion and there is no control over the properties of the beam-pattern possible, such that suitable array layouts can only be obtained by Monte-Carlo simulations. However, a binned randomization of elements around lattice nodes allows to achieve lower sidelobes and allows at the same time to a structured approach to the problem [HAIH01, Hol00, Hen91]. In order to maintain control over the array layout, the extent of randomization has to be small, such that the beam-pattern is only altered in sidelobe structure, but is still closely related to the beam-pattern obtained using strictly lattice nodes. This can be achieved by adding a multivariate random variable with limited support to the position vector of each sensor element such that the randomized position of the i th element is described by

$$\tilde{\mathbf{p}} = \mathbf{p}_i + \boldsymbol{\rho}_i \quad , 1 \leq i \leq \kappa \quad .$$

We assume the random variables to be independent and uniformly distributed on the interval $[-\frac{\lambda}{4}, \frac{\lambda}{4}]$ in the x, z -dimensions. Furthermore, $\boldsymbol{\rho}_i$ is assumed to be deterministic zero in the second dimension, meaning that the sensors are only moved on the aperture surface. In Figure 4.5, we show an example of an array geometry on a (10×10) -aperture with a quincunx grid and binned randomization after several iterations of the proposed algorithm. The added sensors are restricted to lie inside the depicted boxes around the grid nodes as described above. The empty circles denote empty nodes in the quincunx grid which can be used for further iterations by the algorithm. The effect of binned randomization on the SLL is illustrated exemplarily in Figure 4.6 where the expected value and the 2, 5%- and 97.5%-percentiles of distribution of the SLL are shown based on 300 Monte Carlo runs. It can be seen that the average SLL decreases continuously, thus showing the same behavior as in the deterministic case. This is due to the fact that the random variables are distributed with zero mean around the nominal sensor positions. The variance, however increases as expected because more sensors are added to the aperture and the variance of the overall beam-pattern of a binned random array scales with the number of sensors as

$$\text{Var}\{B_S(\mathbf{k})\} = \kappa \left(1 - \text{sinc}^2\left(\frac{N_x k_x}{\kappa 2\pi}\right) \right) \left(1 - \text{sinc}^2\left(\frac{N_z k_z}{\kappa 2\pi}\right) \right) ,$$

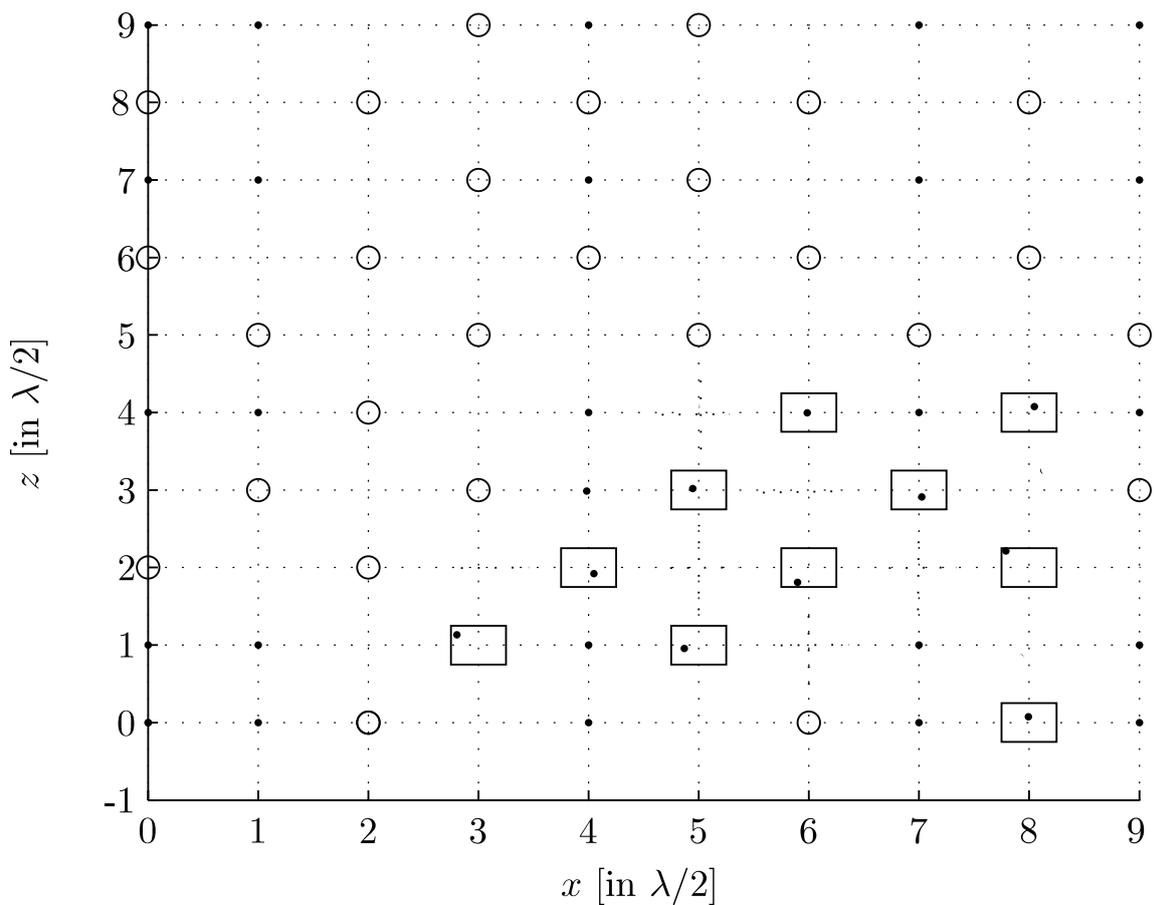


Figure 4.5: Example of an array geometry on a (10×10) -aperture after several iterations of the proposed algorithm. New elements can only be placed around free nodes of a quincunx grid (denoted by circles) and positions are placed using binned randomization. The rectangles denote the bins in which elements have been placed around the grid nodes in previous iterations.

which can be easily derived analogously to the variance of a 1D binned random array (see [Ste72, HAIH01]). Please also note that due to the randomization, there is no distinct change in SLL and it changes only gradually. This is due to the fact that array elements do not align in a regular fashion anymore and their distances are not exact multiples of $\frac{\lambda}{2}$. Thus, the previously described suppression of the first sidelobe into the mainlobe does not occur at a single iteration, but depends on the realizations of the random variable ρ .

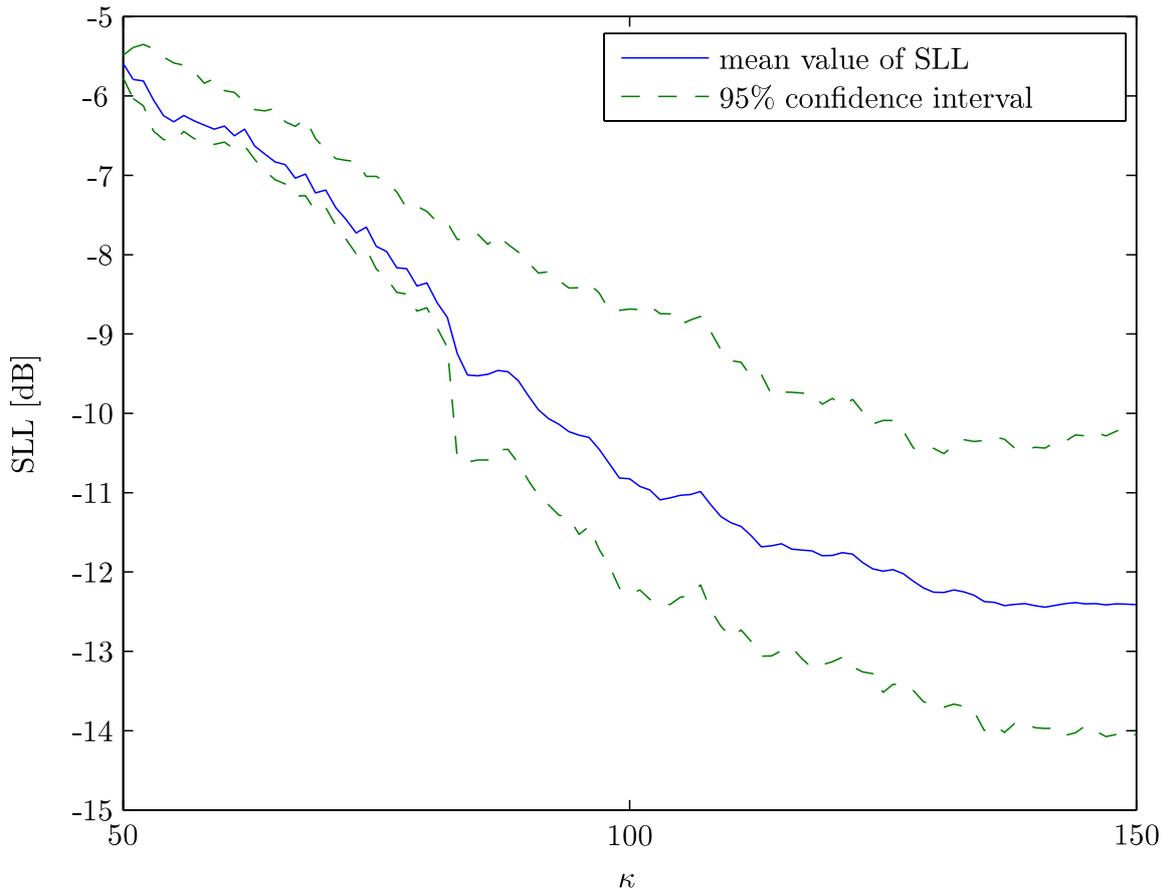


Figure 4.6: Expected value and variance of the SLL using binned randomization.

4.6 Experimental Results

4.6.1 Simulations

The algorithm proposed in Section 4.4 has been applied to various lattice structures and various aperture sizes. We compare the results to a recent approach proposed

in [Kop08] which is also based on different sets. The approach uses a special family of difference sets, called Hadamard sets (H-sets). A library of H-sets is created using cyclic shifting and automorphic transformations of an initial set. From these H-sets, array geometries are constructed and their beam-patterns are evaluated by a search procedure in order to search for solutions which result in a low SLL. In order to compare the results, we choose quadratic apertures, although this is not required by our algorithm. In Table 4.2, the performance of the algorithm is summarized. For results employing binned randomization, we present the mean value and the 95% confidence bounds which have been obtained using Monte-Carlo simulations with 300 runs. When binned randomization is applied, this steep decline of SLL is not present anymore. As mentioned in section 4.5.1, the addition of a random variable on the element positions results in a more irregular sidelobe structure where the first sidelobes are not necessarily the peak sidelobes. As a result, we cannot determine a single iteration where the first sidelobes merge into the mainlobe. Instead, we do now observe a continuous trade-off between high sparsity and low SLL. This is effectively a Pareto-front as it is also created by multi-objective optimization algorithms, though it requires only a low computational complexity due to the simple iterative procedure. Due to the higher restrictions on the quincunx grid, the uniform subarray grows faster on those grids. This leads to the fact that the first sidelobes are suppressed using even fewer elements and result in higher sparsity in the array, if the algorithm is stopped using the same stop criterion. However, due to the possible existence of grating lobes, the randomized quincunx lattices are favorable. As it is depicted in Table 4.2, they exhibit even higher sparsity than the layouts based on rectangular lattices while giving a continuous trade-off between the number of sensors and the resulting SLL.

In Figure 4.7, the sparsity of the resulting arrays, denoted by TDSX, using a rectangular lattice are compared with results from [Kop08], denoted Kop., both in terms of the absolute number of elements κ and the relative Filling-Rate (FR) which is given by $\kappa/(N_X N_Z)$. A change of the SLL of more than 3dB was used as a stop criterion. We see that while κ increases with aperture size for both methods, the FR using our approach is significantly lower, i.e. the arrays exhibit a much higher sparsity, although the SLL is only around 2dB higher.

In Figure 4.8, the mean SLL of a (30×30) -aperture is depicted when rectangular or quincunx lattices are used. It is clearly visible that the quincunx lattice allows to merge the first sidelobe much earlier into the mainlobe such that it can achieve a comparable SLL with a significantly higher sparsity. Although this behavior is less distinct for small apertures, quincunx lattices allow generally a lower SLL for smaller values of κ . However, depending on aperture size, kernel bandwidth and variance of

the binned randomization, we found that the algorithm performed slightly better using rectangular lattices in some configurations.

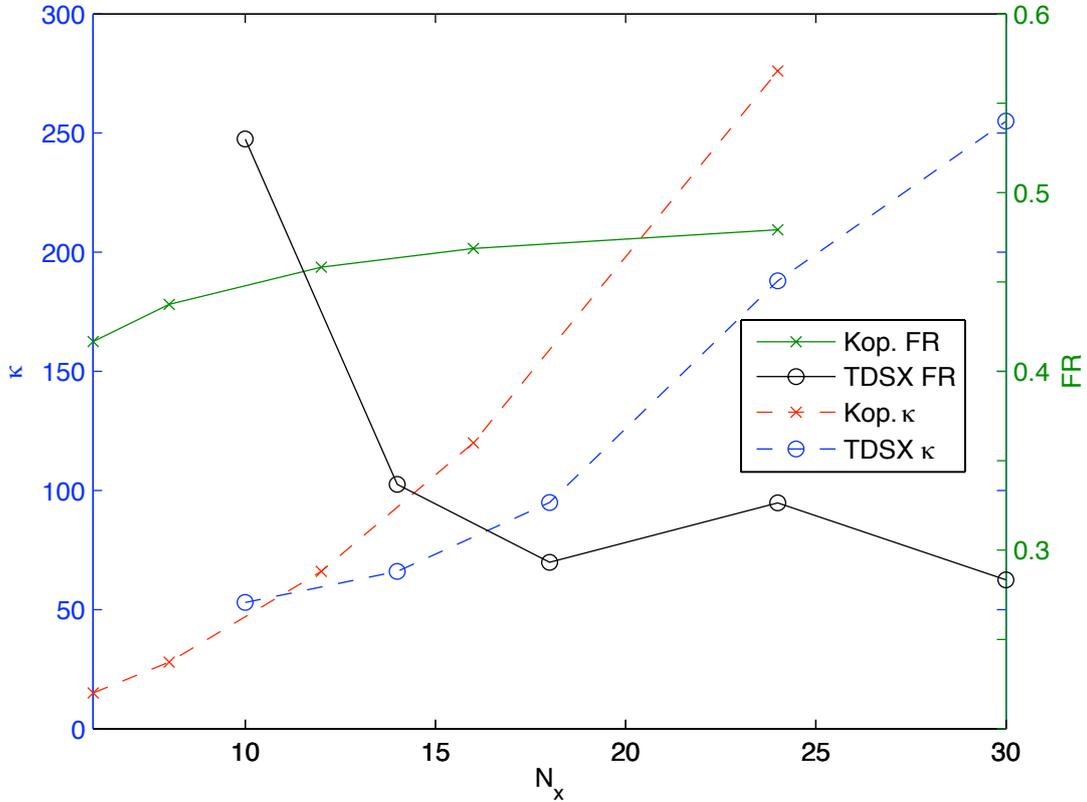


Figure 4.7: Comparison of sparsity of our proposed approach to a design method based on Hadamard matrices for different quadratic apertures (edge length N_x). The left ordinate shows the absolute number of elements κ , the right ordinate the aperture-dependent Filling-Rate (FR).

4.6.2 Kernel Bandwidth

Previously, a cuboid kernel function $\mathbf{W}(x, z)$ was applied using a quadratic base of width $h = 3$ and unity height. Effectively, this weighs all direct neighboring elements around the current position, but is independent of the overall array geometry. With increasing aperture dimensions, the sparsity of the initial as well as the augmented array is higher. Thus, $\mathbf{S}(x_i, z_i)$ will contain less information about the distribution of array elements if the kernel bandwidth is not chosen adaptively. Choosing a larger bandwidth h therefore further improves the performance of the algorithm in terms

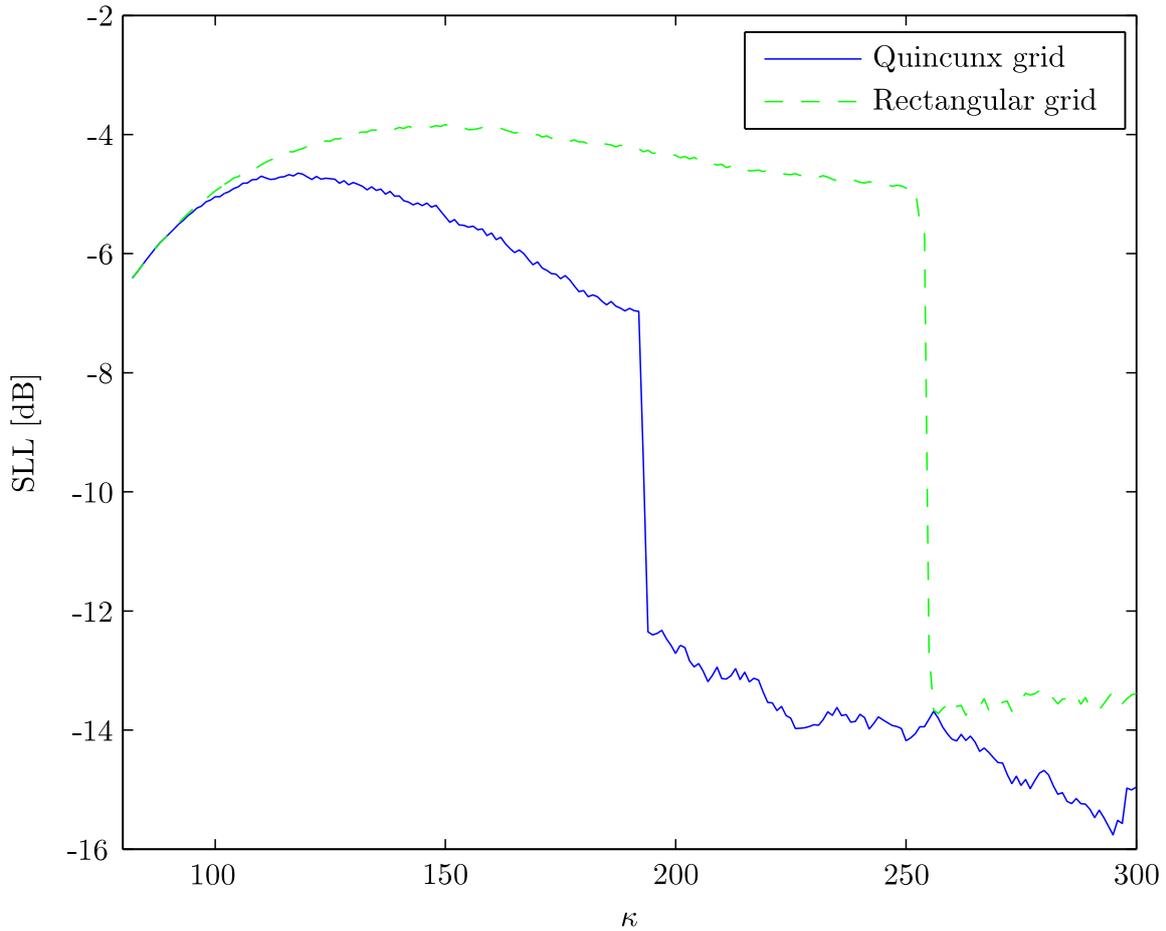


Figure 4.8: Comparison of rectangular and quincunx lattices with respect to sidelobe suppression on a (30×30) -aperture.

of the achieved sparsity, but, as with any kernel function, may also lead to loss of information, if set too large.

Figure 4.9 shows the sparsity of the resulting square array geometries for different values of h over edge length N_x . A rectangular lattice has been applied and the procedure was stopped if the SLL changed more than 3dB. We see that for $h > 3$ sparsity is improved for $N_x > 14$, but does not perform well for small apertures. On the other hand, for $h = 11$, the algorithm only converges for $N_x = 30$, but shows the best sparsity there (19.3%). This testifies that the optimal kernel bandwidth depends on the aperture size and, if set too large, can mislead the element placing such that the first sidelobe is not suppressed.

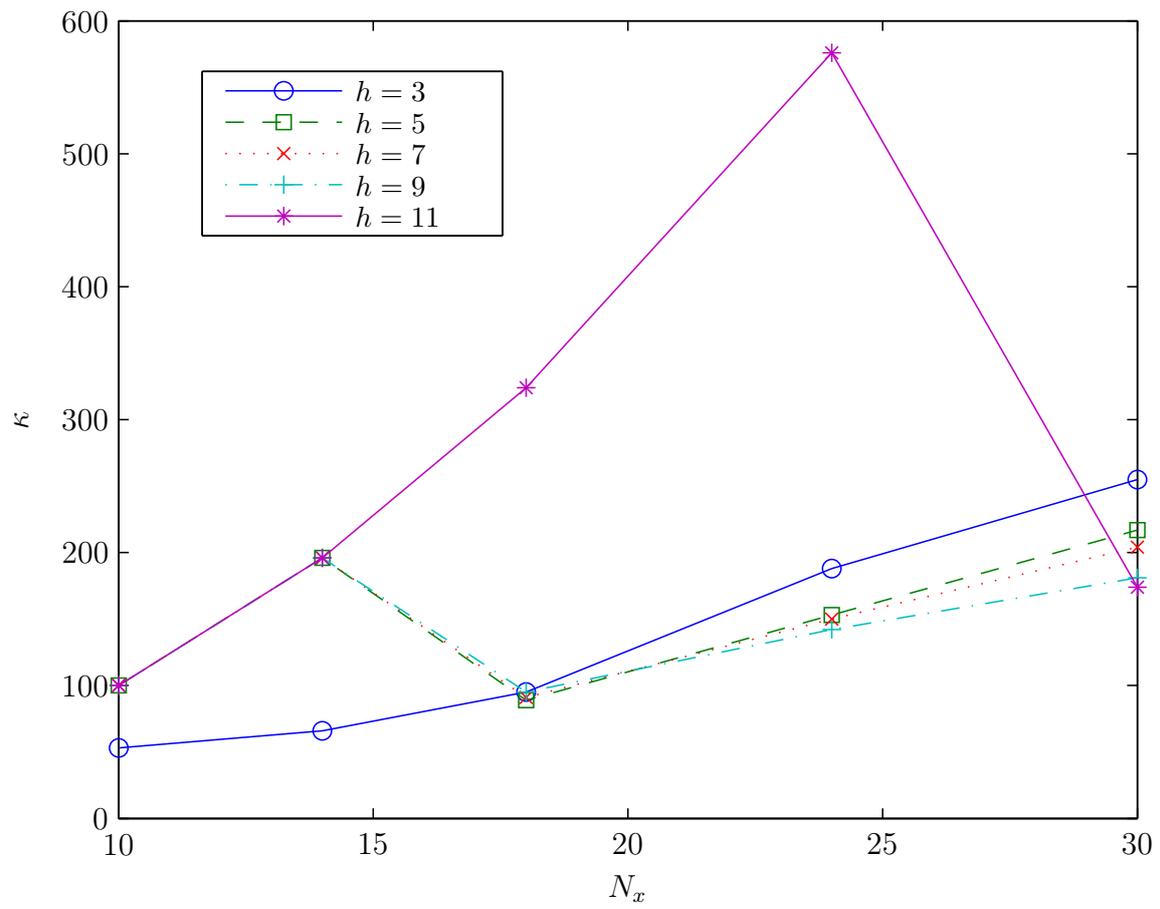


Figure 4.9: Effect of Kernel bandwidth on performance for various square apertures.

4.6.3 Acoustic Imaging

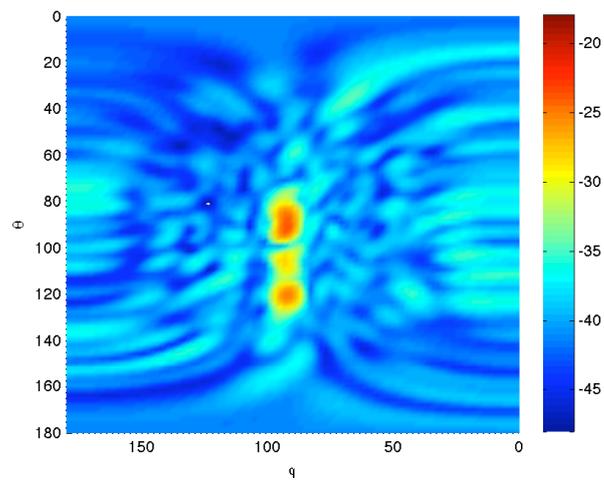
Following the simulation results, we show here how an acoustic imaging system as described in Section 3.1 performs if it is based on sparse array layouts. The images in this section are created by applying the array layouts obtained using the proposed methods to select which data channels were selected from a full 18×18 -array. The objects are illuminated by a narrow-band excitation signal emitted around $f_c = 50$ kHz, i.e., $\lambda \approx 7$ mm. The transmitter is specified to have a beam pattern such that the 3dB-cutoff area is approximately 60° in azimuth and elevation. The reflected echoes are captured by a receiving (18×18)-array of acoustic, omnidirectional sensors (see also Section 3.1 and [MZ07b]). The echoes scattered back from a large pole-object with a rough surface standing in front of the array. The different array layouts are tested by computing the power pattern using only a subset of the full array's sensors. We first compare them based on their performance using Bartlett's beamformer in order to illustrate changes that stem directly from the sparsity. However, since Bartlett beamforming suffers generally from high sidelobes and results in speckle noise in acoustic imaging, the sparse array layouts are also evaluated using Capon's adaptive beamformer, which is the method that is actually used in the imaging system. Figure 4.10 shows the received power over azimuth angle ϕ and elevation angle θ for several array layouts using Bartlett's beamformer. We compare an array with (16×16) -aperture with 120 elements based on Hadamard difference-sets according to [Kop08], a full (18×18) -URA with 324 elements and an array with 95 elements according to the proposed technique based on a rectangular (18×18) -lattice.⁴ It can be seen that all three layouts result in images where the contour of the pole object is visible. We also see how the regular sidelobe pattern of the beam-pattern of the URA results in regular artifacts in the image around the object (Figure 4.10 (b)). Both sparse layouts (Figure 4.10 (a) and (c)) clearly show more severe and more irregular image artifacts due to the sparsity in the arrays. Additionally, the image in Figure 4.10 (c) shows higher sensitivity in end-fire directions. However, since the the imaging system is not operated in end-fire direction, the differences in those regions are not a matter of great interest. In Figure 4.11, we applied Capon's beamformer to the same layouts and data, meaning that the weight vector depends on the signal scenario and information from the covariance matrix estimate is exploited, thus increasing the resolution and noise suppression in the images. Clearly, the image from the full array (Figure 4.11 (b)) shows the best contrast and also the most uniform power reception from the object's surface. In Figure 4.11 (a) and (c), it is shown that the sparsity leads to a slightly reduced contrast of the images. Additionally, both sparse layouts result in a less uniform power reception from the

⁴Unfortunately, we are not able to depict images using layouts based on randomization because the available array data stems from a uniform array.

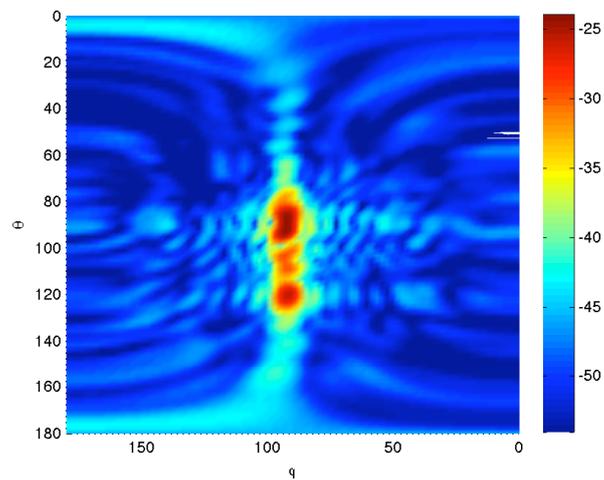
object, which could also be observed using Bartlett's beamformer. The main difference between the two sparse array geometries is the altered reception of the object itself, an artifact that probably stems from the uneven sensor density across the aperture. Note that although the mainlobe of Bartlett's beamformer is quasi-convex for this array geometry, this does not hold true for Capon's beamformer because it calculates the weight vector adaptively. Additionally, the artifacts in end-fire directions are much less visible than in the images obtained using H-set arrays. While this is beneficial in general, the imaging system is not operated in end-fire direction anyway. Thus, the differences in those regions are not a matter of great interest. If the reception of power from the object has to be improved, one can easily augment the array layout by the addition of further elements. Their positions can then be obtained by starting the algorithm initialized with the existing array.

Lattice type	Edge length	κ	$\kappa/(N_X N_Z)$	SLL in dB
rectangular	10	54	0.54	-11.1
	14	67	0.34	-11.2
	18	96	0.30	-11.6
	24	188	0.32	-13.2
	30	255	0.28	-13.7
bin. rnd. rect.	10	54	0.54	-8.9
				-10.8
				-12.0
	14	70	0.36	-10.2
				-11.3
18	113	0.35	-12.2	
bin. rnd. quincunx	18	113	0.35	-11.4
				-12.5
				-13.5
	24	188	0.32	-12.5
				-13.2
30	256	0.28	-13.9	
bin. rnd. quincunx	10	54	0.54	-13.3
				-13.7
				-14.2
	14	70	0.36	-10.7
				-10.8
18	113	0.35	-11.5	
bin. rnd. quincunx	14	70	0.36	-10.6
				-11.9
				-12.9
	18	113	0.35	-11.4
				-12.6
24	152	0.26	-13.8	
bin. rnd. quincunx	24	152	0.26	-11.8
				-13.4
				-14.6
	30	200	0.22	-12.3
				-12.7
30	200	0.22	-13.3	

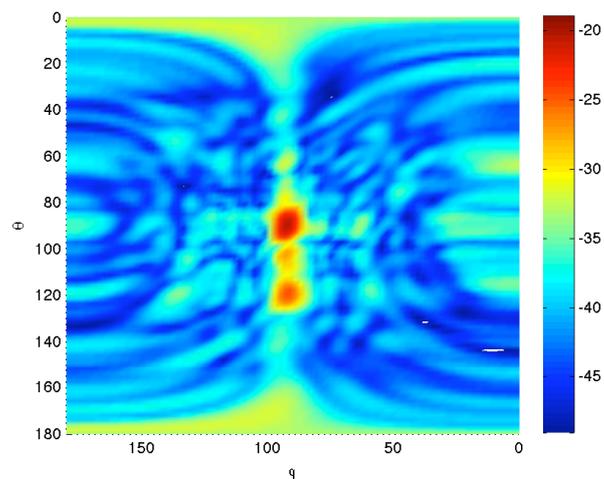
Table 4.2: Side-Lobe-Level and sparsity obtained for various apertures sizes and lattice structures of the designed sparse arrays.



(a) H-set array

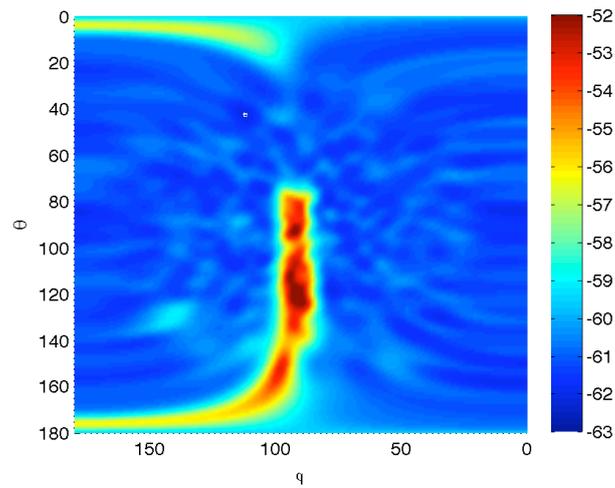


(b) URA

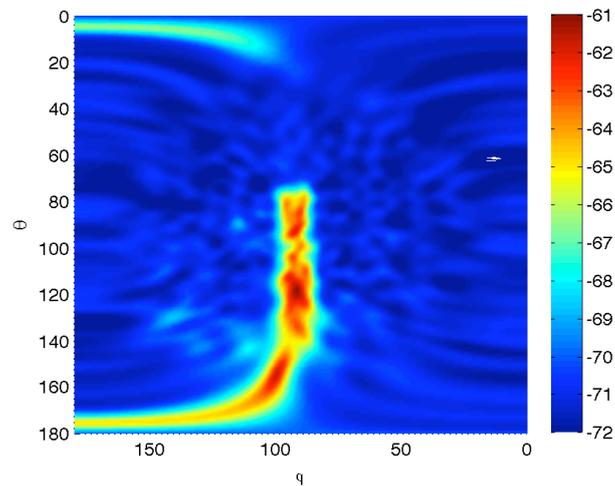


(c) Layout based on proposed approach

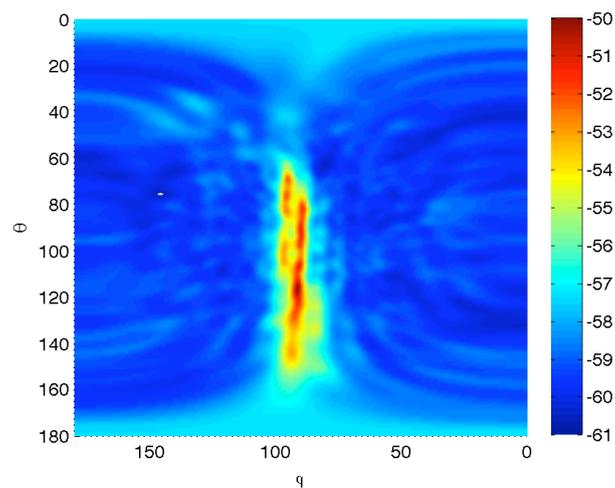
Figure 4.10: Ultrasound images of a PVC pole with rough surface obtained using Bartlett's beamformer on different array geometries. The array layouts are based on (a) a Hadamard difference-set (16x16) with 120 elements, (b) an (18x18)-URA with 324 elements, (c) the proposed approach using a rectangular (18x18)-lattice with 95 elements. All images have a dynamic range of 30dB.



(a) H-set array



(b) URA



(c) Layout based on proposed approach

Figure 4.11: Ultrasound images of a PVC pole with rough surface obtained using Capon's beamformer on different array geometries. The array layouts are based on (a) a Hadamard difference-set (16x16) with 120 elements, (b) an (18x18)-URA with 324 elements, (c) the proposed approach using a rectangular (18x18)-lattice with 95 elements. All images have a dynamic range of 11dB.

Chapter 5

Human Detection and Classification

5.1 Introduction

While the previous chapters of this thesis address the problem of design and calibration of an acoustic imaging systems, this chapter focuses on the question how such a system can be used in robotic applications. While it should be clear that objects can be detected generally by the identification of peaks in the acoustic images, we will show the possibility of detecting a class of objects which is of crucial importance for robotic applications, namely the detection of humans. In robotics, for example, it is not only generally important to detect obstacles in the surroundings of a system, but to specifically become aware of the presence of humans. Detecting persons in the surroundings is crucial for control of robot awareness and navigation [LWLW06, BH09, STI08]. Based on the latter, the trajectory of a person can be estimated and the robot can respond meaningfully, e.g., it can step out of the way of the human, address the person, etc.

Before we go into more detail, we want to motivate the use of acoustic imaging in this context. The benefits in this application will become even clearer after investigation of the limitations and weaknesses of other sensor modalities. In current systems, mainly optical (mono or stereo) cameras and radar are used [YTN05]. While it seems natural to use optical sensors, these suffer from the fact that their performance varies greatly with lighting conditions or employ expensive infrared cameras [LWLW06]. At the same time, optical systems do not always reliably detect range and often assume motion or a specific shape of the persons [CD00, LWLW06]. To mitigate these limitations, they often employ complex multi-layered hierarchies inspired by biology (e.g., see [TBPI10]). Radar-based systems, on the other hand, are currently used, but require expensive hardware and can be unreliable due to the large variability and low intensity of radar reflections from humans [YTN05]. If we were interested in the detection of moving objects only, using an acoustic or electro-magnetic excitation signal and the Doppler effect would be sufficient. However, it would not be possible to detect all still standing objects or persons in the scene.

Acoustic imaging allows for precise angular and range information. Due to the slow wave propagation speed of sound, we can easily discriminate close objects from background and therefore identify obstacles in the surroundings. Additionally, we will show

that the acoustic signature of a human person is quite unique compared to other objects and that this allows to classify the reflecting objects in the scene into human and non-human reflectors. Thus, acoustic sensors provide a powerful, reliable and cheap option for short-range applications in the context of environmental awareness and a human presence detection based on acoustic imaging can greatly enhance the overall system reliability in the aforementioned applications.

In this section, we will firstly show examples of acoustic images of humans and discuss the properties of their acoustic signatures. We then describe the segmentation problem and the technique applied to it in Section 5.2. This is followed by a description of features that can be used for classification in Section 5.3. We present a model that allows to generically parametrize the acoustic images. The model's parameters can be used as features for classification. Additionally, we also describe a model-free feature space based on geometrical and statistical features. After we have described the feature selection mechanism in Section 5.4, we show results obtained from real-data experiments and analyze the performance of several classifiers for the human presence detection and pose classification problems in Section 5.5. As discussed in Chapter 3, the imaging system used for the classification task consists of a single narrowband transmitter and a sparse 2D-array of 30 omnidirectional acoustic receivers. They are positioned in a nonuniform geometry on an area with a diameter of approximately 6cm. The objects in a scene are illuminated by a source signal $s(t)$ and reflect back to the array. The source signal has a center frequency f_c of 48 kHz and wavelength λ . According to the far-field signal model introduced in Section 2.1.1, this results at a given time t in a data vector

$$\mathbf{x}(t) = \mathbf{a}(\theta, \phi, r)T(s(t))e^{-j\frac{2\pi}{\lambda}\tau} + \mathbf{n}(t) \quad , \quad t = 1, \dots, N \quad ,$$

where N is the number of snapshots, τ represents the time delay between transmission and reception, θ is the elevation angle, ϕ is the azimuth angle and r is the distance to the array. The phase differences between the sensors are modeled by the array response vector $\mathbf{a}(\theta, \phi, r)$ and $T(\cdot)$ represents all effects on the target's reflectivity due to object texture and shape. Note that it is sufficient to work under far-field assumptions as long as the objects in the scene have a minimal distance to the array of approximately 1m. The processing chain for the classification task is summarized in Fig. 5.2 and can be described as follows: After some pre-processing and demodulation of $\mathbf{x}(t)$, we apply Capon beamforming on different range gates at distances r to compute a three-dimensional power spectrum estimate $P(\theta, \phi, r)$ from the two-dimensional images

$$P_r(\theta, \phi) = \mathbf{w}^H(\theta, \phi)\hat{\mathbf{R}}_r\mathbf{w}(\theta, \phi)$$

of several range gates (see Fig. 5.1). Here, \mathbf{w} is an adaptive weighting vector as described in Section 2.1.2 and $\hat{\mathbf{R}}_r$ is the sample covariance matrix of the data in range

gate around r . This results in acoustic images that show reflections from objects in the scene which can be used for object detection. Our goal is to find features that discriminate between humans and other objects. Based on a successful detection of human presence, we can further classify the human in terms of the current pose, e.g., whether the person is walking or standing.

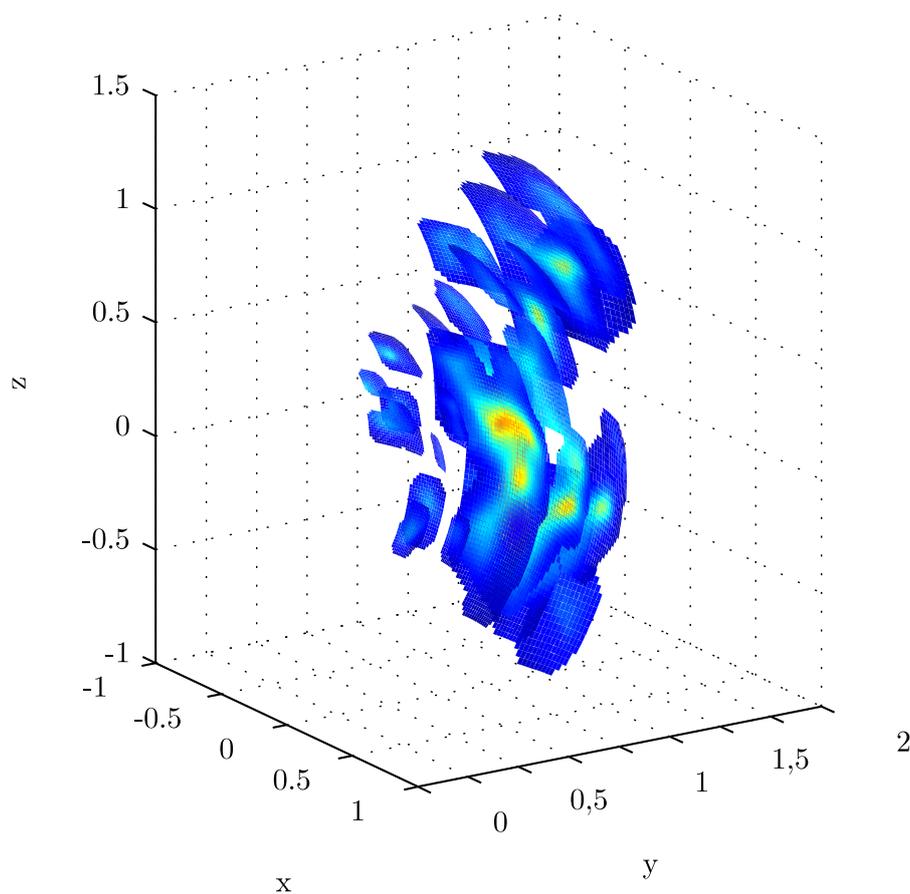


Figure 5.1: Example of a 3D acoustic image of a person facing the sensor array. The coordinate system is given in meters and is centered at the center of the array, which was mounted 93cm above ground. The x -axis denotes cross-range, the y -axis denotes the range.

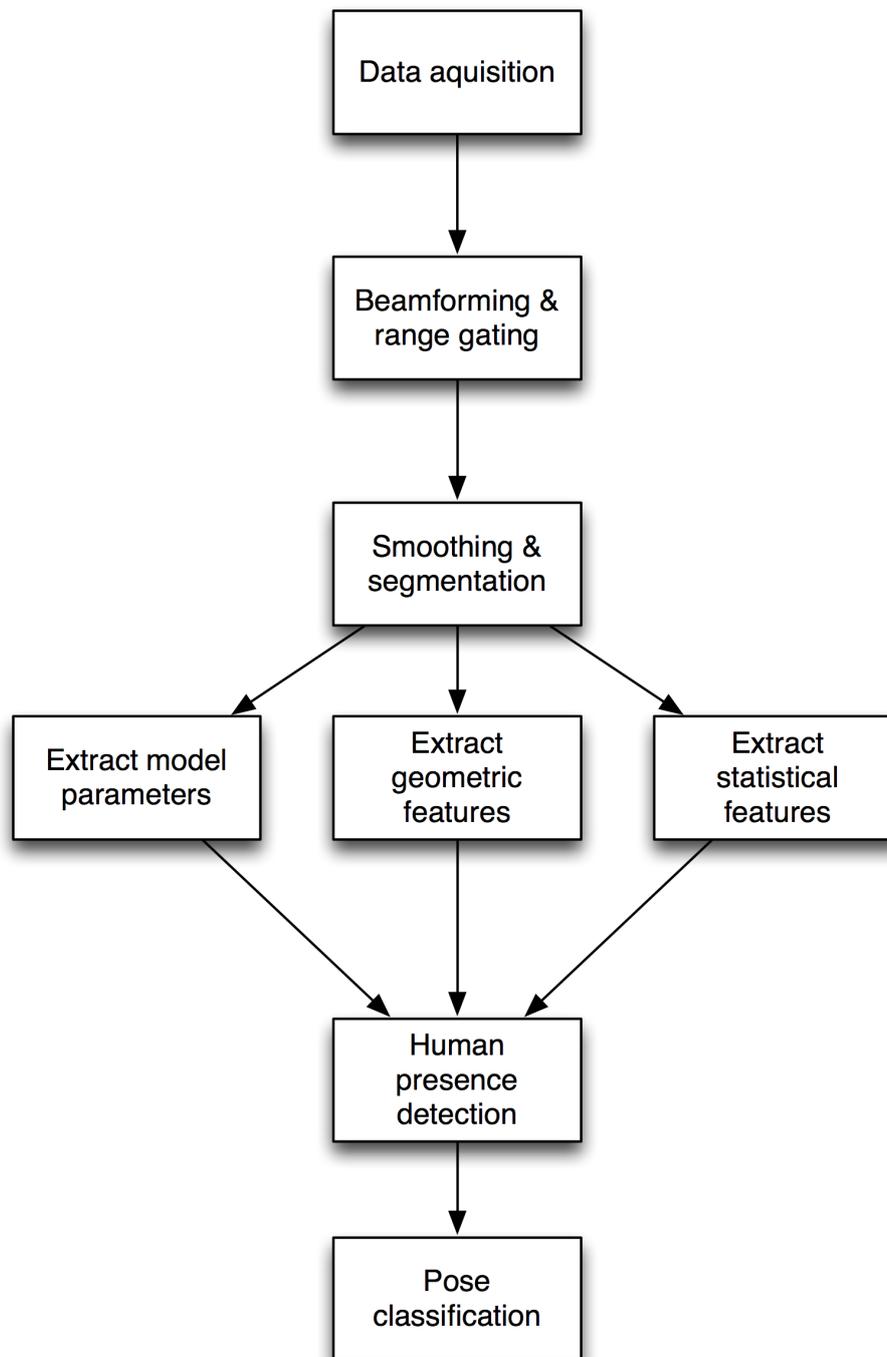


Figure 5.2: Flowchart of the proposed human detection and classification scheme. After the acoustic array records the reflected echoes, an acoustic image is obtained using adaptive beamforming. The foreground region is determined using the EM algorithm and features are extracted from both the image and the time series data. The features are then fed to a previously trained classifier which detects the presence of a human in the scene. If a human is present, its pose can be classified in a subsequent step.

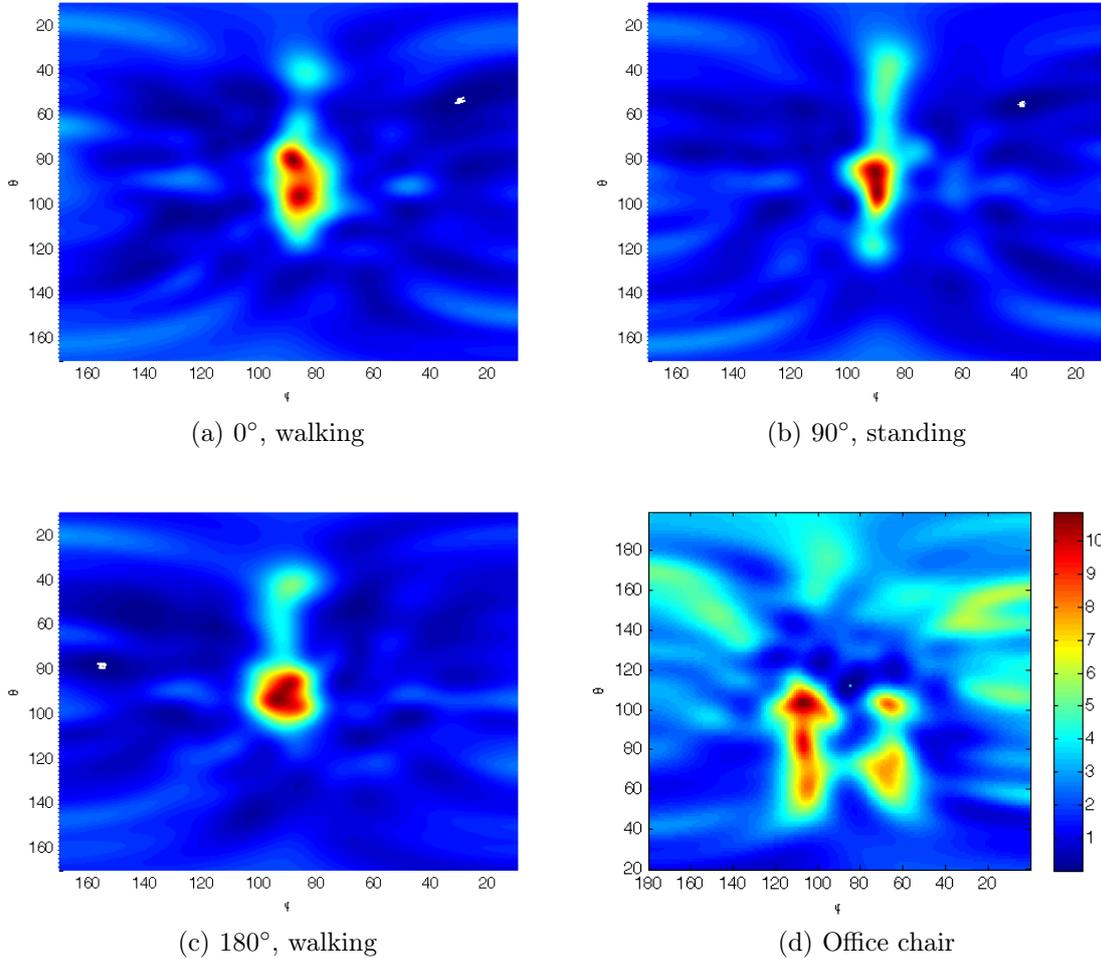


Figure 5.3: Three spatial spectra (in decibel) of humans in azimuth, ϕ , and elevation, θ , in degrees, from different orientations and one spatial spectrum of an office chair. All four images have a dynamic range of 11dB.

In Fig. 5.3 (a)-(d), we show acoustic images from three persons in different orientations and an office chair at a distance of 1.75m away from the array in broadside direction. The persons are facing the array (0°), looking to the side (90°) or looking directly away from the array (180°), and were recorded either walking or standing. The office chair has a diameter of 0.4m, plastic arm rests and a textile seat cover and cushion. As one can see, due to the complex texture of the human body, the excitation signal reliably reflects diffusely back to the array from both the torso and the head. As most objects only show specular reflections, this is quite unique to humans. Large and complex surface textures result in larger reflecting areas. Thus, the occurrence of a torso-shaped reflection is rare and together with another reflector above, the likelihood of human presence is quite high. As mentioned before, we are interested in finding features that exploit this behavior and span a feature space in which we can discriminate not only

between humans and non-human objects, but also classify fundamental poses such as a walking movement in contrast to standing still.

5.2 Segmentation

To solve the classification problem formulated above, we need to segment the image in order to detect the foreground region in a first step. This is followed by a feature extraction where the geometric features are based on properties of this foreground region. Due to the nature of the excitation and reflection of sound, and the fact that the acoustic images are created by scanning the scene using beamforming, the image will not show any sharp edges, but rather smooth contours. The degree of smoothness depends mainly on the object's shape and the beam-width of the applied beamformer, but is inevitable because the array has a limited aperture. Due to these facts, common complex segmentation techniques traditionally used in the image processing community are not required here. For example, active contours [KWT88, BI98, CV01] or min-cut/max-flow techniques [BK04, Set99] were mainly developed to segment highly detailed images, e.g., photographic images or medical images from magnetic resonance brain tomography [CJS93, FD96]. While such methods are undoubtedly very powerful, they do not perform superior to simpler, threshold-based techniques when applied to acoustic images, due to the relative simplicity of the shape of echoes. Thus, it is sufficient to detect acoustic echoes simply based on individual voxel power. To achieve that, we smooth the image $P_r(\theta, \phi)$ in an initial step to reduce potential multimodalities of the torso echoes. We then use the EM algorithm [DLR77, RW84, Bis07] to fit a Gaussian-Mixture-Model (GMM) with two Gaussians $\mathcal{G}(\mu_1, \sigma_1), \mathcal{G}(\mu_2, \sigma_2)$, where $\mu_1 < \mu_2$, to the voxel intensity histogram of the smoothed image. Although the reflected power intensity can also be modeled according to a Rayleigh or Weibull distribution, a sufficient approximation can be obtained using a superposition of Gaussians, which directly results in a low-intensity background and a high-intensity foreground region. The foreground region \mathcal{R} is formed by voxels in the image that have a higher probability to belong to the Gaussian $\mathcal{G}(\mu_2, \sigma_2)$ (see Fig. 5.4). This approach is well-known as a simple segmentation technique (e.g., see [Bis07]). In Figure 5.5, we show an example of the resulting foreground region where the head and torso echoes are not separated into distinct segments.

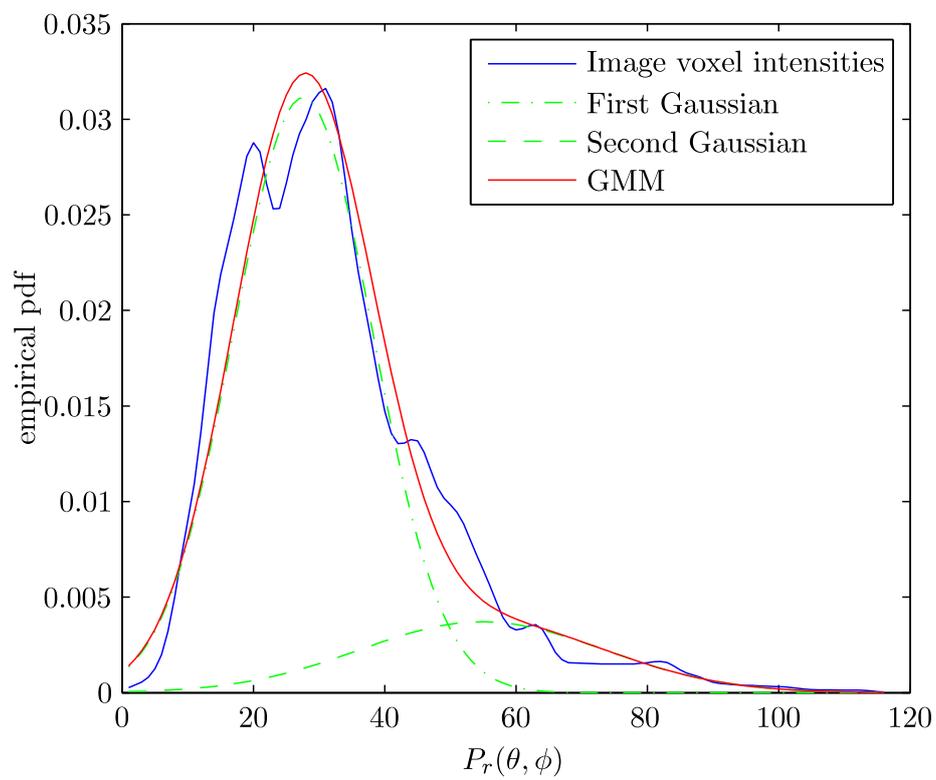


Figure 5.4: Example of an empirical power distribution for an acoustic image. Voxels with high intensity are more likely to belong to $\mathcal{G}(\mu_2, \sigma_2)$ and will be included into the foreground region.

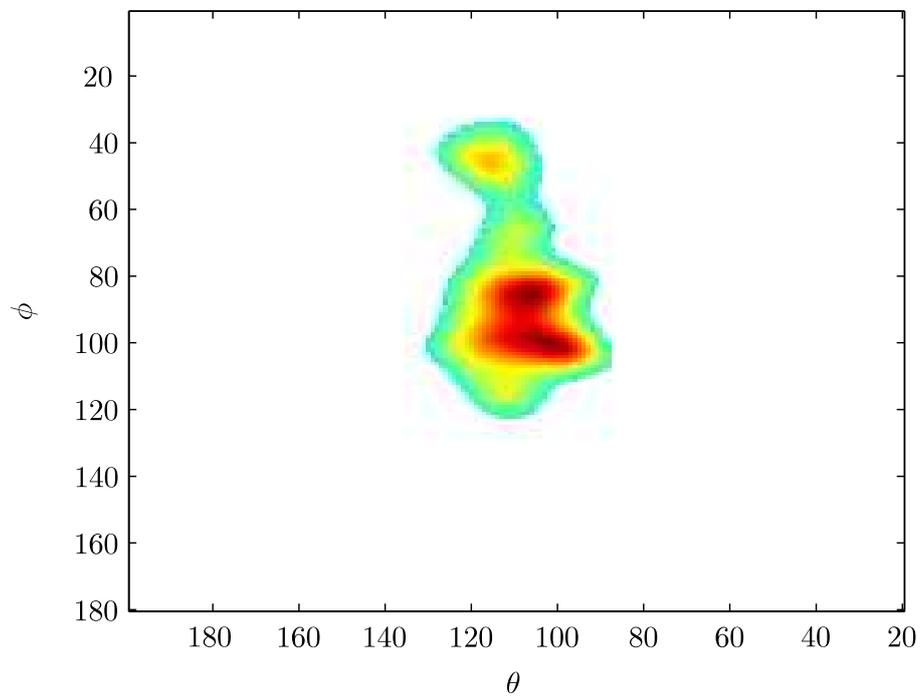


Figure 5.5: Result of the segmentation of an image where the echoes of head and torso are in close distance.

5.3 Feature Extraction

In order to obtain a feature space for human presence detection or classification of a specific pose, we extract features from the acoustic images in three different ways:

1. First, we formulate a generic model for acoustic images and extract the model's parameters as features.
2. Secondly, we extract statistical features from the time series data $\mathbf{x}(t)$ recorded by the array.
3. Finally, we also extract geometric features from the acoustic images $P_r(\theta, \phi)$ obtained from $\mathbf{x}(t)$.

In the following, we describe each of those ways in more detail. The implemented features are chosen such that they correspond to some physical meaning or image parameter that changes with respect to the object in the scene, e.g., the contours of the main echoes or the size of its reflective surface. While it is easily possible to implement many more shape parameters, we restrict the presentation to the features that are used in the classification. Their choice is motivated not only by physical meaning, but also an information-theoretic feature selection procedure described in Section 5.4. This is also validated by the achieved classification performance we describe in Section 5.5. The features are described such that their values are affected by the object's range. However, it is straightforward to make them range-invariant by transforming them to the Cartesian coordinate system and/or normalize them with respect to their TOF.

5.3.1 Modeling the Acoustic Signature

In this section, we describe how the acoustic images can be generically modeled. Our goal is to extract the model parameters as features for further classification of the objects present in the scene. We are interested in two aspects, based on the reliable occurrence of head and torso echoes. First, we want to parametrize the image such that the acoustic signature is preserved and we can establish geometric properties in the image. Secondly, we aim to find clusters in the parameter space that allow us to discriminate between humans and other objects present in the scene. With respect to the detection and classification of *persons*, this means that we want to obtain parameter sets that are unique to the presence of a large torso echo, a weaker

head echo and possibly other echo sources in the scene. We therefore propose to model the spatial power spectra obtained from the acoustic array as a mixture of K two-dimensional Gaussians $\mathcal{G}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, $k = 1, \dots, K$, in the (θ, ϕ) -domain where $\boldsymbol{\mu}_k$ are the mean vectors, and, with $\rho, \sigma_{\theta,k}, \sigma_{\phi,k}$ being the correlation coefficient and the standard deviations in θ and ϕ dimensions, and the covariance matrices

$$\boldsymbol{\Sigma}_k = \begin{pmatrix} \sigma_{\theta,k}^2 & \rho\sigma_{\theta,k}\sigma_{\phi,k} \\ \rho\sigma_{\theta,k}\sigma_{\phi,k} & \sigma_{\phi,k}^2 \end{pmatrix} . \quad (5.1)$$

We then fit the GMM to the image by solving the following optimization problem:

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \epsilon = \arg \min_{\boldsymbol{\beta}} \int_{\mathcal{R}} \|\mathbf{P}_r(\theta, \phi) - f(\boldsymbol{\beta})\|^2 \delta\theta \delta\phi \quad (5.2)$$

$$\text{s.t. } \lambda_{1,k}, \lambda_{2,k} > 0 \quad \forall k = 1, \dots, K \quad (5.3)$$

$$\mathcal{S}_i \not\cap \mathcal{S}_j \quad \forall i \neq j, \quad i, j = 1, \dots, K \quad (5.4)$$

$$\boldsymbol{\mu}_k \in \mathcal{R} \quad \forall k = 1, \dots, K \quad (5.5)$$

where \mathcal{R} is obtained by segmentation (see Section 5.2),

$$\begin{aligned} f(\boldsymbol{\beta}) &= f(w_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1^T, \dots, w_K, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K^T) \\ &= \sum_{k=1}^K w_k \mathcal{G}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) . \end{aligned} \quad (5.6)$$

Here, w_k is a weighting factor and $\boldsymbol{\beta}$ denotes the parameter vector to be estimated, such that

$$\boldsymbol{\beta} = (w_1, \boldsymbol{\mu}_1, \text{vec}\{\boldsymbol{\Sigma}_1\}, \dots, w_K, \boldsymbol{\mu}_K, \text{vec}\{\boldsymbol{\Sigma}_K\})$$

The eigenvalues of $\boldsymbol{\Sigma}_k$ are denoted as $\lambda_{1,k}, \lambda_{2,k}$. The constraint in Eq. (5.3) is required in order to guarantee positive definite covariance matrices for all $\boldsymbol{\Sigma}_k$. The second constraint in Eq. (5.4) is introduced in order to ensure that each echo is modeled only from a single Gaussian. Here, \mathcal{S}_i is an ellipsoid region in the (θ, ϕ) -domain that covers a certain fraction of the volume of the i th Gaussian such that all points (θ, ϕ) in \mathcal{S}_i fulfill

$$((\theta, \phi) - \boldsymbol{\mu}_i) \boldsymbol{\Sigma}_i^{-1} ((\theta, \phi) - \boldsymbol{\mu}_i)^T \leq C(1 - \rho)^2 . \quad (5.7)$$

Here, since we assume two-dimensional Gaussians, C is determined according to the inverse cumulative χ_2^2 distribution such that

$$\int_0^{C^2} \frac{e^{-t/2}}{2} \delta t = F \quad (5.8)$$

is satisfied [AS72].¹ Here, F denotes the fraction of the volume under the Gaussians to be covered by any \mathcal{S}_i . Its value should lie somewhere between 0.5 and 0.9 to control

¹Note that this holds exactly only if $\rho = 0$ and is an approximation otherwise. However, for $\rho > 0$, the ellipsoid is only rotated, thus the relation leads to the desired coverage of the Gaussian.

the separation between the Gaussians effectively. The third constraint in Eq. (5.5) is based on the segmentation result and simply restricts the mean of the Gaussians to lie inside the foreground region. It is not strictly necessary, but prevents divergence of the solving algorithm from reasonable solutions. The problem formulated in Eq. (5.2) can be initialized using knowledge of \mathcal{R} , e.g. the mean vectors $\boldsymbol{\mu}_k, k = 1, \dots, K$ can be defined to be at the locations of the K largest extrema in \mathcal{R} .

To accelerate the convergence properties in the solution space, it is beneficial to reformulate the constraints in Eq. (5.3) and (5.5) into penalty terms of the cost function, e.g. a positive definite $\boldsymbol{\Sigma}_k$ can be favored by a penalty term $\log(\det \boldsymbol{\Sigma}_k)$. The constraint in Eq. (5.5) can be taken into account by penalizing the distance of $\boldsymbol{\mu}_k$ to the center of mass of \mathcal{R} . Thus, the problem from Eq. (5.2) can be reformulated to

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \epsilon = \arg \min_{\boldsymbol{\beta}} \int_{\mathcal{R}} \|\mathbf{P}_r(\theta, \phi) - f(\boldsymbol{\beta})\|^2 \delta\theta \delta\phi \quad (5.9)$$

$$- l_1 \log(\det \boldsymbol{\Sigma}_k) + l_2 \sum_k \log(\det \boldsymbol{\Sigma}_k) \|\boldsymbol{\mu}_k - (\bar{\theta}, \bar{\phi})\|$$

s.t. $\mathcal{S}_i \not\cap \mathcal{S}_j \quad \forall i \neq j$ (5.10)

Here, l_1, l_2 represent constant penalty cost factors that have to be chosen manually and control the weighting of the different terms in the cost function. The point $(\bar{\theta}, \bar{\phi})$ denotes the center of gravity in the foreground region \mathcal{R} and is defined by

$$\bar{\theta} = \frac{\iint_{\mathcal{R}} \theta P_r(\theta, \phi) \delta\theta \delta\phi}{\iint_{\mathcal{R}} P_r(\theta, \phi) \delta\theta \delta\phi}, \quad \bar{\phi} = \frac{\iint_{\mathcal{R}} \phi P_r(\theta, \phi) \delta\theta \delta\phi}{\iint_{\mathcal{R}} P_r(\theta, \phi) \delta\theta \delta\phi} \quad (5.11)$$

The problem in Eq. (5.9) can be solved numerically, e.g. using a Quasi-Newton algorithm, for different K , depending on the target object class, i.e.. As the knees, hands and feet of humans are not always visible in the image, it is sufficient to set $K = 2$ to model only the head and torso echoes, assuming there are not other objects present in the scene. Clearly, one has to set K higher when there is no knowledge about the scene and the number of reflectors available.

Based on the found estimate of the model parameters $\hat{\boldsymbol{\beta}}$, we can then formulate other features that are meaningful to detect humans in the image, e.g. the relative position, distance and angle of the centroids. Additionally, we can expect the head echo to be much smaller than the torso echo, meaning that the ratio of variances is an additional feature. However, the solution of Eq. (5.9) provides not only $\hat{\boldsymbol{\beta}}$, but also the residuals ϵ as a measure for the goodness-of-fit of the model. Therefore, it is theoretically possible to set K as a free parameter and vary the model order. One can then choose how accurately the image should be modeled or whether small reflectors, clutter and

noise should also be modeled or neglected. The classic information criteria can then be used to balance model accuracy versus complexity based on the principle of parsimony (see [SS04, WK85, WZ89]). Together with the segmentation contours, $\hat{\beta}$ and ϵ allow to represent humans in acoustic images quite uniquely and can be supplied to a subsequent classifier. A strong advantage of the method is that head and torso echoes of a human are always explicitly determined, such that it works reliably even when those echo regions are so close that the segmentation does not find disjoint regions anymore. On the other hand, the approach has the severe disadvantage that the formulated optimization problem is based on the least-squares principle and requires complex problem solving routines. The accuracy of the resulting model also depends heavily on the initialization of the problem. We therefore seek to also find more simple features that are easy to compute and do not rely on complex optimization problems

5.3.2 Geometric Features

5.3.2.1 Elliptic Torso Fitting

For the above mentioned reasons, the use of ellipse fitting is a reasonable choice, since it is computationally cheap. Additionally, the torso is not only a reliable source of acoustic reflection of any human, but also shows a significant elliptic shape. We therefore model the contours of the largest echoes in the foreground region \mathcal{R} . We assume that the segment with the largest area represents the torso and extract its contour $\mathbf{C} = (\mathbf{c}_\theta \ \mathbf{c}_\phi)^T$, where $\mathbf{c}_\theta, \mathbf{c}_\phi$ represent the coordinates of the contour pixels and are, thus, of varying length and the superscript T denotes transpose. Now, given the general quadratic expression of an ellipse as

$$\mathbf{C}_e^T \boldsymbol{\Sigma}^{-1} \mathbf{C}_e + c = 0, \quad (5.12)$$

we can transform the data by subtraction of its center of mass and rotation to obtain a standardized ellipse of the form

$$\tilde{\mathbf{C}}_e^T \mathbf{S} \tilde{\mathbf{C}}_e + \tilde{c} = 0.$$

With λ_1, λ_2 being the eigenvalues of $\boldsymbol{\Sigma}^{-1}$, $a = \sqrt{-\frac{\tilde{c}}{\lambda_1}}, b = \sqrt{-\frac{\tilde{c}}{\lambda_2}}$ are the ellipse's axes and, since λ_1, λ_2 are invariant to translation and rotation, $\mathbf{S} = \text{diag}(\lambda_1, \lambda_2)$. We then fit the ellipse parameters to the data by minimizing the *geometric* distances of the ellipse to \mathbf{C} , meaning that we minimize the sum of the squares of the distances to the data points [GGS94]. This can be obtained using the constraint $\lambda_1 + \lambda_2 = 1$ (or $\text{trace}(\mathbf{S}) = 1$). This has been shown to be more robust than minimizing the algebraic distances [GGS94, FPF99, FF95] and results in a nonlinear (total) least squares

problem which we solve using a Gauss-Newton procedure following [GG94]. The initial estimates of the parameters are found using the *Bookstein constraint* $\lambda_1^2 + \lambda_2^2 = 1$, which is more efficient, but not as accurate as the trace constraint. In [FPF99], an alternative approach is given and the authors show that their strategy is even more robust for the general problem of conic fitting. However, we found that for the acoustic images, using the trace constraint as described above is sufficient and the results do not differ largely.

5.3.2.2 Generic Shape Parameters

In addition to the elliptic parameters, we also use more generic shape parameters of the acoustic images. The most simple feature used here is the total area A of the largest continuous segment in \mathcal{R} . Based on that, we also compute a measure of smoothness of the contours. We implemented several features such as the model order of a polygonal fit to the contour, the eccentricity of the contour in polar coordinates and also more sophisticated transformations of the contour (see also [ZL04]). However, we found that a rather simple measure performed best from this category of features. To obtain it, we determine the smallest convex polygon that fully contains the echo segment and determine its area A_c . Then, simply the *convexity*

$$cv = A_c/A$$

shows how smooth the contour of an echo segment is. The feature selection described in Section 5.4 showed that this convexity dominates other shape parameters than A for the problem of pose classification.

5.3.3 Statistical Features

In addition to features extracted from the acoustic images, we also analyze the time series data of the array channels in $\mathbf{x}(t)$. They contain direct information about echo distribution in depth as well as information about the object's general pose. An analysis of the distribution of the data shows changes within the classes. More specifically, the extreme values as well as the percentile spreads in the empirical distributions are correlated with the reflecting surface of the objects in the scene, suggesting that the tails of the distributions change depending on the human pose. This is likely due to the fact that a walking human rotates the body slightly. The received power will therefore vary in signal strength because the reflections stem from reflective surfaces that are not orthogonal to the direct line between array and person. For these reasons, we extract features which estimate the tails of the time series as well as measures of power concentration along the range.

5.3.3.1 Hill Estimator

There exist several indices that measure the extent to which a distribution is heavy-tailed (see [PDM98] for an overview). Here, we use the *Hill estimator* [Hil75] which is defined as follows: Let $X^{(1)}, X^{(2)}, \dots, X^{(n)}$ be the order statistics, obtained by $X^{(1)} \geq X^{(2)} \geq \dots \geq X^{(n)}$, of a distribution $F(x)$ and $x^{(1)} \geq x^{(2)} \geq \dots \geq x^{(N)}$ realizations thereof. $F(x)$ can be of any form, but its (upper) tail should follow a general form such that for large x , $1 - F(x) \propto x^{-\alpha}$. A value of $\alpha = 1$ corresponds to the Cauchy distribution and distributions with $\alpha \rightarrow 2$ approach the normal distribution [PDM98]. The Hill estimator of the tail index of $F(x)$ is then given by

$$\alpha_H^{(m)} = \frac{m+1}{\sum_{i=1}^m i \cdot \ln \left(\frac{x^{(i)}}{x^{(i+1)}} \right)},$$

where m is the number of order statistics taken into account in the estimation. The choice of m influences the statistical properties of $\alpha_H^{(m)}$, e.g., a large m will result in a small variance, but a large bias of the estimator. An iterative selection procedure to choose m optimally is described in [Hsi99]. Applied to this problem, it results in approximately $m = 0.2N$. Clearly, the samples $x^{(n)}$ could be taken from any single channel of the array. At the same time, it is advisable to make use of *all* the data in $\mathbf{x}(t)$, not only a single channel. However, one should not simply merge data from different channels in order to increase the sample size as the channels are highly correlated and the Hill estimator was developed based on the assumption of independent and identically distributed statistics. Alternatively, we average the absolute values of the data across all channels in the base-band to obtain a pseudo-channel $\tilde{x}(t)$, $t = 1, \dots, N$ with order statistics $\tilde{X}^{(i)}$, which greatly reduces the variance of $\alpha_H^{(m)}$ compared to using a single channel.

5.3.3.2 Power-related Tail Parameters

Since $\alpha_H^{(m)}$ is a *shape* parameter that does not contain information about the strength of the received echoes, we additionally look at other measures based on the samples above the 80-percentile of the empirical distribution, denoted by $\tilde{x}_{0.8}^{(i)} = \tilde{x}^{(m)}, \dots, \tilde{x}^{(1)}$, $m = \lfloor 0.2N \rfloor$ to obtain more information about the maximally returned power. We then estimate both the sample mean and the sample variance in the upper tail as

$$\hat{\mu}_{0.8} = \frac{1}{m} \sum_{i=1}^m \tilde{x}^{(i)}$$

and

$$\hat{\sigma}_{0.8} = \frac{1}{m-1} \sum_{i=1}^m (\tilde{x}^{(i)} - \hat{\mu}_{0.8})^2.$$

These features measure the amount of power and the concentration in the tail, which consistently shows to possess discriminative power especially for the pose classification (see Section 5.5.4).

5.3.3.3 Depth-related Parameters

As the reflected echoes from the objects vary across range depending on the shape of an object or the pose of a human, it is reasonable to also measure the concentration of power along the distance to the array. While walking humans will most likely swing their arms during movement, a standing human is expected to hold them closer to their bodies most of the time. This results in the fact that walking humans on average reflect echoes across a larger range than standing humans. To measure this effect, we estimate central geometric moments in the pseudo-channel, i.e., the moment of first order by $m_1 = \sum_{i=1}^N \tilde{x}(i)(i - t_p)^2$, where t_p denotes the location of the point of mass along the samples. If a human is present, t_p will correspond to the depth at which the torso reflects the largest echo. To obtain an even stronger emphasis on the power received, we also estimate the second central geometric moment, which is simply $m_2 = \sum_{i=1}^N \tilde{x}^2(i)(i - t_p)^2$.

5.4 Feature Selection

To select a good feature subset for the human presence detection, it is reasonable to select features that are heavily affected by the presence of a large torso echo from the human. Thus, we manually choose to select the ellipse's axis parameters \hat{a}, \hat{b} obtained from the elliptic fit of the largest echo. Together with the reflective area A , we obtain a three-dimensional feature space that is well separated (see Section 5.5.3). However, the selection of features is more complex for the problem of pose classification. While it is straightforward to identify several features that are clearly correlated with the class label, the selection of a reasonable feature subspace for that problem is more difficult and a statistical analysis of the feature space is necessary to obtain good classification performance.

There exist numerous approaches for feature selection such as sequential searches, genetic algorithms, branch-and-bound search (see [JZ97] for an overview). They all have

in common that they need to define a measure of merit in order to evaluate a specific feature space for some training data. The most common measure here is the classification error rate, which determines how accurate a specific classifier performs in separating data into a number of classes. However, it can only be calculated by choosing a certain classifier which has to be retrained for every feature subset. While this is clearly possible, there is also the possibility to analyze the feature space and possible subspaces without the need of a specific classifiers. We use the minimal-Redundancy-Maximal-Relevance (mRMR) method of [PLD05], which is proven to give results that maximize the dependency of the class labels on the selected features. The method relies on the mutual information between two random variables X and Y , defined as

$$I(X, Y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy .$$

The relevance of a set \mathcal{S} of features is then measured by the averaged mutual information of all features with the class labels as

$$D(\mathcal{S}, \mathcal{C}_n) = \frac{1}{|\mathcal{S}|} \sum_{f_i \in \mathcal{S}} I(f_i, \mathcal{C}_n)$$

and the redundancy of \mathcal{S} is measured by the average mutual information between two features f_i, f_j as

$$R(\mathcal{S}) = \frac{1}{|\mathcal{S}|^2} \sum_{f_i, f_j \in \mathcal{S}} I(f_i, f_j) .$$

The algorithm maximizes the relevance while simultaneously minimizing the redundancy in \mathcal{S} . This is performed by maximizing the difference $D(\mathcal{S}, \mathcal{C}_n) - R(\mathcal{S})$ with respect to single features in \mathcal{S} . Thus, a set with high discriminative power is created using only a small number of features that complement one another. Kernel density estimation is used to obtain estimates of D and R . Gaussian kernels have been found to be suitable. The features are then selected using an incremental search that adds one feature at a time. In [PLD05], this "first-order incremental selection" is shown to be optimal in that it maximizes the dependency of the class label on the resulting feature subspace. The features selected with this procedure are the ones described in the previous section. They also validate the manual selection of the elliptical parameters and the torso area as sufficient for the human detection problem because additional features only have a small effect on $D(\mathcal{S}, \mathcal{C}_n)$ while increasing $R(\mathcal{S})$. While it is sufficient to use only geometric features for the detection problem, only the combination of features from statistical and geometric categories allows a reasonable classification performance for the pose classification problem. The GMM parameters can be used for a successful detection, but are inferior to the simpler geometric features described in section 5.3.2 due to their complexity. We therefore show the results using the model-based approach, but do not further include those parameters in the feature space for the classification results.

5.5 Results

In this section, we present the experimental setup used to record the real data on which the results in this section are based. We will compare the classification performance of several methods for both the human presence detection and the pose classification problem. All features are scaled before any feature selection, training or testing to avoid numerical dominance of any feature. The scaling factors were chosen such that all features lie in the interval from 0 to 1. We apply four different classifiers to all classification problems:

- a simple LDA classifier that assumes homoscedastic (equal covariance matrices) distributions for the classes,
- a QDA classifier that makes no assumptions about the covariances of the classes,
- SVMs using a linear kernel,
- and SVMs using a Gaussian radial basis function as kernel.

The performance is compared based on the same training and test data. The SVMs were tested with several penalty terms and function parameters (e.g. variances of the Gaussian kernels). The described results are obtained using the best choice of parameters from a logarithmic grid search.

5.5.1 Experimental Setup

The imaging system we use consists of a single narrow-band acoustic transmitter with a center-frequency of 48 kHz. The receiving 2D-array consists of 30 omnidirectional microphones which are positioned in a nonuniform sparse geometry. It is mounted at a height of 93cm. The transmitter is specified to have a beam pattern such that the 3dB-cutoff area is approximately 60° in azimuth and elevation and is mounted close to the array. The system was placed in an indoor scenario and we recorded several data sets from humans and non-human objects. The class of non-human objects consists of several pieces of typical office furniture, such as an office chair, a desk, a room plant, etc. All objects were put in front of the array at different orientations relative to the array and the scene was illuminated by the single transmitter. The recorded echos were then processed using adaptive beamforming as described in Sections 2.1.2 and 3.1.

5.5.2 Modeling Results

We demonstrate the approach described above and apply it to data from two different scenes, each containing a single object. In the first scene, a human was standing in front of the array at a distance of 1.75m, facing the array. In the second scene, the office chair from Figure 5.3 was placed at the same position. In Figure 5.6, we can see the reconstructed images based on the segmentation region \mathcal{R} and the estimated parameters $\hat{\beta}$ for a GMM with $K = 2$. The resulting image for the first scene (Fig. 5.6a) clearly shows the Gaussian that models the strong torso echo ($w_1 = 8.296$) while the head echo, modeled by a significantly weaker Gaussian above ($w_2 = 3.637$), is less strongly visible, but reliably located above the torso as expected. Thus, although the two echo sources were not well separated, the image was successfully parametrized. In the second scene, the two strongest echoes are reflected from both one of the arm rests and a corner of the seat cushion. This is also modeled in the reconstructed image (Fig. 5.6b), where only two smaller, almost equally strong reflections are modeled at both corresponding regions in the image. We see that modeling the acoustic image by a GMM allows to represent the spatial information about the position of reflecting surfaces in a parameter vector. As described previously, the approach can be extended by application of the model with a varying order K if it is desired to model the image in more detail, resulting in a model-order selection problem.

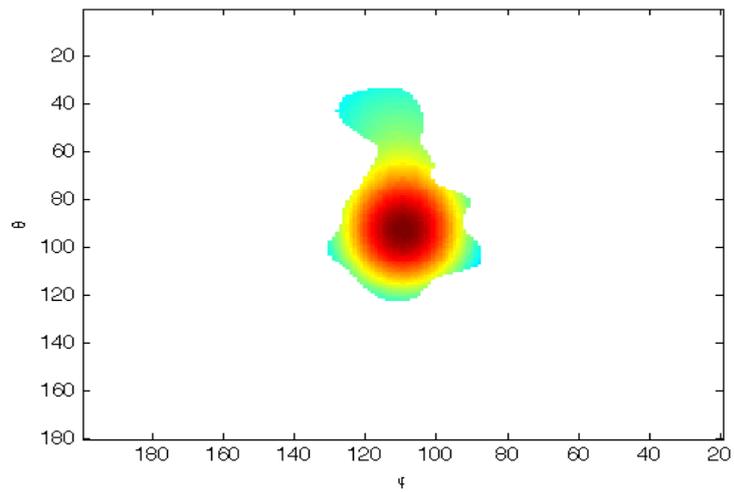
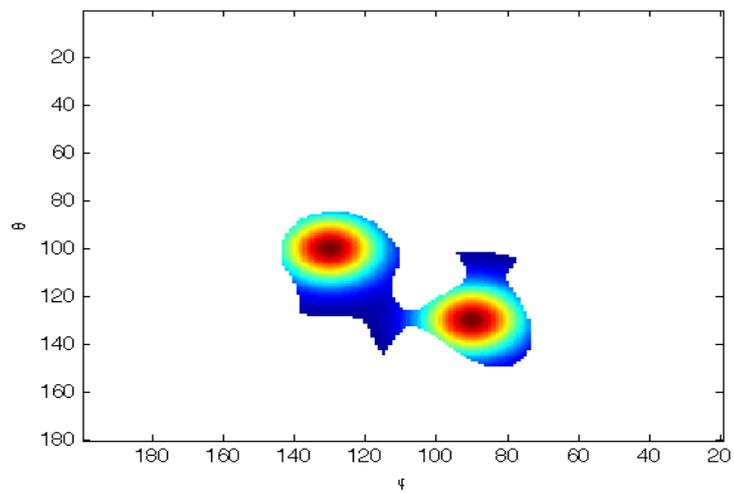
(a) Walking person, orientation 180° (b) Office chair, orientation 0°

Figure 5.6: Reconstructed images based on the estimated parameters for a person (a) and an office chair (b).

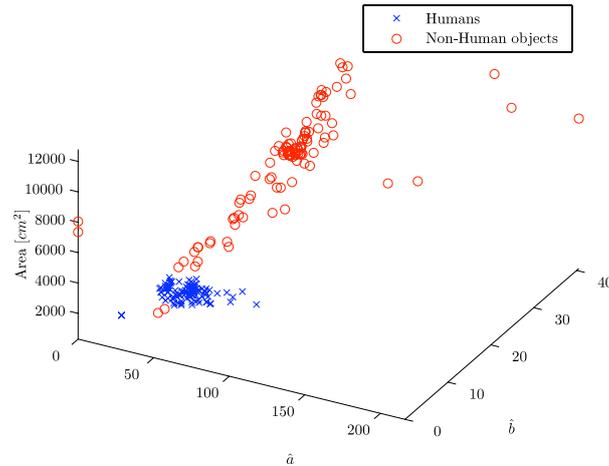


Figure 5.7: Three-dimensional feature space for the two classes of human and non-human objects. The figure shows both estimated ellipse parameters \hat{a} , \hat{b} of the torso echo segment as well as the overall area of the echo in cm^2 . Humans are represented by crosses, non-human objects by circles.

5.5.3 Human Detection Performance

Due to the unique echoes returned from the human body, especially the torso, we can detect reliably the presence of a human in the image by simply using the ellipse parameters and other simple geometric features. In Figure 5.7, we show the resulting three-dimensional feature space for the two estimated ellipse shape parameters as well as the overall echo region in cm^2 . Since all objects were placed at a distance of 1.75m and with a pixel resolution of 0.5° , a single pixel in the image represents roughly 1.5 cm^2 . In the Figure, we can clearly distinguish the two classes from each other and the large majority of the data is well separated in the two classes. Exceptions occur for two scenes in which non-human objects lie close to the main cluster of human scenes as well as two non-human scenes where the ellipse fitting failed to produce an ellipse due to the complex concave structure of the returned echo contours. However, the Figure shows that the two classes can be easily distinguished from each other, e.g., by LDA or QDA. To test the performance in this two-class problem, we use a data set of 107 humans and 101 non-human objects. We create a training set of 70% of the data by randomly choosing 74 humans (H) and 70 non-human objects (NH) from the full data set. The remaining 64 scenes are then used to evaluate different classifiers. Using LDA, we can correctly classify 61 objects or 95.31% of the test set. Three scenes with humans were not detected correctly and no false alarms are present. When using QDA, the correct classification rate, also called *total classification accuracy*, improves to 96.88%, although one human image is not detected correctly. However,

			classified as	
Method	total accuracy	class	H	NH
LDA	95.31%	H	33 (100%)	0 (0%)
		NH	3(9.68%)	28(90.32%)
QDA	96.88%	H	32 (96.97%)	1 (3.03%)
		NH	1 (3.23%)	30 (96.77%)
SVMs	96.88%	H	33(100%)	0(0%)
		NH	2(6.45%)	29(93.55%)

Table 5.1: Classification performance of Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and the Support Vector Machines (SVMs) for the two-class problem of human presence detection. A number of 144 images (70%) of the data was used for training and 64 images for testing. Both sets contain equal proportions of both classes.

			classified as	
Method	total accuracy	class	H	NH
LDA	92.79%	H	100%	0%
		NH	14.85%	85.15%
QDA	97.60%	H	98.13%	1.87%
		NH	2.97%	97.03%
SVMs	97.60%	H	100%	0%
		NH	4.95%	95.05%

Table 5.2: Classification performance of Linear Discriminant Analysis, Quadratic Discriminant Analysis and a Support Vector Machines for the same problem using cross-validation.

the classification improves for both classes. Clearly, the assumption of equal covariance matrices of the classes made by LDA is not realistic and QDA does take this into account, resulting in higher classification accuracy. When using SVMs, either with a linear kernel or radial basis functions, the results are similar to the ones obtained with QDA. The correct classification rate is equal, the single class performance differs only slightly. The obtained confusion matrices are depicted in Table 5.1. The classification rates are robust to the selection of training data, due to the well-separated clusters in the feature space. To evaluate the classification performance using the largest possible training sets with K scenes, we also employ a leave-one-out cross-validation, where we repeatedly train the classifiers based on $K - 1$ scenes and test performance for the remaining one, such that all scenes belong $K - 1$ times to the training set and are used once as a test scene. Table 5.2 shows the results obtained from the cross-validation procedure. Clearly, LDA performs worse in terms of false alarms, it detects 14.85% of the non-human objects as humans, but maintains the perfect detection of

all humans. In summary, LDA results in a classifier that is biased towards humans. It is not generally recommendable due to its simple linear decision rule and assumptions about the equal covariance of the classes. Both QDA and SVMs perform better in total classification accuracy. However, the results using SVMs indicate also a bias towards humans, although on a much smaller scale. The quadratic classifier shows a more balanced result, at the cost of a slightly reduced detection of humans. Its total classification accuracy is again identical to the one obtained using SVMs.

5.5.4 Pose Classification Performance

To analyze the classification performance for standing and walking humans, we recorded another data set of 73 scenes in which one of 4 different persons is present at a time. The subjects were recorded both while walking and standing and looking at different directions relative to the array ($0^\circ, 90^\circ, 180^\circ, 270^\circ$). The persons were recorded during six poses while walking a full sequence of steps which forms the basis of a periodic walking movement. In the following, we describe the recorded poses with an example of a walking movement that starts with the right leg:

1. Initially, the person stands on one foot (e.g. the left) and extends the other to step forward.
2. In the second pose, the person has shifted weight to the other (right) foot and stands on it while the heel of the initial (left) foot is lifted from the ground.
3. The next sample is recorded when the person's initial (left) foot is moved forward and is parallel to the standing foot. Weight is still on the other (right) foot. The knee of the left foot is slightly bent, the foot above ground.
4. Afterwards, the person extends the left leg forward, placing the heel on the ground, while body weight is still mainly on the right foot.
5. Again, body weight is shifted during the step such that the person lifts the heel of the right foot and is now standing on the left foot.
6. In the last pose of the sequence, the person pulls the right leg forward. The recorded pose is symmetric to pose 3.

A walking person will periodically follow this sequence of poses. Due to the similarity of pose 3 and 6 to a standing pose, it is hard to decide binary between a walking and

a standing person based on a single image, even for a human using an optical image, let alone an automatic classification based on an acoustic image. We therefore exclude poses 3 and 6 in our analysis and restrict the problem to discriminating between persons standing still and persons in poses 1, 2, 4 and 5. We trained several classifiers on the two classes of walking and standing humans using all features described in Section 5.3. From the data set available, we used 24 images from pose 1 and 4 together with 22 images from poses 2 and 5 for the class of walking persons (W) and 27 images of standing humans (S).

While we generally can obtain classifiers with a satisfactory classification accuracy, the results depend more heavily on the selection of the training set than in the human detection problem. In Figure 5.8, we show the classification accuracy of different classifiers when the training and test sets are chosen randomly for a Monte Carlo simulation based on 100 runs. Clearly, the accuracy varies significantly between 55% and values above 95%. The mean accuracy is 77.26% for LDA, 84.17% for QDA, 76.83% for the SVM with a linear kernel and 74.13% for the SVM using Gaussian radial basis functions. However, we see that QDA outperforms the other classifiers in that it has not only a higher mean accuracy, but its distribution is skewed to the left and its mass is concentrated in the high values of accuracy. The reason for the lower accuracy of the SVMs is that the data set also contains outliers in the feature space, e.g., due to some aberrant shapes in the images. SVMs are known to be sensitive to outliers in the training set because they try to minimize the penalty terms of the soft margins and, thus, the decision boundary is greatly affected by aberrant points. A more detailed analysis of the confusion matrices shows that the classification of walking humans is relatively robust to the selection of the training data. Figure 5.9 shows that for the large majority of the Monte Carlo runs, the class is correctly identified with an accuracy of more than 90%. It is the class of standing humans that is more sensitive to the selection of training data. The resulting classifiers are therefore all biased towards the class of walking humans. We see that again QDA is not only robust, but achieves on average the highest classification accuracy. It was also the only classifier that achieved a perfect classification in a single Monte Carlo run.

Table 5.3 shows the performance of the classifiers under the use of cross-validation. We can see that the total accuracy ranges from 79% to more than 90%, with the QDA being again the best classifier. Also the LDA performs reasonable, both classes are correctly identified with comparable rates. Again, the SVMs perform worse which is again caused by the presence of outliers in the feature space. This effect is present for different block lengths in the cross-validation. Since the number of available data from the standing persons is rather small, this effect influences the overall performance of the SVMs. However, the performance can be improved by postfiltering the results

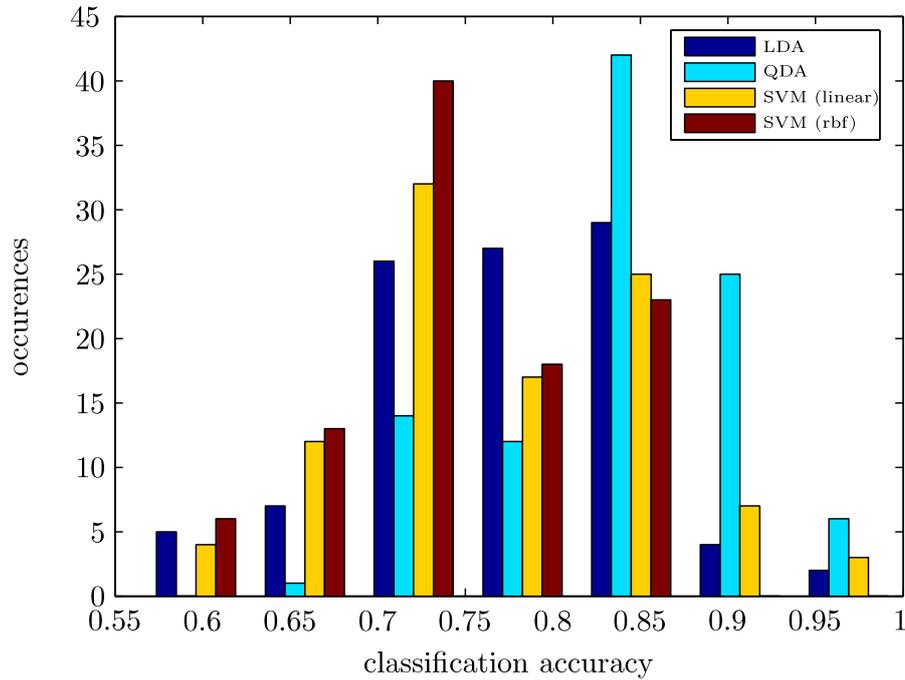


Figure 5.8: Histogram of the classification accuracy for 100 randomly selected training (70% of the data) and test (30% of the data) sets.

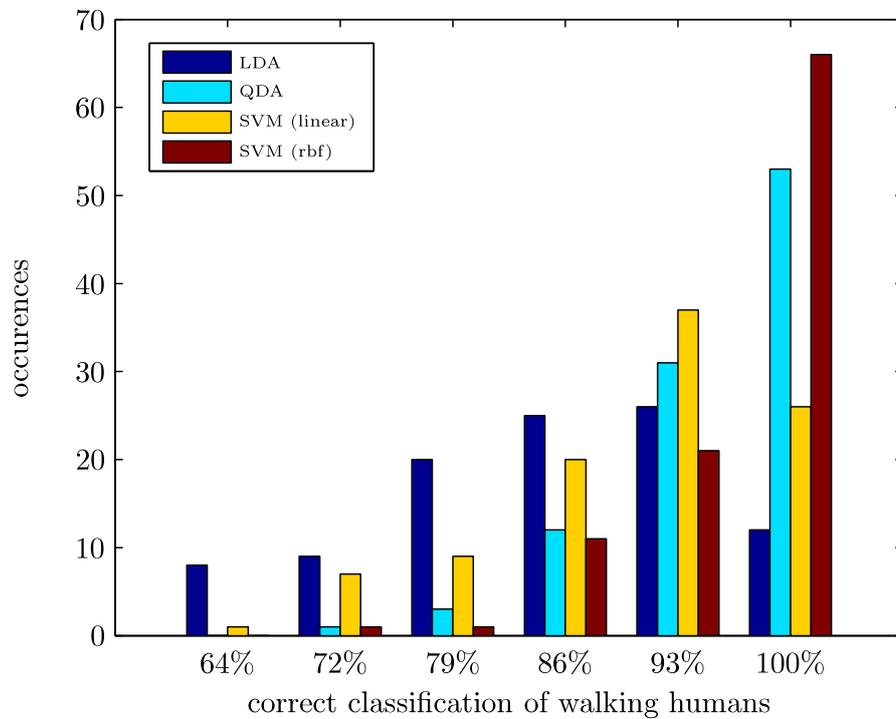


Figure 5.9: Histogram of the classification accuracy for 100 randomly selected training (70% of the data) and test (30% of the data) sets for all walking humans. The majority of the classifiers result in a correct classification of more than 90% for this class.

Method	total accuracy	class	classified as	
			W	S
LDA	87.31%	W	86.30%	13.70%
		S	10.96%	89.04%
QDA	91.89%	W	100%	0%
		S	21.92%	78.08%
SVM (linear)	79.21%	W	86.30%	13.70%
		S	32.88%	67.12%
SVM (rbf)	79.47%	W	93.15%	6.85%
		S	43.84%	56.16%

Table 5.3: Classification performance of Linear Discriminant Analysis, Quadratic Discriminant Analysis and Support Vector Machines with linear and Gaussian kernels for the classes of walking and standing humans using cross-validation.

of the elliptic fitting in order to detect aberrant shapes prior to the classification. Alternatively, using more training data will also improve the performance of the SVMs such that they perform equally good or even better than the other classifiers.

Chapter 6

Conclusions and Outlook

This chapter presents an overview of the conclusions drawn from results in the preceding chapters as well as an outlook to possible future work.

6.1 Conclusions

In this thesis, we have discussed the design of active acoustic imaging systems using acoustic arrays for general object detection tasks as well as object classification in the context of robotic applications. Acoustic imaging can greatly increase the reliability of scene analysis by offering an angular and spatial resolution and direct range measurements independently of any other sensor entity. Furthermore, the echoes depend mainly on the shape of an object, not its material. It is therefore possible to compensate limitations and weaknesses of other sensors such as cameras, radar or lidar. Due to their cheap costs and reliability even in most severe conditions, acoustic sensors can be used for imaging and are a valuable tool in all situations where objects have to be detected and analyzed. We have discussed basic principles of active acoustic imaging systems that operate in air and presented an approach to calibrate acoustic arrays which are affected by position errors and other, direction-independent errors using a method that overcomes some limitations of previous methods and is computationally simple. This allows the calibration of acoustic arrays using only a low number of calibration sources. Due to the cost-sensitive nature of some applications, we also emphasized on the problem of sparse array design. We presented approaches to design highly sparse arrays with good resolution and noise suppression capabilities and demonstrated their effectiveness by both simulations and application to real data measurements. The images obtained using those sparse layouts show the significant echoes and shapes of objects in the scene. For large reflecting areas, the power reception can be altered. However, this is not affecting the ability to detect objects and model their shapes in a 3D fashion. On the functional level, we have developed features which allow a reliable detection of persons and distinguishes them from other objects. The obtained classifiers allow correct classification of human and non-human objects with a rate of almost 97 percent. Furthermore, we have shown that it is possible to classify the pose of a person even further, e.g., we presented a feature set which allows to distinguish between walking and standing persons using not only geometrical features, but also

statistical features obtained from the time-series data of the system. The features have been chosen using a feature selection method based on mutual information criteria. Based on this feature set, a correct classification rate of 87–90 percent can be achieved.

6.2 Outlook

The system used throughout this work was so far limited to narrowband excitation signals. Therefore, it is worthwhile to consider broadband excitation which is expected to allow even higher spatial resolution, if needed. Additionally, the obtained echoes from the objects could be analyzed for different frequencies which extends the possibilities to infer information about texture and shape of the objects. Another intuitive step to take of the presented work is the extension to dynamic scenes, i.e., to track objects over time. This would allow infer more information about more complicated objects or estimate and predict trajectories in a scene. By analyzing echoes obtained from different viewing angles, it could be possible to obtain a more precise model of the reflecting objects and model the surrounding scene of the platform more accurately. With respect to sparse array design, a natural extension of the presented methods would be to apply the design concepts to 3D arrays. In that context, the pressure of reducing the number of sensors is even higher, and we expect similar gains in sparsity by applying the theory of array redundancy and controlled manipulation thereof. It would also be worthwhile to analyze the relation between the non-uniformity of arrays and their performance in application-specific figures of merit. For example, some applications of acoustic imaging require a reliable detection not only of the object it self, but also its shape. Simply using the mean square error (MSE) in such situations is not sufficient and it would be beneficial to have a an effective way of measuring what the effect of an increased sparsity of the array is in terms of reduced quality of shape representation. On the other hand, in DOA estimation problems, the effect of sparsity can not be measured by the Cramer-Rao-Bound, but more general bounds have to be taken into account. With a more general figure of merit, one could more easily design arrays with maximal sparsity, if the desired estimation accuracy is known.

The presented results on human presence detection and classification can be extended to other classes. For example, it should be possible to classify the pose of a human not only by movement, but also by the orientation relative to the array, e.g., whether the person is facing the platform or looking to left, to the right or away from the platform. Moreover, it might be possible to even recognize simple gestures in 3D and therefore improve the gesture recognition of a robot by combining acoustic imaging with optical cameras. As stated above, also the classification of dynamic scenes is possible

and the classifier could be integrated in a framework to obtain reliable detection and classification results over a period of time.

Bibliography

- [AGGS96] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, and N. K. Spencer. Comparison of doa estimation performance for various types of sparse antenna array geometries. In *Proceedings of the EUSIPCO*, 1996.
- [AGGS98] Y.I. Abramovich, D.A. Gray, A.Y. Gorokhov, and N.K. Spencer. Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays. I. fully augmentable arrays. *IEEE Trans. Signal Process.*, 46(9):2458–2471, 1998.
- [AH02] A. Austeng and S. Holm. Sparse 2-D arrays for 3-D phased array imaging - design methods. *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, 49(8):1073–1086, Aug 2002.
- [AHW⁺97] A. Austeng, S. Holm, P.K. Weber, N. Aakvaag, and K. Iranpour. 1D and 2D algorithmically optimized sparse arrays. In *Proceedings of the IEEE Ultrasonics Symposium*, volume 2, pages 1683–1686 vol.2, 1997.
- [Aka74] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Autom. Control*, AC 19:716–723, 1974.
- [AS72] M Abramowitz and I A Stegun. Handbook of mathematical functions with formulas, graphs, and mathematical tables, 1972.
- [ASG99a] Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov. Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays. II. partially augmentable arrays. *IEEE Trans. Signal Process.*, 47(6):1502–1521, 1999.
- [ASG99b] Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov. Resolving manifold ambiguities in direction-of-arrival estimation for nonuniform linear antenna arrays. *IEEE Trans. Signal Process.*, 47(10):2629–2643, 1999.
- [AW89] A.S. Acampora and J.H. Winters. Three-dimensional ultrasonic vision for robotic applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(3):291–303, 1989.
- [BB99] Dimitri P. Bertsekas and Dimitri P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition, September 1999.
- [Bed86] S.D. Bedrosian. Nonuniform linear arrays: Graph-theoretic approach to minimum redundancy. *Proc. of the IEEE*, 74(7):1040–1043, 1986.
- [BH09] Nicola Bellotto and Huosheng Hu. Multisensor-based human detection and tracking for mobile service robots. *IEEE Trans. Syst., Man, Cybern. B*, 39(1):167–181, 2009.
- [BI98] A. Blake and M. Isard. *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion*. Springer-Verlag New York, Inc., New York, USA, 1998.

- [Bis07] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1st edition, October 2007.
- [BK04] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(9):1124–1137, 2004.
- [Boh88] D.A. Bohn. Environmental effects on the speed of sound. *J. of Audio Engineering*, 36(4):223–231, April 1988.
- [BV98] K.L. Bell and H.L. Van Trees. Adaptive beamforming for spatially spread sources. In *Proceedings of the Ninth IEEE SP Workshop on Statistical Signal and Array Processing*, pages 1–4, 1998.
- [Cab08] Cabinet Office of Japan. Strategy for innovative technology. Technical report, Council for Science and Technology Policy, 2008.
- [Cap69] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969.
- [CD00] R. Cutler and L.S. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):781–796, 2000.
- [CJS93] Z.-H. Cho, J. P. Jones, and M. Singh. *Foundations of Medical Imaging*. Wiley, New York, 1993.
- [CST00] N. Cristianini and J. Shawe-Taylor. *An introduction to Support Vector Machines*. Cambridge University Press, March 2000.
- [CV01] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Trans. Image Process.*, 10(2):266–277, 2001.
- [CZO87] H. Cox, R. Zeskind, and M. Owen. Robust adaptive beamforming. *IEEE Trans. Acoust., Speech, Signal Process.*, 35(10):1365–1376, Oct 1987.
- [DD94a] M.A. Doron and E. Doron. Wavefield modeling and array processing, part i - spatial sampling. *IEEE Trans. Signal Process.*, 42(10):2549–2559, Oct 1994.
- [DD94b] M.A. Doron and E. Doron. Wavefield modeling and array processing, part ii - algorithms. *IEEE Trans. Signal Process.*, 42(10):2560–2570, Oct 1994.
- [DD94c] M.A. Doron and E. Doron. Wavefield modeling and array processing, part iii - resolution capacity. *IEEE Trans. Signal Process.*, 42(10):2571–2580, Oct 1994.
- [DHS01] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley, New York, 2001.
- [DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

- [DLX00] K.R. Dandekar, H. Ling, and G. Xu. Smart antenna array calibration procedure including amplitude and phase mismatch and mutual coupling effects. In *IEEE International Conference on Personal Wireless Communications*, pages 293–297, 2000.
- [EJS82] J.E. Evans, J.R. Johnson, and D.F. Sun. High resolution angular spectrum estimation techniques for terrain scattering analysis and angle of arrival estimation in atc navigation and surveillance system. Technical Report 582, MIT Lincoln Lab, Lexington, MA, 1982.
- [FD96] A. Fenster and D.B. Downey. 3-D ultrasound imaging: a review. *IEEE Engineering in Medicine and Biology Magazine*, 15(6):41–51, 1996.
- [fE10] United Nations Economic Commission for Europe. Statistical database - road accidents by accident type, light condition, country and year, 2010.
- [FF95] Andrew Fitzgibbon and Robert B. Fisher. A buyer’s guide to conic fitting. In *In British Machine Vision Conference*, pages 513–522, 1995.
- [FPF99] A. Fitzgibbon, M. Pilu, and R.B. Fisher. Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(5):476–480, 1999.
- [FW92] B. Friedlander and A.J. Weiss. Direction finding using spatial smoothing with interpolated arrays. *IEEE Trans. Aerosp. Electron. Syst.*, 28(2):574–587, 1992.
- [GCB⁺08] S. Gidel, P. Checchin, C. Blanc, T. Chateau, and L. Trassoudaine. Pedestrian detection method using a multilayer laserscanner: Application in urban environment. In *International Conference on Intelligent Robots and Systems*, pages 173–178, 2008.
- [GGS94] W. Gander, G.H. Golub, and R. Strebler. Least-squares fitting of circles and ellipses. *BIT Numerical Mathematics*, 43(4):558–578, Dec 1994.
- [HAIH01] S. Holm, A. Austeng, K. Iranpour, and J.-F. Hopperstad. *Nonuniform Sampling - Theory and Practice*, chapter 19, pages 787–833. Kluwer, New York, 2001.
- [HB91] Y.D. Huang and M. Barkat. Near-field multiple source localization by passive sensor array. *IEEE Trans. Antennas Propag.*, 39(7):968–975, 1991.
- [Hen91] W.J. Hendricks. The totally random versus the bin approach for random arrays. *IEEE Trans. Antennas Propag.*, 39(12):1757–1762, 1991.
- [Hil75] B.M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, Sep. 1975.
- [HJOK85] S. Haykin, J.H. Justice, N.L. Owsley, and A.C. Kak. *Array Signal Processing*. Prentice-Hall, New Jersey, USA, 1985.

- [HMJG06] M. Heckmann, M. Moebus, F. Joublin, and C. Goerick. Speaker independent voiced-unvoiced detection evaluated in different speaking styles. In *Proc. of Interspeech 2006, International Conference on Spoken Language Processing*, Pittsburgh, USA, September 2006.
- [Hol00] S. Holm. Sparse and irregular sampling in array processing. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 3850–3853 vol.6, 2000.
- [Hsi99] P.-H. Hsieh. Robustness of tail index estimation. *Journal of Computational and Graphical Statistics*, 8(2):318–332, 1999.
- [Hun00] E.K.L. Hung. Matrix-construction calibration method for antenna arrays. *IEEE Trans. Aerosp. Electron. Syst.*, 36(3):819–828, 2000.
- [ISO10] ISO 17386:2010: Transport information and control systems – manoeuvring aids for low speed operation (MALSO) – performance requirements and test procedures, 2010.
- [JP AE09] Graziella Jost, Marco Popolizio, Richard Allsop, and Vojtech Eksler. 3rd road safety annual report - 2010 on the horizon. Technical report, European Transport Safety Council (ETSC), 2009.
- [JRS07] N. Jin and Y. Rahmat-Samii. Advances in particle swarm optimization for antenna designs: Real-number, binary, single-objective and multiobjective implementations. *IEEE Trans. Antennas Propag.*, 55(3):556–567, 2007.
- [JZ97] A. Jain and D. Zongker. Feature selection: evaluation, application, and small sample performance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(2):153–158, 1997.
- [Kop92] L.E. Kopilovich. New approach to constructing two-dimensional aperture synthesis systems. *IEE Proc. F, Radar and Signal Processing*, 139(5):365–368, 1992.
- [Kop08] L. E. Kopilovich. Square array antennas based on Hadamard difference sets. *IEEE Trans. Antennas Propag.*, 56(1):263–266, Jan 2008.
- [KS91] L.E. Kopilovich and L.G. Sodin. Two-dimensional aperiodic antenna arrays with a low sidelobe level. *IEE Proc. H, Microwaves, Antennas and Propagation*, 138(3):233–237, 1991.
- [KV96] H. Krim and M. Viberg. Two decades of array signal processing research. *IEEE Signal Process. Mag.*, 13(4):67–94, 1996.
- [KWT88] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *INTERNATIONAL JOURNAL OF COMPUTER VISION*, 1(4):321–331, 1988.
- [Lee56] J. Leech. On the representation of 1, 2, ..., n by differences. *J. of the London Mathematical Society*, 31:160–169, 1956.

- [Lee99] D.G. Leeper. Isophoric arrays-massively thinned phased arrays with well-controlled sidelobes. *IEEE Trans. Antennas Propag.*, 47(12):1825–1835, Dec 1999.
- [Lie09] Bernd Liepert. The strategic research agenda for robotics in europe. Technical report, European Robotics Technoloy Platform, 2009.
- [Lin92] D.A. Linebarger. A fast method for computing the coarray of sparse linear arrays. *IEEE Trans. Antennas Propag.*, 40(9):1109–1112, 1992.
- [Lit09] Gudrun Litzenberger. World robotics. Technical report, International Federation of Robotics, 2009.
- [LLV06] M. Lanne, A. Lundgren, and M. Viberg. Calibrating an array with scan dependent errors using a sparse grid. In *Fortieth Asilomar Conference on Signals, Systems and Computers (ACSSC)*, pages 2242–2246, 2006.
- [Lo63] Y. Lo. A probabilistic approach to the design of large antenna arrays. *IEEE Trans. Antennas Propag.*, 11(1):95–96, 1963.
- [Lo64] Y. Lo. A mathematical theory of antenna arrays with randomly spaced elements. *IEEE Trans. Antennas Propag.*, 12(3):257–268, 1964.
- [LST93] D.A. Linebarger, I.H. Sudborough, and I.G. Tollis. Difference bases and sparse sensor arrays. *IEEE Trans. Inf. Theory*, 39(2):716–721, 1993.
- [LSW03] J. Li, P. Stoica, and Z. Wang. On robust capon beamforming and diagonal loading. *IEEE Trans. Signal Process.*, 51(7):1702–1715, 2003.
- [LTB98] G.R. Lockwood, J.R. Talman, and S.S. Brunke. Real-time 3-D ultrasound imaging using sparse synthetic aperture beamforming. *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, 45(4):980–988, 1998.
- [LVL05] B.K. Lau, M. Viberg, and Y.-H. Leung. Data-adaptive array interpolation for DOA estimation in correlated signal environments. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages iv/945–iv/948 Vol. 4, 2005.
- [LVL07] M. Lanne, M. Viberg, and A. Lundgren. Adaptive beamforming using calibration vectors with unknown gain and phase. *IEEE Antennas and Propagation Society International Symposium*, pages 4208–4211, June 2007.
- [LWLW06] Z. Li, K. Wang, L. Li, and F.-Y. Wang. A review on vision-based pedestrian detection for intelligent vehicles. In *Vehicular Electronics and Safety, 2006. ICVES 2006. IEEE International Conference on*, pages 57–62, 2006.
- [Mar86] F. Marvasti. Signal recovery from nonuniform samples and spectral analysis on random nonuniform samples. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 11, pages 1649–1652, 1986.
- [MC90] F.A. Marvasti and L. Chuande. Parseval relationship of nonuniform samples of one- and two-dimensional signals. *IEEE Trans. Acoust., Speech, Signal Process.*, 38(6):1061–1063, 1990.

- [MD01] Y. Meurisse and J.-P. Delmas. Bounds for sparse planar and volume arrays. *IEEE Trans. Inf. Theory*, 47(1):464–468, 2001.
- [Mof68] A. Moffet. Minimum-redundancy linear arrays. *IEEE Trans. Antennas Propag.*, 16(2):172–175, 1968.
- [MT94] V. Murino and A. Trucco. Underwater 3D imaging by FFT dynamic focusing beamforming. In *Image Processing, 1994. Proc. ICIP-94., IEEE International Conference*, volume 1, pages 890–894 vol.1, 1994.
- [MT00] V. Murino and A. Trucco. Three-dimensional image generation and processing in underwater acoustic vision. *Proc. of the IEEE*, 88(12):1903–1948, 2000.
- [MVZ10] M. Moebus, M. Viberg, and A.M. Zoubir. Parametrization of acoustic images for the detection of human presence by mobile platforms. In *Proc. of the IEEE Int’l Conference on Acoustics, Speech and Signal Processing*, 2010.
- [MWW96] Yan Meng, Kon Max Wong, and Qiang Wu. Direction finding for point and dispersed sources: VEC-MUSIC and its performance. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 2908–2911, 1996.
- [Myo09] Yumiko Myoken. Research and development for next generation service robots in japan. Technical report, British Embass Japan, Science and Innovation Sectio, Jan 2009.
- [MZ07a] M. Moebus and A.M. Zoubir. A comparison of beamforming algorithms for 3D ultrasound imaging in air. In *Proceedings of the 3rd Biennial on DSP for In-Vehicle and Mobile Systems*, Istanbul, Turkey, June 2007.
- [MZ07b] M. Moebus and A.M. Zoubir. Three-dimensional ultrasound imaging in air using a 2D array on a fixed platform. In *Proc. of the IEEE Int’l Conference on Acoustics, Speech and Signal Processing*, volume 2, pages II–961–II–964, 2007.
- [MZ08a] M. Moebus and A.M. Zoubir. 2D nonuniform array design for imaging applications. In *Proc. of the Int’l ITG Workshop on Smart Antennas (WSA)*, pages 180 – 183, Darmstadt, Germany, Februar 2008.
- [MZ08b] M. Moebus and A.M. Zoubir. *In-Vehicle Corpus and Signal Processing for Driver Behavior*, chapter 13, pages 147–159. Springer, 2008.
- [MZ08c] M. Moebus and A.M. Zoubir. On the design of sparse arrays using difference sets. In *Proc. of the 5th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 247–251, Darmstadt, Germany, 2008.
- [MZ09a] M. Moebus and A.M. Zoubir. Local array calibration using parametric modeling of position errors and a sparse calibration grid. In *Proc. of the Int’l IEEE Workshop on Statistical Signal Processing*, Cardiff, Wales, UK, Sep 2009.

- [MZ09b] M. Moebus and A.M. Zoubir. Verfahren zum Entwerfen einer Anordnung von Sensoren, 2009. in application phase, app.-no. 10 2009 044 230.8.
- [MZ10a] M. Moebus and A.M. Zoubir. Detection and classification of humans using acoustic imaging. *Pattern Recognition*, 2010. under review.
- [MZ10b] M. Moebus and A.M. Zoubir. Ultrasound object detection using 2D sparse arrays with low sidelobes. *Signal Processing*, 2010. under review.
- [NS96] B. C. Ng and C. M. S. See. Sensor-array calibration using a maximum-likelihood approach. *IEEE Trans. Antennas Propag.*, 44(6):827–835, 1996.
- [NW99] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer, New York, 2nd edition, July 1999.
- [OM05] U. Oktel and R.L. Moses. A Bayesian approach to array geometry design. *IEEE Trans. Signal Process.*, 53(5):1919–1923, May 2005.
- [OV91] B. Ottersten and M. Viberg. Local modeling and robust estimation for high-resolution direction finding. In *Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1005–1009, 1991.
- [PBNH85] S.U. Pillai, Y. Bar-Ness, and F. Haber. A new approach to array geometry for improved spatial spectrum estimation. *Proceedings of the IEEE*, 73(10):1522–1524, Oct. 1985.
- [PDM98] O.V. Pictet, M.M. Dacorogna, and U.A. Müller. *A practical guide to heavy tails: statistical techniques and applications*, chapter 3, pages 283–310. Birkhäuser Boston Inc., Cambridge, MA, USA, 1998.
- [PH96] D.J. Powell and G. Hayward. Flexible ultrasonic transducer arrays for nondestructive evaluation applications. I. The theoretical modeling approach. *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, 43(3):385–392, 1996.
- [PK91] J. Pierre and M. Kaveh. Experimental performance of calibration and direction-finding algorithms. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1365–1368 vol.2, 1991.
- [PLD05] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(8):1226–1238, 2005.
- [PPL90] D. Pearson, S.U. Pillai, and Y. Lee. An algorithm for near-optimal placement of sensor elements. *IEEE Trans. Inf. Theory*, 36(6):1280–1284, 1990.
- [Ric07] C.D. Richmond. Cross coherence and joint pdf of the bartlett and capon power spectral estimates. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages II–901–II–904, 2007.

- [Ric10] C.D. Richmond. Statistical analysis of capon-bartlett 2-d cross spectrum. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010.
- [RL75] W. Richards and Yuen Lo. Antenna pattern synthesis based on optimization in a probabilistic sense. *IEEE Trans. Antennas Propag.*, 23(2):165–172, 1975.
- [Ruf93] C.S. Ruf. Numerical annealing of low-redundancy linear arrays. *IEEE Trans. Antennas Propag.*, 41(1):85–90, 1993.
- [RW84] Richard A. Redner and Homer F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, Apr., 1984.
- [SC08] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer, 2008.
- [Sch86] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.*, 34(3):276–280, 1986.
- [Set99] J. A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press, June 1999.
- [SK85] T.-J. Shan and T. Kailath. Adaptive beamforming for coherent signals and interference. *IEEE Trans. Acoust., Speech, Signal Process.*, 33(3):527 – 536, Jun 1985.
- [SK92] A. L. Swindlehurst and T. Kailath. A performance analysis of subspace-based methods in the presence of model errors. I. The MUSIC algorithms. *IEEE Trans. Signal Process.*, 40(7):1758–1774, July 1992.
- [SK93] A. L. Swindlehurst and T. Kailath. A performance analysis of subspace-based methods in the presence of model error. II. Multidimensional algorithms. *IEEE Trans. Signal Process.*, 41(9):2882–2890, Sept. 1993.
- [SS98] A. L. Swindlehurst and P. Stoica. Maximum likelihood methods in radar array signal processing. *Proc. IEEE*, 86(2):421–441, Feb. 1998.
- [SS04] P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Process. Mag.*, 21(4):36–47, 2004.
- [Ste72] B. Steinberg. The peak sidelobe of the phased array having randomly located elements. *IEEE Trans. Antennas Propag.*, 20(2):129–136, 1972.
- [Ste00] S. Stergiopoulos. *Advanced Signal Processing Handbook: Theory and Implementation for Radar, Sonar, and Medical Imaging Real-Time Systems*. CRC Press, Inc., 2000.

- [STI08] Go Shingu, Kenichi Takizawa, and Tetushi Ikegami. Human body detection using mimo-uwb radar sensor network in an indoor environment. *International Conference on Parallel and Distributed Computing Applications and Technologies*, 0:437–442, 2008.
- [SWK85] T.-J. Shan, M. Wax, and T. Kailath. On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Trans. Acoust., Speech, Signal Process.*, 33(4):806 – 811, Aug 1985.
- [Tap02] M. Tapio. Direction and spread estimation of spatially distributed signals via the power azimuth spectrum. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, volume 3, pages 3005–3008, 2002.
- [TBPI10] F.H.C. Tivive, A. Bouzerdoum, S.L. Phung, and K.M. Iftexharuddin. Adaptive hierarchical architecture for visual recognition. *Appl. Opt.*, 49(10):B1–B8, April 2010.
- [TDMO04] M. Tons, R. Doerfler, M.-M. Meinecke, and M.A. Obojski. Radar sensors and sensor platform used for pedestrian protection in the ec-funded project save-u. In *Intelligent Vehicles Symposium, 2004 IEEE DOI - 10.1109/IVS.2004.1336489*, pages 813–818, 2004.
- [Tru99] A. Trucco. Thinning and weighting of large planar arrays by simulated annealing. *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, 46(2):347–355, 1999.
- [Van02] H. L. Van Trees. *Detection, Estimation and Modulation Theory Part IV - Optimum Array Signal Processing*, volume 4. Wiley, 2002.
- [Vib10] M. Viberg. Error modeling and calibration for high resolution doa estimation. *Signal Processing*, 90(5):279–282, May 2010.
- [VO91] M. Viberg and B. Ottersten. Sensor array processing based on subspace fitting. *IEEE Trans. Signal Process.*, 39(5):1110–1121, 1991.
- [VS94a] M. Viberg and A. L. Swindlehurst. Analysis of the combined effects of finite samples and model errors on array processing performance. *IEEE Trans. Signal Process.*, 42(11):3073–3083, Nov. 1994.
- [VS94b] M. Viberg and A.L. Swindlehurst. A Bayesian approach to auto-calibration for parametric array signal processing. *IEEE Trans. Signal Process.*, 42(12):3495–3507, 1994.
- [VVB88] B.D: Van Veen and K.M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Mag.*, 5(2):4–24, 1988.
- [WF93] A.J. Weiss and B. Friedlander. Performance analysis of spatial smoothing with interpolated arrays. *IEEE Trans. Signal Process.*, 41(5):1881–1892, 1993.
- [WHO09] Global status report on road safety, 2009.

- [WK85] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust., Speech, Signal Process.*, 33(2):387 – 392, 1985.
- [WWMR94] Q. Wu, K.M. Wong, Y. Meng, and W. Read. DOA estimation of point and scattered sources - VET-MUSIC. In *IEEE Seventh SP Workshop on Statistical Signal and Array Processing*, pages 365–368, 1994.
- [WZ88] M. Wax and I. Ziskind. Detection of fully correlated signals by the mdl principle. *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 5:2777 – 2780, 1988.
- [WZ89] M. Wax and I. Ziskind. Detection of the number of coherent signals by the mdl principle. *IEEE Trans. Acoust., Speech, Signal Process.*, 37(8):1190 – 1196, 1989.
- [YTN05] N. Yamada, Y. Tanaka, and K. Nishikawa. Radar cross section for pedestrian in 76GHz band. In *European Microwave Conference*, volume 2, Oct. 2005.
- [ZL04] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.

Curriculum Vitae

Education & Work Experience

- Since Oct. 2010 Project Leader Advanced Active Safety
General Motors Europe, Adam Opel GmbH
Rüsselsheim, Germany
- July 2009 Visiting researcher at the division "Signal Processing & Antennas",
Department of Signals & Systems, Chalmers University of Technology
Göteborg, Sweden
- Apr. 2006 - Sep. 2010 Research Associate at the Signal Processing Group,
Institute of Telecommunications, Technische Universität Darmstadt
Darmstadt, Germany
- Feb. 2006 - Mar. 2006 Visitor at the Communications and Signal Processing Lab,
Curtin University of Technology
Perth, Australia
- Oct. 2005 Diploma (Dipl.-Wirtsch.-Ing.) for the combined studies of
electrical engineering and business administration,
Technische Universität Darmstadt
Darmstadt, Germany
Grade "very good" (1.4)
- Apr. 2006 - Sep. 2010 Diploma thesis on "Speaker-independent Voiced/Unvoiced
Detection in Spontaneous and Read Speech"
Honda Research Institute Europe, Offenbach, Germany
- Apr. 2006 - Sep. 2010 Study thesis on "Finance problems in project portfolio management" ,
at Campana & Schott, Frankfurt, Germany
- Dec. 2003 - Apr. 2004 Study thesis on "Detection of Digital Watermarks in
Geodetic Map Data Using the Bootstrap"
Institute of Telecommunications, Technische Universität Darmstadt
Darmstadt, Germany
- Okt. 1999 - Mar 2002 Study of electrical engineering and business administration
Technische Universität Darmstadt, Darmstadt, Germany
(Vordiplom "good", 2nd best of class)
- Apr. 2006 - Sep. 2010 Abitur at Elly-Heuss-Gymnasium, Wiesbaden
Grade 1.1

Publications

A.1 Internationally Refereed Publications

- M. Heckmann, M. Moebus, F. Joublin, and C. Goerick. Speaker independent voiced-unvoiced detection evaluated in different speaking styles. In *Proc. of Interspeech 2006, International Conference on Spoken Language Processing*, Pittsburgh, USA, September 2006
- M. Moebus and A.M. Zoubir. Three-dimensional ultrasound imaging in air using a 2D array on a fixed platform. In *Proc. of the IEEE Int'l Conference on Acoustics, Speech and Signal Processing*, volume 2, pages II-961–II-964, 2007
- M. Moebus and A.M. Zoubir. A comparison of beamforming algorithms for 3D ultrasound imaging in air. In *Proceedings of the 3rd Biennial on DSP for In-Vehicle and Mobile Systems*, Istanbul, Turkey, June 2007
- M. Moebus and A.M. Zoubir. 2D nonuniform array design for imaging applications. In *Proc. of the Int'l ITG Workshop on Smart Antennas (WSA)*, pages 180 – 183, Darmstadt, Germany, Februar 2008
- M. Moebus and A.M. Zoubir. *In-Vehicle Corpus and Signal Processing for Driver Behavior*, chapter 13, pages 147–159. Springer, 2008
- M. Moebus and A.M. Zoubir. On the design of sparse arrays using difference sets. In *Proc. of the 5th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pages 247–251, Darmstadt, Germany, 2008
- M. Moebus and A.M. Zoubir. Local array calibration using parametric modeling of position errors and a sparse calibration grid. In *Proc. of the Int'l IEEE Workshop on Statistical Signal Processing*, Cardiff, Wales, UK, Sep 2009
- M. Moebus, M. Viberg, and A.M. Zoubir. Parametrization of acoustic images for the detection of human presence by mobile platforms. In *Proc. of the IEEE Int'l Conference on Acoustics, Speech and Signal Processing*, 2010
- M. Moebus and A.M. Zoubir. Ultrasound object detection using 2D sparse arrays with low sidelobes. *Signal Processing*, 2010. under review
- M. Moebus and A.M. Zoubir. Detection and classification of humans using acoustic imaging. *Pattern Recognition*, 2010. under review

A.2 Filed Patent Applications

- M. Moebus and A.M. Zoubir. Verfahren zum Entwerfen einer Anordnung von Sensoren, 2009. in application phase, app.-no. 10 2009 044 230.8

