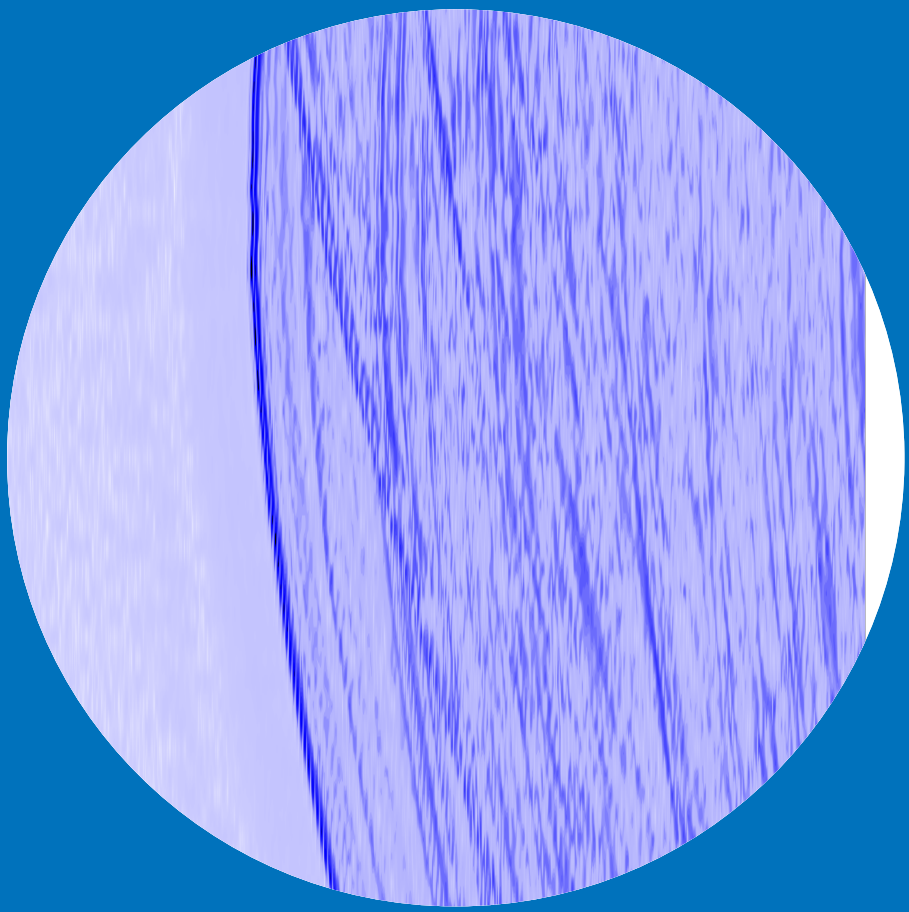
Department of Media Technology

Localization and tracing of early acoustic reflections

Sakari Tervo



Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 143/2011

Localization and tracing of early acoustic reflections in enclosures

Sakari Tervo

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the School of Science for public examination and debate in Auditorium T2 at the Aalto University School of Science (Espoo, Finland) on the 13th of January 2012 at 12 noon.

Aalto University
School of Science
Department of Media Technology

Supervisor
Professor Lauri Savioja

Instructor
Adjunct professor Tapio Lokki

Preliminary examiners
Prof. Angelo Farina, University of Parma, Italy
Prof. Dr. ir. Emanuël A.P. Habets, International Audio Laboratories Erlangen, Germany

Opponent
Associate professor Ramani Duraiswami, University of Maryland, the United States of America

Aalto University publication series
DOCTORAL DISSERTATIONS 143/2011

ISBN 978-952-60-4437-8 (printed)
ISBN 978-952-60-4438-5 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Unigrafia Oy
Helsinki 2011

Finland

The dissertation can be read at http://lib.tkk.fi/Diss/

Publication orders (printed book):
http://lib.tkk.fi/Diss/

441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Sakari Tervo

Name of the doctoral dissertation
Localization and tracing of early acoustic reflections in enclosures

Publisher School of Science

Unit Department of Media Technology

Series Aalto University publication series DOCTORAL DISSERTATIONS 143/2011

Field of research Acoustics

Manuscript submitted 23 August 2011 **Manuscript revised** 9 November 2011

Date of the defence 13 January 2012 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

Objective room acoustic studies are conducted by measuring room impulse responses. The standard techniques include the use of an omni-directional source and, in most cases, one omni-directional microphone. This approach is well defined when measuring the standard room acoustic parameters.

Recently, early reflections, the first arriving sound waves in the room impulse response after the direct sound, have gained attention in research. The spatial location of the early reflections, i.e., the location of the image-source, can be used in room acoustic studies, auralization, room geometry inference, and in-situ measurement of acoustic properties of surfaces from room impulse responses. The location, however, cannot be obtained from the standard room impulse response measurement. Therefore, special microphone array techniques have been used for spatial analysis of room impulse responses.

This thesis studies the localization of early reflections. Firstly, a measurement technique of room impulse responses with directional loudspeakers is proposed. This allows better spatial and temporal separability between the reflections than the standard omni-directional loudspeaker.

Secondly, the use of microphone array techniques on the localization of early reflections is studied. Several techniques used in other localization tasks are transformed and applied for the localization of early reflections. In detail, the combination of time of arrival and time difference of arrival is researched. Moreover, interpolation of the time difference of arrival estimation function is proposed. The use of sound intensity vector based localization is also considered.

Finally, novel ad-hoc localization techniques for early reflections are proposed. Results for theoretical, simulation, and real data experiments are presented.

Keywords Early reflections, localization, room acoustics

ISBN (printed) 978-952-60-4437-8 **ISBN (pdf)** 978-952-60-4438-5

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Espoo **Location of printing** Helsinki **Year** 2012

Pages 202 **The dissertation can be read at** http://lib.tkk.fi/Diss/

Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Sakari Tervo

Väitöskirjan nimi
Varhaisten akustisten heijastusten paikannus huoneissa

Julkaisija Perustieteiden korkeakoulu

Yksikkö Mediatekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 143/2011

Tutkimusala Akustiikka

Käsikirjoituksen pvm 23.08.2011 **Korjatun käsikirjoituksen pvm** 09.11.2011

Väitöspäivä 13.01.2012 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Huoneakustiset tutkimukset suoritetaan mittaamalla huoneimpulssivasteita. Standardin mukaan lähteen tulisi olla ympärisäteilevä. Suurimmassa osassa mittauksista käytetään yhtä mikrofonia. Tämä menetelmä soveltuu hyvin standardin mukaisten huoneakustisten parametrien mittaukseen.

Viimeaikoina tarve ymmärtää varhaisten heijastusten vaikutusta akustiikassa on lisääntynyt. Varhaisten heijastusten tiedetään vaikuttavan havaittuun tilaääneen merkittävästi, etenkin konserttisalissa. Ensimmäinen askel varhaisten heijastusten tutkimisessa on niiden paikantaminen. Tarkasti tiedetty paikka auttaa heijastuksen piirteiden tarkastelussa. Heijastuksen paikkaa ei voi kuitenkaan arvioida tai mitata standardin mukaisista huoneimpulssivastemittauksista. Siksi heijastuksen paikannukseen ja huoneakustiikan tila-analyysiin on käytetty erilaisia monimikrofonimenetelmiä.

Tässä väitöskirjassa tutkitaan heijastusten paikannusta erilaisilla signaalinkäsittely-, monimikrofoni- ja kaiutintekniikoilla. Ensiksi suuntaavia kaiuttimia ehdotetaan käytettäväksi ympärisäteilevän kaiuttimen sijasta sillä niillä voidaan saavuttaa parempi tilallinen ja ajallinen erottelu heijastuksien välille kuin ympärisäteilevällä kaiuttimella.

Toiseksi erilaisia monimikrofonitekniikoita tutkitaan ja sovelletaan heijastusten paikannukseen. Näitä paikannustekniikoita on aikaisemmin käytetty erilaisissa paikannustehtävissä monilla eri tieteenaloilla. Erityisesti tutkitaan aikaviiveiden ja aikaviiveerojen yhdistämistä paikan arvioinnissa. Lisäksi ehdotetaan interpolaatio-menetelmää parantamaan paikannuksen tarkkuutta. Ääni-intensiteetin käyttöä heijastusten paikannuksessa tutkitaan myös.

Lopuksi ehdotetaan erityismenetelmiä juuri akustisten heijastusten paikannukseen.

Avainsanat Heijastus, paikannus, huoneakustiikka

ISBN (painettu) 978-952-60-4437-8 **ISBN (pdf)** 978-952-60-4438-5

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Espoo **Painopaikka** Helsinki **Vuosi** 2012

Sivumäärä 202 **Luettavissa verkossa osoitteessa** http://lib.tkk.fi/Diss/

Preface

The research work for this thesis has been carried out in the Department of Media Technology in Helsinki University of Technology 2008-2009, and in Aalto University 2010-2011. Part of the research was initially started during a research visit to Philips research in 2007. Moreover, some parts of the thesis were written during another research visit to University of York in 2010. The work has been supported by Helsinki Doctoral Programme in Computer Science, the Finnish Foundation for Technology Promotion, and the Nokia Foundation. The work has also received funding from the Academy of Finland, project no. [119092], the European Research Council under the European Community's Seventh Framework Programme / ERC grant agreement no. [203636], and European Cooperation in Science and Technology [TD0804].

Firstly, I wish to acknowledge my supervisor, Prof. Lauri Savioja, and my instructor, Dr. Tapio Lokki, for discussions, guidance and the inspiring atmosphere in our research group.

I thank Prof. Angelo Farina and Prof. Dr. ir. Emanuël Habets for the pre-examination of the thesis. Their expertise and valuable comments helped to improve the quality of the thesis. A special thanks goes to my co-authors Dr. Teemu Korhonen and Dr. Jukka Pätynen for the collaboration in research, and to Mr. Philip Robinson for proofreading the manuscript.

I wish to thank all the people that I have had the pleasure working or discussing with during the course of my work. Thanks to all my co-workers in our department, Dr. Jukka Pätynen, Dr. Sampo Vesa, Dr. Samuel Siltanen, Dr. Alex Southern, Mr. Raine Kajastila, Mr. Hannes Gamper, Mr. Antti Kuusinen, Mr. Heikki Vertanen, Mr. Philip Robinson, Mr. Robert Albrecht, and Dr. Timo Tossavainen, it has been a joy working with you. I also thank my colleagues at Acoustics laboratory, especially Dr. Ville Pulkki, Mr. Marko Hiipakka, Mr. Jukka Ahonen, and Mr. Mikko-Ville Laitinen. Moreover, I would like to thank my instructors at Philips Research, Dr. Aki Härmä and Dr. Steven van de Par, and my former co-workers at Tampere University of Technology, especially Dr.

Teemu Korhonen, Dr. Pasi Pertilä, and Dr. Tuomo Pirinen. In addition, I thank Dr. Damian Murphy for collaboration during my visit to University of York.

Thanks to all my friends for keeping me entertained outside of work. Finally, I wish to express my gratitudes towards my family for supporting me during the making of this thesis.

Helsinki, November 29, 2011,

Sakari Tervo

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Sakari Tervo, Jukka Pätynen, and Tapio Lokki. Acoustic Reflection Path Tracing Using A Highly Directional Loudspeaker. In *Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009) p. 245–248*, New Paltz, NY, USA, October 18-21 2009.

II Sakari Tervo, Lauri Savioja, and Tapio Lokki. Maximum Likelihood Estimation of Loudspeaker Locations from Room Impulse Responses . *Journal of the Audio Engineering Society*, (In press) 14 pages , 2012.

III Sakari Tervo. Direction Estimation Based on Sound Intensity Vectors. In *Proceedings of the 2009 European Signal Processing Conference (EUSIPCO 2009), p. 700–704*, Glasgow, Scotland, August 24-28 2009.

IV Sakari Tervo and Teemu Korhonen. Estimation of Reflective Surfaces from Continuous Signals. In *Proceedings of the 35th International Conference on Acoustics, Speech, and Signal Processing, (ICASSP 2010), p. 153–156*, Dallas, TX, USA, March 14-19 2010.

V Sakari Tervo and Tapio Lokki. Interpolation Methods for the SRP-PHAT Algorithm. In *Proceedings of the 11th International Workshop for Acoustic Echo and Noise Control (IWAENC 2008), Article ID 9037*, Seattle, WA, USA, September 14-17 2008.

VI Sakari Tervo, Teemu Korhonen, and Tapio Lokki. Estimation of Reflections from Impulse Responses. *Journal of Building Acoustics*, Volume 18, Number 1-2, p. 159–174, March 2011.

Author's Contribution

Publication I: " Acoustic Reflection Path Tracing Using A Highly Directional Loudspeaker"

A measurement technique for the investigation of early reflections is developed. The method is based on measuring the impulse responses with a directional loudspeaker spanned over a set of angles. In addition, the direction of arrival is estimated from the sound intensity vectors. The results show that the use of directional loudspeakers provides better spatial and temporal resolution than the standard omni-directional loudspeaker for early reflections.

The present author invented the measurement technique, wrote 90 % of the article and implemented all of the experiments. Dr. Jukka Pätynen, and Dr. Tapio Lokki assisted in taking measurements for this article.

Publication II: "Maximum Likelihood Estimation of Loudspeaker Locations from Room Impulse Responses "

A method for calibrating loudspeaker locations in acoustic measurements is developed. The method uses time of arrival and time difference of arrival obtained from the direct sound of room impulse response measurements. Results show that the method outperforms previously proposed methods in the loudspeaker localization task in theory and in practical situations.

The present author invented the method, wrote 90 % of the article, and implemented all of the experiments.

Publication III: "Direction Estimation Based on Sound Intensity Vectors"

Direction estimation methods that use sound intensity vectors are compared in real situations. The comparison reveals that the mixture model-based direction estimation methods outperforms the direct average based methods. In addition, phase information is found to provide more reliable direction estimation than when the amplitude and phase information are both used.

The present author is the sole author of the article.

Publication IV: "Estimation of Reflective Surfaces from Continuous Signals"

An inverse mapping of the first order reflections in the time difference of arrival framework is given. The mapping is used together with acoustic source localization to deduce the surface location and normal from a speech signal. The inverse mapping proposed in this article is developed by Dr. Teemu Korhonen. The method is demonstrated in an auditorium.

The present author designed and conducted the experiments that validate the method in real situations, and wrote 50 % of the article. Additional results produced with the method are found in the doctoral thesis of Dr. Korhonen.

Publication V: "Interpolation Methods for the SRP-PHAT Algorithm"

An interpolation method is developed for a popular acoustic source localization algorithm, the steered response power phase transform (SRP-PHAT). The interpolation is done to the cross correlation functions which are then used by the SRP-PHAT algorithm. Experiments are conducted in a concert hall environment and results are compared to the standard Fourier-interpolation method. The results indicate that the developed method outperforms the standard method.

The present author invented the method under the supervision of Dr. Aki Härmä and Dr. Steven van De Par at Philips Research. The present author wrote 95 % of the article.

Publication VI: “Estimation of Reflections from Impulse Responses”

An overview of acoustic localization techniques for the localization of early reflections is given and a visualization example is presented. The performance of the methods for direction estimation of the early reflections is studied. In addition, approaches for diffusion estimation of the early reflections are proposed and studied.

The present author wrote 90 % of this article and implemented all of the experiments.

List of Abbreviations

2-D	Two-dimensional
3-D	Three-dimensional
CRLB	Cramér-Rao Lower Bound
FFT	Fast Fourier Transform
SIRR	Spatial Impulse Response Rendering
MSE	Mean Squared Error
MMSE	Minimum Mean Squared Error
MLE	Maximum Likelihood Estimation
PL	Pseudo-Likelihood
SRP	Steered Response Power
LS	Least squares
GCC	Generalized cross correlation
PHAT	Phase Transform
CC	Direct Cross Correlation
ASDF	Average Squared Difference Function
TDOA	Time Difference of Arrival
TOA	Time of Arrival, Time of Flight
CM	Combined Method

Some mathematical notations and symbols

a	Scalar
$\|a\|$	Absolute value of a
$j = \sqrt{-1}$	Imaginary unit
a^*	Complex conjugate
$\mathbf{a}$	Vector
$\|\mathbf{a}\|$	Euclidean distance
$\mathbf{A}$	Matrix
$\mathbf{A}^{-1}$	Inverse of $\mathbf{A}$
$\mathbf{A}^{\mathrm{T}}$	Transpose of $\mathbf{A}$
$\mathrm{trace}\{\mathbf{A}\}$	Trace of $\mathbf{A}$
$\mathbf{I}$	Identitity matrix
$\mathbf{J}(\boldsymbol{\theta})$	Fisher information matrix for $\boldsymbol{\theta}$
θ	Parameter
$\boldsymbol{\theta}$	Parameter vector
$\hat{\chi}$	Measurement
$\hat{\boldsymbol{\chi}}$	Measurement vector
ϕ, θ, φ	Angle
t	Time
f	Frequency
ω	Angular frequency
$F\{\cdot\}$	Fourier transform
$F^{-1}\{\cdot\}$	Inverse Fourier transform
$p(t)$	Time domain signal
$P(\omega)$	Frequency domain signal
$G_{p,p}(\omega)$	Auto spectral density of $p(t)$
$\mathrm{E}\{\cdot\}$	Expected value (over time)
$\mathrm{var}\{\cdot\}$	Measured variance
$\hat{\cdot}$	Estimate
$\mathrm{cov}\{\cdot\}$	Covariance
Σ	Covariance matrix
$\mathcal{N}(\mu, \sigma)$	Gaussian distribution with mean μ and variance σ^2
$\mathcal{U}(a_1, a_2)$	Uniform distribution from a_1 to a_2
$\mathcal{R}(a)$	Rayleigh distribution with parameter a
$\mathcal{E}(a)$	Exponential distribution with parameter a

Some mathematical notations and symbols continued

$p(\cdot\|\cdot)$	Probability density function
$L(\cdot\|\cdot)$	Likelihood
$\lambda(\cdot\|\cdot)$	Log-likelihood
x, y, z	Cartesian coordinates
c	Speed of sound
α	Absorption coefficient
β	Reflection coefficient
$\mathbf{s}$	Source position
$\mathbf{x}$	Source candidate position
$\mathbf{r}$	Receiver candidate position
$\triangleq$	Denote
(x, y, z)	3D coordinate location
$\sim$	Distributed according to
$\{\cdot\}$	Set
$x \in [a_1, a_2]$	x Belongs to a closed interval from a_1 to a_2
$x \in (a_1, a_2)$	x Belongs to an open interval from a_1 to a_2

1. Introduction

The location of acoustic reflections, i.e., the image-sources, is a useful piece of information in room acoustic studies, auralization, room geometry inference, and in-situ measurement of acoustic properties of surfaces from room impulse responses. In spatial room impulse response rendering [1,2] the locations of the reflections are used in spatial reproduction. Incorrect or inaccurate reflection localization will lead to incorrect auralization of the space. Moreover, the locations of the reflections can be used together with the source location to deduct the normals and the locations of the reflective surfaces [3–5], that is, to infer the room geometry. In addition, the location of the reflection is needed for accurate time windowing of the reflection from the room impulse response when estimating, for example, the absorption coefficient of the surface from in-situ measurements [6, 7].

The standardized way of studying room acoustics is to measure an impulse response using a sound source in the performance area and a microphone in the audience area [8]. The impulse response is considered to consist of three parts that have their distinct features. The direct sound arrives first, then the early reflections, followed by the late reverberation. The important difference between early reflections and late reverberation is that late reverberation or reverberation refers to the part of the impulse response, which has some specific statistical properties [9–11]. The early reflections are the discrete events before the late reverberation which do not have these statistical features.

The topic of this thesis is the objective localization of early reflections and the direct sound, using measurement devices and related applied mathematics. Instead of a mono room impulse response, a spatial room impulse response is preferred when studying the location of reflections. The spatial impulse response is measured with a microphone array instead of a single microphone. Special microphone arrays and techniques

are presented and applied to this problem [12, 13]. Typically the spatial impulse response measurement is done with techniques such that the auralization of the enclosure is also possible. In our studies [14], the auralization is based on sound intensity vector analysis and synthesis [1,2], and therefore a specially designed open spherical microphone array is used.

1.1 Scope

This thesis studies localization and tracing of early reflections, as well as calibration of measurement system, and measurement of room impulse responses. All of the analysis is based on measured spatial room impulse responses. Figure 1.1 shows the subtasks required in the localization of reflections. Reflection locations can be used in several applications, for example in speech source localization [15].

Initially, the main motivation for this study was to better explain some objective properties of the acoustics of the concert halls together with the subjective evaluations, as in [16]. This is not yet completed and it is the future work of the author.

The contributions of this thesis are shown in Table 1.1. In detail, the contributions are:

1. **Room impulse response measurement**

   A measurement technique that improves the spatial and temporal separability of reflections has been developed. The method is based on the use of highly directional loudspeaker. The method was demonstrated with a Panphonics panel loudspeaker in Publication I. Comparison between the standard omni-directional, and two directional loudspeakers is given in Chapter 4.

2. **Localization methods**

   The study of the theoretical and practical performance, and the development of localization methods in the acoustic reflection localization task is done in Publications IV, V, and VI. Some results for the theoretical performance are presented in Publication II and in Chapter 6. Additional results for practical situations are presented in Chapter 8.

Table 1.1. Contributions of this thesis to various subtasks of acoustic reflection localization.

Measurement of room impulse responses	Automatic calibration
- Sparse impulse response technique [I]	- Source position estimation [II]
Localization of reflections	**Visualization**
- Comparison of methods [VI] - Interpolation methods [V] - Sound intensity based direction estimation [III] - Localization of reflective surfaces from speech [IV]	- Tracing of reflections I

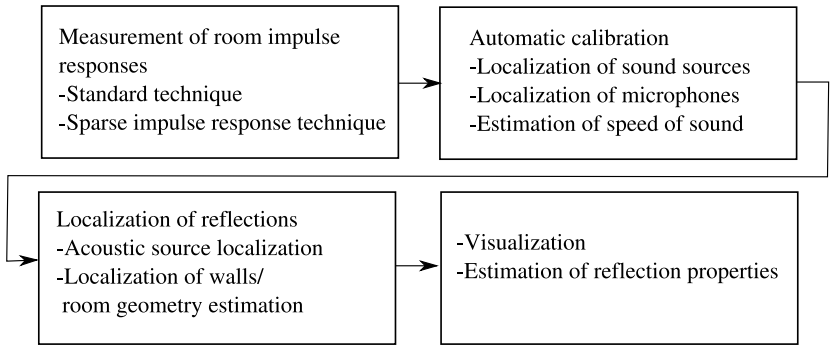


Figure 1.1. Subtasks in the localization of acoustic reflections.

3. **Calibration of the measurement system**

A method, robust with respect to noise, to be used in acoustic measurements for the calibration of the loudspeaker and microphone array positions is developed in Publication II.

4. **Visualization of reflections**

A technique for visualization of early reflections is presented in Publication I. The method is based on inversely using the ray-tracing approach. In the Appendix, a comparison between the different visualization techniques is given.

1.2 Organization

This thesis presents 6 publications and related background information. Chapter 2 gives some basic information about signal processing techniques and room acoustics. Chapter 3 lists the research related to the reflection localization. In Chapter 4, the standard measurement and the proposed room impulse response measurement techniques are presented. Relevant localization methods are reviewed in Chapter 5, theoretical and practical performance of the methods are presented in Chapters 6 and 7, respectively. Visualization examples of early reflections are provided in the Appendix. A summary of the work is given in Chapter 8.

2. Background

The goal of this thesis is to study estimation and methods related to localization of acoustic reflections. This chapter outlines the background on estimation theory, sound, and acoustics, as related to the localization of reflections in the context of this thesis.

2.1 Estimation theory

The measurement or estimation of some physical phenomenon always includes a random error. This error is due to unideal conditions in real situations and is commonly referred to as noise.

In the scope of this thesis the noise is always considered to be additive. That is, if the parameter to be measured is θ, then the measurement or the estimation can be given as [17]:

$$\hat{\theta} = \theta + \varepsilon, \tag{2.1}$$

where $\hat{\cdot}$ denotes an estimate, ε is the error term. A set of logical operations and calculations which produce the estimate is called the *estimator*. The estimator is unbiased if in overall it produces the correct value, i.e.:

$$\mathrm{E}\{\hat{\theta} - \theta\} = 0, \tag{2.2}$$

where $\mathrm{E}\{\cdot\}$ denotes the expectation. Usually the error is assumed to be normally distributed with zero mean. Within this assumption the random error term can be described by only one term, the variance:

$$\sigma_e^2 = \mathrm{var}(\varepsilon) = \mathrm{E}\{[\varepsilon - \mu_\varepsilon]^2\}, \tag{2.3}$$

where $\mu_\varepsilon = \mathrm{E}\{\varepsilon\}$. Perhaps a more intuitive quantity describing the error variance is the signal-to-noise ratio (SNR)

$$\mathrm{SNR} = \theta^2/\sigma_e^2, \tag{2.4}$$

which is given in decibel-scale as

$$\text{SNR [dB]} = 10\log_{10}\{\theta^2/\sigma_e^2\}\text{[dB]}. \tag{2.5}$$

In a typical estimation task, instead of a single parameter θ, a parameter vector

$$\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]^{\mathrm{T}} \in \mathbb{R}^K$$

is estimated, and the estimation vector is then given as

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \mathbf{e}. \tag{2.6}$$

In that case also the noise term is a K-dimensional vector

$$\mathbf{e} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_K]^{\mathrm{T}} \in \mathbb{R}^K.$$

Again, if the estimator is unbiased

$$\mathrm{E}\{\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\} = \mathbf{0}. \tag{2.7}$$

The error vector is described by the error covariance matrix,

$$\Sigma = \mathrm{E}\left\{[\mathbf{e} - \boldsymbol{\mu_e}][\mathbf{e} - \boldsymbol{\mu_e}]^{\mathrm{T}}\right\}, \tag{2.8}$$

where

$$\boldsymbol{\mu_e} = \mathrm{E}\{\mathbf{e}\}.$$

The individual components of the error covariance matrix are given as

$$\mathrm{cov}(\varepsilon_x, \varepsilon_y) = \mathrm{E}\left\{[\varepsilon_x - \mu_{\varepsilon_x}][\varepsilon_y - \mu_{\varepsilon_y}]^{\mathrm{T}}\right\} \tag{2.9}$$

In the case studied in this thesis, the parameter vector $\boldsymbol{\theta}$ is the 3-D location of the reflection.

Often the parameters cannot be measured directly. Instead some other variable $\hat{\chi}$ is measured, which is then related to the estimated parameter by a linear or non-linear model, i.e. $\chi(\boldsymbol{\theta})$.

2.1.1 Maximum likelihood estimation

The parameter $\boldsymbol{\theta}$ can be estimated in several ways. One of the most popular methods is the maximum likelihood estimation (MLE) method. The MLE can be considered as two-step estimation approach. Firstly, the measurements

$$\hat{\boldsymbol{\chi}} = [\hat{\chi}_1, \hat{\chi}_2, \ldots, \hat{\chi}_N], \hat{\boldsymbol{\chi}} \in \mathrm{R}^N,$$

are assumed to have an error probability density function $f(\hat{\chi}_i; \chi_i(\boldsymbol{\theta}))$ where the true values of the variable are related to the parameter,

$$\boldsymbol{\chi}(\boldsymbol{\theta}) = [\chi_1(\boldsymbol{\theta}), \hat{\chi}_2(\boldsymbol{\theta}), \ldots, \hat{\chi}_N(\boldsymbol{\theta})], \boldsymbol{\chi}(\boldsymbol{\theta}) \in \mathrm{R}^{N \times K}.$$

The joint probability density function for the variables $\boldsymbol{\chi}(\boldsymbol{\theta})$ given the measurements $\hat{\boldsymbol{\chi}}$ is formed by *multiplying* the individual density functions [17]

$$\mathcal{L}(\boldsymbol{\chi}(\boldsymbol{\theta}); \hat{\boldsymbol{\chi}}) = f(\hat{\boldsymbol{\chi}}; \boldsymbol{\chi}(\boldsymbol{\theta})) = \prod_{i=1}^{N} f(\hat{\chi}_i; \chi_i(\boldsymbol{\theta})) \tag{2.10}$$

This joint density function is referred to as likelihood, and it is denoted with $\mathcal{L}(\cdot;\cdot)$. Assuming the normal distributions in Eq. (2.10) give a multivariate normal distribution [17]

$$\mathcal{L}(\boldsymbol{\chi}(\boldsymbol{\theta}); \hat{\boldsymbol{\chi}}) = \frac{\exp\left(-\frac{1}{2}[\hat{\chi}_1 - \chi_1(\boldsymbol{\theta}), \ldots, \hat{\chi}_N - \chi_N(\boldsymbol{\theta})]\Sigma^{-1}[\hat{\chi}_1 - \chi_1(\boldsymbol{\theta}), \ldots, \hat{\chi}_N - \chi_N(\boldsymbol{\theta})]^{\mathrm{T}}\right)}{(2\pi)^{(N)/2}\sqrt{\det(\Sigma)}}, \tag{2.11}$$

where Σ is the covariance matrix that includes the variances of the individual error probability functions and their covariances. In the case of independent variables, Σ is a diagonal matrix with diagonal components corresponding to the error variances σ^2. In the dependent case, the covariance matrix is symmetric and it includes information on the correlation between the variables.

In the second part of MLE, the likelihood is maximized. However, it is often more common to use the log-likelihood instead

$$\lambda(\boldsymbol{\chi}(\boldsymbol{\theta}); \hat{\boldsymbol{\chi}}) \stackrel{\triangle}{=} \log\{\mathcal{L}(\boldsymbol{\chi}(\boldsymbol{\theta}); \hat{\boldsymbol{\chi}})\} = \sum_{i=1}^{N} \log\{f(\hat{\chi}_i; \chi_i(\boldsymbol{\theta}))\}. \tag{2.12}$$

The argument that maximizes the likelihood function is called as the maximum likelihood estimate

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{\lambda(\boldsymbol{\chi}(\boldsymbol{\theta}); \hat{\boldsymbol{\chi}})\}, \tag{2.13}$$

where $\hat{\boldsymbol{\theta}}$ is the N-dimensional estimated parameter vector.

2.1.2 Gauss-Markov theorem

The Gauss-Markov theorem states that [18, p. 217], in the case when the noise variances are equal $\mathrm{var}\{\varepsilon_i\} = \sigma^2$, zero mean $\mathrm{E}\{\varepsilon_i\} = 0$, and the noise terms are uncorrelated, i.e., $\mathrm{cov}\{\varepsilon_i, \varepsilon_j\} = 0$, the best linear unbiased estimator (BLUE) is the ordinary least squares estimator, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{N} (\hat{\chi}_i - \chi_i(\boldsymbol{\theta}))^2 \right\}. \tag{2.14}$$

This is also called the minimum mean squared error estimator (MMSE). It is straightforward to show that Eq. (2.14) is a direct result of Eq. (2.12) with the given assumptions.

2.1.3 Monte-Carlo simulations and error metrics

Monte-Carlo simulations are a useful tool for inspecting the performance of an estimator. In the simulations the modeled process is simulated N times, with selected models for the signal and error. The output of the estimator is then observed, and the estimator variance can be calculated directly from the output values. Often, instead of the variance, root mean squared error (RMSE) of the estimator is calculated

$$\mathrm{RMSE}(\hat{\boldsymbol{\theta}}) = \sqrt{\mathrm{MSE}(\hat{\boldsymbol{\theta}})} = \sqrt{\mathrm{E}\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\}}. \tag{2.15}$$

Other alternatives for the error measure are the mean absolute error or median error. These measures do not weight large errors as heavily as RMSE.

Another metric used in the estimation is the number of anomalous estimates or the anomaly percentage. It is defined as the ratio between the estimates that have an error greater than some threshold and the total number of estimates

$$\mathrm{AN}(\hat{\boldsymbol{\theta}}) = \mathrm{E}\left\{\mathbf{1}\left\{\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\| > \varepsilon\right\}\right\} \tag{2.16}$$

where $\mathbf{1}\{\cdot\} = 1$ if the condition is true and 0 otherwise.

2.1.4 Cramér-Rao lower bound

The lower bound for the estimator covariance is given by the Cramér-Rao lower bound (CRLB). In the multivariate case, it is given by the matrix inverse of the Fisher information matrix $\mathbf{J}(\boldsymbol{\theta})$ [17, Ch. 3]

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \geq \mathbf{J}(\boldsymbol{\theta})^{-1}. \tag{2.17}$$

In the single variable case, the Fisher information is one dimensional and the covariance is simply variance. The Fisher information matrix is defined as the squared derivative of the log-likelihood of the estimate probability density function, and it is given in the single parameter case as [17, Ch. 3]

$$J(\theta) = \mathrm{E}\left\{\left[\frac{\partial \lambda(\chi(\theta); \hat{\chi})}{\partial \theta}\right]^2\right\}, \tag{2.18}$$

and in the multivariate case as

$$\mathbf{J}(\boldsymbol{\theta}) = \mathrm{E}\left\{ \left[\frac{\partial \lambda(\boldsymbol{\chi}(\boldsymbol{\theta}); \hat{\boldsymbol{\chi}})}{\partial \boldsymbol{\theta}} \right] \left[\frac{\partial \lambda(\boldsymbol{\chi}(\boldsymbol{\theta}); \hat{\boldsymbol{\chi}})}{\partial \boldsymbol{\theta}} \right]^{\mathrm{T}} \right\}. \tag{2.19}$$

The mean squared error is limited by the CRLB

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}) \leq \mathrm{trace}\{\mathbf{J}(\boldsymbol{\theta})^{-1}\} \tag{2.20}$$

If the estimator achieves the CRLB and is unbiased, it is called as an efficient estimator. The CRLB may not be achieved by any estimator. Especially, if the measured variable is not an injection, an efficient estimator does not exist [19] and thus the CRLB is not achieved by any estimator.

2.2 Sound

A sound source emits sound energy in a medium. The sound energy causes the fluid particles of the medium to move from their initial state. The movements of the particles are described by the instantaneous particle velocity. On the other hand, the pressure of the medium changes due to different densities introduced by the particle movements. That is, the sound pressure is the effect of the sound power emitted by a sound source. This pressure is often referred to as acoustic pressure. The sound field has certain characteristics that are different in the near-field and the far-field. The sound field in the near-field is called active and in the far-field reactive [20]. In this thesis, the source is always considered to be in the far-field.

2.2.1 Sound pressure

The total sound pressure is the superposition of atmospheric pressure p_0 and the acoustic pressure p [11, 21–23]:

$$p_{\mathrm{tot}} = p_0 + p; \tag{2.21}$$

Often, in acoustics, a quantity called the sound pressure level is used instead of the total sound pressure. It is given as the relative change in the acoustic pressure respective to the just audible hearing threshold (2×10^{-5} Pa) [11].

2.2.2 The wave equation

Using Newton's laws of motion, and assuming that the air has no net velocity, i.e. the air does not move, sound pressure can be expressed using the wave equation [11, 23]

$$\bigtriangledown^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0, \tag{2.22}$$

where t is time and c is the speed of sound. In this case, the sound pressure is a four-dimensional scalar function consisting of three coordinate components and time, i.e., $p = p(x, y, z, t)$.

2.2.3 Sound intensity

Particle velocity $\mathbf{v}$ describes the speed of the air (fluid) particle movements. Together with sound pressure they define the instantaneous sound intensity [21–23],

$$\boldsymbol{I} = p\mathbf{v}. \tag{2.23}$$

Note that the sound intensity is a vector quantity as is the particle velocity. Sound intensity is perhaps best described as the flow of energy or the sound power per area.

2.3 Measurement of sound pressure and intensity

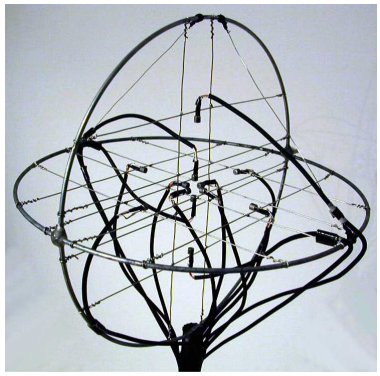
Sound pressure is measured with a pressure microphone. The microphones that are used in this thesis, translate the mechanical vibration of the diaphragm (membrane) of the microphone into electric current using capacitance change, or electromagnetic induction. Although there exist special intensity sensors, such as the ones Microflown has developed [24], here the intensity is measured using pressure microphone pairs.

2.3.1 Fourier transform and spectral density

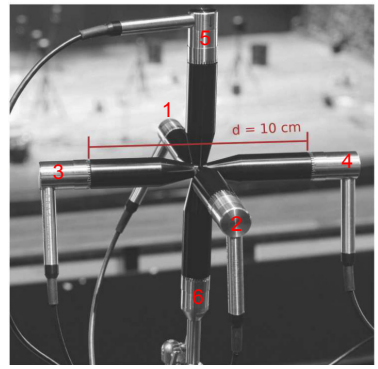
The pressure signal recorded with a microphone is denoted with $p(t)$. The Fourier transform of the continuous time signal $p(t)$ is given as [25, 26]

$$\tilde{P}(\omega) = \lim_{T\to\infty} TP(\omega) \tag{2.24}$$

$$= \int_{-\infty}^{\infty} p(t)e^{-j\omega t}dt \tag{2.25}$$



(a) TKK-3-D microphone array



(b) G.R.A.S. sound intensity probe

Figure 2.1. Microphone arrays used in this thesis. TKK 3-D microphone array has 12 microphones equally spaced on two spheres with diameters of 10 mm and 100 mm. G.R.A.S. array has a 6 microphones on a single sphere with diameter of 100 mm. Spacing d_{spc} is equal for microphone pair on a single axis on a single sphere. See Table 2.1 for the locations of the microphones in the array.

where $\omega = 2\pi f$ is the angular frequency and the discrete Fourier transform of the signal of length T is given by

$$P(k) = \frac{1}{T}\int_{-T/2}^{T/2} p(t)e^{-jkt\omega_\Delta}dt \tag{2.26}$$

with $\omega_\Delta = \frac{2\pi}{T}$. The discrete signal has a power spectral density which is equal to [25, 26]

$$\mathrm{E}[P(k)P^*(k)] = \frac{1}{T}G_{p,p}(k), \tag{2.27}$$

where $\cdot^*$ denotes the complex conjugate. The spectral density of the continuous signal approaches [25, 26]

$$\mathrm{E}[\tilde{P}(\omega)\tilde{P}^*(\omega)] = \tilde{G}_{p,p}(\omega). \tag{2.28}$$

2.3.2 Sound intensity measurement using microphone pairs

Throughout this thesis, the microphone array design that is used is an open spherical microphone array. Examples are shown in Fig. 2.1. The microphones are omni-directional. This setup is the optimal six-microphone-setup for localization, as shown in [27]. The use of this kind of array makes it possible to measure sound intensity on 3-D coordinate system. Other microphone configurations can be used as well to obtain the 3-D sound intensity [28]. Since sound intensity can be measured, auralization using spatial impulse response rendering technique (SIRR) is possible [1, 12]. In SIRR the features relevant for human perception are analyzed from the sound intensity vectors.

Table 2.1. Origin centered coordinates for the microphone arrays. Spacing d_{spc} is equal for each microphone pair on a single axis.

Microphone No.	X [m]	Y [m]	Z [m]
1	$d_{\text{spc}}/2$	0	0
2	$-d_{\text{spc}}/2$	0	0
3	0	$d_{\text{spc}}/2$	0
4	0	$-d_{\text{spc}}/2$	0
5	0	0	$d_{\text{spc}}/2$
6	0	0	$-d_{\text{spc}}/2$

On a certain axis x, the instaneous reactive sound intensity is given in the frequency domain as

$$I_{\mathrm{x}}(\omega) = \Re\{P^*(\omega)U_{\mathrm{x}}(\omega)\}, \tag{2.29}$$

where $P(\omega)$ and $U_a(\omega)$ are the frequency presentations of the sound pressure and of the particle velocity with angular frequency ω [12]. In addition, $\Re\{\cdot\}$ is the real part of a complex number and $(\cdot)^*$ denotes the complex conjugate.

The pressure in the middle of the array, shown in Fig. 2.1, can be estimated as the average pressure of the microphones [12, 29]:

$$P(\omega) \approx \frac{1}{6}\sum_{n=1}^{6} P_n(\omega). \tag{2.30}$$

In the frequency domain, the particle velocity is estimated for the x-axis as:

$$U_{\mathrm{x}}(\omega) \approx \frac{-j}{\omega\rho_0 d}[P_1(\omega) - P_2(\omega)], \tag{2.31}$$

where d is the distance between the two receivers, j is the imaginary unit, and, for example, with the speed of sound $c = 343$ m/s, the median density of air is $\rho_0 = 1.204$ kg/m^3. The particle velocity is calculated similarly for y-axis

$$U_{\mathrm{y}}(\omega) \approx \frac{-j}{\omega\rho_0 d}[P_3(\omega) - P_4(\omega)], \tag{2.32}$$

and for z-axis

$$U_{\mathrm{z}}(\omega) \approx \frac{-j}{\omega\rho_0 d}[P_5(\omega) - P_6(\omega)]. \tag{2.33}$$

The overall sound intensity vector for a frequency ω is then noted with $\boldsymbol{I}(\omega) = [I_x(\omega), I_y(\omega), I_z(\omega)]$. The sound intensity estimation with microphone pair technique is limited by the distance between the microphones.

Frequencies above

$$f > \frac{c}{d} \tag{2.34}$$

are spatially aliased and the sound intensity for them cannot be properly estimated using the above equations. The low frequency limit is typically set by the properties of the pressure microphones.

The estimation of sound intensity vectors using Eqs. (2.30) and (2.31) is shown to be biased [29]. The bias is described by the equation [29]

$$g(\theta) = \mathbf{arc}\tan\left(\frac{\sin(\omega d \sin(\theta)/(2c))}{\sin(\omega d \cos(\theta)/(2c))}\right). \tag{2.35}$$

The unbiased estimate θ_{unb} is obtained via the inverse function as $\theta_{\text{unb}} = g^{-1}(\theta)$. The bias is caused by the fact that the pressure gradient is a sinusoidal one instead of the assumed constant. The bias cannot be corrected for frequencies [29]

$$f > \frac{1}{\sqrt{2}}\frac{c}{d}, \tag{2.36}$$

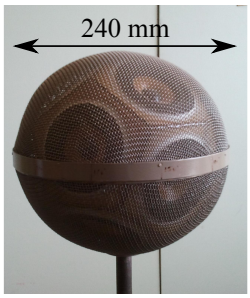
which is much lower than the previously set threshold by the spatial aliasing.

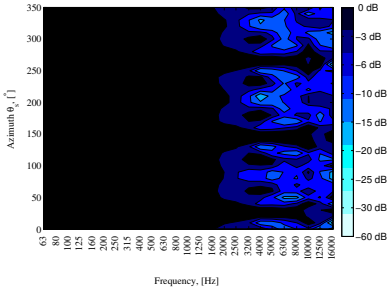
In this thesis, the bias correction is not used since the highest frequency in the experiments is selected to be so low that the bias can be neglected.

2.4 Directivity of the sources

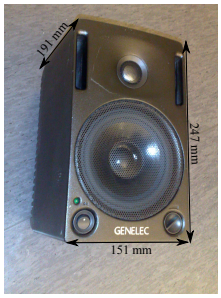
Assuming a homogeneous medium and a free path between the source and the sensor, the direct sound wave arriving at a sensor depends on the characteristic of the sound source. The most used characterization is the directivity of the source [30]. It is a measure of how much energy the source emits to a certain angle at a certain distance. It is measured in free-field conditions: in an anechoic chamber, or in a room where the reflective surfaces are sufficiently far so that windowing can be applied to isolate the direct sound from the reflections. The more measurements made around the source, the more accurate estimation of the directivity is achieved. The acoustic power can be estimated with a surface integral over the directivity measurement, defined by an ISO-standard [31]. Figure 2.2 shows examples of the directivities of three loudspeakers.

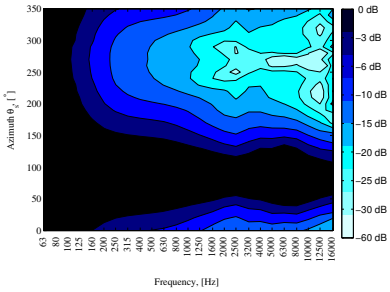
The directivity and acoustic power measurements assume that the sound source is a point source. This is not true with real sound sources. For example, a violin has a vibrating body which emits energy in addition to that emitted from the f-holes [30]. Also, loudspeakers are not point sources.



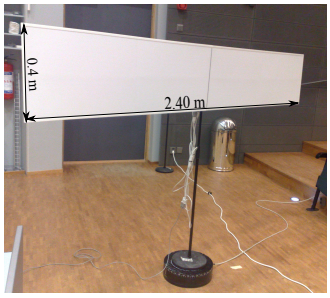


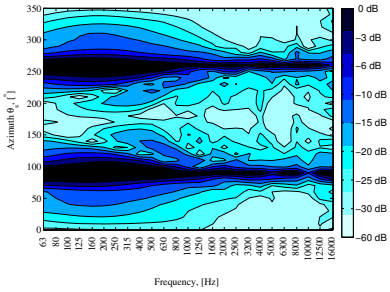
(a) Standard omni-directional





(b) Genelec 1029A





(c) Panphonics panel loudspeaker

Figure 2.2. Pictures and dimensions of three loudspeakers, and their directivity measured in 1/3-octave bands, at 12 m distance at every 10 degrees azimuth. The speaker is facing the microphone when azimuth angle is 90 degrees.

For instance, a widely used monitor loudspeaker Genelec 1029A has two elements, the bass-element and the tweeter, which both emit sound energy. The bass-element reacts more slowly to the changes that the coil passes on than the tweeter. For this reason, and due to the different locations of the elements, the high frequencies arrive at a sensor placed in front of the loudspeaker earlier than the low frequencies, as shown in Fig. 2.3 where the impulse response is filtered at the cross-over frequency of the loudspeaker. The fact that a loudspeaker consists of several sound sources affects the phase of the received signal. When the single location of the sound source is wanted, the acoustic center of the source is used,

which is the weighted average of the sound energy over an area.

Near-field acoustic holography [32, 33] is a useful tool for describing the sound source. Acoustic holography is concerned with the inverse problem of what the sound source has emitted given the observations of sound pressure at some distance. Typically, a grid of sensors is placed in the vicinity of the source, and the Kirchhoff-Helmholtz integral is used to inversely to predict where the energy is distributed on a hologram plane [34]. It can be used, for example, in noise source measurements or to investigate which parts of an instrument emit sound energy.

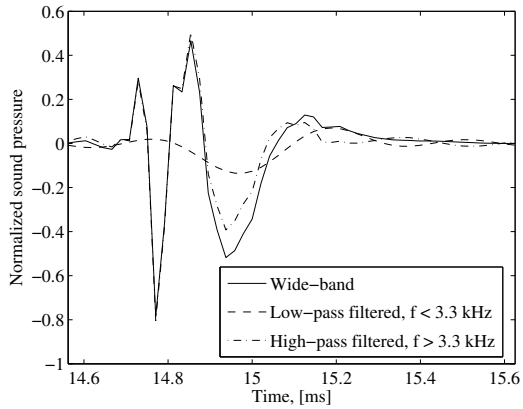
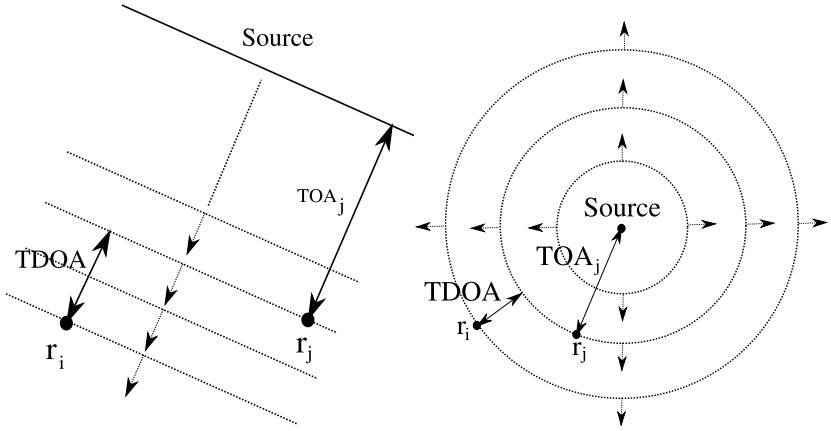


Figure 2.3. Impulse response of Genelec 1029A, measured at approximately 5.1 m distance, in front of the loudspeaker. The low frequencies arrive later at the microphone than the high frequencies.

2.5 Geometrical quantities

Useful geometrical quantities in acoustic source localization are time of arrival (TOA) and time difference of arrival (TDOA). The calculation of these quantities depend on the selected wave propagation model. Two commonly used wave propagation models are the spherical and the plane wave propagation models. Fig. 2.4 illustrates the principles of these two models in 2D. The plane wave propagation model is usually assumed and used, if the intra-sensor distances are small and the source is far away from the sensors.



(a) Plane wave propagation **(b)** Spherical wave propagation

Figure 2.4. Plane and spherical wave propagation models. Time of arrival (TOA) and time difference of arrival (TDOA) are presented also for both cases.

2.5.1 Time of arrival

Time of arrival (TOA), often also referred to as time of flight, is the time that the sound wave takes to travel from the source to the receiver. In the case of spherical wave propagation model it is given as:

$$t(\mathbf{r}_n; \mathbf{x}) = c^{-1} \|\mathbf{r}_n - \mathbf{x}\| \tag{2.37}$$

and for plane wave propagation model as

$$t(\mathbf{r}_n; \mathbf{x}) = |c^{-1} \mathbf{n}^{\mathrm{T}} (\mathbf{r}_n - \mathbf{x})|, \tag{2.38}$$

where c is the speed of sound and $\mathbf{n}$ is the direction of the plane wave.

2.5.2 Time difference of arrival

Time difference of arrival (TDOA) for a spherical wave propagation model is the difference of two TOAs:

$$\tau(\mathbf{r}_i, \mathbf{r}_j; \mathbf{x}) = c^{-1} (\|\mathbf{r}_i - \mathbf{x}\| - \|\mathbf{r}_j - \mathbf{x}\|), \tag{2.39}$$

where c is again the speed of sound and $(\cdot)^{\mathrm{T}}$ denotes vector transpose. For the plane wave propagation model, the TDOA formulates into

$$\tau(\mathbf{r}_i, \mathbf{r}_j; \mathbf{x}) = c^{-1} \mathbf{n}^{\mathrm{T}} (\mathbf{r}_i - \mathbf{r}_j). \tag{2.40}$$

2.6 Propagation of sound in enclosures in short

The inspection of sound phenomena are now restricted to room conditions. In a room environment, when a wave confronts a surface S, the reflected wave depends on the features of the surface. The surfaces considered here are impenetrable, rigid, or porous. An impenetrable surface does not transmit any waves to the other side of the surface [22]. A rigid surface is stationary, i.e. does not move, and a porous wall is not necessarily rigid or impenetrable [22]. A porous surface can transmit some of the arriving energy through refraction [35].

2.6.1 Speed of sound

Particle velocity describes the speed of the particle movements. However, the more interesting quantity in room acoustics is the speed of the propagating sound pressure wave, commonly known as the speed of sound. In room air, the most prominent factors that affect the speed of sound are the temperature, relative humidity, barometric pressure, and carbon dioxide content [36].

Several approximations, all derived from fluid theory, exist for the speed of sound calculation [36]. Throughout this thesis the speed of sound is calculated using the approximation presented in [36, p. 1046], and assuming that the carbon dioxide content and the barometric (atmospheric) pressure are 0 % and 1013 hPa, respectively. The relative humidity of the air and the temperature are measured using commercially available equipment. Based on measurements by the author, during an acoustic measurement, for example in a concert hall, these factors change over time. In this thesis, it is assumed that the air in the enclosure is homogeneous during each measurement.

2.6.2 Attenuation and air absorption

In general, the amplitude of the sound pressure decreases in relation to $1/r$, where r is the distance from the source, for spherical waves, and by $1/\sqrt{r}$ for cylindrical waves. This is caused by the fact that the energy is spread over a bigger area, thereby attenuated.

In addition to attenuation, the air absorbs some of the energy of the sound wave [35, 37, 38]. Air absorption is a function of frequency, and in general it depends on distance and the same physical quantities as the

speed of sound [38].

2.6.3 Specular reflections

An impenetrable surface S can be stationary or vibrating. Consider a point $\mathbf{x}_S$ on the surface S with velocity of the (moving) surface $\mathbf{v}_S = d\mathbf{x}_S/dt$ near the point $\mathbf{x}_S$. The velocity of the fluid $\mathbf{v}$ at the boundary has to be equal to the velocity of the particles near the boundary, i.e. [22]:

$$\mathbf{v} \cdot \mathbf{n}_S = \mathbf{v}_S \cdot \mathbf{n}_S, \tag{2.41}$$

where n_S is the normal component of the surface at $\mathbf{x}_S$.

On stationary surfaces, the surface does not move ($\mathbf{v}_S = 0$), and one has $\mathbf{v} \cdot \mathbf{n}_S = 0$ [22]. This implies that the particle velocity at the boundary is 0. Therefore, a plane wave at a flat rigid surface is reflected according to the law of mirrors (also included in Snell's law), i.e. the reflected wave is the mirrored angle with respect to the normal of the surface as shown in Fig. 2.5. The specularly reflected wave can be modeled conveniently using the image-source principle [39], shown in Fig. 2.6.

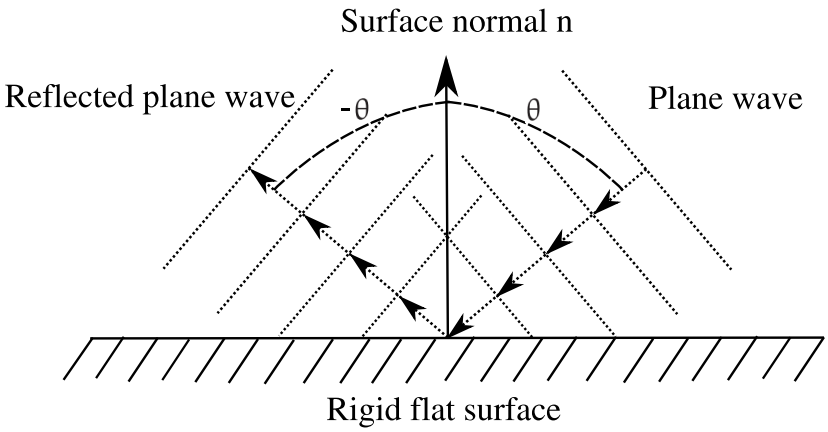


Figure 2.5. A plane wave at flat rigid surface is reflected according to the law of mirrors. After [23].

2.6.4 Specific acoustic impedance and absorption

A boundary condition where the surface is not necessarily rigid or impenetrable is described by the specific acoustic impedance. Specific acoustic impedance Z is the relation between sound pressure p and particle veloc-

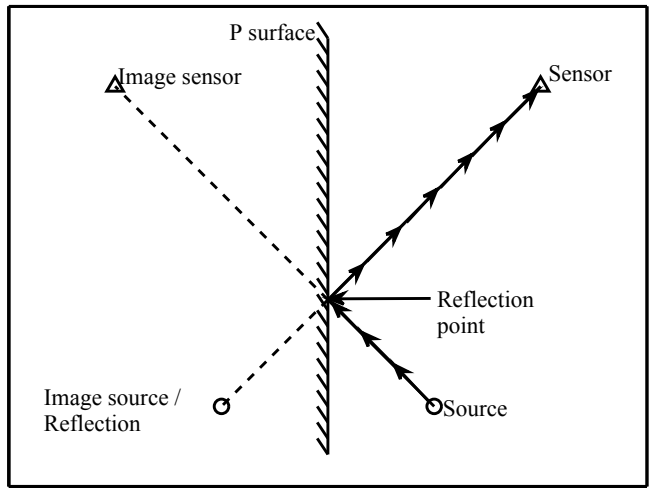


Figure 2.6. Concepts of image-source and image-sensor.

ity v on the surface S, i.e. [22]:

$$Z(\omega) = \frac{p(\omega)}{v(\omega)}. \tag{2.42}$$

This is analogous to the electrical circuits, i.e. the relation between impedance, current, and voltage. Note that in this case, the particle velocity is not written in the vector form because the measurement is considered at one connection point. The specific acoustic impedance consists of specific acoustic reactance and resistance, which are the real and imaginary parts of $Z(\omega)$, respectively. The resistance can be seen as the part where energy is lost, and reactance as the part where energy is stored.

A closely related quantity to the specific acoustic impedance is the pressure-amplitude reflection coefficient β which describes the relation between pressures of the incident arriving waveform and the reflected wave. Through some theoretical examination (see [22] or [23] for details) the relation to specific acoustic impedance is given:

$$\frac{Z(\omega)}{\rho c} = \frac{1 + \beta(\theta, \phi, \omega)}{1 - \beta(\theta, \phi, \omega)}, \tag{2.43}$$

where ρc is the characteristic specific impedance of air, and θ is the angle of the incident wave. So, $\beta(\theta, \phi, \omega)$ depends on the angle of incident and frequency. For a plane wave at a flat rigid surface the reflection coefficient is independent of the angle of incidence.

If the reflection coefficient is less than 1, the material absorbs energy. The absorption coefficient is defined as [22]:

$$\alpha(\theta, \phi, \omega) = 1 - |\beta(\theta, \phi, \omega)|^2. \tag{2.44}$$

Measurement of the absorption coefficient can be done using an impedance tube measurement [40, 41], various in-situ measurement methods [6, 7, 42], or the reverberation chamber technique [43].

2.6.5 Diffraction

Diffraction occurs when a sound wave confronts an edge. A practical example of this is confronted in everyday-life: a person is able to hear what someone is speaking on the other side of a corner. There are three regions around the corner where different waves besides the diffracted wave occurs. In the first region, only reflected wave is possible, in the second region there is only direct wave and no reflected wave, and in the third region there is only diffracted wave. The formal definitions for these cases are given in [23]. It is found to be important to model the diffraction for auralization purposes [44].

2.6.6 Scattered reflections or diffusion

When the surface is rough or someway uneven, the measurement or the modeling of specular reflections becomes difficult. In this case, scattering and diffusion coefficients are a useful way to describe the behavior of the sound field [43]. The phenomenon that causes diffuse reflection is the diffraction in very small scale [45]. Scattering and diffusion coefficients describe the reflection from a surface that is not perfectly specular. For example, the scattering coefficient is calculated by dividing the reflection in to two components: the specular reflection, and the scattered reflections [46]. Several measurement approaches and different definitions for the coefficient exist for diffusion and scattering [43, 47].

2.6.7 Definitions of the diffuse sound field

A sound field is perfectly diffuse if the directional energy density inside a volume is equal for each point and direction [22]. In practice this means that direction and the phase of the sound field are uniformly distributed and the amplitude is equally distributed for each point. Thus, the sound

field is spatially homogeneous and isotropic. Another definition for diffuse sound field is that the net energy flow over the volume is zero, i.e. the sound intensity over the surface S of the volume V

$$\int_S \boldsymbol{I} d\mathbf{S} = 0, \tag{2.45}$$

where $d\mathbf{S}$ is a surface element. A way of constructing a diffuse sound field is by superpositioning infinite number of plane waves with random phases in the volume. In practice, a finite number of plane waves, e.g., 1145, will produce a diffuse sound field [48].

The diffuseness of a sound field can be measured with spatial correlation function [48–50], its variations [51–53], spatial coherence [54], its variations [53], or spatial uniformity of the sound field [51].

2.6.8 Measurement of instantaneous diffusion

All of the above methods measure the diffuseness of a sound field over a large set of measurements. A more interesting method in the context of this thesis is the one that can describe the diffusion of a part of the room impulse response. Examples of this kind of method is the diffuseness analysis used in SIRR [1, 12, 55]. Other methods are presented in Publication VI and [56].

2.7 The room impulse response

When a sound wave propagates in an enclosure, it is affected by the phenomena listed above. The signal received in the sensor is therefore a modified version of the signal emitted by the source. If the source signal is a single impulse, the signal arriving to a sensor is called the impulse response. In the context of this thesis, the impulse response can be presented as

$$h(t) = \sum_{k=1}^{K} h_k(t) + w(t)d\omega + w(t) \tag{2.46}$$

where

$$h_k(t) = \int \alpha_k(\omega) e^{j\omega(t-t_k+\phi_k(\omega))} d\omega \tag{2.47}$$

is a single reflection, $\alpha_k(\omega)$ is the frequency dependent attenuation factor for each sound wave k, t_k is the time delay related to the distance of the path of a reflection, and $w(t)$ is measurement noise that is assumed to be

independent and normally distributed. The attenuation factor $\alpha_k(\omega)$ is dependent on the properties of the surface and air absorption [11]. Quite often in real situations the phase term $\phi_k(\omega)$ is dependent on the frequency. Here the frequency dependency of the phase term $\phi_k(\omega)$ is acknowledged, but the analysis of the room impulse responses assumes that with early reflections the phase is independent of the frequency, i.e. $\phi_k(\omega) = 0, \forall k$.

The first arriving sound wave in Eq. (2.46) is referred to as the direct sound. The sound waves arriving after the direct sound are called the early reflections, up to a time instant called a mixing time t_m [10]. The early reflections are considered to be discrete events, with only small deviations in the phase of the sound wave. After the mixing time, the impulse response is called late reverberation. The impulse response, especially the late reverberation, exhibits some statistical behavior [9–11]. The reflections cannot therefore be identified or localized from the late reverberation. Figure 2.7 illustrates the three parts of the impulse response.

2.7.1 Modal and echo density

The modal density, the number of modes, i.e., resonance frequencies, at a frequency f is given as [11, p. 61]

$$\frac{dN_f}{df} = 4\pi V \frac{f^2}{c^3}, \tag{2.48}$$

where V is volume, c is speed of sound, and N_f is number of modes. The echo density, the number of reflections arriving at time t is [11, p. 92]

$$\frac{dN_r}{dt} = 4\pi \frac{c^3 t^2}{V}, \tag{2.49}$$

where N_r is the number of reflections. Both of these equations apply to rooms with arbitrary shape [11, p. 92]. When frequency increases, modal density becomes large and when time increases, echo density becomes large.

2.7.2 Central limit theorem

In the discrete time domain, the samples $\{h_k^{(i)}\}, i \in \{1 \dots L\}$ of a reflection n arriving within the time window dt are considered random variables with mean $\mathrm{E}\{h_k\} = \mu$, variance $\mathrm{var}\{h_k\} = \sigma^2$, and some unknown probability density function. According to the central limit theorem, as K approaches infinity, the mean of the samples approaches a normal distri-

bution [57, p. 357]:

$$\frac{1}{K}\sum_{k=1}^{K}\{h_k^{(i)}\} \xrightarrow{d} \mathcal{N}(\mu_r,\ \sigma_r^2), \tag{2.50}$$

with some mean μ_r and variance σ_r^2. Thus, the sum of an infinite number of reflections can be considered normally distributed in the discrete time domain. Note that if $\{h_k^{(i)}\}_{i=1}^{L}$ is normally distributed, then the mean of the reflections is always normal. If $\{h_k^{(i)}\}_{i=1}^{L}$ is not normally distributed, then it takes $K = \infty$ reflections to achieve normality, as stated by Cramér's theorem.

In practice, it is not required to have infinite number of reflections to achieve normality. The number of reflections at which the average of them is normally distributed depends on the reflection signals h_n. However, no matter what the reflection signal h_n is, it is inevitable that after a certain number of reflections the time distribution of their average is normal.

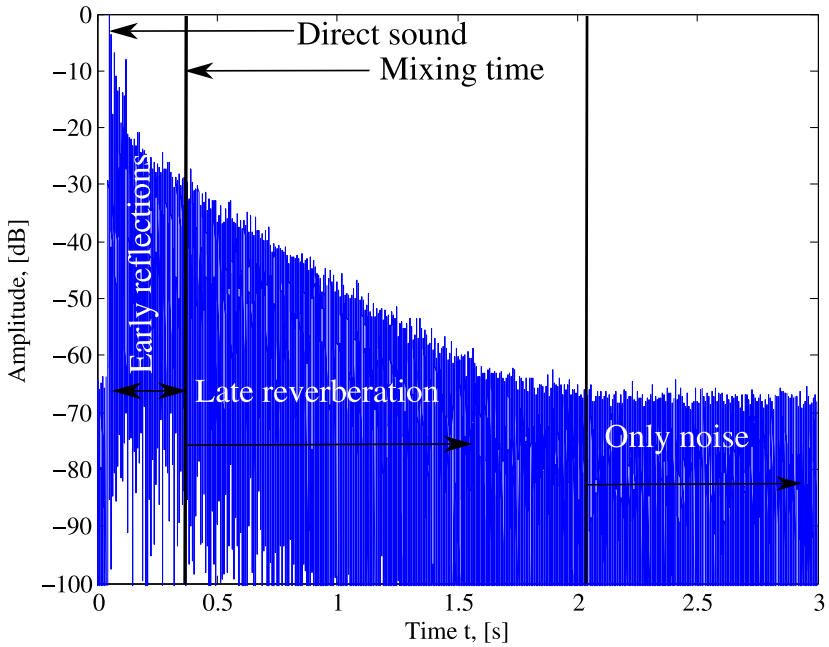
Since the modal density and echo density are differential measures, the frequency and time intervals in them are infinitesimal, respectively. Then, in those infinitesimal intervals the distributions of the time domain and frequency domain pressure signals are normally distributed when the number of modes and reflections is high enough. This model is introduced by Schroeder [9] and later complemented by Polack [58] and they are summarized in the following section.

2.7.3 Statistical models of the room impulse response

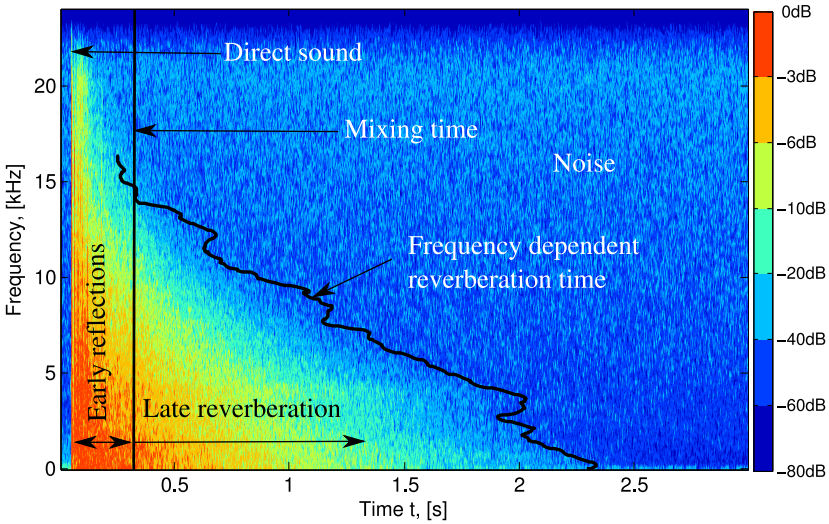
For a given room impulse response, when considered in the frequency domain, if the distance from the source to the receiver is sufficient, and if enough room modes are excited simultaneously, then the real and imaginary parts can be considered independent Gaussian processes [9]. The amplitude of the frequency domain room impulse response $H(f)$ therefore follows the Rayleigh distribution, i.e. $\|H(f)\| \sim \mathcal{R}(\sigma_f^2)$, where σ_f^2 is the standard deviation in the frequency domain [9, 59]. This applies for frequencies above the Schroeder frequency [9],

$$f_{\text{Schroeder}} \approx 2000\sqrt{\frac{T_{60}}{V}}, \tag{2.51}$$

where T_{60} is the reverberation time and V is the volume of the room. It should be noted that the energy follows exponential distribution, i.e., $\|H(f)\|^2 \sim \mathcal{E}(\lambda_e)$, where $\lambda_e > 0$ is a parameter, which can be calculated from σ_f.



(a) Time-domain



(b) Frequency-time-domain

Figure 2.7. Impulse response from a concert hall. Early reflections appear before mixing time and late reverberation.

In the time domain, a statistical model of the room impulse response can be applied after the mixing time [10]. Discussion about the mixing time is presented in the next subsection. The statistical model in the time domain is given as [10, 60]:

$$h(t) = b(t)f(t), \tag{2.52}$$

where $f(t)$ is a monotonically decaying function, and $b(t)$ is zero mean normally distributed noise, i.e. $b(t) \sim N(0, \sigma_t^2)$, and σ_t is the fixed standard deviation. The decaying function is dependent on the reverberation time T_{60} of the room determined by the following relation [10]:

$$f(t) = e^{-\delta t}, \tag{2.53}$$

where $\delta = 3\log(10)/T_{60}$ is the damping factor. The reader is reminded that the total distribution of $h(t)$ is not normal since it is multiplied with the decaying function $f(t)$. However, the frequency domain transform $F\{h(t)\}$, where $F\{\cdot\}$ is the Fourier transform, follows the Rayleigh distribution with variance

$$\sigma_f^2 = \sigma_t^2 \frac{\mathrm{E}\{f^2(t)\}}{2} \tag{2.54}$$

where $\mathrm{E}\{\cdot\}$ denotes the expected value.

In real rooms, the reverberation time and the decaying function are a function of frequency, i.e. $\delta(f) = 3\log(10)/T_{60}(f)$. Figure 2.7, shows an example of the frequency dependent reverberation time in a concert hall, estimated as proposed in [10]. A generalization of Eq. (2.52) to the frequency dependent case (originally suggested by J.-D. Polack according to [10]) is given by the ensemble average of the Wigner-Ville distribution:

$$< W(t,f) >= \|H(f)\|^2 \, e^{-2\delta(f)t} \tag{2.55}$$

where $\|H(f)\|^2$ is the power spectral density. That is, the average of the Wigner-Ville distribution over a set of time instants and frequencies has an exponentially decaying shape in the time domain multiplied by the power spectral density.

The Wigner-Ville distribution itself is defined as [10]:

$$W(t,f) = \int_{-\infty}^{\infty} h(t-\tau/2)h(t+\tau/2)e^{-j2\pi f\tau}dt. \tag{2.56}$$

The Wigner-Ville distribution has the following properties. The integration of Eq. (2.56) with respect to frequency produces the temporal energy density

$$h^2(t) = \int_{-\infty}^{\infty} W(t,f)df, \tag{2.57}$$

and integration with respect to time gives the spectral energy density

$$\|H(f)\|^2 = \int_{-\infty}^{\infty} W(t, f)dt. \tag{2.58}$$

Equation (2.56) allows the attenuation factor δ in Eq. (2.52) to be dependent on frequency.

2.8 Mixing time

Mixing time appears in various applications and research studies that use or study room impulse responses [10, 37, 61–73]. Generally, in impulse response analysis and synthesis, the mixing time is used as the time after which the impulse response can be approximated by an appropriate model. This is generally much more efficient than modeling all of the reflections in the impulse response. Consequently, the use of statistical models can save considerably on computation time and/or required system memory, which are important aspects, for example, in auralization applications [37, 61, 62, 68, 69], particularly if real-time interaction and dynamic source and receiver positioning are required.

Traditionally, the mixing time is subjectively defined simply to be 80 ms [8, 74]. Furthermore, values from 50 to 200 ms have been suggested for the mixing time from the human hearing point of view [75–77]. Although these figures are reasonable as a subjective parameter, they might not correspond to the objective mixing time. That is, as the objective mixing time is dependent on the physical properties of a concert hall, it is not reasonable to assume that these physical properties do not change between concert halls. Therefore, there is a need to estimate the mixing time from a room impulse response directly.

2.8.1 Formal definitions

Echo density, i.e. Eq. (2.49), is related to the room volume through the billiard theory [10, 60]. A sufficiently large echo density should also indicate the mixing time of a room. Different values for the sufficiently large echo density have been proposed, varying from 1000 to 10000, according to [62].

Several authors define the mixing time, as the time instant when 10 or more reflections overlap in a time window of 24 ms [10, 58, 78, 79]. This

corresponds to approximately [10]

$$t_m \approx \gamma\sqrt{V} \tag{2.59}$$

where $\gamma = 1 \times 10^{-3}$ [s/m^3] is a normalizing constant.

Another approach is to define the mixing time through energy. In [67], the mixing time as the time when the energy of the impulse response has decreased a certain amount from the overall energy level. Values from -20 dB to -15 dB are used [67].

2.8.2 Estimation methods

The estimation of mixing time is an ungrateful research area, because absolute reference, i.e. ground truth, for the mixing time does not exist. Yet, several research articles about the topic exist [59, 63–66, 70–73]. The approximation in Eq. (2.59) is based on theoretical developments and has not been verified by experimental results from real data. In addition, it is debatable whether the mixing time should be given as a transition time zone, rather than a strict transition time. Therefore, all that can be done is to compare the output of the methods in different situations, as done, for example, in [72].

There exist several methods that estimate the mixing time of a room impulse response based on statistical assumptions of the properties of the signal. Mixing time is estimated as the time when the kurtosis and standard deviation ratio are close to that of a Gaussian distribution [63]. The same approach is used for separating the late reverberation of impulse responses in order that the spatial coherence and correlation functions of impulse responses might be examined [53, Fig 5.]. In addition, the echo density, for some reason, is estimated with the standard deviation ratio [80]. However, the actual relation between standard deviation ration and echo density is not shown.

The relation in Eq. (2.59) suggests that the room volume or echo density can be used to calculate the mixing time. Hence, if the echo density or the volume of the room is estimated from a single impulse response, as in [81], then the mixing time is also estimated.

In [64, 70], the mixing time is estimated from the phase of the impulse response, assuming that the phase of the impulse response is linear when the early reflections are dominant and non-linear when the late reverberation starts. From this non-linearity the mixing time can then be determined. Theoretical relation between the non-linearity of the phase of the

impulse response and the mixing time has not been presented in [64, 70].

It is suggested in [65, 66] that matching pursuit can be used for finding the reflections within a room impulse response to estimate the mixing time. Matching pursuit, in this case, is essentially the same as calculating the cross correlation between a prototype of the direct sound signal and to the rest of the impulse response. The time instant when the number of reflections no longer follows a predefined cubic model of the echo density, given in Eq. (2.49), is then the mixing time.

According to [59], mixing time can be identified when the correlation between the amplitude of certain frequencies of the whole impulse response and late sound is sufficiently low. This is proposed as the definition of mixing time and the relevance of this definition and its relation to other acoustic parameters is discussed. It is found that mixing time and reverberation time have the highest correlation out of the studied acoustic parameters [59].

The temporal overlap of reflections is used to define the mixing time in [71]. The basic assumption is that the original emitted sound wave from the sound source widens after each reflection. The width of these reflections is compared to the time differences between the reflections to deduce the mixing time.

In [73] the room's free path temporal distribution is considered to be an indicator of the mixing time. The free path temporal distribution is obtained by ray tracing and it describes the energy of the reflections at each time instant. In ergodic rooms, the energetic average of the path lengths converges rapidly after the mixing and the free path value becomes independent of the time.

3. Related research

Previous work and related research on localization of reflections are described. Measurement of room impulse responses is presented. The most relevant localization methods are discussed, and different approaches and setups for localization of reflections used in the previous research are reviewed. Possible application areas for reflection localization are listed in the end of the chapter.

3.1 Room impulse response measurement

The standard way of studying room acoustics is to measure a room impulse response [8]. The standard states that an omni-directional source and also, in most of the cases, an omni-directional microphone are to be used in the measurement.

Recently, advanced microphone array techniques have been applied for room impulse response measurements [1, 13, 82–91]. The advantage over traditional omni-directional microphone measurement is that spatial analysis of the impulse response can be applied. In addition, auralization of the space is made possible [1, 2, 83].

3.2 Localization methods

Source localization methods are based on time of arrival estimation (TOA), time difference of arrival (TDOA), or directly on the signals.

TDOA estimation is a far more popular topic than TOA estimation. This is due to the fact that time of arrival (TOA) cannot be directly measured with unknown source signals. More than ten methods have been developed for the TDOA problem over the last decades [25, 92–98]. One of the

most popular approaches is the generalized correlation [25]. Other methods include time domain difference function [96], and the use of some additional information such as the fundamental frequency [93]. The theoretical performance of TDOA estimation is well known in theory for the case when additive noise is present [99, 100]. Lower bounds, such as Cramér-Rao, describe the variance of the TDOA estimation in the case of additive noise [99]. The accuracy of TDOA based localization is limited by the sampling frequency. In [101] parabolic fit and in [102] exponential fit are proposed for interpolating the TDOA estimate.

The most straightforward algorithm for TOA estimation is a simple peak-picking algorithm [65,66,103]. In addition, it has been proposed that statistical features, such as kurtosis, can be used to detect peaks [104]. Other methods are based on correlation or some other similarity measure and they usually require a priori knowledge of the signal [66, 103]. In principle, the onset detection methods used in music signal analysis could be used here [105]. The theoretical performance of TOA estimation is not studied extensively under additive noise to the knowledge of the present author. TOA accuracy can be improved by basic Fourier-interpolation or by assuming a shape for the estimation function, similarly as in TDOA estimation.

When two- or three-dimensional localization is desired, the TDOAs, TOAs, or the signals are combined spatially using an acoustic source localization function. Popular acoustic source localization functions are the maximum likelihood estimation (MLE) function [106], steered response power (SRP) functions [106, 107], and pseudo-likelihood functions [108].

MLE methods have been formulated for TOA [109], TDOA [106, 109–111], and signal models [112–117]. Advancements in MLE for a signal model come from an updated noise model [117] or an updated signal model [112].

The MLE for TDOA, with certain assumptions, can be presented as a least squares (LS) problem. The TDOA LS problem has gained lot of attention in research [109, 118–126, 126–136], mostly because the LS solution can be given in closed form by making first some assumption on the error or on the signal. The LS solutions and problems are so addressed in research that several textbooks deal with them (e.g. [137, 138]).

Also, the MLE for TOA can be presented as a LS problem. Closed form solutions for the TOA LS problem have also been applied [139–142].

The SRP-Phase Transform algorithm has been studied extensively [106,

107, 116, 143–146] and it has been followed by various modifications and optimizations [116, 144, 145, 147–150]. The SRP method is shown to be equivalent to basic beamforming [151, 152].

The performance of the localization can be studied with CRLB [17], dilution of precision, which is a special case of CRLB [153, Ch.3.3], or similar variance analysis [128, 154].

The direction of arrival of the sound wave can be estimated using the sound intensity vectors [1,12,29,155–159]. These vectors can be measured using a special microphone, such as first order B-format microphone. The location of the source can be estimated as the average of the intensity vectors over time or frequency.

3.3 Localization of reflections and room geometry estimation

A relevant topic to the localization of reflections is the localization of the reflective surfaces, or the blind estimation of room geometry. Namely, the estimation of reflective surfaces is equivalent to localization of first order reflections. The localization of reflections and estimation of room geometry from room impulse responses have been studied in several research articles [1,3–5,12,13,85–87,160–165]. The approaches are based on TOA, TDOA, and direction of arrival (DOA) estimation. TOA estimation requires that the loudspeakers and microphones are time-synchronized, and the TDOA and DOA based methods do not require synchorinization.

In [1, 12] a technique called spatial impulse response rendering (SIRR) is developed. The analysis part of SIRR inspects the direction of arrival of the reflection and the diffuseness of the sound field. Since the analysis is done in short time windows, the location of the reflections can be deduced using the a priori knowledge of speed of sound, the time of arrival and the estimated DOA which is calculated from sound intensity vectors.

In another study, a spherical microphone array with an integrated video camera is used in [13, 85, 160] for visually inspecting the reflections. The energy of the spherical beamformer output that is applied for an impulse response that is divided into short time windows is overlayed on top of a panorama video image from the center of the microphones. The location of the reflection is then inspected visually for each frame. The maximum of the beamformer output corresponds to the DOA of the reflection and the distance to the reflection is calculated from the time stamp of the current

frame.

In [161] the reflections are localized using TDOA estimation with a microphone array that consists of 8 microphones. The method is demonstrated in an auditorium.

In [5] the room geometry is estimated by rotating a B-format microphone around a loudspeaker, directed towards the microphone. The estimation is based on the TOA and the DOA of the first arriving reflection in each direction. For each direction a single TOA and DOA estimate is obtained. In the post-processing phase the TOA and DOA measurements are grouped using hierarchical clustering to avoid estimating the same plane multiple times.

The reflecting plane parameters are estimated by rotating an omnidirectional microphone around a loudspeaker which is directed towards to microphone in [3]. The impulse responses are transformed into an acoustic localization map from where the local maxima correspond to the plane locations. As the source position is known, the plane parameters can be calculated.

In [4] the reflecting plane parameters are estimated with a common tangent algorithm in two dimensional space. The problem is first formulated into quadractic equation that describes the relation between the TOAs and plane parameters and source location. For a single reflection the solution of this quadratic equation provides the parameters of a single plane. The solution is called the common tangent algorithm (COTA). For multiple planes, the estimated TOAs are first grouped using the generalized Hough transform and then the plane for each group is solved using the COTA. The generalized Hough transform detects the TOAs that describe the same plane. The approaches in [4] are extended to three dimensions in [164]. Moreover, a closed form solution for the plane parameter estimation from the quadratic equation is presented in [165].

COTA is applied in [162] for the estimation multiple plane parameters in two dimensional space. Whereas in [4] the grouping was done with the generalized Hough trasnform in [162] the grouping is done with an iterative search. The iteration proceeds as follows. First the parameters of the closest plane are estimated. Then the TOAs associated with the first plane are removed and the search is performed again. This iteration is performed as many times as there are a priori known planes.

In [163] a closed form solution to the above mentioned quadratic equation that describes the relation between the TOA and plane parameters is

presented for the 2-D case. In the solution, two planes are selected where the cost function is inhomogeneous. Then, the gradients of the cost function on these planes are solved analytically. The minimum of the obtained solutions corresponds to the plane parameters. Moreover, the generalized Hough transform is applied to improve the estimation of the parameters of a single plane.

The room geometry has also been estimated from continuous signals [15, 166–169]. The advantage of these approaches is that they are blind, i.e. there is no need for controlled source signal.

Inverse mapping of the multi-path propagation problem for first order reflections in TDOA framework is presented in [15]. The mapping is used together with acoustic source localization to estimate reflective surfaces from speech signals in meeting rooms.

In [167] a circular microphone array is used around a loudspeaker to estimate the room geometry. A constrained room model and L1-regularized least-squares method is used to obtain the locations of walls. This method can be considered as semi-blind since it requires the knowledge of the number of walls.

Acoustic imaging for finding room geometry and other acoustic properties of enclosure is applied in [86, 87]. Acoustic imaging is based on the inverse extrapolation of the Kirchoff-Helmholtz and Rayleigh integrals. An acoustic image can be created by measuring multiple impulse responses, for example, on a line grid with B-format microphone [86, 87].

In [166], the location of the reflections is found by beamforming a speech signal. The direction of the source is found from the maximum direction and the direction of the reflections corresponds to smaller local maxima in the beamformer output. The TDOA between the reflection and the direct sound can be estimated from the beamformer output. From the directions and the TDOA the location of the reflector can then be deduced.

The location of a planar reflector is estimated in two dimensions from direction of arrival estimates in [168]. An unconstrained least squares solution is developed for quadratic constraints that represent the reflection path parameters.

In [169] the location of planar reflector is estimated in two dimensions using a white noise source and spherical beamforming [169]. A very similar approach is used in [170] where reflectors are localized in three dimensions using music signals and spherical beamforming. The difference is that a spherical microphone array is used in [170] and circular in [169].

The basic idea in [169, 170] is identical to the one presented in [166], the difference is in the beamforming techniques and in the TDOA estimation method.

3.4 Automatic calibration

In principle, any general localization method can be used to calibrate the positions of the loudspeakers and the microphones in the measurement system. Previously, at least MLE for TOA or TDOA [109–111,171,172], LS for TOA [142], Multi-PHAT for TDOA [15], and beamforming [173] have been used to calibrate some parts of the measurement system. The requirement for the number of microphones and/or loudspeakers are given for different calibration cases in [109].

3.5 Visualization of reflections

The visualization of the reflections is an important step in studying them. A good visualization enables intuitive and quick inspection of the reflections and their properties. The reflections can be illustrated by overlaying them on top of an image as in [13, 85, 160, 174]. In [1] the directions of the reflections are plotted on top of the spectrogram of the impulse response.

3.6 Application areas

Concert hall acoustics can be studied effectively by subjective listening tests [16, 175]. The methodology used in [16] and [175] allows the comparison between objectively measured physical features of the concert halls and subjectively elicited attributes. It is not yet fully understood which physical properties of the concert hall acoustics explain the subjective perception of the acoustics [175].

The main motivation for the studies in this thesis is that it is thought that some properties of the acoustics of concert halls and other musical performance spaces can be explained by the features of the early reflections. As an example, the importance of temporal envelope preserving early reflections has been recently demonstrated in concert halls [176].

These reflections are reflected from flat surfaces.

If some feature of the reflection is to be extracted, the location of the reflection is needed. Although the location can be calculated from the computer aided design schemes obtained from architectural design, this might be cumbersome if the geometry is complex. In addition, the architectural design schemes of the enclosure are not always available. Since the spatial room impulse responses are measured in the acoustic studies anyway, the localization of the reflections from them is a natural choice.

In addition to the main motivation of this thesis, the location of the reflections are useful to know, for example, in acoustic source localization that utilizes reflections [15, 177–184]. Overall, these methods exhibit better performance than the traditional acoustic source localization methods when strong enough specular reflections are present.

4. Room impulse response measurement

In a room environment, sound $s(t)$, emitted from the sound source at position $\mathbf{s}$, and received at receiver n at position $\mathbf{r}_n$, is affected by the impulse response $h(t; \mathbf{r}_n, \mathbf{s})$:

$$p(t; \mathbf{r}_n, \mathbf{s}) = h(t; \mathbf{r}_n, \mathbf{s}) * s(t) + w(t), \tag{4.1}$$

where $*$ denotes convolution and, $w(t)$ is the measurement noise, independent and identically distributed for each receiver. For simplicity, the impulse response measured at receiver n is noted with $h_n(t)$ in this section.

4.1 Standard measurement technique

In the majority of cases, the impulse response is measured by playing back a signal $s(t)$ from a loudspeaker and recording it with a microphone. The most popular signals for the source excitation are the sine-sweep [185, 186], the maximum length sequence [187], and the optimized time-streched pulse [188]. An estimate of the impulse response is then obtained by deconvolution of the received signal and the source signal

$$\hat{h}(t; \mathbf{r}_n, \mathbf{s}) = p(t; \mathbf{r}_n, \mathbf{s}) *^{-1} s(t) + \tilde{w}(t), \tag{4.2}$$

where $*^{-1}$ is the deconvolution operator and $\tilde{w}(t)$ is the (i.i.d.) noise term. As well-known, deconvolution corresponds to division in the frequency domain. The signal to noise ratio in the impulse response measurement is defined by the ratio

$$\text{SNR [dB]} = 10 \log_{10} \left(\frac{\hat{h}^2(t_{\text{dir}})}{\text{E}\{\tilde{w}^2(t)\}} \right), \tag{4.3}$$

where t_{dir} is the time of arrival of the direct sound. The noise variance (the energy) can be approximated from the beginning of the impulse response, before the direct sound, or from the end, where there is no signal.

The ISO-standard for room acoustic parameters states that an omnidirectional source is to be used [8]. By definition, the omnidirectional source emits equal amount of energy to all directions. In reality, according to the standard [8], small variations up to 6 dB are allowed in different frequency bands.

4.1.1 Sine-sweep technique

The sine sweep source signal is given as [186]:

$$s(t) = \sin\left(\frac{\omega_1 T}{\log\{\omega_2/\omega_1\}}\left(\exp\left(\frac{t}{T}\log\{\omega_2/\omega_1\}\right) - 1\right)\right), \tag{4.4}$$

where ω_1 and ω_2 are the lower and upper frequency of the sweep, and T is the total length of the sweep. The advantage of the sine-sweep signal over the maximum length sequence is that the harmonic distortion of the loudspeaker can be removed from the impulse response as pointed out by Farina, e.g. in [186]. Sometimes the sine-sweep with Eq. (4.4) is referred to as logarithmic sine sweep, since in logarithmic scale the frequency changes linearly. The SNR achieved with the sine-sweep technique is approximately from 60 to 90 dB in the measurements taken for this thesis.

4.1.2 On the use of natural sound sources

Sometimes balloon bursts and gunshots [66, 70] are used as the source signal. In this case, the exact source signal is unknown, and therefore the emitted sound is usually assumed to resemble an impulse closely enough. However, at least balloon bursts have been shown not to fulfill the ISO-standard on the directivity of an omnidirectional source [189]. In addition, the balloon burst has a poor repeatability if the balloon type, the pressure, or the bursting technique changes [189].

4.2 The sparse impulse response technique

The ISO-standard measurement is well suited for the estimation of the traditional room acoustic parameters. However, here the interest is in the early reflections and their properties. With the omni-directional source, if the length of the reflection path, that is the path from the source via the reflections to the receiver, is equal with two or more reflections, then they

arrive at the receiver at the same time and can not be localized properly. In practical situations, a short time window is used in the analysis of the reflections. Then the reflection paths need only to be approximately the same when they already overlap in the analysis window and interfere with the directional analysis. In addition, since the reflections in real situations are often not discrete events, they tend to spread over time and overlap with each other even more. Moreover, due to the physical limitations of the loudspeakers in dimensions and on the frequency band, even the emitted sound field is not a perfect Dirac-impulse, especially if the loudspeaker consists of several elements as shown in Fig. 2.3.

Recently a novel measurement technique, the sparse impulse response technique, for the investigation of early reflections was developed in Publication I. The technique takes advantage of directional loudspeakers. A directional loudspeaker emits more sound in some directions than others. When a room impulse response is measured with such a directional loudspeaker, some reflections are excited with more energy than others. This way, some reflections have a better signal-to-noise or signal-to-interference ratio than others and should be more separable in the impulse response.

4.2.1 Measurement

The impulse response measured with a directional loudspeaker that is directed to an angle $\{\theta_s, \phi_s\}$ at time instant t is denoted with $h(t, \theta_s, \phi_s)$ and it is named in Publication I, as *a sparse impulse response*. Here, the loudspeaker is only rotated with respect to the z-axis, therefore ϕ_s is no longer used in the notation. In theory, if the loudspeaker has an infinitely narrow directivity, all the reflections that do not have exactly the same reflection path length should be separable in time and space. The idea is analogous to the ray tracing method [190], used, for example, in room acoustics simulations, where the rays are first sent from the source position and then observed in the receiver position. However, since infinitely narrow directionality is not practically achieved with loudspeakers, the idea is more analogous to beam-tracing [191] than ray-tracing.

In the case of unequal reflection paths, and ideal reflections, the rotation angle that produces the largest absolute pressure at some time instant t gives the direction to which the loudspeaker was directed to produce the sound pressure observed in the receiver. Thus, the direction of the loudspeaker can be estimated as the maximum argument of the absolute

pressure values of the sparse responses, i.e.:

$$\hat{\theta}_s(t) = \arg\max_{\theta_s}\{|h(t,\theta_s)|\}. \tag{4.5}$$

When $\hat{\theta}_s(t)$ is used as an argument in the sparse impulse response as $h(t,\hat{\theta}_s(t))$, an impulse response that includes the reflections from the strongest direction of the loudspeaker is formed. This response is named as the *compound sparse response*. The separability in space, and the knowledge of the geometry of the enclosure, allows the tracing of reflections at each time instant from the source to the receiver.

In theory, by using only one microphone and an infinitely directive loudspeaker that produces Dirac-impulses, and having only ideal perfectly specular reflections, Eq. (4.5) will produce all the reflections that do not have equal path lengths. However, since in reality, the impulses are not perfect Dirac-impulses, not all the reflections are separable in real situations.

4.2.2 Comparison to other techniques and discussion

Other authors have also spanned a directional loudspeaker around its axis to achieve more spatial separation. Günel was the first to present this idea in room acoustic measurements [5]. In [5] a loudspeaker is directed to different angles around its z-axis and a B-format microphone is at a fixed length and direction with respect to the loudspeaker. Antonacci *et al.* also span a directional loudspeaker around its z-axis [3, 4]. The setup is otherwise the same as in Günel's method but the B-format microphone is replaced with an omnidirectional microphone.

The difference of the proposed method and other methods, is that the sparse impulse response and the compund sparse impulse response can be measured with any loudspeaker and microphone setup whereas other methods are designed for setup where a microphone and a loudspeaker are interconnected. Moreover, Günel's method only considers one impulse response in one direction at a time, whereas the presented method considers all the directions simultaneously in the compund sparse phase presented in Eq. (4.5). Thus, the presented method is designed to replace the traditional single source impulse response measurement, when the other presented methods with loudspeaker spanning are specially designed for a certain measurement task, e.g. room geometry estimation as in [5].

The spatio-temporal separability of the reflections can be achieved by using directional microphone or directional loudspeakers. Here, the ad-

vantages or disadvantages of the directional loudspeakers over the directional microphones are not studied. A comparison is made only with the traditional omni-directional and the directional loudspeakers measurements.

4.3 Experiments

The goal is to compare the proposed technique with two loudspeakers to the standard measurement technique in same conditions. The loudspeakers are Genelec 1029 A, Panphonics panel loudspeaker, and a standard omni-directional loudspeaker. The directivities of the sources is discussed and depicted in Section 2.4. The impulse responses are measured using the sine-sweep technique at 48 kHz, and the frequency band is from 40 Hz to 24 kHz.

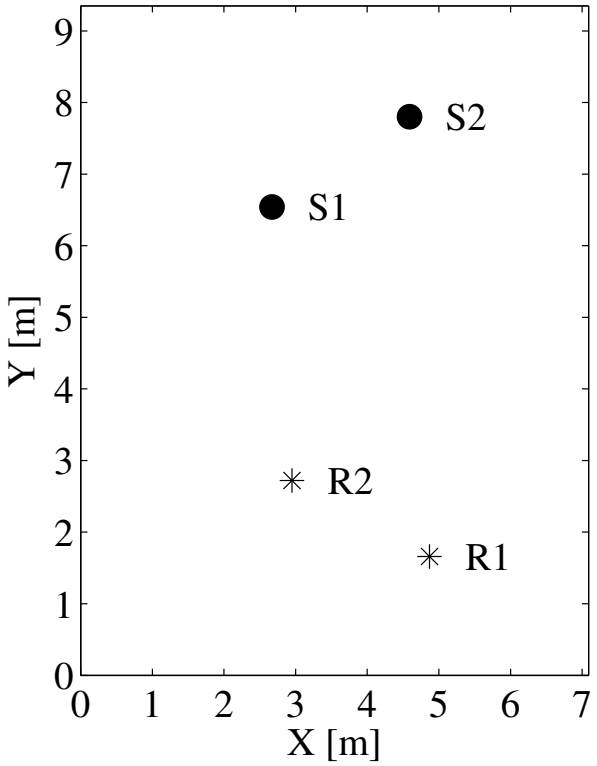
4.3.1 Setup

Experiments are conducted in two auditoria illustrated in Fig. 4.1. Auditorium 1 has a volume of 250 m^3. The auditorium was stripped of all furniture and has a shoebox shape. The acoustic center of the source and location of the microphone array center are shown in Fig. 4.1(a). In addition, the height for both the source and the array was set to 1.4 m. The G.R.A.S microphone array with 6 microphones, shown in Fig. 2.1(b) is used in both receiver locations.

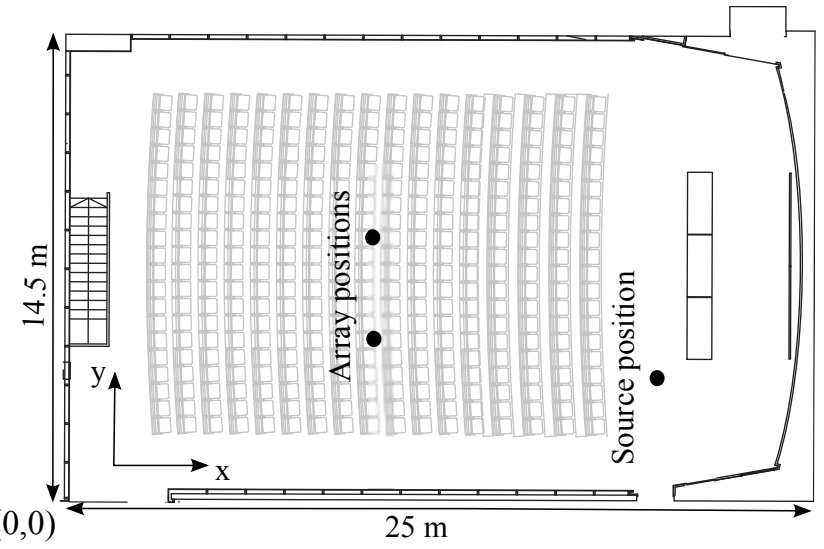
In Auditorium 2, shown in Fig. 4.1(b), the audience area has an inclination of about 10 degrees, as the height of the auditorium decreases from about 8 m to 5 m, leading to a volume of 1800 m^3. One source position and two array positions were used in the experiments of Auditorium 2. The height of the source and the array in this auditorium were about 1.2 m from the floor level. Each of the receiver locations in Auditorium 2 has the TKK-3D 12 -microphone array illustrated in Fig. 2.1(a).

4.3.2 Results

Figure 4.2 shows examples of the sparse impulse responses measured at every 10 degrees in azimuth angle with Genelec 1029A and the Panphonics loudspeaker from Auditorium 2. The corresponding compound impulse responses are shown below the sparse responses, in Fig. 4.2. Visual in-



(a) Auditorium 1



(b) Auditorium 2

Figure 4.1. Array (R) and source (S) positions and the floorplans of the auditoriums.

spection shows that, when compared to a response measured with an omnidirectional source in the same position, shown in the bottom plot of Fig. 4.2, the proposed measurement technique provides higher peaks that can be easily recognized with both tested loudspeakers. Although the sparse response for the Panphonics loudspeaker is shown from 0 to 350 degrees in Fig. 4.2, only the angles from 0 to 170 degrees are used in the analysis with the Panphonics loudspeaker due to the dipole directivity pattern.

In addition to the visual inspection, the performance of the impulse response measurement can be verified by counting the number of recognizable reflections within the impulse response. A good impulse response for the reflection tracing task is the one that has more identifiable reflections. Here, the number of recognizable reflections is calculated using the local energy ratio [81].

The identification of a reflection is based on the relation between the absolute sound pressure in a small analysis window and the current absolute sound pressure. The local energy is calculated for the directional sources from the compound sparse responses as:

$$E_{\text{loc}}(t) = \frac{1}{T_{\text{loc}}} \int_{\tau=t-T_{\text{loc}}/2}^{\tau=t+T_{\text{loc}}/2} \text{w}_H(t)|h(\tau, \hat{\theta}_s(t))|d\tau \tag{4.6}$$

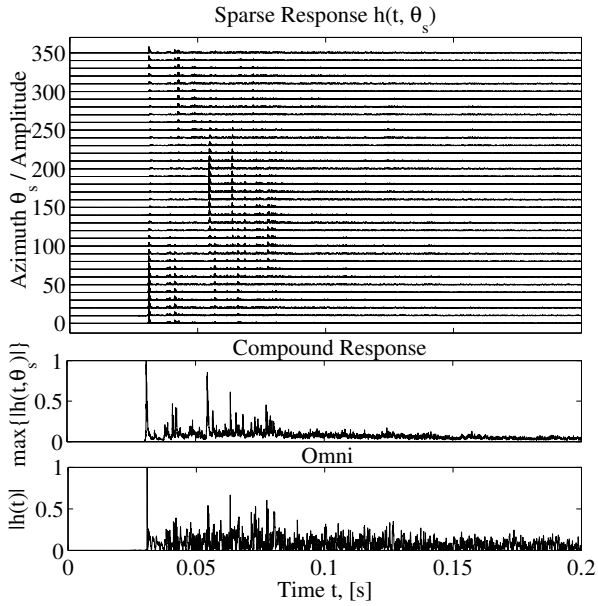
where $\text{w}_H(t)$ is a Hanning window function of length $T_{\text{loc}} = 128$ samples (2.67 ms). Note that, unlike in [81], here a Hanning windowing function is used. The decision whether the sample is a reflection or not, is given by [81]:

$$h_{\text{refl}}(t) = \begin{cases} 1, & \text{if} \quad |h(t, \hat{\theta}_s(t))| > \varepsilon E_{\text{loc}}(t) \\ 0, & \text{otherwise,} \end{cases} \tag{4.7}$$

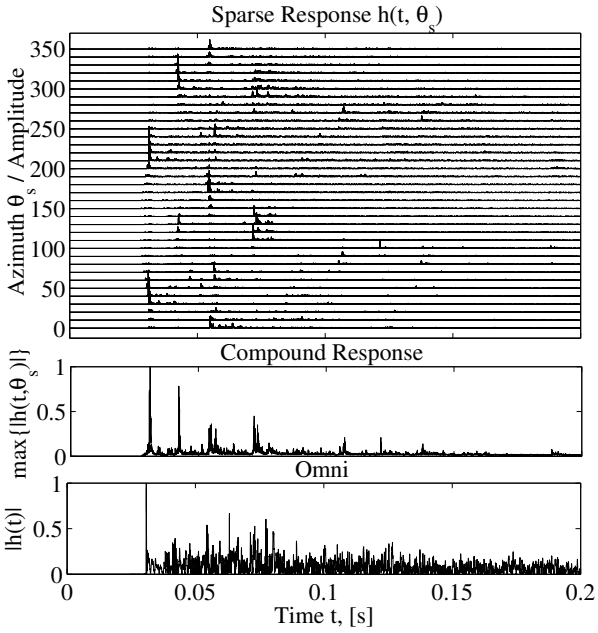
where ε is the threshold value for the detection. For the omnidirectional source, the detection procedure is the same with the exception that the standard impulse response is used instead of the compound sparse response.

The number of identified reflections, noted here with K, with respect to the threshold ε is shown in Fig. 4.3. The results are averaged over all the measured impulse responses for each auditorium. That is, for Auditorium 1 and 2, the results are averaged over 24 measurements for each loudspeaker type. In Auditorium 1 the 24 measurements consists of a single source position and 12 microphones of the TKK-3D array in two different receiver locations. In Auditorium two the 24 measurements include two source positions and six microphones of the G.R.A.S. microphone array in two receiver positions.

The results indicate that the proposed measurement technique provides more recognizable reflections than the standard measurement technique. With an arbitrarily selected threshold of $\varepsilon = 4$, the omnidirectional source, Genelec 1029A, and the Panphonics panel loudspeaker give 2, 61, and 132, reflections for the Auditorium 1, and 3, 101, and 169, for Auditorium 2, respectively. Thus, the more directional the loudspeaker is, the more individual reflections can be identified. The number of identified reflections depends strongly on the threshold. However, the order of the number of detected reflections with different loudspeakers stays the same, no matter what threshold value is selected. In addition, as expected, the larger space (Auditorium 2) has more identifiable reflections. As the distance between the individual reflections is longer in a larger space the reflections become more separable.

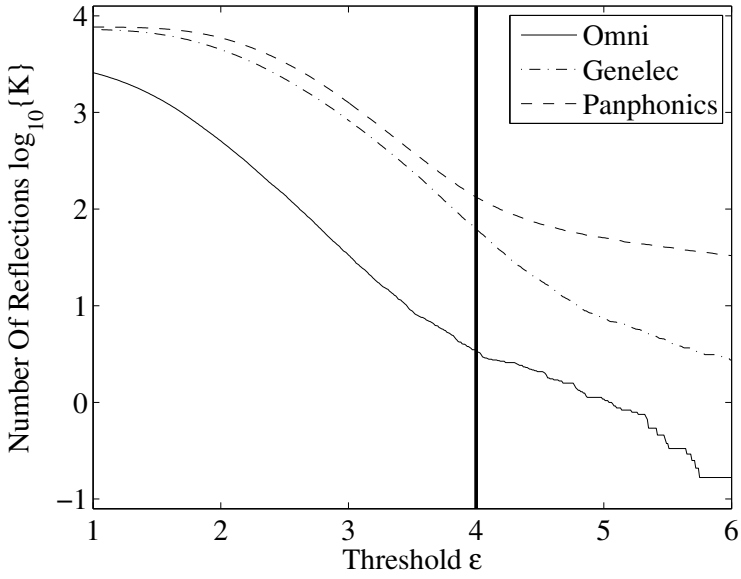


(a) Genelec 1029A

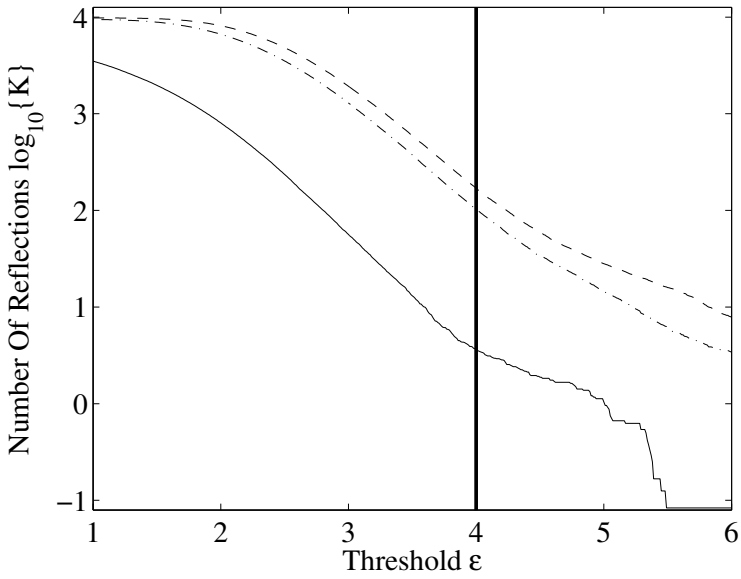


(b) Panphonics

Figure 4.2. Sparse impulse responses with (a) Genelec 1029A and (b) the Panphonics loudspeaker on a wide band from 40 Hz to 24 kHz in Auditorium 2. The panphonics loudspeaker provides sharper peaks to the sparse impulse response due to its higher directionality. The maximum is normalized to 0 dB for the compound sparse responses and the response measured with the standard method.



(a) Auditorium 1



(b) Auditorium 2

Figure 4.3. Number of identified reflections versus the local energy ratio threshold. The selected threshold ($\varepsilon = 4$) is depicted with a thick black line. The proposed measurement technique with Genelec 1029A and the Panphonics loudspeaker provides more spatial separability than the standard measurement technique with omnidirectional source since more reflections are found.

5. Localization Methods

This chapter presents methods from earlier research that are applied in this thesis for the localization of reflections. Also some novel ad-hoc localization functions are proposed. It should be noted, that the analysis does not differentiate between a reflection and other acoustic phenomena, but all the sound waves arriving at the receivers are considered reflections and treated in the same manner.

5.1 Signal Model

The spherical wave propagation model (see Chapter 2) is assumed. The formulations are the same for plane wave propagation model with the exception that the TOA and TDOA terms are replaced by those given in Eqs. (2.38) and (2.40), respectively. The only exception is the sound intensity vector based localization which is only capable of direction of arrival estimation and assumes always plane wave propagation model.

The assumed signal model is the following

$$\begin{aligned} h_1(t) &= a_1 s(t-t_1) + w_1(t) \\ h_2(t) &= a_2 s(t-t_2) + w_2(t) \\ &\vdots \quad\quad \vdots \\ h_N(t) &= a_N s(t-t_N) + w_N(t), \end{aligned} \tag{5.1}$$

where the noises $w_i(t), i = 1, \cdots, N$ are normally distributed and uncorrelated with each other and with the loudspeaker impulse response $s(t)$.

Then, in the frequency domain, the signal model is given as

$$\begin{aligned} H_1(\omega) &= A_1(\omega)S(\omega)e^{-j\omega t_1} + W_1(\omega) \\ H_2(\omega) &= A_2(\omega)S(\omega)e^{-j\omega t_2} + W_2(\omega) \\ \vdots &\quad \vdots \\ H_N(\omega) &= A_N(\omega)S(\omega)e^{-j\omega t_N} + W_N(\omega), \end{aligned} \tag{5.2}$$

where the signal, noise, and received signal have spectral densities $G_{s,s}(\omega) = \mathrm{E}[S(\omega)S^*(\omega)]$, $G_{w_1,w_1}(\omega) = \mathrm{E}[W_1(\omega)W_1^*(\omega)]$, and $G_{h_1,h_1}(\omega) = \mathrm{E}[H_1(\omega)H_1^*(\omega)]$, respectively. The amplitudes $A_n(\omega), n = 1, \cdots, N$ are dependent on the distance from the source to the microphones, the directivity of the source and of the microphones, the properties of the reflective surfaces, and the air absorption. Here the amplitudes are assumed to be equal to unity, i.e.,

$$A_n(\omega) = 1, \forall n, \omega.$$

This model is assumed for simplicity in the cases studied in this thesis since omnidirectional microphones are used, the aperture size of the microphone array is small, and the loudspeaker is in the far-field. Moreover, it is assumed that the reflections can be windowed from the spatial impulse responses.

5.2 Time difference of arrival estimation

In the TDOA estimation, the task is to estimate the time delay $\tau_{i,j} = t_i - t_j$ between two received signals $h_i(t)$ and $h_j(t)$. The maximum argument of the estimation function $R_{h_i,h_j}(\tau)$ is the TDOA estimate, i.e.,

$$\hat{\tau}_{i,j} = \arg\max_{\tau}\{R_{h_i,h_j}(\tau)\}. \tag{5.3}$$

Next, two approaches used in previous research for TDOA estimation are formulated.

5.2.1 Generalized correlation method

The most used TDOA estimation approach is the generalized correlation method [25]. The generalized cross correlation (GCC) function between two received impulse responses h_i and h_j is calculated as [25]:

$$\mathrm{R}^{\mathrm{GCC}}_{h_1,h_2}(\tau) = \mathcal{F}^{-1}\{\mathcal{W}(\omega)\hat{G}_{h_1,h_2}(\omega)\}, \tag{5.4}$$

where $\mathcal{W}(\omega)$, and $\mathcal{F}^{-1}$, are the weighting function, and inverse Fourier transform, respectively.

Maximum likelihood estimation

The well-known maximum likelihood weighting is given as [25]

$$\mathcal{W}_{h_1,h_2}^{\mathrm{MLE}}(\omega) = \frac{1}{|G_{h_1,h_2}(\omega)|}\frac{C_{h_1,h_2}(\omega)}{[1-C_{h_1,h_2}(\omega)]} \tag{5.5}$$

where

$$C_{h_1,h_2}(\omega) = \frac{|G_{h_1,h_2}(\omega)|^2}{G_{h_1h_1}(\omega)G_{h_2h_2}(\omega)} \tag{5.6}$$

is the magnitude squared coherence function. For the derivation of the MLE weighting function see [25]. Since the noises are assumed to be uncorrelated, the true spectral densities can be written as [25]

$$G_{h_1,h_2}(\omega) = G_{s,s}(\omega)e^{-j\omega\tau_{i,j}}, \tag{5.7}$$

$$G_{h_1,h_1}(\omega) = G_{s,s}(\omega) + G_{w_1,w_1}(\omega), \text{and} \tag{5.8}$$

$$G_{h_2,h_2}(\omega) = G_{s,s}(\omega) + G_{w_2,w_2}(\omega) \tag{5.9}$$

Then, by using these equivalences in Eq. (5.5), one has

$$\mathcal{W}_{h_1,h_2}^{\mathrm{MLE}}(\omega) = \frac{G_{s,s}(\omega)}{G_{w_1,w_1}(\omega)G_{w_2,w_2}(\omega) + G_{s,s}(\omega)G_{w_1,w_1}(\omega) + G_{s,s}(\omega)G_{w_2,w_2}(\omega)}. \tag{5.10}$$

In practical situation, since the signal is an impulse response, it is easy to estimate the noise auto power spectral density $G_{w_1,w_1}(\omega)$ from the beginning of the impulse response. Then, the auto spectral density of the source signal is obtained from Eq. (5.8), e.g., $G_{s,s}(\omega) = G_{h_1,h_1}(\omega) - G_{w_1,w_1}(\omega)$. The MLE weighting then formulates to

$$\begin{aligned}\mathcal{W}_{h_1,h_2}^{\mathrm{MLE}}(\omega) = \quad & G_{h_1,h_2}(\omega)\times \\ & \{G_{w_1,w_1}(\omega)G_{w_2,w_2}(\omega)+ \\ & [G_{h_2,h_2}(\omega) - G_{w_1,w_1}(\omega)]G_{w_1,w_1}(\omega)+ \\ & [G_{h_1,h_1}(\omega) - G_{w_2,w_2}(\omega)]G_{w_2,w_2}(\omega)\}^{-1}.\end{aligned} \tag{5.11}$$

By assuming that the spectral densities of the noise signals are equal $G_{w,w}(\omega) = G_{w_2,w_2}(\omega) = G_{w_1,w_1}(\omega)$, one has

$$\mathcal{W}_{h_1,h_2}^{\mathrm{MLE}}(\omega) = \frac{G_{h_1,h_2}(\omega)}{(G_{h_2,h_2}(\omega) + G_{h_1,h_1}(\omega))G_{w,w}(\omega) - G_{w,w}^2(\omega)}. \tag{5.12}$$

Note that there are three options for estimating $G_{s,s}(\omega)$ and two options for estimating $G_{w,w}(\omega)$. One possibility is to estimate $G_{s,s}(\omega)$ as the (weighted) average over the different estimates, and insert them in to

$$\mathcal{W}_{h_1,h_2}^{\mathrm{MLE}}(\omega) = \frac{1}{2G_{w,w}(\omega) + G_{w,w}^2(\omega)/G_{s,s}(\omega)}. \tag{5.13}$$

If the noise can not be estimated, the first version of the MLE weighting in Eq. (5.5) can be used, but the coherence should then be estimated using for example Welch's approach [26, 192]. Coherence estimation can be problematic for non-stationary signals [92]. In addition, since it includes additional computational load, it is not used in this thesis.

Other weighting functions

Practical weighting functions that do not require estimation of the noise auto power spectral densities exist. In this thesis, direct cross correlation (CC) weighting [25]

$$\mathcal{W}^{\mathrm{CC}}(\omega) = 1 \tag{5.14}$$

and phase transform (PHAT) are used

$$\mathcal{W}^{\mathrm{PHAT}}_{h_i,h_j}(\omega) = 1/\|G_{h_i,h_j}(\omega)\|. \tag{5.15}$$

5.2.2 Average square difference function

Similar to the generalized correlation method, are the difference function based methods [96]. In these methods, two signals are subtracted from each other, while the other signal is delayed by the TDOA. Here, the average squared difference function (ASDF) is also tested [95, 96]:

$$\mathrm{R}^{\mathrm{ASDF}}_{h_i,h_j}(\tau) = \int_{-T/2}^{T/2} [h_i(t) - h_j(t-\tau)]^2 dt, \tag{5.16}$$

where T is the length of the integration window. With ASDF, instead of the maximum, the minimum argument of the estimation function is the TDOA estimate

$$\hat{\tau}_{i,j} = \arg\min_{\tau} \{\mathrm{R}^{\mathrm{ASDF}}_{h_i,h_j}(\tau)\}. \tag{5.17}$$

5.3 Time of arrival estimation

In time of arrival estimation, the delay t_n of a signal is estimated. In a short time window the maximum argument of the estimation function $D_n(t)$ is the TOA estimate

$$\hat{t}_n = t_{\mathrm{start}} + \arg\max_{t} \{D_n(t)\}, \tag{5.18}$$

where t is limited by the starting point, and the ending point of the time window, i.e., $t_{\mathrm{start}} < t < t_{\mathrm{end}}$.

Since the problem is similar to TDOA estimation, also the TDOA estimation methods introduced above can be applied for TOA estimation. This requires the knowledge of the source signal.

5.3.1 Auto correlation method

This method requires a priori information of the sound source used. First, a reference $s(t)$ is measured for the sound source in free-field conditions: in an anechoic chamber, or it can be windowed from an in-situ impulse response. The reference represents the waveform of the emitted impulse response from the source. The reference is then correlated with the impulse response

$$D^{\mathrm{AC}}_{s,h_1}(t) = \int_{-T/2}^{T/2} s(\xi)h_1(\xi + t)d\xi, \tag{5.19}$$

where AC denotes auto correlation, and T ms is the length of the short time analysis window. Defrance *et al.* use similar auto correlation approach for detecting reflections from a single impulse response [65, 66]. In addition, similar auto correlation method has been used to detect the TOA of a reflection as a preliminary task before absorption coefficient calculations [42].

Maximum likelihood estimation

The autocorrelation function can be given in the frequency domain as the generalized correlation function

$$\mathrm{D}^{\mathrm{AC}}_{s,h_1}(\tau) = \mathcal{F}^{-1}\{\mathcal{W}_{s,h_1}(\omega)G_{s,h_1}(\omega)\} \tag{5.20}$$

By definition, the maximum likelihood weighting also for this method is given by Eq. (5.5). Since the other signal is the true signal without noise the spectral densities can be written as

$$G_{s,h_1}(\omega) = G_{s,s}(\omega)e^{-j\omega t_1}, \text{and} \tag{5.21}$$

$$G_{h_1,h_1}(\omega) = G_{s,s}(\omega) + G_{w_1,w_1}(\omega). \tag{5.22}$$

Then, the MLE weighting for the auto-correlation method is given as

$$W^{\mathrm{MLE-AC}}_{s,h_1}(\omega) = \frac{1}{|G_{s_1,h_1}(\omega)|}\frac{C_{s,h_1}(\omega)}{[1 - C_{s,h_1}(\omega)]} \tag{5.23}$$

$$= \cdots \tag{5.24}$$

$$= \frac{1}{G_{w_1,w_1}(\omega)} \tag{5.25}$$

where $C_{s,h_1}(\omega)$ is the magnitude squared coherence between s and h_1.

The analogy between the AC method for TOA and the generalized correlation method for TDOA is obvious. The difficulty with the AC method is that, a real loudspeaker emits different impulses in different directions.

Thus, the method requires the response of the loudspeaker in each direction as a priori knowledge. This can be artificially done using the sparse impulse response technique as in Publication I.

5.3.2 Maximum absolute pressure

Peak detection is a straightforward method to detect the TOA of a sound wave. It is assumed that the arriving sound wave introduces an impulse, a local maximum or minimum, that can be detected. The maximum argument is then the estimated TOA

$$\hat{t}_n = \arg\max_t \{|h_n(t)|\}. \tag{5.26}$$

This may also include some windowing or filtering.

5.3.3 Other methods

The statistical features of impulse response differ when there is a reflection present in the analysis window [53, 63, 70, 104]. One way of measuring the statistical difference is the kurtosis [104]. Other option is to detect the peak from a local absolute pressure ratio between the current absolute pressure and its surroundings [81]. Here, these statistical approaches are no longer pursued in the TOA estimation.

5.4 Localization functions

When robust 3-D or 2-D localization is required, the TOA or TDOA information is combined spatially over several microphones and microphone pairs, respectively. Three commonly used state-of-the-art acoustic source localization functions are formulated next for TOA, TDOA, and their combination. This leads to nine different localization functions in total. That is, for each dataset (TOA, TDOA, or their combination) three methods are formulated. In addition, the methods are compared to a MLE function designed for the signal model. Also some least-squares localization approaches and sound intensity vector based methods are discussed.

For each method, the maximum argument of the localization function $P(\mathbf{x})$ is the location estimate, i.e.

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \{P(\mathbf{x})\}. \tag{5.27}$$

For notational convenience, a TOA, $t(\mathbf{r}_n;\mathbf{x})$, is denoted by $t_n(\mathbf{x})$, where $n = 1 \dots N$, and N is the number of microphones. In addition, a TDOA, $\tau(\mathbf{r}_i, \mathbf{r}_j;\mathbf{x})$, is denoted by $\tau_m(\mathbf{x})$, where $m = \{i, j\} = 1 \dots M$ is a tuple, and M is the number of microphone pairs. The TDOA estimates are denoted with $\hat{\tau}_m$, and the TDOA estimation function $R_{h_i,h_j}(\tau)$ with $R_m(\tau)$. In this thesis, the number of microphones is $N = 6$, and the number of microphone pairs is $M = 15$. Then, the microphone pairs m from 1 to 15 are $\{\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{1,6\},\{2,3\},\{2,4\},\{2,5\},\{2,6\},\{3,4\},\{3,5\},\{3,6\},\{4,5\},\{4,6\},\{5,6\}\}$.

5.4.1 Maximum likelihood estimation for time of arrival and time difference of arrival

The MLE function for TDOA is given as the joint probability density function [109]

$$P_{\text{MLE-TDOA}}(\mathbf{x}) = \prod_{m=1}^{M} p(\hat{\tau}_m; \tau_m(\mathbf{x})) = \tag{5.28}$$

$$= \frac{\exp(-\frac{1}{2}[\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}(\mathbf{x})]\Sigma^{-1}[\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}(\mathbf{x})]^{\mathrm{T}})}{(2\pi)^{(M)/2}\sqrt{\det(\Sigma_{\text{TDOA}})}}, \tag{5.29}$$

where $p(\hat{\tau}_m; \tau_m(\mathbf{x}))$ is the normal error probability density function for a TDOA estimate,

$$\hat{\boldsymbol{\tau}} = [\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_M], \tag{5.30}$$

$$\boldsymbol{\tau}(\mathbf{x}) = [\tau_1(\mathbf{x}), \tau_2(\mathbf{x}), \dots, \tau_M(\mathbf{x})], \tag{5.31}$$

$$\Sigma_{\text{TDOA}} = \mathbf{I}\sigma^2_{\text{TDOA}}, \tag{5.32}$$

with σ_{TDOA} as the standard deviation of the error and $\tau_m(\mathbf{x})$ is given by Eq. (2.39).

The MLE function for TOAs, assuming normally distributed errors is given as [109]

$$P_{\text{MLE-TOA}}(\mathbf{x}) = \prod_{n=1}^{N} p(\hat{t}_n; t_n(\mathbf{x})) = \tag{5.33}$$

$$= \frac{\exp(-\frac{1}{2}[\hat{\boldsymbol{t}} - \boldsymbol{t}(\mathbf{x})]\Sigma^{-1}[\hat{\boldsymbol{t}} - t(\mathbf{x})]^{\mathrm{T}})}{(2\pi)^{(N)/2}\sqrt{\det(\Sigma_{\text{TOA}})}}, \tag{5.34}$$

where $p(\hat{t}_n; t_n(\mathbf{x}))$ is the normal error probability density function for a TDOA estimate,

$$\hat{\boldsymbol{t}} = [\hat{t}_1, \hat{t}_2, \dots, \hat{t}_M], \tag{5.35}$$

$$\boldsymbol{t}(\mathbf{x}) = [t_1(\mathbf{x}), t_2(\mathbf{x}), \dots, t_M(\mathbf{x})], \tag{5.36}$$

$$\Sigma_{\text{TOA}} = \mathbf{I}\sigma^2_{\text{TOA}}, \tag{5.37}$$

with σ_{TOA} as the error standard deviation and $\tau_m(\mathbf{x})$ is given by Eq. (2.37).

For combining the TOA and TDOA information with MLE an assumption is made, that the TDOA and TOA have independent errors. Then, the MLE function for CM is given as the multiplication of MLE-TOA and MLE-TDOA functions:

$$P_{\text{MLE-CM}}(\mathbf{x}) = P_{\text{MLE-TOA}}(\mathbf{x}, \sigma_{\text{TOA}}) P_{\text{MLE-TDOA}}(\mathbf{x}, \sigma_{\text{TDOA}}). \tag{5.38}$$

If different error variances σ^2_{TOA} and σ^2_{TDOA} are assumed for TOA and TDOA, respectively, the MLE-TOA and MLE-TDOA functions have different weightings. In Publication II, it is found that $\sigma^2_{\text{TOA}} = \sigma^2_{\text{TDOA}}$ is a reasonable choice.

The measurement errors of the TDOAs and TOAs can be highly correlated if certain TOA and TDOA estimation methods are used. As a consequence, the covariance matrix of the combined method is no longer a diagonal matrix as assumed above. A further investigation should be conducted to study which of the estimators produce errors that correlate. In Publication II the maximum absolute pressure ise used for the TOA estimation and the direct cross correlation for the TDOA estimation and it is found that in most of the cases the errors do not correlate, i.e. the covariance matrix is diagonal. If the TDOAs are directly calculated from the estimated TOAs then the combined method will have the same performance as the MLE-TOA and will not gain any advantage.

5.4.2 Maximum likelihood estimation for the signal model

Earlier, the maximum likelihood estimation was formulated with respect to TOA and TDOA estimates. It is also possible to formulate the MLE directly with respect to the source signal and the measurement noise [112–117]

$$\begin{aligned} P(\mathbf{x}) &= \prod_{\omega} p(\boldsymbol{H}(\omega); \mathbf{x}) = \\ &= \prod_{\omega} \frac{\exp(-1/2[\boldsymbol{H}(\omega) - \boldsymbol{D}(\omega, \mathbf{x})S(\omega)]\boldsymbol{Q}^{-1}(\omega)[\boldsymbol{H}(\omega) - \boldsymbol{D}(\omega, \mathbf{x})S(\omega)])}{(2\pi)^{N/2}\sqrt{\det(\boldsymbol{Q}(\omega))}}, \end{aligned} \tag{5.39}$$

where

$$\boldsymbol{H}(\omega) = [H_1(\omega), H_2(\omega), \ldots, H_N(\omega)]^{\mathrm{T}}, \tag{5.40}$$

$$\boldsymbol{D}(\omega, \mathbf{x}) = [e^{-j\omega t_1(\mathbf{x})}, e^{-j\omega t_2(\mathbf{x})}, \ldots e^{-j\omega t_N(\mathbf{x})}]^{\mathrm{T}}, \tag{5.41}$$

$$\boldsymbol{Q}(\omega) = \mathbf{I}\sigma_F^2. \tag{5.42}$$

where $\sigma_F^2 = \mathrm{E}\{G_{w,w}(\omega)\}$ is the expected noise variance and it is assumed constant for all frequencies. Under the assumption on the independent errors, and using the log-likelihood leads to a maximization function [112, 114, 116, 117]

$$L_{\text{MLE-S}}(\mathbf{x}) = \int_{\omega} |\sum_{n=1}^{N} H_n(\omega) e^{j\omega t_n(\mathbf{x})}/\sigma_F|^2 d\omega. \tag{5.43}$$

This approach is denoted with MLE-S and it stands for MLE for the signal model.

5.4.3 Steered response power

A popular family of TDOA-based acoustic source localization functions is the SRP methods. In these methods, the acoustic source localization likelihood is evaluated as a spatial combination of cross correlation functions $\mathrm{R}_m(\tau)$ for each location candidate, denoted with x [106, 107]:

$$P_{\text{SRP-TDOA}}(\mathbf{x}) = 1/M \sum_{m=1}^{M} \mathrm{R}_m(\tau_m(\mathbf{x})). \tag{5.44}$$

The SRP using generalized correlation method with PHAT weighting is commonly referred to as SRP-PHAT function, introduced originally in [107].

The signals can be similarly steered using TOAs, as the TDOA estimation functions were steered using TDOAs. In steered beamforming the signals are artificially steered by delaying them towards a location. The sum-and-delay beamformer is considered as the most basic case of beamforming [193]. When the sum-and-delay beamformer output is squared the output is SRP [106]

$$P_{\text{SRP-TOA}}(\mathbf{x}) = \int \left| 1/N \sum_{n=1}^{N} h_n(t - t_n(\mathbf{x})) \right|^2 dt. \tag{5.45}$$

This function is the same as MLE with the signal model in Eq. (5.43) without the variance term. However, if Eq. (5.46) is implemented in the frequency domain, the TOA information is lost, since SRP-TOA becomes the same as SRP-TDOA with an additional (constant) energy term [151, 152].

Since the room impulse responses are already directly mapped into the TOAs, the time variable becomes $t = 0$. The time integral over dt then has no effect on the localization function and Eq. (5.45) is written as

$$P_{\text{SRP-TOA}}(\mathbf{x}) = \left| 1/N \sum_{n=1}^{N} h_n(t_n(\mathbf{x})) \right|^2, \tag{5.46}$$

which is computationally more efficient implementation of the SRP-TOA than the first one.

The TOA and TDOA information can be both used to measure the position of a reflection. Intuitively, the next step is to combine both TOA and TDOA information. The SRP function, when TDOA and TOA information are both used, is here proposed to be calculated as

$$P_{\text{SRP-CM}}(\mathbf{x}) = (1-W)P_{\text{SRP-TOA}}(\mathbf{x}) + WP_{\text{SRP-TDOA}}(\mathbf{x}), \tag{5.47}$$

where CM stands for combined method, and $0 < W < 1$ is a weighting factor, included in this function since the steered response is effectively used twice in SRP-CM.

5.4.4 Maximum pseudo-likelihood

Recently it was shown in [108] and [153] that the use of multiplication instead of addition is advantageous in the steering function. This leads to a pseudo-likelihood function [108, 152, 153]

$$P_{\text{PL-TDOA}}(\mathbf{x}) = \prod_{m=1}^{M} \mathrm{R}_m(\tau_m(\mathbf{x})), \tag{5.48}$$

where PL stands for pseudo-likelihood. It should be noted that thresholding and shaping has to be done for the TDOA estimation functions so that they are non-negative pseudo-likelihoods [15]. It is straightforward to show that, if the maximum of TDOA estimation function is modeled with a probability density function, PL-TDOA and MLE-TDOA methods are the same methods.

Here it is proposed that the PL function for TOA is formed by multiplying the individual TOA estimation functions, i.e.,

$$P_{\text{PL-TOA}}(\mathbf{x}) = \prod_{n=1}^{N} D_n(t_n(\mathbf{x})). \tag{5.49}$$

Thresholding and shaping can be done for the TOA estimation functions so that they are non-negative pseudo-likelihoods. In the simplest case, the TOA estimation function is the absolute maximum of the room impulse response:

$$P_{\text{PL-TOA}}(\mathbf{x}) = \prod_{n=1}^{N} |h_n(t_n(\mathbf{x}))|. \tag{5.50}$$

The analogy between PL-TOA and MLE-TOA is the same as with TDOAs. If only one maximum is selected in PL-TOA from the impulse response,

and the corresponding TOA is assigned with an error probability density function, PL-TOA and MLE-TOA are the same methods.

The combined maximum pseudo-likelihood is here proposed to be the multiplication of the PL-TOA and PL-TDOA functions

$$P_{\text{PL-CM}}(\mathbf{x}) = P_{\text{PL-TOA}}(\mathbf{x})P_{\text{PL-TDOA}}(\mathbf{x}). \tag{5.51}$$

As in MLE-CM, also in PL-CM, weighting can be applied for PL-TOA and PL-TDOA functions. If the shaping functions for PL-TOA and PL-TDOA are selected as probability density functions, then PL-CM is equal to the MLE function. Note that here the weighting of PL-TOA or PL-TDOA similarly as SRP-TOA and SRP-TDOA in SRP-CM has not effect, since the weighting will not change the maximum of PL-CM. However, although the PL-CM cannot be weighted, the logarithmic version of it can be, i.e.,

$$\lambda_{\text{PL-CM}}(\mathbf{x}) = (1-W)\log\{P_{\text{PL-TOA}}(\mathbf{x})/N\} + W\log\{P_{\text{PL-TDOA}}(\mathbf{x})/M\}, \tag{5.52}$$

where the log-pseudo-likelihoods of TOA and TDOA are normalized with N and M, respectively. The weighting W is ad-hoc weighting and does not correspond to anything in theory.

5.4.5 Least squares localization approaches

When independent and normally distributed errors are assumed for the MLE-TOA, MLE-TDOA, or MLE-CM, it follows from the properties of the normal distribution, that the mean square error function (MMSE) of the estimates is also the MLE function of estimates [106, 194]:

$$P_{\text{MMSE}}(\mathbf{x}) = \sum_{m=1}^{M} (\hat{\theta}_m - \theta_m(\mathbf{x}))^2, \tag{5.53}$$

With MMSE the minimum argument is the position estimate instead of the maximum. The solution in Eq. (5.53) is of least squares form. Possibly the most straightforward solution for TOA and TDOA data is the unconstrained least squares (ULS).

Table 5.1 lists some of the optimization methods and closed form solutions used for the least squares problem of TOA and TDOA. Possibly some other optimization methods have also been proposed for the problem, but the main focus in this work is not in the optimization methods. In principle, any well behaving global optimization algorithm can be used for the problem, as long as the initial guess given for the algorithm is good enough.

As shown in Table 5.1, interestingly, the ULS solution for TOA has not been presented for planar waves. However, since the plane wave equations are linear, the ULS solutions are trivial to formulate. The ULS solution for spherical wave propagation model with TOA and TDOA can be formulated by integrating solutions in [139] and [118].

5.4.6 Sound intensity vector based localization

Sound intensity measurement assumes plane wave propagation model. Therefore, with the microphone array used here, only the direction of the arriving sound can be achieved. The direction of the arriving sound wave can be estimated as the spherical mean (SME) of the sound intensity vectors over a frequency band [199]

$$\hat{\mathbf{n}} = \frac{S}{\|S\|}, \tag{5.54}$$

where

$$S = \sum_{\omega_1}^{\omega_2} \overline{\boldsymbol{I}}(\omega) \tag{5.55}$$

with $\overline{\boldsymbol{I}}(\omega) = \boldsymbol{I}(\omega)/\|\boldsymbol{I}(\omega)\|$, which is the amplitude normalized version of the discretized sound intensity vector. The length of each sound intensity vector is first normalized to unity based on the results in Publication III, where the normalized vectors are found to provide more noise robust results than the unnormalized ones.

In Publication III four other possibilities for estimating the direction of arrival from the sound intensity vectors are presented and discussed. Although the methods in Publication III are given in 2-dimensions they can be extended to 3 dimensional data by using spherical probability density functions instead of circular. It is shown in Publication VI that the sound intensity vector based methods do not perform as well in the direction estimation of the reflections as the TDOA based methods. This is due to the limited frequency band, that is a feature of sound intensity vector based direction estimation.

5.5 Examples of the localization maps

Examples of localization maps with different methods are provided in Fig. 5.1. The data is a simulated perfect reflection with no noise at (2,11,1.5) m, and the array is at (0,0,0) m. As can be seen the TDOA based methods

Table 5.1. Some of the least squares localization approaches that have been applied for time of arrival (TOA) and time difference of arrival (TDOA)-based localization.

Wave model	Data	Reference	Solution / Optimization Method
Spherical			
	TDOA	[118]	ULS
		[119]	SI, SX, PX
		[120]	SX
		[121]	TWLS
		[124]	EULS
		[125]	CLS
		[126][1]	ULS
		[127]	WCLS
		[109, 129]	LM
		[130]	ALS
		[131][1]	TWLS
		[132]	PSO
		[133]	SDP
		[134]	CRM
	TOA	[139]	ULS, NCLS
		[140]	ALS
		[109]	LM
		[141][2]	TWLS, MMA
	TOA & TDOA	None	ULS
		[195]	-
		[196]	ALS
		[197][3]	-
Planar			
	TDOA	[198]	ULS
	TOA	None	ULS
	TOA & TDOA	None	ULS

ALS: Approximate least squares through Taylor-Series expansion, CLS: Constrained least squares, CRM: Convex relaxation methods, EULS: Extended unconstrained least squares, LM: Levenberg–Marquardt method, MMA: Min-max algorithm, NCLS: Non-convex constrained least squares, PSO: Particle swarm optimization, PX: Plane intersection, SI: Spherical interpolation, SX: Spherical intersection, SDP: Semi-definite programming, TWLS: Two-step weighted least squares, ULS: Unconstrained least squares, WCLS: Weighted constrained least squares, [1]: Used for joint speed of sound and position estimation, [2]: TOA and unknown time term, [3]: Review article.

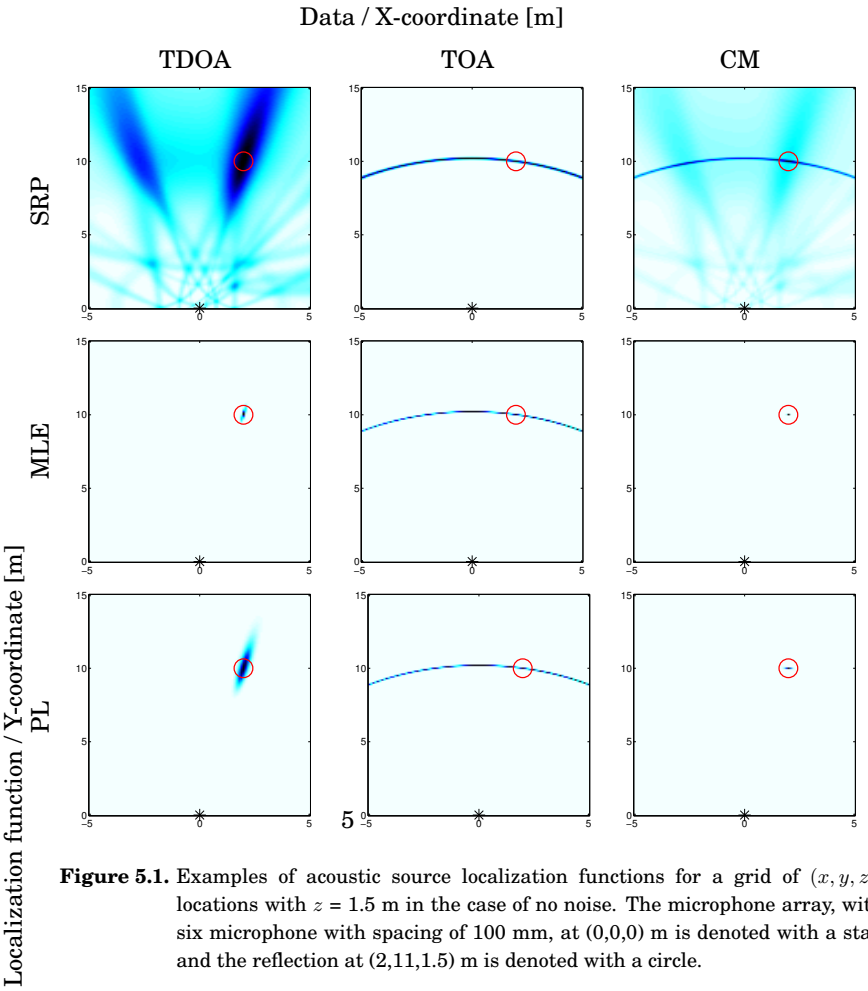


Figure 5.1. Examples of acoustic source localization functions for a grid of (x, y, z)-locations with z = 1.5 m in the case of no noise. The microphone array, with six microphone with spacing of 100 mm, at (0,0,0) m is denoted with a star, and the reflection at (2,11,1.5) m is denoted with a circle.

provide good information about the direction where as the TOA based methods seem to work well in the distance estimation. When the TOA and TDOA are combined a better localization method is made. As seen in Fig. 5.1 SRP methods have more "ghosts" than other methods, i.e., local maxima that do no correspond to the true reflection location. In this example, the simplest search of the maximum is presented. That is, the maximum can be found using a predefined grid of locations. However, this is often not very efficient, therefore some other methods for the search of the maximum are discussed next.

5.6 Search of the extremum

Basically any global optimization method can be used for the search of the extremum. In general, there is no way of ensuring that the global optimization method will converge to the global extremum since localization with spherical wave propagation model is a non-linear problem. Therefore there is usually a need for Monte-Carlo simulations to validate the optimization method for a certain problem. Since the literature on optimization methods is extensive, only some selected methods used for localization are discussed here.

In addition to the closed form solutions and optimization methods listed in Table 5.1 used for TOA and TDOA based methods, other optimization methods have been proposed for other ASL functions. Especially the search of the maximum of the SRP-PHAT function has been of interest [144, 147–150, 200, 201].

The most naive and straightforward method for the search of the maximum is to use a (predefined) grid of location candidates. The drawbacks of this approach is the slowness of the computation when the grid size is large. Namely, 3D grid of a volume of say a concert hall, the number of data points becomes very large, thus the estimation meets the curse of dimensionality. The number of data points naturally depends on the selected grid spacing. However, since the evaluation of the ASL function is the same at each selected time instant for any data point, the process can be parallelized as in [202]. Using parallel computation decreases the time used for the evaluation in total, but requires special implementation considerations and special equipment, such as the general purpose graphic processing unit.

Specially designed sequential Monte-Carlo methods, a.k.a. particle filtering, can be used to track speech and other sources [15, 108, 152, 203–205]. The advantage of particle filtering is that only a small subset of samples is needed to represent the underlying probability distribution. For reflection localization, particle filtering approaches are not useful since the reflections are not moving targets but discrete events in the spatial room impulse response. However, particle swarm optimization [206] has similar features as particle filtering, i.e. it includes a randomization step, and it has been used, for example, with the LS approach [132] and with the MLE [207]. It could also be applied to other ASL functions.

In this thesis, the well-known Nelder-Mead method is used to find the

extremum in the ASL functions [208]. The Nelder-Mead method requires a proper initial guess in the source localization problem for the parameters to be estimated.

5.7 Automatic calibration of the loudspeaker and microphone positions

When the room impulse responses have been measured, the calibration of the loudspeaker and/or microphone positions, and the estimation of speed of sound in the measurement system can be done from the direct sound, which is the first event in the impulse response. In principle, any of the above methods can be used to localize the loudspeakers and/or the microphones. Raykar *et al.* have listed the number of required microphones and loudspeakers in different calibration schemes [109].

5.8 Localization of reflections

After the direct sound, the rest of the events in the room impulse response are reflections. The processing of the spatial impulse response measured with a compact microphone array is done in short time windows [1, 12, 13, 81, 160], and [Publication I]. The analysis window size is selected so that it includes as few reflections as possible but it is still possible to do some processing for the data in the window. Using proper time windowing, the reflections can be temporally and spatially separated. Since the maximum intra-sensor distance in the microphone array is 10 cm, the minimum time window length is about 0.3 ms. Based on previous knowledge [1, 12, 81], and [Publication I], a good window size for the analysis of early reflections is approximately from 1 ms to 4 ms.

Naturally, the number of reflections arriving within one window depends on the echo density defined by Eq. (2.49). Echo density states that the larger the room, the larger the temporal and spatial spacing between the reflections. In addition, the smaller the time interval, the less reflections within a time window.

In this work, it is assumed that there is only one reflection present per analysis window. This is generally true for the first order reflections with the suggested 4 ms analysis window in large spaces, such as auditoriums

and concert halls. Using Eq. (2.49), on average, only one individual reflection should be present in an analysis window, when

$$t \leq \sqrt{\frac{dN_r}{dt}\frac{V}{4\pi c^3}}. \tag{5.56}$$

For example, if $V = 1800$ m^3, $dt = 0.004$ s, and $c = 345$ m/s, then less than two reflections $dN_r = 2$ are present in the analysis window until about $t < 0.042$ s after the direct sound, which corresponds to about 14.4 m in distance. In practice, the number of reflections within a window greatly depends on the location of the source and of the receiver, Eq. (5.56) can be seen as a guideline.

The case where there are more than one reflection present within one analysis window is left for future research. In principle, it is the same problem as the multiple source localization problem, and some of the methods used for that problem, e.g. [209], should also be applicable here.

With the assumption of only one reflection per analysis window, the measurement noise is the only aspect corrupting the localization results. A recognizable feature, also shown in Fig. 2.7, is the fact that the signal-to-noise ratio (SNR) decreases as the time increases. Thus, the reflections that arrive later in time have lower SNR.

5.9 Computational complexity of the localization methods

Although reflection localization within the framework of this thesis is always an offline task, some comparison between the complexity of the methods is provided. The complexity is compared with the 'Big O notation', $\mathcal{O}(\cdot)$.

For basic beamforming the complexity is built up from the number of ASL function evaluations E, the length of the signal L, and the number of the microphones N. For cross correlation the complexity of the estimation function is $\mathcal{O}(L\log\{L\})$ and since all the microphones are used twice in the calculation of the ASL function the complexity increases by $\mathcal{O}(L^2)$. [15]

Moreover, the complexity of the TOA estimation with the simple peak picking method is $\mathcal{O}(L)$. For TOA estimation with AC approach the complexity is $\mathcal{O}(L\log\{L\})$, but that approach is not used here. Since the MLE-S method calculates the ASL function over a frequency band, its complexity is increased by the number of frequencies used $\mathcal{O}(F)$.

Table 5.2 lists the computational complexity of the methods introduced

Table 5.2. Computational complexity of the localization methods in the reflection localization task.

Data	Method	Complexity
TOA	MLE-S	$\mathcal{O}(EL\log\{L\}NF)$
TOA	SRP, PL, & MLE	$\mathcal{O}(NL+NE)$
TDOA	SRP, PL, & MLE	$\mathcal{O}(N^2L\log\{L\}+EN^2)$
TOA & TDOA	SRP, PL, & MLE	$\mathcal{O}(NL+NE+N^2L\log\{L\}+EN^2)$

E: Number of ASL function evaluations, L: The length of the signal, N: The number of the microphones, and F: The number of the frequency bins

in this chapter. The TOA-based methods have lower computational complexity than the other methods since the room impulse responses are directly mapped into the TOAs.

As the number of evaluations increases, the computational complexity and time of MLE-S increases. This results was also pointed out by Korhonen for the time domain beamformer [15]. However, when the number of evaluations increases, the computational complexity of the time domain beamformer (SRP-TOA) does not increase as rapidly as the computational complexity of the conventional time-domain beamformer. This is due to direct mapping of impulse response to TOAs, which does not require additional calculations.

5.10 Interpolation Methods

Due to the limited sampling frequency, interpolation is required in practical situations in the TDOA and TOA based localization. Namely, the sampling frequency sets an upper limit for the spatial resolution that can be achieved. Here, three possibilities for interpolation are presented. The first one interpolates the received signal, the second one interpolates TDOA or TOA estimates by making assumption on the shape of the estimation function. These estimates can be directly used in TOA and TDOA based MLE methods. The third approach extends the function fitting for TOA and TDOA estimation function. These interpolated estimation functions can then be used in SRP and PL methods.

5.10.1 Signal

The most straightforward way of interpolation is to upsample the signals by Fourier-interpolation. Upsampling the signals with Fourier-interpolation consists of two parts. First zeros are added, and then the signal is low-pass filtered [210].

5.10.2 Time difference of arrival estimate

In traditional TDOA estimation, the interpolation is done usually by fitting a parabola [101] or an exponential function [102] to the maximum peak of the TDOA function. TDOA and its interpolation leads to a single time delay estimate. These interpolated values can then be used in the MLE methods. Similarly the TOA estimates can be interpolated by assuming some shape for the energy or the pressure of the room impulse response. This would require a priori knowledge of the impulse shape, as does the interpolation of the TDOA estimate.

5.10.3 Time difference of arrival estimation function

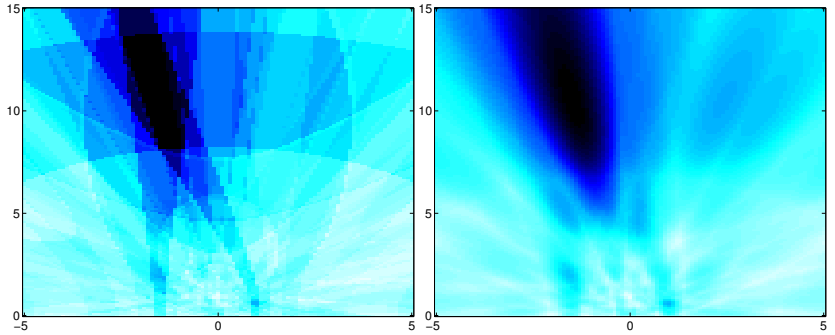
In the SRP and PL methods, the spatial response is built on the TDOA and TOA estimation functions. The above TDOA interpolation methods can not be used directly for interpolating the TDOA estimation functions for the SRP or PL methods. Therefore, an algorithm for using the function fitting approaches in the steered response function is developed in Publication V. Although this approach is designed for TDOA estimation functions, it can be directly applied also for TOA estimation functions. Here, for clarity it is formulated for TDOA estimation functions.

The algorithm makes an assumption on the TDOA estimation function shape near the maxima. Throughout this thesis, the exponential shape is used:

$$f_l(\tau) = a_1 e^{-b_l(\tau - c_l)^2}, \tag{5.57}$$

where $a_1, b_l,$ and c_l are the coefficients and f_l is the function for l^{th} local maximum. Other possibility is the parabolic shape, but it is shown to perform worse than the exponential shape in the interpolation task of the cross correlation function in [102] and in Publication V.

The interpolation of a TDOA estimation function is described by the following steps. Firstly, the TDOA estimation function is normalized so



(a) SRP-PHAT at 48 kHz, no interpolation **(b)** SRP-PHAT at 48 kHz, exponential interpolation

Figure 5.2. Example of the interpolation of SRP-PHAT function with exponential fitting applied to the cross correlation vectors. The microphone array is at (0,0,0) m.

that it is positive. Secondly, the local maxima are searched from the TDOA function in the region of interest. Thirdly, the coefficients in Eq. (5.57) are solved using the local maximum and two neighboring points on both sides of the maximum. This leads to a function $f_l(\tau)$ for each local maximum l. Finally, as a result, the interpolated TDOA function can be evaluated at any time delay τ:

$$\mathrm{R}_{\mathrm{interpolated}}(\tau) = \max_{l} f_l(\tau). \tag{5.58}$$

If the number of the local maxima is reduced similarly as in [145], the method will be more efficient in terms of computational time. In addition, an advantage of the proposed algorithm over e.g. the Fourier-interpolation is that the TDOA function is presented with a limited number of coefficients, when in the Fourier-interpolation the number of samples increases with the sampling frequency. The interpolation method is suitable for other TDOA estimation functions than cross correlation function as well and the shape assumption is not limited to the ones presented here. Figure 5.2 shows an example of the interpolation with exponential assumption for SRP-PHAT function.

6. Theoretical performance

This chapter presents a theoretical performance limits in the acoustic reflection localization framework. Different localization approaches are compared in the theoretical framework.

6.1 Overview

The positions of the sensors and the source as well as the signal and the noise have an effect on the localization variance. These effects can be theoretically measured using Cramér-Rao lower bound (CRLB) [17] analysis.

The theoretical boundaries given in this section use the assumption that the source signal and noise signals are white Gaussian noise. This assumption is necessary and required to make the signal model in Eq. (5.2) mathematically tractable [25, 26].

6.2 Time difference of arrival estimation

The theoretical performance bounds for TDOA estimation have been a topic of various research studies [26, 99, 100, 211–213]. In addition to CRLB, other performance bounds have been presented. For example, the Ziv-Zakai lower bound is of interest in the presence of large errors [26, 213]. Here only CRLB is considered.

The Fisher information for TDOA estimation is given as [26, 213]

$$J(\tau) = \frac{2T}{2\pi} \int_0^\infty (\omega)^2 \frac{C_{h_1,h_2}(\omega)}{1 - C_{h_1,h_2}(\omega)} d\omega \tag{6.1}$$

where T is the window length and the magnitude squared coherence is related to the SNR via [26]

$$\frac{C_{h_1,h_2}(\omega)}{1 - C_{h_1,h_2}(\omega)} = \frac{\mathrm{SNR}^2(\omega)}{1 + 2\mathrm{SNR}(\omega)}. \tag{6.2}$$

Note that the Fisher information above is independent of τ. Setting the power spectral densities flat as

$$G_{s,s}(\omega) = \begin{cases} G_{s,s} & ,|\omega| \in [\omega_c - B/2, \omega_c + B/2] \\ 0 & ,\text{otherwise} \end{cases} \tag{6.3}$$

with center frequency ω_c. Assuming also that the noises are equal $G_{n_1,n_1}(\omega) = G_{n_2,n_2}(\omega) = G_{n,n}(\omega)$, the Fisher information formulates into [213]

$$J(\tau) = \frac{\mathrm{SNR}^2}{1 + 2\mathrm{SNR}} \frac{T}{\pi}(B\omega_c^2 + B^3/12). \tag{6.4}$$

This analysis is valid only for $T \gg 2\pi/B$ and for sufficiently large SNR values, in detail [100, 213]

$$\mathrm{SNR} > \frac{12\omega_c^2}{\pi T B^3}\left[\Phi^{-1}\Big(\frac{1}{24}\frac{B^2}{\omega_c^2}\Big)\right]^2 \tag{6.5}$$

where $\Phi^{-1}(x)$ is the inverse of the exponential integral

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\mu^2/2} d\mu. \tag{6.6}$$

6.3 Time of arrival estimation

The Fisher information for TOA estimation is given by through the derivation of the MLE function in Eqs. (5.20) and (5.23), and it is equal to

$$J(t) = \frac{2T}{2\pi} \int_0^\infty \omega^2 \frac{C_{s,h_1}(\omega)}{1 - C_{s,h_1}(\omega)} d\omega \tag{6.7}$$

where

$$\frac{C_{s,h_1}(\omega)}{1 - C_{s,h_1}(\omega)} = \frac{G_{s,s}(\omega)}{G_{w,w}(\omega)} = \mathrm{SNR}(\omega). \tag{6.8}$$

With the same assumptions on the spectral densities as with in TDOA estimation, the Fisher information becomes

$$J(t) = \mathrm{SNR}\frac{T}{\pi}(B\omega_c^2 + B^3/12). \tag{6.9}$$

Since SNR > 0, it can be seen that the CRLB is always smaller for TOA estimation since Fisher information in TOA estimation is higher.

6.4 Localization

The log-likelihood of the localization with respect to signal model is given by Eq. (5.39). The Fisher information matrix is formulated as [113, 114,

214]

$$\mathbf{J}(\mathbf{x}) = 2\Re[\mathbb{H}(\mathbf{D}(\omega,\mathbf{x}))^H Q^{-1}\mathbb{H}(\mathbf{D}(\omega,\mathbf{x}))], \tag{6.10}$$

where

$$\begin{aligned}\mathbb{H}(\omega,\mathbf{D}(\mathbf{x})) &= \left[\frac{\partial S(\omega)D_1(\omega,\mathbf{x})}{\partial\mathbf{x}},...,\frac{\partial S(\omega)D_N(\omega,\mathbf{x})}{\partial\mathbf{x}}\right] \\ &= S(\omega)\left[\frac{\partial e^{-j\omega t_1(\mathbf{x})}}{\partial\mathbf{x}},...,\frac{\partial e^{-j\omega t_N(\mathbf{x})}}{\partial\mathbf{x}}\right].\end{aligned}$$

For a single microphone and frequency the differential with respect to location x is given by

$$\frac{\partial S(\omega)e^{-j\omega t_n(\mathbf{x})}}{\partial\mathbf{x}} = -S(\omega)j\omega\frac{\partial t_n(\mathbf{x})}{\partial\mathbf{x}}e^{-j\omega t_n(\mathbf{x})}, \tag{6.11}$$

where

$$\frac{\partial}{\partial\mathbf{x}}t_n(\mathbf{x}) = c^{-1}\left(\frac{\mathbf{x}-\mathbf{r}_n}{\|\mathbf{x}-\mathbf{r}_n\|}\right). \tag{6.12}$$

When assuming independent errors and equal error variances, the Fisher information matrix can be expressed as

$$\mathbf{J}(\mathbf{x}) = \left(\frac{2}{2\pi}\int_{\omega=0}^{\infty}(\omega\|S(\omega)\|)^2\,df\right)[H_{\text{TOA}}^{\text{T}}(\boldsymbol{t}(\mathbf{x}))\mathbf{Q}^{-1}H_{\text{TOA}}(\boldsymbol{t}(\mathbf{x}))] \tag{6.13}$$

where a design matrix is given for TOAs as

$$H_{\text{TOA}}(\boldsymbol{t}(\mathbf{x})) = \begin{bmatrix}\frac{\partial}{\partial\mathbf{x}}t_1(\mathbf{x})\\ \frac{\partial}{\partial\mathbf{x}}t_2(\mathbf{x})\\ \vdots\\ \frac{\partial}{\partial\mathbf{x}}t_N(\mathbf{x})\end{bmatrix}. \tag{6.14}$$

Moreover, when constant spectral densities for noise and signal are assumed on a certain frequency band B, and within some time window of length T, the Fisher information formulates into

$$\begin{aligned}\mathbf{J}(\mathbf{x}) &= \left(\frac{2T}{2\pi}\int_{\omega=0}^{\infty}\omega^2\frac{\|S(\omega)\|^2}{\sigma_F^2}df\right)[H_{\text{TOA}}^{\text{T}}(\boldsymbol{t}(\mathbf{x}))H_{\text{TOA}}(\boldsymbol{t}(\mathbf{x}))] &(6.15)\\ &= \left(\frac{2T}{2\pi}\int_{\omega=0}^{\infty}\omega^2\text{SNR}(\omega)d\omega\right)[H_{\text{TOA}}^{\text{T}}(\boldsymbol{t}(\mathbf{x}))H_{\text{TOA}}(\boldsymbol{t}(\mathbf{x}))] &(6.16)\\ &= \text{SNR}\frac{T}{\pi}(B\omega_c^2+B^3/12)[H_{\text{TOA}}^{\text{T}}(\boldsymbol{t}(\mathbf{x}))H_{\text{TOA}}(\boldsymbol{t}(\mathbf{x}))]. &(6.17)\end{aligned}$$

6.5 Time difference of arrival based localization

The probability density function for TDOAs is given in Eq. (5.28). The Fisher information matrix for TDOA is given by [121, 153]:

$$\mathbf{J}(\mathbf{x}) = E\left[\left(\frac{\partial}{\partial\mathbf{x}}\log p(\boldsymbol{\tau};\mathbf{x})\right)\left(\frac{\partial}{\partial\mathbf{x}}\log p(\boldsymbol{\tau};\mathbf{x})\right)^{\text{T}}\right]_{\mathbf{x}=\mathbf{x}_0}, \tag{6.18}$$

where $\mathbf{x}_0$ is the true source position.

The partial derivation with respect to the source position $\mathbf{x}$ is

$$\frac{\partial}{\partial \mathbf{x}} \log p(\boldsymbol{\tau}; \mathbf{x}) = -\left(\frac{\partial}{\partial \mathbf{x}} \boldsymbol{\tau}(\mathbf{x}) \right)^{\mathrm{T}} \Sigma_{\mathrm{TDOA}}^{-1} (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}(\mathbf{x})), \tag{6.19}$$

where for a single TDOA the partial derivate of Eq. (2.39) is:

$$\frac{\partial}{\partial \mathbf{x}} \tau_m(\mathbf{x}) = c^{-1} \left(\frac{\mathbf{x} - \mathbf{r}_i}{\|\mathbf{x} - \mathbf{r}_i\|} - \frac{\mathbf{x} - \mathbf{r}_j}{\|\mathbf{x} - \mathbf{r}_j\|} \right) \tag{6.20}$$

Following [153], the partial derivates can be re-formulated in to a matrix

$$H_{\mathrm{TDOA}}(\boldsymbol{\tau}(\mathbf{x})) = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} \tau_1(\mathbf{x}) \\ \frac{\partial}{\partial \mathbf{x}} \tau_2(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial \mathbf{x}} \tau_M(\mathbf{x}) \end{bmatrix}. \tag{6.21}$$

The Fisher information matrix is then given by [153]:

$$I(\mathbf{x}) = H_{\mathrm{TDOA}}^{\mathrm{T}} \Sigma_{\mathrm{TDOA}}^{-1} H_{\mathrm{TDOA}} \tag{6.22}$$

and the Cramer-Rao lower bound is calculated using Eq. (2.17). The minimum variance that TDOA estimation can achieve is given by Eq. (6.4). The covariance matrix can be replaced by this information which yields

$$\mathbf{J}(\mathbf{x}) = \frac{1}{\sigma_{\mathrm{TDOA}}^2} H_{\mathrm{TDOA}}^{\mathrm{T}} H_{\mathrm{TDOA}} = H_{\mathrm{TDOA}}^{\mathrm{T}} H_{\mathrm{TDOA}} J(\tau) \tag{6.23}$$

since $\min \sigma_{\mathrm{TDOA}}^2 = 1/J(\tau)$, and $\mathbf{J}(\boldsymbol{\tau}) = \mathbf{I} \times J(\tau)$ due to the independence assumption.

6.6 Time of arrival based localization

The probability density function of the error is given in Eq. (5.33). The calculation of CRLB for TOA proceeds as previously for TDOAs. The difference is that the partial derivation in Eq. (6.20) for TOAs has the form given in Eq. (6.12). The partial derivates are re-formulated into a matrix, which has the form given in Eq. (6.14).

The Fisher information matrix is then given as in Eq. (6.22) by replacing H_{TDOA} with H_{TOA}, and the Cramer-Rao lower bound is calculated using Eq. (2.17).

When the partial derivates of TOAs are substituted to the Fisher information matrix in Eq. (6.22), and the minimum variance of the TOA

estimation given in Eq. (6.9) is used as the variance for the covariance matrix one has

$$\mathbf{J}(\mathbf{x}) = H_{\text{TOA}}^{\text{T}} \Sigma_{\text{TOA}}^{-1} H_{\text{TOA}} = \frac{1}{\sigma_{\text{TOA}}^2} H_{\text{TOA}}^{\text{T}} H_{\text{TOA}} = J(t) H_{\text{TOA}}^{\text{T}} H_{\text{TOA}}, \tag{6.24}$$

which is exactly the same as Eq. (6.17). That is, in theory, localization using time of arrival estimation, SRP-TOA or MLE-S function have the same performance.

6.7 Combination of time difference and time of arrival information based localization

When the errors are independent the covariance matrix for the combination of TOA and TDOA estimates is given as

$$\Sigma_{\text{CM}} = \text{diag}(\sigma_{\text{TOA}}^2, \ldots, \sigma_{\text{TOA}}^2, \sigma_{\text{TDOA}}^2, \ldots, \sigma_{\text{TDOA}}^2),$$

where the first values are TOA variances and the rest are TDOA variances.

For notational convenience, it is of use to define a measurement vector including both TOA and TDOA measurements $\hat{\boldsymbol{\chi}} = [\hat{\chi}_1, \hat{\chi}_2, ... \hat{\chi}_{N+M}] = [\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_N, \hat{\tau}_1, \hat{\tau}_2, \ldots, \hat{\tau}_M]$. That is, with the 6 microphones used in this thesis, the 6 first values of the vector are TOAs and the rest TDOAs. Then the combined design matrix is given as

$$H_{\text{CM}}(\boldsymbol{\chi}(\mathbf{x})) = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} \chi_1(\mathbf{x}) \\ \frac{\partial}{\partial \mathbf{x}} \chi_2(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial \mathbf{x}} \chi_{N+M}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} t_1(\mathbf{x}) \\ \frac{\partial}{\partial \mathbf{x}} t_2(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial \mathbf{x}} t_N(\mathbf{x}) \\ \frac{\partial}{\partial \mathbf{x}} \tau_1(\mathbf{x}) \\ \frac{\partial}{\partial \mathbf{x}} \tau_2(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial \mathbf{x}} \tau_M(\mathbf{x}) \end{bmatrix} \tag{6.25}$$

The Fisher information matrix is then given by

$$\mathbf{J}(\mathbf{x}) = H_{\text{CM}}^{\text{T}} \Sigma_{\text{CM}}^{-1} H_{\text{CM}} = H_{\text{CM}}^{\text{T}} \text{diag}(J(t), \ldots, J(t), J(\tau), \ldots, J(\tau)) H_{\text{CM}}. \tag{6.26}$$

6.8 Theoretical comparison

In this theoretical comparison the frequency and the temporal parameters are fixed to $\omega_c/(2\pi) = 12$ kHz, $B/(2\pi) = 24$ kHz, $T = .004$ s. This corresponds to a situation where full bandwidth at 48 kHz sampling frequency and 4 ms time window is used in the analysis. The idea is to compare the localization methods in the same conditions.

Figure 6.1 presents the CLRB for TOA and TDOA against SNR. In addition, CRLB for TDOA that is calculated as the difference of two TOA estimates is presented. TOA estimation has smaller CLRB than TDOA estimation, which is not surprising, since in TOA estimation it is assumed that both source and noise signals are known. The CRLB of the traditional TDOA estimation approaches the CRLB of the TDOA estimation which is calculated as the difference of two TOAs, as expected from their equations.

In Fig. 6.2 CRLB for TOA, TDOA, and CM are shown with parameters at location (10.5, 8.2, 2) m. The microphone array is the one given in Table 2.1 with $d_{\text{spc}} = 100$ mm. As mentioned, the CRLB for signal model is the same as CRLB for TOA. Clearly, CM has the smallest CRLB and TOA the second smallest. Interestingly around -25 dB, TOA and CM have the same performance. This is caused by the increment in the variance of TDOA, shown in Fig. 6.1.

Figure 6.3 shows an example of the CRLB for x, y, and z components with TOA, TDOA, and CM data with SNR$= 30$ dB. It can be seen that CM has the smallest CRLB in all conditions, and TOA the second smallest. Thus it is expected that CM and TOA will perform well in the reflection localization with the given setup.

In the next chapter, the theoretical performance bounds are compared with Monte-Carlo simulation results.

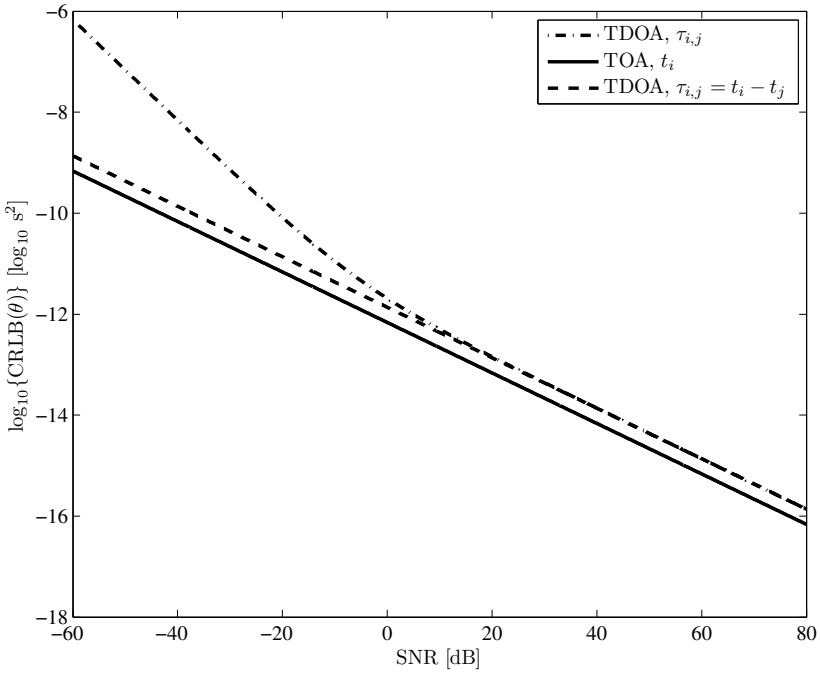


Figure 6.1. Cramer-Rao lower bound versus signal-to-noise ratio (SNR) for TDOA and TOA.

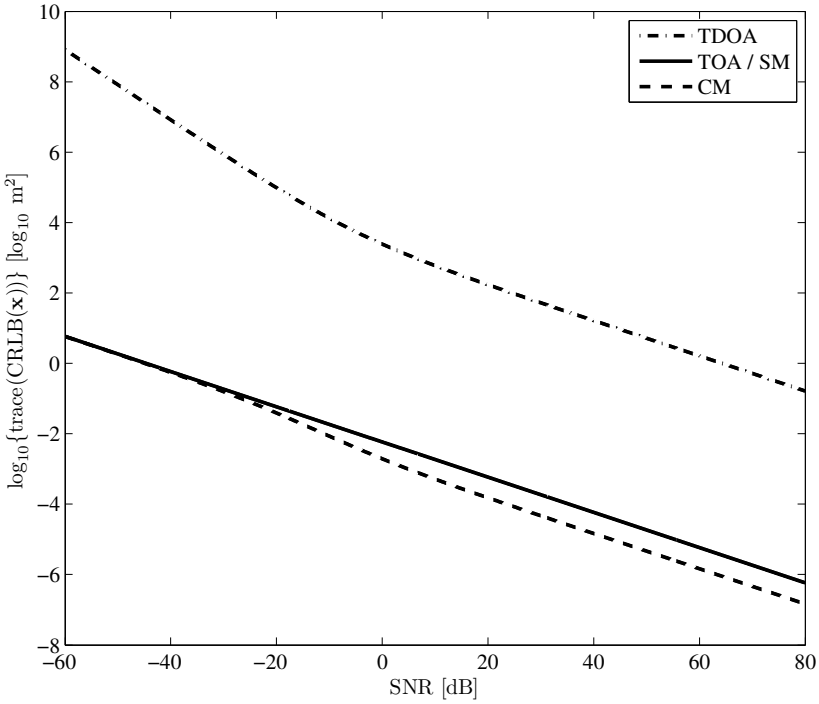


Figure 6.2. Cramer-Rao lower bound versus signal-to-noise ratio (SNR) for localization using TDOA, TOA, and CM at (10.5, 8.2, 2) m.

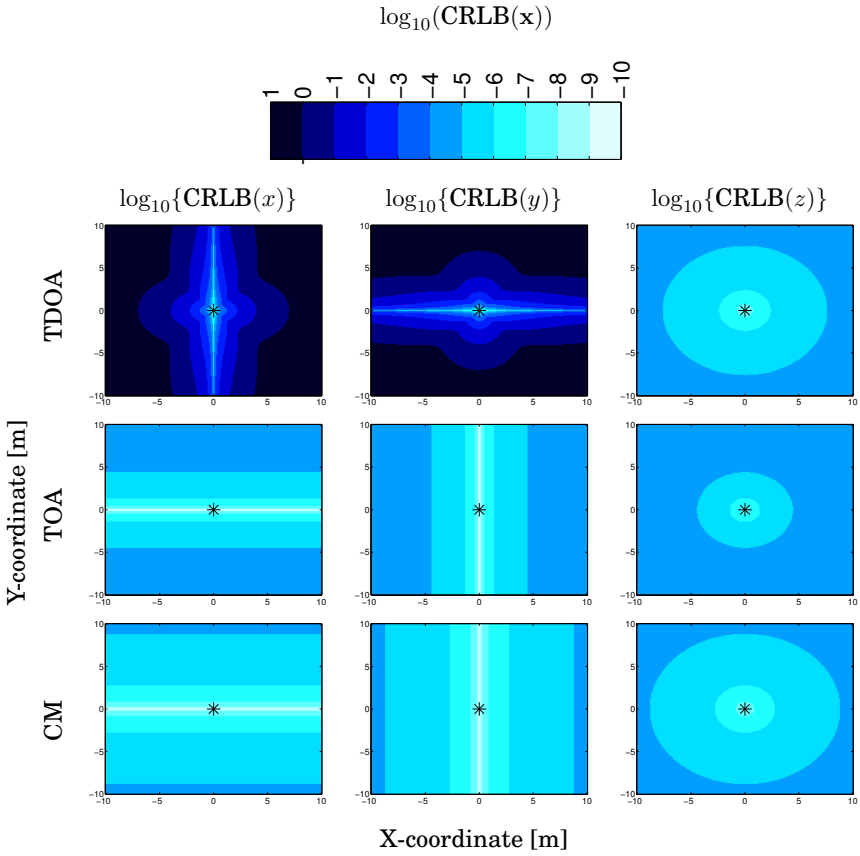


Figure 6.3. Cramér-Rao lower bound presented for three diagonal components of Eq. (2.17) for a grid of $\{x, y, z\}$-locations (with $z = 0$). The signal-to-noise ratio is 30 dB. The array is depicted with a star. The maps are presented in 10-base logarithmic scale to enhance the differences between them. The color scale is the same for all maps.

7. Experiments

This chapter presents simulated and real data experiments. The performance of TOA, TDOA, and localization methods is under investigation. The CRLB for each estimation task is also presented.

7.1 Monte-Carlo simulations

The reflection signal model in the following Monte-Carlo simulations is of exponential form

$$s_n(t|t_n(\mathbf{x}), \sigma^2) = e^{-(t-t_n(\mathbf{x}))^2/\sigma^2}. \tag{7.1}$$

Throughout the simulations the 'variance' parameter of the reflection signal is set to $\sigma = 2/f_s$, where $f_s = 10,000$ Hz is the sampling frequency. The TOA $t_n(\mathbf{x})$ is calculated assuming the spherical wave propagation model.

Since the assumed reflection signal is exponential, the exponential fitting for the TDOA and TOA estimates and for TDOA and TOA estimation functions presented in [102], and in Publication V, respectively, are applied. As an example, in the case of no noise the direct cross correlation of two exponential functions is an exponential function. This result is well known for the example with normal distributions.

7.1.1 Time difference of arrival estimation

Time difference of arrival estimation methods, introduced in Section 5.2 are compared against signal-to-noise-ratio. The length of the time window is set to 4 ms in this experiment and the reflection signal in Eq. (7.1) is used. The TDOAs are randomized from a uniform distribution between -1 and 1 ms, i.e. $\mathcal{U}(-1, 1)$ ms.

The results of 10,000 Monte-Carlo samples are presented in Fig. 7.1. As expected, the MLE is the most robust against noise having the small-

est number of anomalous estimates. ASDF has the smallest number of anomalous estimates when SNR $<$ 20 dB, but this is due to its limitations in the TDOA estimation. That is, the maximum TDOA error with ASDF is half of that of the other methods.

The most accurate method is MLE when SNR $<$ 60 dB. When 60 dB $<$ SNR $<$ 80 dB, CC and ASDF, are the most accurate and when SNR $>$ 80 dB, ASDF is the most accurate.

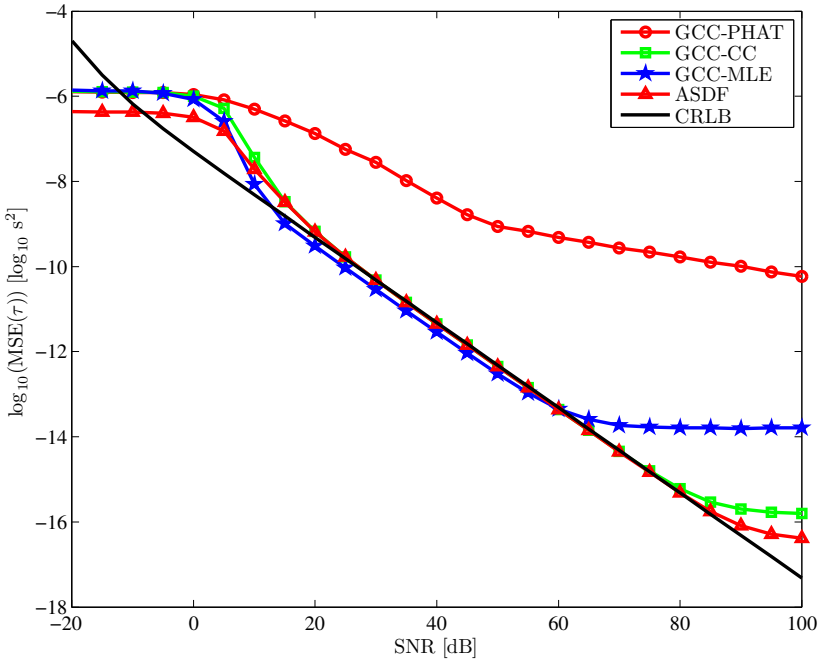
As shown in Fig. 7.1, ASDF and GCC-CC achieve CRLB when 25 dB $<$ SNR $<$ 75 dB. Moreover, GCC-MLE is lower than the CLRB when 15 dB $<$ SNR $<$ 55 dB. This result indicates that the GCC-MLE TDOA estimation is biased. The bias is a result of the exponential fitting. With very high SNR values the CRLB does not predict the MSE of the methods. This behaviour was also noticed in [95]. The reason for this behaviour is the truncated window size [95]. The two different windows include two different peaks that have different samples [95]. True zero delay value can therefore only be achieved with autocorrelation and zero noise level.

Direct cross correlation (CC) is the most reasonable selection for TDOA estimation for reflection localization since it does not require a priori information about the noise as MLE does. Moreover, CC performs well when compared to the other methods, and the calculation is straightforward and computationally light.

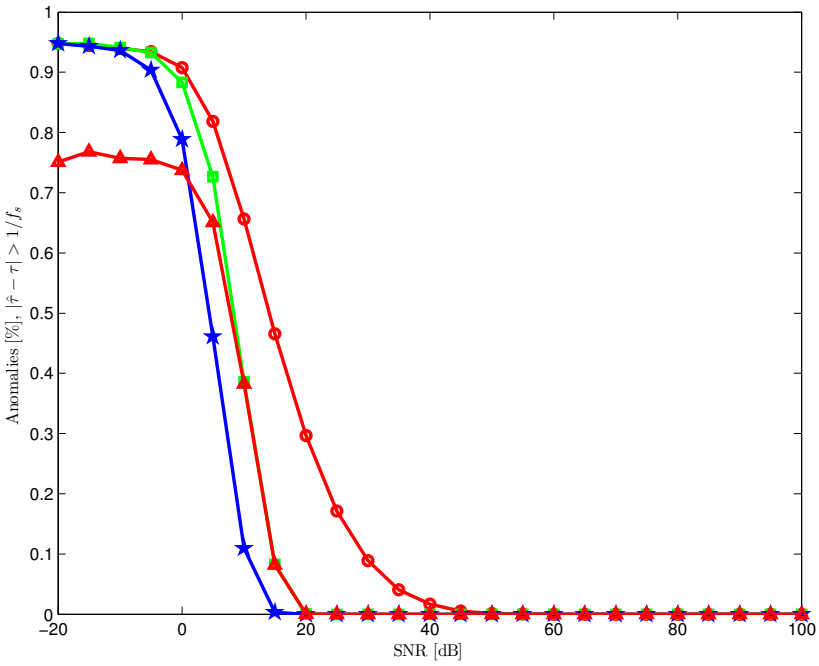
7.1.2 Time of arrival estimation

Time of arrival estimation methods, introduced in Section 5.3 are tested against signal-to-noise-ratio. The length of the time window is set to 4 ms. The TOAs are randomized from a uniform distribution between -1 and 1 ms, i.e. $\mathcal{U}(-1, 1)$ ms.

The results of 10,000 Monte-Carlo samples are presented in Fig. 7.2. The simple peak picking method is noted with $\arg\max\{h(t)\}$ in the results of Fig. 7.2. ASDF and CC are the most accurate methods for the TOA estimation. MLE is the most robust against noise, but loses accuracy, due to the fact that the exponential fit does not describe the MLE function shape. The peak picking method, that does not require any a priori knowledge about the source signal or the noise signal, performs in general better than PHAT and has smaller variance than MLE when SNR $>$ 20 dB. As in TDOA estimation, also here the maximum TOA errors for ASDF are half of the maximum error of the other methods.



(a) Variance



(b) Anomaly %

Figure 7.1. Results for TDOA estimation against signal to noise ratio.

As shown in Fig. 7.2, ASDF and AC-CC achieve CRLB when 15 dB $<$ SNR $<$ 75 dB. When SNR $<$ 15 dB, the estimation is saturated as the large number of anomalies suggests. As with the TDOA estimation, also here the MSEs of the methods does not achieve CRLB with very high SNR values. The explanation for this behaviour is the same as earlier for TDOA estimation.

The TOA estimation with GCC-MLE is not realistic, since it would require the knowledge of both source and noise signals. Here, the focus is on the blind methods that do not require a priori information. Since the peak picking method is the only blind method and has a performance that is comparable to the other methods, it is the most reasonable choice in the general case for the estimation of TOAs.

7.1.3 Localization

Nine different localization methods are tested. In detail, SRP, MLE, and PL with TOA, TDOA, and CM data are used for localization of reflections. The formulation for the methods is given in Section 5. Direct cross correlation and direct peak picking methods with exponential fitting provided in Sections 3 and 4 are used for TDOA and TOA estimation, respectively. Since MLE-S will lead to the same localization result as SRP-TDOA, as shown in [151], it is not tested here.

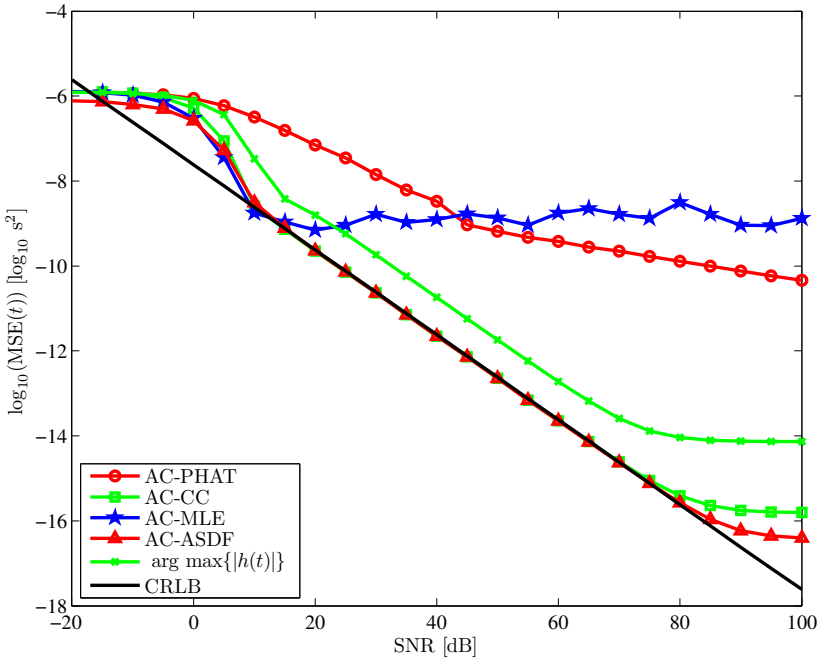
The reflection location is drawn 1,000 times from a 3-D uniform distribution between -20 and 20 m, i.e. $x \sim \mathcal{U}(-20, 20)$ m, $y \sim \mathcal{U}(-20, 20)$ m and $y \sim \mathcal{U}(-20, 20)$ m. The microphone array is set to (0,0,0) and the reflection signal is windowed with 4 ms time window around the TOA between the reflection location and (0,0,0).

The reflection signal model is the one presented in Eq. (7.1). The location is searched from the localization function using the Nelder-Mead simplex method implemented in MATLAB's `fminsearch`. The initial location value for the optimization method is set to the vicinity of the true location.

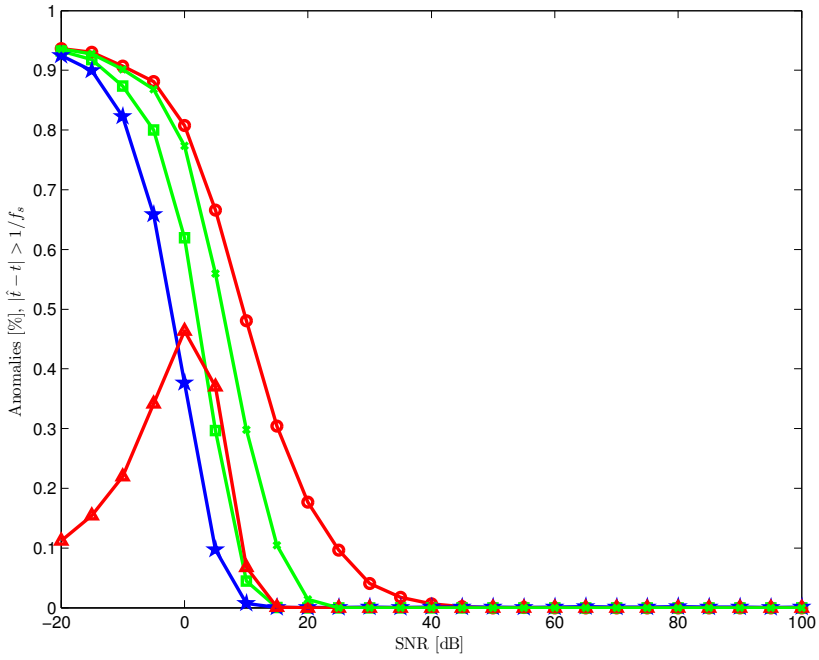
Optimization of the parameters

The weighting parameters for the combined methods are optimized. The question is, which weight produced the best result for each method? For MLE the weighting factor κ is defined as the relation between the TOA and TDOA variance, as

$$\kappa = \frac{\sigma_{\text{TDOA}}}{\sigma_{\text{TOA}}} \tag{7.2}$$



(a) Variance



(b) Anomaly %

Figure 7.2. Results for TOA estimation against signal to noise ratio.

This selection sets the following limitations as shown in Publication II:

$$\lim_{\kappa\to\infty} P_{\text{MLE-CM}}(\mathbf{x}) = P_{\text{MLE-TOA}}(\mathbf{x}), \tag{7.3}$$

$$\lim_{\kappa\to 0} P_{\text{MLE-CM}}(\mathbf{x}) = P_{\text{MLE-TDOA}}(\mathbf{x}). \tag{7.4}$$

For PL-CM and SRP-CM the weighting is limited to $0 < W < 1$. This gives the following obvious limits for SRP-CM function

$$\lim_{W\to 0} P_{\text{SRP-CM}}(\mathbf{x}) = P_{\text{SRP-TOA}}(\mathbf{x}) \tag{7.5}$$

$$\lim_{W\to 1} P_{\text{SRP-CM}}(\mathbf{x}) = P_{\text{SRP-TDOA}}(\mathbf{x}). \tag{7.6}$$

and for PL-CM

$$\lim_{W\to 0} \lambda_{\text{PL-CM}}(\mathbf{x}) = \log\{P_{\text{PL-TOA}}(\mathbf{x})/N\} \tag{7.7}$$

$$\lim_{W\to 1} \lambda_{\text{PL-CM}}(\mathbf{x}) = \log\{P_{\text{PL-TDOA}}(\mathbf{x})/M\}. \tag{7.8}$$

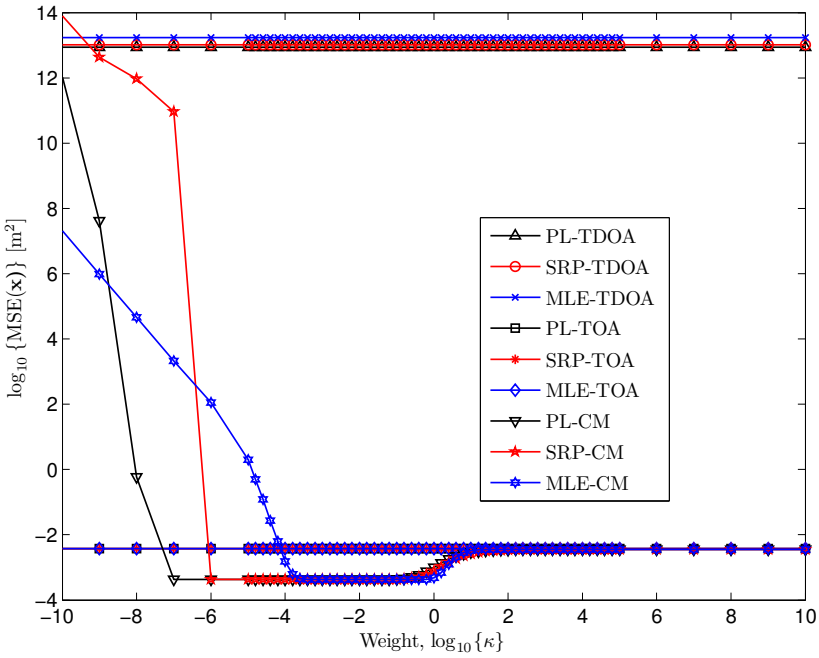
The weight factor κ is changed from $\log_{10}\{\kappa\} = -10, \ldots, 10$. For MLE-CM the variance of the TOA error is set to $\sigma^2_{\text{TOA}} = 1$, and the variance of the TDOA error is altered as $\sigma_{\text{TDOA}} = \kappa\sigma_{\text{TOA}}$. The weight for SRP-CM and PL-CM is $0 < W < 1$, and it is calculated as $W = 1/(10^{\kappa} + 1)$.

The results of this experiment are shown in Fig. 7.3. Also shown are the performance of TOA and TDOA based methods. All the combined methods achieve the same performance with some weighting.

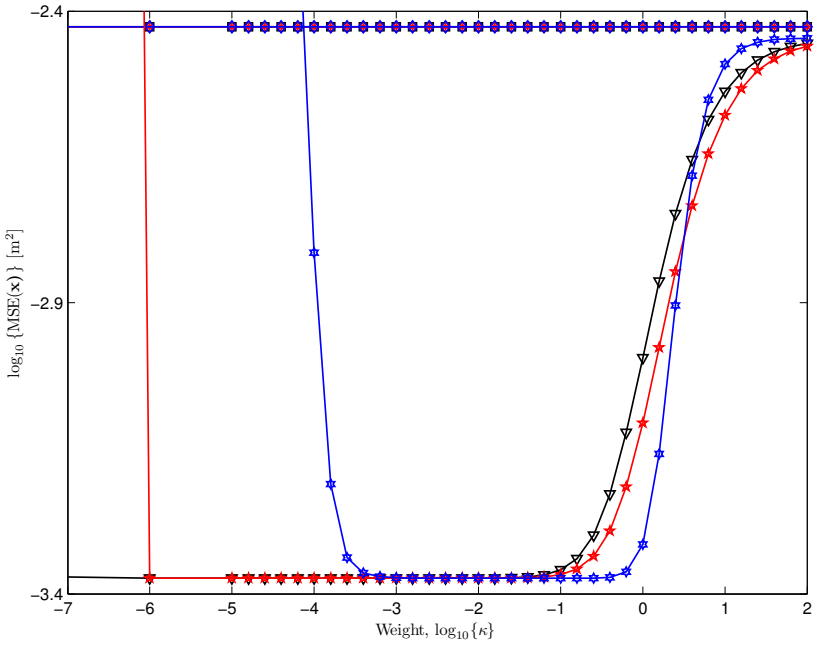
As shown in Fig. 7.3, the optimal weight for SRP-CM is $W \in \left(\frac{1}{10^{-6}+1}, \frac{1}{10^{-1.2}+1}\right)$, for PL-CM $W \in \left(\frac{1}{10^{-7}+1}, \frac{1}{10^{-1.4}+1}\right)$, and for MLE-CM $\log_{10}\{\kappa\} \in (-3.4, -0.2)$. A reasonable choice for MLE-CM weighting factor is $\log_{10}(\kappa) = -2$ since it is close to the middle region of the optimal values. For MLE-CM this means that TOA variance is about 100 times the TDOA variance. For SRP-CM, and PL-CM the value $W = 1/(10^{-2} + 1) = 0.99$ for the weight is a good choice because this is in middle region of the optimal values. This means that SRP-TOA has a weight $W = .01$. and the SRP-TDOA has the weight $W = 0.99$. The same applies for PL-TOA and PL-TDOA in the PL-CM method. These optimized values are used in the experiments in the following experiments.

Simulation results for localization methods

As can be seen from Fig. 7.4 the CM-based methods have the smallest RMSE, MLE-CM having the smallest and SRP-CM the highest, out of the combined data methods. At 15 dB MLE-CM has smaller RMSE than MLE-CM. This is due to the fact that at 15 dB, the probability of anomalous estimate grows quite large for the TDOA estimation, which is weighted heavily in SRP-CM and PL-CM.



(a) Mean squared error



(b) Details of (a)

Figure 7.3. Optimization results against weighting parameter κ with signal-to-noise ratio of 60 dB and with 1,000 Monte-Carlo Samples for each SNR condition.

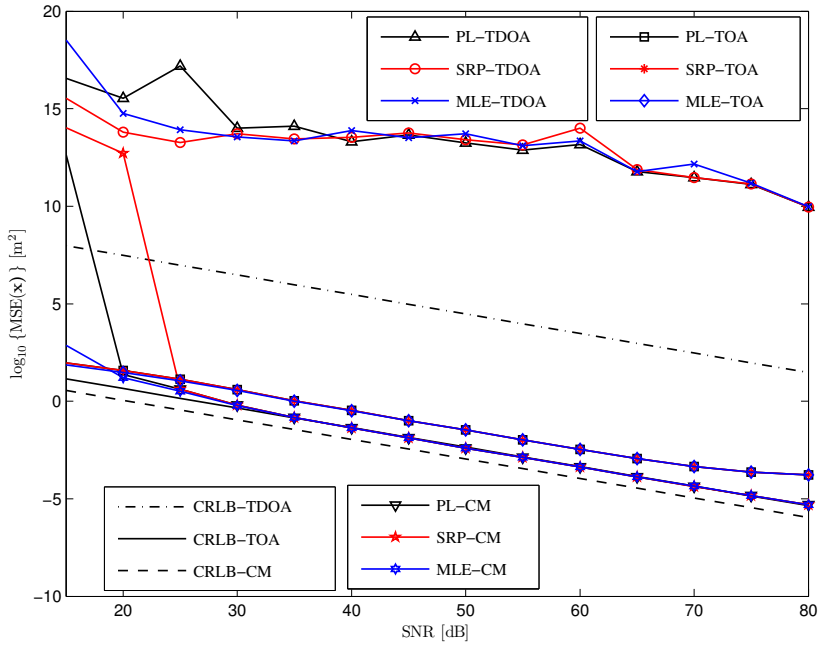
TOA based methods have clearly smaller RMSE than TDOA based methods. Again, MLE-TOA has the smallest RMSE and SRP-TOA the highest out of the TOA-based methods. The results thus indicate that combining TOA and TDOA data is advantageous in the localization of reflections in the current framework. Moreover, methods based only on TDOA information do not perform well in the reflection localization task with the given setup.

As shown in Fig. 7.4, the methods achieve CRLB for TOA but not the CRLB for CM. This is due to the selection of the TOA and TDOA estimation methods. Since no a priori information of the source signal or the noise signal is used in the localization, the CRLB-CM cannot be achieved. As with TOA and TDOA estimation, the CRLB is best achieved when 15 dB $<$ SNR $<$ 75 dB. It is evident from the results that combining the TOA and TDOA estimation benefits the localization, since without any a priori information, the same performance can be achieved as when the source, and noise signal would be known.

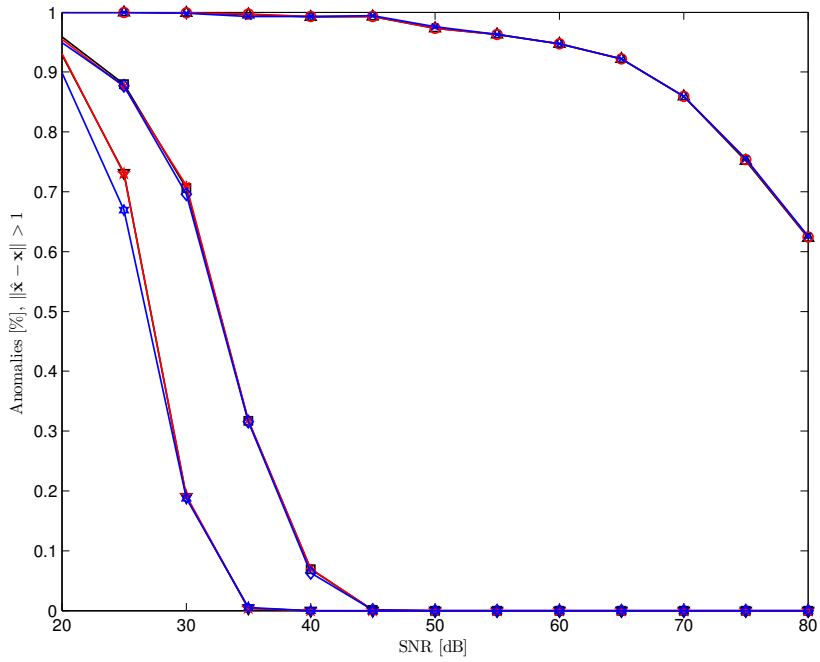
7.2 Real data experiments

Real data experiments were conducted in Lahti concert hall. The measurement setup is depicted in Fig. 7.5. One source location on the stage and one receiver location in the audience area was used. The loudspeaker on the stage was of type Genelec 1029A, and the G.R.A.S microphone array, introduced in Section 2.2, with $d_{\text{spc}} = 100$ mm spacing, was used in the receiver location. The height of the loudspeaker and the microphone array from the stage level was about 1.2 m, and 1.0 m, respectively. The sampling frequency was set to 48 kHz in the measurements. The impulse responses were measured using the sine-sweep technique with a 6 s long source signal with bandwidth from 40 Hz to 24 kHz.

Three reflections are windowed from the room impulse responses based on the source and receiver positions and the geometry of the hall. The estimated traces of the reflections are shown in Fig. 7.5. The time domain signals and frequency responses of the reflections in microphone no. 1 (-x direction) of the microphone array are shown in Fig. 7.6. The first two reflections, illustrated in Fig. 7.6, are from the curved side walls. The third reflection is a second order reflection via the same curved walls and it is already disturbed by another reflection arriving 1.2 ms before it. This



(a) Mean squared error



(b) Anomaly %

Figure 7.4. Results for localization against signal-to-noise ratio (SNR) from 1,000 Monte-Carlo samples.

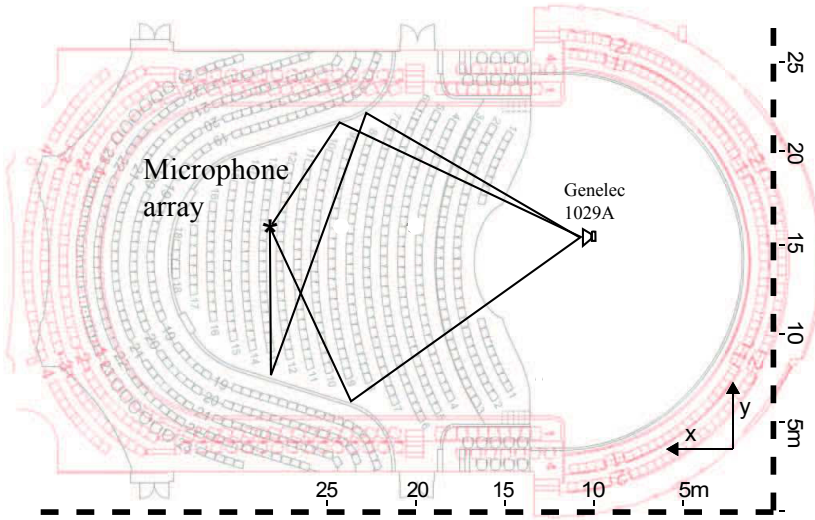


Figure 7.5. The setup in the real experiments. A loudspeaker of type Genelec 1029 A is located in the stage area, and the G.R.A.S. microphone array in the audience area. The reflections used in the experiments are illustrated with lines.

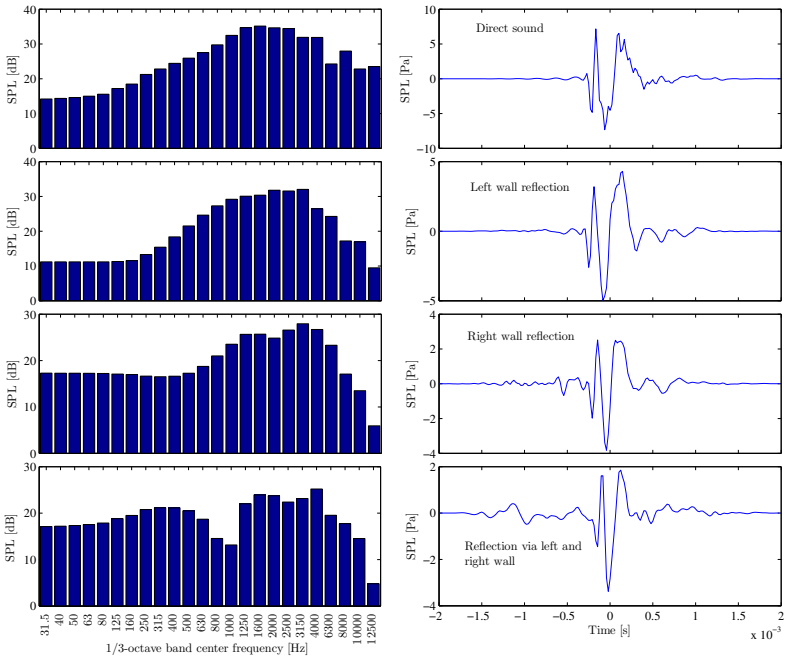


Figure 7.6. Time domain presentation of the reflections (right) and their frequency responses (left) on 1/3-octave bands. Please note that the y-axis scaling changes between the subplots and the sound pressure levels are relative. A typical impulse response of Genelec 1029A has two peaks and is visible in all the time domain impulses.

shows as emphasized low frequency content in the signal. As can be seen from Fig. 7.6, the wall reflections have a quite similar shape as the direct sound. This is due to the fact that the directionality of the loudspeaker stays similar in the frontal plane and the wall materials are highly reflective. Namely, the curved side walls in the inner stalls, characteristic to the Lahti concert hall, are of painted concrete which has a reflection coefficient of about 0.99 over the audible frequencies [43].

The measurement noise is removed from the impulse responses using spectral subtraction method [215]. The spectral subtraction will not benefit the localization accuracy. The spectral subtraction is made so that it can be assumed that the noise level is 0, and the SNR can be calculated in a more precise manner. The localization result after the spectral subtraction is chosen as the reference in these experiments. White noise is added to the clean signals as earlier in the simulations. The setup corresponds to the situation that was simulated earlier in this chapter, the difference is that here the reflection signals are measured in real situation.

7.2.1 Results

The localization results for the real reflections are shown in Fig. 7.7. In overall, the performance is clearly worse in the real situation than in the simulated situation. This is due to the fact that the real signals are not as easily localized as the simulated ones since their frequency content is not constant and they include several peaks instead of a single peak. This makes the TOA-based localization especially difficult. The secondary peak in the reflection signal causes the localization to vary between several locations. This is visible as an increase in the RMSE in Fig. 7.7, when 35 dB $<$ SNR $<$ 70 dB.

The real experiments reveal the weaknesses of MLE-TOA. When SNR$<$70 dB, MLE-TOA has worse performance than the other TOA-based localization methods. This is due to the fact that the time domain impulse response has two peaks. In the TOA estimation, only the maximum is selected. Since both of the peaks are almost equally strong, it is very probable that when additive noise is present the wrong one is selected.

MLE-CM and PL-CM have the best performance in the localization of real reflections. SRP-CM has clearly worse performance than other combined methods when SNR$<$70 dB. The reason for the weak performance of the SRP-CM is thought to be the fact that the competing maxima in

the TOA estimation functions induce even more ghosts to the localization functions than with a single peak.

7.3 Discussion

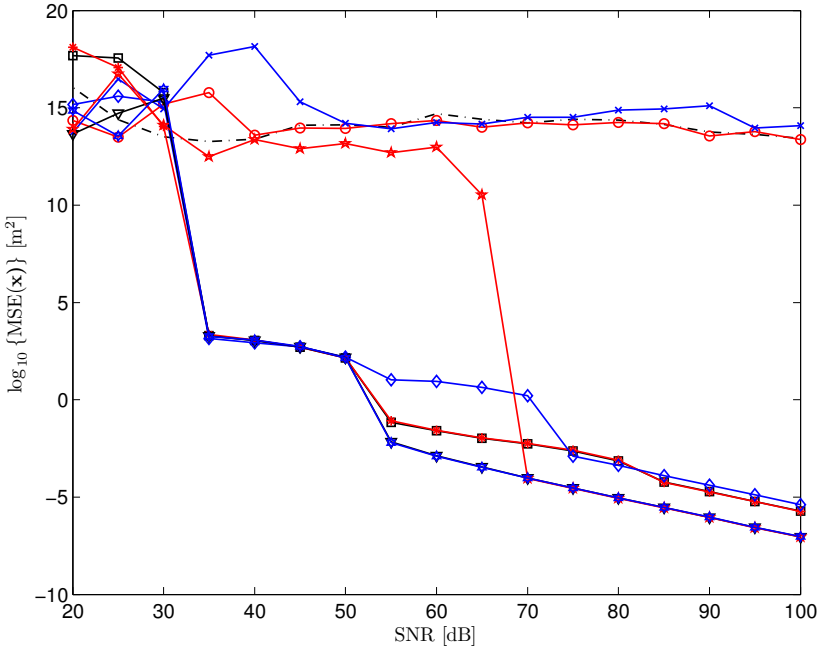
The errors in the localization with all the methods might be caused by other acoustic phenomena, for example the diffraction from chairs in the enclosure.

The case where there are more than one reflection present within one analysis window was not studied in this thesis. In principle, it is the same problem as the multi-source localization problem, and some of the methods used for that problem, e.g. the one presented in [209], should also be applicable for this problem. The MLE method presented in this thesis can not be directly applied for the multi-reflection localization problem. However, the PL method is directly applicable. Therefore in future work the PL method is preferred.

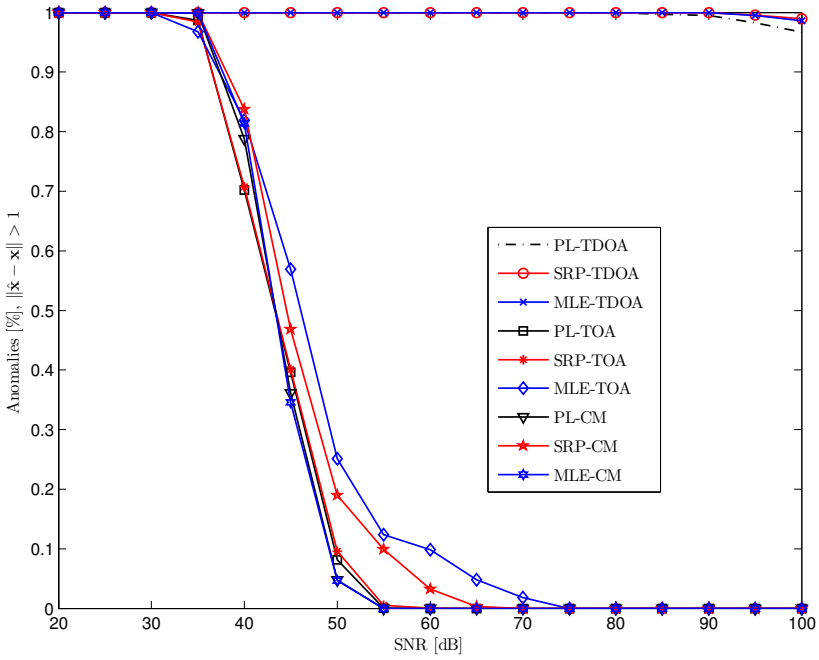
Since the SRP-CM adds the squared impulse responses and TDOA estimation functions, it is possible that the true reflection location gets less evidence than a "ghost" or a competing reflection. This behavior is also recognized in speech source localization [108]. The problem is not present in the PL-CM method, as shown in Fig. 5.1, since the ghosts are effectively downsized. Therefore, PL-CM outperforms SRP-CM in real situations.

One reason for the anomalous estimates with all the methods is that the arriving sound wave from the direction of the reflection is not as "impulse-like" as the sound wave in front of the loudspeaker. Thus, the magnitude of emitted sound wave in the direction of the reflections is lower, and does not contain as much high frequency energy as the impulse in front of the loudspeaker.

Moreover, the impulse response of the loudspeaker consists of two impulses instead of one, as shown in Fig. 7.6. In this case, the reflections do not introduce sharp peaks in the localization function with the TOA methods, and the intersection of the spheres is "blurred". By analyzing the two impulses of the loudspeaker impulse response with linear filtering, it is revealed that the first peak consists of frequencies that are above approximately 3.3 kHz, which is the cut-off frequency between loudspeaker elements, and the second peak for the frequencies below 3.3 kHz. Thus, the lower frequencies arrive about 0.3 ms later than the high frequencies, in



(a) Mean squared error



(b) Anomaly %

Figure 7.7. Results for localization against signal-to-noise ratio (SNR) with the real reflection signals.

front of the loudspeaker.

The two peaked impulse response of Genelec 1029A is caused by two issues. Firstly, the loudspeaker consists of two elements that are separated by approximately 10 cm. This causes some differences in the delays for low and high frequencies, depending on the direction of the loudspeaker with respect to the microphone. Secondly, the low-frequency-element of the loudspeaker has a higher mass, thus it does not respond to the voltage changes in the coil as quickly as the tweeter, thus causing the low frequencies to be delayed. All of the above, makes accurate localization of the loudspeaker quite difficult using only TOA information. One can also ask: What is then the location of a two-way loudspeaker? The methods in this thesis, assume that it is the acoustic center of the loudspeaker.

One possibility to get around the above problems related to the loudspeaker non-idealities is to use only the phase information of the signal. However, this decreases the SNR in the frequencies that have a low magnitude and as a result decreases the performance, as seen in the simulations with PHAT which uses only the phase information.

Another possibility to obtain more accurate TOA information is to measure the impulse response of the loudspeaker to a grid of directions in free-field conditions. Then the impulse response of the loudspeaker can be compensated from the impulse response by deconvolving the reflection with the free-field impulse response in the corresponding direction. This however would require a large data space of a priori measurements of the loudspeaker. The accuracy could be further improved if a one-way loudspeaker would be used.

TOA estimation can be also improved by applying the sparse impulse response technique presented in Publication I. The higher the directionality of the loudspeaker is, the better the TOA estimation accuracy is, when the sparse impulse response technique is used.

8. Summary

This thesis presented techniques for localizing early reflections from room impulse responses. A measurement technique for the investigation of early reflections was proposed and studied. Several localization methods were proposed. The performance of the localization methods was studied in theory, as well as in simulated, and in realistic situations.

8.1 Main results

The main results of this thesis can be summarized as follows:

- When studying the early reflections, a directional loudspeaker should be preferred because better spatial and temporal spacing can be achieved. The more directional the loudspeaker is, the more separability is achieved.

- One way loudspeaker is preferred in the localization of early reflections. Each element in the loudspeaker produces a peak in the impulse response. Therefore multi-element loudspeakers cause multi-peaked impulses in many cases, which then complicate the localization.

- Localization of the reflections should use both the time of arrival and the time difference of arrival information. The combination of these two pieces of information was shown to provide better performance than when only time of arrival or time difference of arrival was used.

- Simple direct cross correlation and peak-picking are good-enough-methods for TDOA estimation and TOA estimation in the reflection localization, respectively. Although better performing methods exist for both TDOA and TOA estimation, they require a priori knowledge of the source or of

the noise signals.

- Localization methods that use pressure signals directly should be preferred over the sound intensity vector based methods in the reflection localization task.

- Maximum pseudo-likelihood and maximum likelihood estimation methods should be preferred over steered response power methods in the localization of reflections, since they have better performance. The decrement in the performance of the steered response power methods was considered to be due to the ghosts in the localization functions.

- Interpolation is needed to achieve better spatial resolution. The proposed interpolation method is found to provide a clear improvement to the baseline method. The method is based on assuming the shape of the local maxima of the time difference of arrival or time of arrival estimation functions.

- In addition to room impulse responses, it is possible to localize reflections from speech or other continuous signals, without any a priori knowledge of the source signal. The localization of a reflection with speech sources has a worse performance than with impulse responses since the signal-to-noise ratio is typically lower for speech than for impulse responses.

8.2 Future work

Future work in the area of reflection localization includes:

- The development of an algorithm that can deal with multiple reflections arriving during the same time window. This thesis considered the case when a reflection arrives during a short time window.

- Theoretical performance of the localization of reflections when directional loudspeakers are used.

- The use of superdirectional microphone arrays along with superdirec-

tional loudspeakers should be investigated. This could be applied, for example, in the in-situ measurement of absorption coefficients.

Bibliography

[1] J. Merimaa and V. Pulkki. Spatial Impulse Response Rendering I: Analysis and Synthesis. *The Journal of the Audio Engineering Society*, 53(12):1115–1127, 2005.

[2] V. Pulkki and J. Merimaa. Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests. *The Journal of the Audio Engineering Society*, 54(1-2):3–20, 2006.

[3] D. Aprea, F. Antonacci, A. Sarti, and S. Tubaro. Acoustic reconstruction of the geometry of an environment through acquisition of a controlled emission. In *17th European Signal Processing Conference*, pages 710–714, 2009.

[4] F. Antonacci, A. Sarti, and S. Tubaro. Geometric Reconstruction of the Environment from its Response to Multiple Acoustic Emissions. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2822–2825, 2010.

[5] B. Günel. Room shape and size estimation using directional impulse response measurements. In *3rd EAA Congress on Acoustics, Forum Acusticum*, 2002.

[6] E. Mommertz. Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique. *Applied Acoustics*, 46(3):251–263, 1995.

[7] C. Nocke. In-situ acoustic impedance measurement using a free-field transfer function method. *Applied Acoustics*, 59(3):253–264, 2000.

[8] ISO Standard 3382-1. Acoustics – measurement of room acoustic parameters – part 1: Performance spaces, 2009.

[9] M.R. Schroeder. Statistical parameters of the frequency response curves of large rooms. *The Journal of the Audio Engineering Society*, 35(5):299–305, 1987.

[10] J.M. Jot, L. Cerveau, and O. Warusfel. Analysis and synthesis of room reverberation based on a statistical time-frequency model. In *103th Audio Engineering Society Convention*, 1997. Paper number 4629.

[11] H. Kuttruff. *Room acoustics, 4th Ed.* Spon Press, NY, NY, USA, 2000.

[12] J. Merimaa. *Analysis, Synthesis, and Perception of Spatial Sound–Binaural Localization Modeling and Multichannel Loudspeaker Reproduction*. PhD thesis, Helsinki University of Technology, 2006.

[13] A. O'Donovan, R. Duraiswami, and D. Zotkin. Imaging concert hall acoustics using visual and audio cameras. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5284–5287, 2008.

[14] T. Lokki, L. Savioja, et al. Auralization, url: `auralization.tkk.fi`, (Last accessed) July 2011.

[15] T. Korhonen. *Acoustic Source Localization Utilizing Reflective Surfaces*. PhD thesis, Tampere University of Technology, 2010.

[16] T. Lokki, H. Vertanen, A. Kuusinen, J. Pätynen, and S. Tervo. Concert hall acoustics assessment with individually elicited attributes. *The Journal of the Acoustical Society of America*, 130(2):835–849, Aug. 2011.

[17] S.M. Kay. *Fundamentals of Statistical signal processing, Volume 1: Estimation theory*. Prentice-Hall, New Jersey, USA, 1998.

[18] J. Fox. *Applied regression analysis, linear models, and related methods*. Sage Publications, Inc, London, UK, 1997.

[19] A. Høst-Madsen. On the existence of efficient estimators. *IEEE Transactions on Signal Processing*, 48(11):3028–3031, 2000.

[20] F. Jacobsen. *Springer handbook of acoustics*, chapter 25 Sound Intensity, pages 1053–1075. Springer, NY, NY, USA, 2007.

[21] F. Fahy. *Sound intensity (2nd ed.)*. E&FN Spon, Chapman & Hall, London, UK, 1995.

[22] A.D. Pierce. *Acoustics: an introduction to its physical principles and applications*. Acoustical Society of America, 1994.

[23] A.D. Pierce. *Springer handbook of acoustics*, chapter 3. Basic Linear Acoustics. New York: Springer, 2007.

[24] H.E. de Bree. An overview of microflown technologies. *Acta acustica united with Acustica*, 89(1):163–172, 2003.

[25] C. Knapp and G. Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4):320–327, 1976.

[26] G.C. Carter. Coherence and time delay estimation. *Proceedings of the IEEE*, 75(2):236–255, 1987.

[27] B. Yang and J. Scheuing. Cramer-rao bound and optimum sensor array for source localization from time differences of arrival. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 961–964, 2005.

[28] R. Hickling and A. W. Brown. Determining the direction to a sound source in air using vector sound-intensity probes. *The Journal of the Acoustical Society of America*, 129(1):219–224, 2011.

[29] M. Kallinger, F. Kuech, R. Schultz-Amling, G. del Galdo, J. Ahonen, and V. Pulkki. Enhanced Direction Estimation Using Microphone Arrays for Directional Audio Coding. In *Hands-Free Speech Communication and Microphone Arrays*, pages 45–48, 2008.

[30] J. Pätynen and T. Lokki. Directivities of Symphony Orchestra Instruments. *Acta Acustica united with Acustica*, 96(1):138–167, 2010.

[31] ISO Standard 3745-1. Determination of sound power levels of noise sources using sound pressure – Precision methods for anechoic and hemi-anechoic rooms, 2009.

[32] J.D. Maynard, E.G. Williams, and Y. Lee. Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH. *The Journal of the Acoustical Society of America*, 78:1395–1413, 1985.

[33] W.A. Veronesi and J.D. Maynard. Nearfield acoustic holography (NAH) II. Holographic reconstruction algorithms and computer implementation. *The Journal of the Acoustical Society of America*, 81:1307–1322, 1987.

[34] K. Yang-Hann. *Springer Handbook of Acoustics*, chapter 26. Acoustic Holography. Springer-Verlag, New York, NY, USA, 2007.

[35] M. McPherson M.A. Breazeale. *Springer handbook of Acoustics, (Ed. Rossing, Thomas D.)*, chapter 6. Physical Acoustics, pages 209–237. Springer, New York, NY, USA, 2007.

[36] G.S.K. Wong. *Springer handbook of acoustics*, chapter 24 Microphones and Their Calibration, pages 1024–1048. Springer, NY, NY, USA, 2007.

[37] T. Lokki. *Physically-based Auralization-Design, Implementation, and Evaluation*. PhD thesis, Helsinki University of Technology, 2002.

[38] C.M. Harris. Absorption of sound in air versus humidity and temperature. *The Journal of the Acoustical Society of America*, 40:148–159, 1966.

[39] J.B. Allen and D.A. Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950, 1979.

[40] Y. Liu and F. Jacobsen. Measurement of absorption with a pu sound intensity probe in an impedance tube. *The Journal of the Acoustical Society of America*, 118:2117–2120, 2005.

[41] W.T. Chu. Transfer function technique for impedance and absorption measurements in an impedance tube using a single microphone. *Journal of the Acoustical Society of America*, 80(2):555–560, 2010.

[42] P. Robinson and N. Xiang. On the subtraction method for in-situ reflection and diffusion coefficient measurements. *Journal of the Acoustical Society of America, Express Letters*, pages EL99 – EL104, 2010.

[43] T.J. Cox and P. D'Antonio. *Acoustic absorbers and diffusers: theory, design, and application*. London and New York: Spon Press, 2004.

[44] R.R. Torres, U.P. Svensson, and M. Kleiner. Computation of edge diffraction for more accurate room acoustics auralization. *The Journal of the Acoustical Society of America*, 109:600–610, 2001.

[45] B.I. Dalenbäck, M. Kleiner, and P. Svensson. A macroscopic view of diffuse reflection. *Journal of the Audio Engineering Society*, 42(10):793–807, 1994.

[46] M. Vorländer and E. Mommertz. Definition and measurement of random-incidence scattering coefficients. *Applied Acoustics*, 60(2):187–199, 2000.

[47] T.J. Cox, B.I.L. Dalenback, P. D'Antonio, J.J. Embrechts, J.Y. Jeon, E. Mommertz, and M. Vorlander. A tutorial on scattering and diffusion coefficients for room acoustic surfaces. *Acta Acustica united with Acustica*, 92(1):1–15, 2006.

[48] B. Rafaely. Spatial-temporal correlation of a diffuse sound field. *The Journal of the Acoustical Society of America*, 107:3254–3258, 2000.

[49] R.V. Waterhouse. Statistical Properties of Reverberant Soundfields. *The Journal of the Acoustical Society of America*, 43:1436–1443, 1968.

[50] C.T. Morrow. Point-to-point correlation of sound pressures in reverberation chambers. *Journal of Sound and Vibration*, 16(1):29–42, 1971.

[51] H. Nelisse and J. Nicolas. Characterization of a diffuse field in a reverberant room. *The Journal of the Acoustical Society of America*, 101(6):3517–3524, 1997.

[52] W.K. Blake and R.V. Waterhouse. The use of cross-spectral density measurements in partially reverberant sound fields. *Journal of Sound and Vibration*, 54(4):589–599, 1977.

[53] M. Kuster. Spatial correlation and coherence in reverberant acoustic fields: Extension to microphones with arbitrary first-order directivity. *The Journal of the Acoustical Society of America*, 123:154–162, 2008.

[54] F. Jacobsen and T. Roisin. The coherence of reverberant sound fields. *The Journal of the Acoustical Society of America*, 108:204–210, 2000.

[55] J. Ahonen and V. Pulkki. Diffuseness estimation using temporal variation of intensity vectors. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 285–288, 2009.

[56] O. Thiergart, G. Del Galdo, and E.A.P. Habets. Diffuseness estimation with high temporal resolution via spatial coherence between virtual first-order microphones. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 217–220, 2011.

[57] P. Billingsley. *Probability and measure*. John Wiley & Sons, New York, NY, USA, 1995.

[58] J.-D. Polack. *La transmission de l'énergie sonore dans les salles*. PhD thesis, Université du Maine, Le Mans, 1988.

[59] T. Hidaka, Y. Yamada, and T. Nakagawa. A new definition of boundary point between early reflections and late reverberation in room impulse responses. *The Journal of the Acoustical Society of America*, 122:326–332, 2007.

[60] J.-D. Polack. Modifying chambers to play billiards, the foundations of reverberation theory. *Acustica*, 76(6):257–272, 1992.

[61] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. Creating interactive virtual acoustic environments. *The Journal of the Audio Engineering Society*, 47(9):675–705, 1999.

[62] K. Meesawat and D. Hammershøi. An investigation on the transition from early reflections to a reverberation tail in a BRIR. In *International Conference on Auditory Display*, 2002.

[63] R. Stewart and M. Sandler. Statistical Measures of Early Reflections of Room Impulse Responses. In *10th International Conference on Digital Audio Effects (DAFx-07)*, pages 59–62, 2007.

[64] G. Defrance and J.-D. Polack. Measuring the mixing time in auditoria. In *Acoustics 08*, pages 3869–3874, 2008.

[65] G. Defrance, L. Daudet, and J.-D. Polack. Detecting arrivals within room impulse responses using matching pursuit. In *11th International Conference on Digital Audio Effects (DAFx-08)*, pages 1–4, 2008.

[66] G. Defrance, L. Daudet, and J.-D. Polack. Using matching pursuit for estimating mixing time within room impulse responses. *Acta Acustica united with Acustica*, 95(6):1082–1092, 2009.

[67] E. Lehmann and A. Johansson. Diffuse reverberation model for efficient image-source simulation of room impulse responses. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(8), 2009.

[68] C. Borß. A Novel Approach for Optimally Matching a Late Reverberation Model to an Image Source Model-Or: What Does a Football Have to Do With Shoebox Shaped Rooms? In *EAA Symposium on Auralization*, 2009.

[69] G. Kearney, C. Masterson, S. Adams, and F. Boland. Towards Efficient Binaural Room Impulse Response Synthesis. In *EAA Symposium on Auralization*, 2009.

[70] G. Defrance and J.-D. Polack. Estimating the mixing time of concert halls using the eXtensible Fourier Transform. *Applied Acoustics*, 71:777–792, 2010.

[71] Cheol-Ho Jeong, Jonas Brunskog, and Finn Jacobsen. Room acoustic transition time based on reflection overlap. *The Journal of the Acoustical Society of America*, 127(5):2733–2736, 2010.

[72] A. Lindau, L. Kosanke, and S. Weinzierl. Perceptual Evaluation of Physical Predictors of the Mixing Time in Binaural Room Impulse Responses. In *128th Audio Engineering Society Convention*, 2010. Paper number 8089.

[73] Alexis Billon and Jean-Jacques Embrechts. Numerical evidence of mixing in rooms using the free path temporal distribution. *The Journal of the Acoustical Society of America*, 130(3):1381–1389, 2011.

[74] A.C. Gade. *Springer handbook of Acoustics, (Ed. Rossing, T.D.)*, chapter 9. Acoustics in Halls for Speech and Music, pages 301–353. Springer, New York, NY, USA, 2007.

[75] K.H. Kuttruff. Auralization of impulse responses modeled on the basis of ray-tracing results. *The Journal of the Audio Engineering Society*, 41:876–876, 1993.

[76] L.L. Beranek. Concert hall acoustics–1992. *The Journal of the Acoustical Society of America*, 92:1–39, 1992.

[77] T. Hidaka, L.L. Beranek, and T. Okano. Interaural cross-correlation, lateral fraction, and low-and high-frequency sound levels as measures of acoustical quality in concert halls. *The Journal of the Acoustical Society of America*, 98:988–1007, 1995.

[78] L. Cremer, H.A. Müller, and T.D. Northwood. *Principles and applications of room acoustics*. Applied Science, London, UK, 1982.

[79] W. Reichardt and U. Lehmann. Raumeidruck als oberbegriff von raumlichkeit und halligkeit. *Acoustica*, 40:174–183, 1978.

[80] J. Abel and P. Huang. A Simple, Robust Measure of Reverberation Echo Density. In *121st Audio Engineering Society Convention*, 2006. Paper number 6985.

[81] M. Kuster. Reliability of estimating the room volume from a single room impulse response. *The Journal of the Acoustical Society of America*, 124:982–993, 2008.

[82] A. Farina and R. Ayalon. Recording concert hall acoustics for posterity. In *24th Audio Engineering Society Conference on Multichannel Audio, Banff, Canada*, pages 26–28, 2003.

[83] A. Farina, A. Capra, L. Conti, P. Martignon, and F.M. Fazi. Measuring spatial impulse responses in concert halls and opera houses employing a spherical microphone array. In *19th International Congress on Acoustics (ICA), Madrid*, 2007. Paper number RBA-07-010.

[84] A. Farina, P. Martignon, A. Capra, and S. Fontana. Measuring impulse responses containing complete spatial information. In *22nd Audio Engineering Society UK Conference*, 2007.

[85] A. Farina, A. Amendola, A. Capra, and C. Varani. Spatial analysis of room impulse responses captured with a 32-capsules microphone array. In *130th Audio Engineering Society Convention, London, 13-16 May 2011*, 2011. Paper number 8400.

[86] M. Kuster, D. de Vries, E.M. Hulsebos, and A. Gisolf. Acoustic imaging in enclosed spaces: Analysis of room geometry modifications on the impulse response. *The Journal of the Acoustical Society of America*, 116:2126–2137, 2004.

[87] M. Kuster and D. de Vries. Modelling and Order of Acoustic Transfer Functions Due to Reflections from Augmented Objects. *EURASIP Journal on Advances in Signal Processing*, 2007. Article ID 30253.

[88] H. Okubo, M. Otani, R. Ikezawa, S. Komiyama, and K. Nakabayashi. A system for measuring the directional room acoustical parameters. *Applied Acoustics*, 62(2):203–215, 2001.

[89] A. Omoto and H. Uchida. Evaluation method of artificial acoustical environment: Visualization of sound intensity. *Journal of Physiological Anthropology and Applied Human Science*, 23(6):249–253, 2004.

[90] B.N. Gover, J.G. Ryan, and M.R. Stinson. Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array. *The Journal of the Acoustical Society of America*, 116(4):2138–2148, 2004.

[91] M. Park and B. Rafaely. Sound-field analysis by plane-wave decomposition using spherical microphone array. *The Journal of the Acoustical Society of America*, 118:3094, 2005.

[92] M.S. Brandstein and H.F. Silverman. A robust method for speech signal time-delay estimation in reverberant rooms. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 375–378, 1997.

[93] M.S. Brandstein. Time-delay estimation of reverberated speech exploiting harmonic structure. *The Journal of the Acoustical Society of America*, 105:2914, 1999.

[94] J. Benesty. Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. *The Journal of the Acoustical Society of America*, 107:384–391, 2000.

[95] G. Jacovitti and G. Scarano. Discrete time techniques for time delay estimation. *IEEE Transactions on Signal Processing*, 41(2):525–533, 1993.

[96] J. Chen, J. Benesty, and Y. Huang. Performance of GCC-and AMDF-based time-delay estimation in practical reverberant environments. *EURASIP Journal on Applied Signal Processing*, 2005:25–36, 2005.

[97] J. Chen, J. Benesty, and Y. Huang. Time delay estimation in room acoustic environments: an overview. *EURASIP Journal on Applied Signal Processing*, 2006(1), 2006. Article ID 26503.

[98] Eric A. Lehmann. Particle filtering approach to adaptive time-delay estimation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 1129–1132, 2006.

[99] A. Weiss and E. Weinstein. Fundamental limitations in passive time delay estimation–Part I: Narrow-band systems. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 31(2):472–486, 1983.

[100] E. Weinstein and A. Weiss. Fundamental limitations in passive time-delay estimation–Part II: Wide-band systems. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 32(5):1064–1078, 1984.

[101] X. Lai and H. Torp. Interpolation methods for time delay using cross-correlation for blood velocity measurement. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 46(2):277–290, 1999.

[102] L. Zhang and X. Wu. On cross correlation based discrete time delay estimation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume IV, pages 981–984, 2005.

[103] C. Falsi, D. Dardari, L. Mucchi, and M.Z. Win. Time of arrival estimation for uwb localizers in realistic environments. *EURASIP Journal on Applied Signal Processing*, 2006:152–152, 2006.

[104] John Usher. An improved method to determine the onset timings of reflections in an acoustic impulse response. *The Journal of the Acoustical Society of America, Express Letters*, 127(4):EL172–EL177, 2010.

[105] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M.B. Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, 2005.

[106] J.H. DiBiase, H.F. Silverman, and M.S. Brandstein. *Microphone arrays: signal processing techniques and applications*, chapter 8 Robust Localization in Reverberant Rooms, pages 157–180. Springer Verlag, New York, NY, USA, 2001.

[107] M. Omologo, P. Svaizer, and R. De Mori. *Spoken dialogues with computers (Ed. E. De Mori)*, chapter Acoustic Transduction, page 61. Academic Press, London, UK, 1998.

[108] P. Pertilä, T. Korhonen, and A. Visa. Measurement combination for acoustic source localization in a room environment. *EURASIP Journal on Audio, Speech, and Music Processing*, 2008. Article ID 278185.

[109] V.C. Raykar, I.V. Kozintsev, and R. Lienhart. Position calibration of microphones and loudspeakers in distributed computing platforms. *IEEE Transactions on Speech and Audio Processing*, 13(1):70–83, 2005.

[110] V.C. Raykar, I. Kozintsev, and R. Lienhart. Self localization of acoustic sensors and actuators on distributed platforms. In *International Workshop on Multimedia Technologies in E-Learning and Collaboration*, 2003.

[111] V.C. Raykar, I. Kozintsev, and R. Lienhart. Position calibration of audio sensors and actuators in a distributed computing platform. In *ACM international conference on Multimedia*, pages 572–581, 2003.

[112] I. Ziskind and M. Wax. Maximum likelihood localization of multiple sources by alternating projection. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(10):1553–1560, 1988.

[113] J.C. Chen, R.E. Hudson, and K. Yao. A maximum-likelihood parametric approach to source localizations. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 3013–3016, 2001.

[114] J.C. Chen, R.E. Hudson, and K. Yao. Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field. *IEEE Transactions on Signal Processing*, 50(8):1843–1854, 2002.

[115] B. Mungamuru and P. Aarabi. Joint sound localization and orientation estimation. In *IEEE International Conference on Information Fusion*, pages 81–85, 2003.

[116] B. Mungamuru and P. Aarabi. Enhanced Sound Localization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(3):1526–1540, 2004.

[117] C. Zhang, Z. Zhang, and D. Florêncio. Maximum likelihood sound source localization for multiple directional microphones. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 125–128, 2007.

[118] H. Schau and A. Robinson. Passive source localization employing intersecting spherical surfaces from time-of-arrival differences. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(8):1223–1225, 1987.

[119] J. Smith and J. Abel. Closed-form least-squares source location estimation from range-difference measurements. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(12):1661–1669, 1987.

[120] J. Abel and J. Smith. The spherical interpolation method for closed-form passive source localization using range difference measurements. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 12, pages 471–474, 1987.

[121] Y.T. Chan and K.C. Ho. A simple and efficient estimator for hyperbolic location. *IEEE Transactions on Signal Processing*, 42(8):1905–1915, 1994.

[122] K. Yao, R.E. Hudson, C.W. Reed, D. Chen, and F. Lorenzelli. Blind beamforming on a randomly distributed sensor array system. *IEEE Journal on Selected Areas in Communications*, 16(8):1555–1567, 1998.

[123] C.W. Reed, R. Hudson, and K. Yao. Direct joint source localization and propagation speed estimation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 1169–1172, 1999.

[124] M.D. Gillette and H.F. Silverman. A linear closed-form algorithm for source localization from time-differences of arrival. *IEEE Signal Processing Letters*, 15:1–4, 2008.

[125] Y. Huang, J. Benesty, G.W. Elko, and RM Mersereati. Real-time passive source localization: A practical linear-correction least-squares approach. *IEEE Transactions on Speech and Audio Processing*, 9(8):943–956, 2001.

[126] A. Mahajan and M. Walworth. 3D position sensing using the differences in the time-of-flights from a wave source to various receivers. *IEEE Transactions on Robotics and Automation*, 17(1):91–94, 2002.

[127] H.C. So and S.P. Hui. Constrained location algorithm using TDOA measurements. *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, 86(12):3291–3293, 2003.

[128] V.C. Raykar and R. Duraiswami. Approximate expressions for the mean and the covariance of the maximum likelihood estimator for acoustic source localization. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 73–76, 2005.

[129] T. Ajdler, I. Kozintsev, R. Lienhart, and M. Vetterli. Acoustic source localization in distributed sensor networks. In *IEEE 38th Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1328–1332, 2004.

[130] P. Stoica and J. Li. Lecture notes-source localization from range-difference measurements. *IEEE Signal Processing Magazine*, 23(6):63–66, 2006.

[131] J. Zheng, K.W.K. Lui, and H.C. So. Accurate three-step algorithm for joint source position and propagation speed estimation. *Signal Processing*, 87(12):3096–3100, 2007.

[132] K.W.K. Lui, J. Zheng, and HC So. Particle swarm optimization for time-difference-of-arrival based localization. In *European Signal Processing Conference*, pages 414–417, 2007.

[133] K.W.K. Lui, W.K. Ma, HC So, and F.K.W. Chan. Semi-definite programming algorithms for sensor network node localization with uncertainties in anchor positions and/or propagation speed. *IEEE Transactions on Signal Processing*, 57(2):752–763, 2009.

[134] K. Yang, G. Wang, and Z.-Q. Luo. Efficient convex relaxation methods for robust target localization by a sensor network using time differences of arrivals. *IEEE Transactions on Signal Processing*, 57:2775–2784, July 2009.

[135] H. Jwu-Sheng and Y. Chia-Hsin. Estimation of Sound Source Number and Directions under a Multisource Reverberant Environment. *EURASIP Journal on Advances in Signal Processing*, 2010:Article ID 870756, 2010.

[136] M. Walworth and A. Mahajan. 3D position sensing using the difference in the time-of-flights from a wave source to various receivers. In *8th International Conference on Advanced Robotics*, pages 611–616, 1997.

[137] Å. Björck. *Numerical methods for least squares problems*. Society for Industrial and Applied Mathematics, Amsterdam, Holland, 1996.

[138] C.L. Lawson and R.J. Hanson. *Solving least squares problems*, volume 15. Society for Industrial and Applied Mathematics, Amsterdam, Holland, 1995.

[139] K.W. Cheung, H.C. So, W.K. Ma, and Y.T. Chan. Least squares algorithms for time-of-arrival-based mobile location. *IEEE Transactions on Signal Processing*, 52(4):1121–1130, 2004.

[140] W. Kim, J.G. Lee, and G.I. Jee. The interior-point method for an optimal treatment of bias in trilateration location. *IEEE Transactions on Vehicular Technology*, 55(4):1291–1301, 2006.

[141] E. Xu, Z. Ding, and S. Dasgupta. Source Localization in Wireless Sensor Networks from Signal Time-of-Arrival Measurements. *IEEE Transactions on Signal Processing*, –(Early Access):1–11, 2011.

[142] P. Pertilä, M. Mieskolainen, and M.S. Hämäläinen. Closed-form self-localization of asynchronous microphone arrays. In *Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, pages 139–144, 2011.

[143] P. Aarabi. The Fusion of Distributed Microphone Arrays for Sound Localization. *EURASIP Journal on Applied Signal Processing*, 2003(4):338–347, 2003.

[144] D.N. Zotkin and R. Duraiswami. Accelerated speech source localization via a hierarchical search of steered response power. *IEEE Transactions on Speech and Audio Processing*, 12:499–508, 2004.

[145] J.M. Peterson and C. Kyriakakis. Hybrid algorithm for robust, real-time source localization in reverberant environments. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages iv/1053–iv/1056, 2005.

[146] A. Johansson, G. Cook, and S. Nordholm. Acoustic direction of arrival estimation, a comparison between Root-MUSIC and SRP-PHAT. In *TENCON, IEEE Region 10*, volume B, pages 629–632, 2004.

[147] H. Do and H.F. Silverman. A fast microphone array srp-phat source location implementation using coarse-to-fine region contraction (CFRC). In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 295–298, 2007.

[148] H. Do, H. F. Silverman, and Y. Yu. A real-time srp-phat source location implementation using stochastic region contraction (src) on a large-aperture microphone array. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 121–124, 2007.

[149] H. Do and H.F. Silverman. A method for locating multiple sources from a frame of a large-aperture microphone array data without tracking. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 301–304, 2008.

[150] H. Do and H. F. Silverman. Stochastic particle filtering: A fast SRP-PHAT single source localization algorithm. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 213–216, 2009.

[151] J.M. Valin, F. Michaud, and J. Rouat. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems*, 55(3):216–228, 2007.

[152] E.A. Lehmann. *Particle Filtering Methods for Acoustic Source Localisation and Tracking*. PhD thesis, Australian National University, 2004.

[153] P. Pertilä. *Acoustic Source Localization in a Room Environment and at Moderate Distances*. PhD thesis, Tampere University of Technology, 2009.

[154] M.S. Brandstein, J.E. Adcock, and H.F. Silverman. Microphone-array localization error estimation with application to sensor placement. *The Journal of the Acoustical Society of America*, 99(6):3807–3816, 1996.

[155] B. Günel, H. Hacıhabiboğlu, and A.M. Kondoz. Acoustic Source Separation of Convolutive Mixtures Based on Intensity Vector Statistics. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 16(4):748–756, 2008.

[156] M. Hawkes and A. Nehorai. Wideband source localization using a distributed acoustic vector-sensor array. *IEEE Transactions on Signal Processing*, 51(6):1479–1491, 2003.

[157] A. Nehorai and E. Paldi. Acoustic vector-sensor array processing. *IEEE Transactions on Signal Processing*, 42(9):2481–2491, 1994.

[158] D. Levin, E.A.P. Habets, and S. Gannot. On the angular error of intensity vector based direction of arrival estimation in reverberant sound fields. *The Journal of the Acoustical Society of America*, 128:1800–1811, 2010.

[159] D. Levin, S. Gannot, and E.A.P. Habets. Direction-of-arrival estimation using acoustic sensor-vectors in the presence of noise. In *IEEE International Conference of Acoustics, Speech, and Signal Processing*, pages 105–108, 2011.

[160] A. O'Donovan, R. Duraiswami, and J. Neumann. Microphone arrays as generalized cameras for integrated audio visual processing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[161] Eric Van Lancker. Localization of reflections in auditoriums using time delay estimation. In *108th Audio Engineering Society Convention*, 2000. Paper number 5168.

[162] J. Filos, E.A.P. Habets, and P.A. Naylor. A two-step approach to blindly infer room geometries. In *International Workshop on Acoustic Echo and Noise Cancellation*, 2010.

[163] A. Canclini, M. R. P. Thomas, A. Antonacci, F. Sarti, and P. A. Naylor. Robust inference of room geometry from acoustic impulse responses. In *19th European Signal Processing Conference*, pages 161–165, 2011.

[164] E. A. Nastasia, F. Antonacci, A. Sarti, and S. Tubaro. Localization of planar acoustic reflections through emission of controlled stimuli. In *19th European Signal Processing Conference*, pages 156–160, 2011.

[165] A. Canclini, F. Antonacci, M. R. P. Thomas, J. Filos, A. Sarti, P. A. Naylor, and Tubaro S. Exact localization of acoustic reflectors from quadratic constraints. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 17–20, 2011.

[166] A.E. O'Donovan, R. Duraiswami, and D.N. Zotkin. Automatic matched filter recovery via the audio camera. In *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 2826–2829, 2010.

[167] D. Ba, F. Ribeiro, C. Zhang, and D. Florencio. L1 regularized room modeling with compact microphone arrays. In *35th IEEE International Conference Acoustics, Speech and Signal Processing*, pages 157–160, 2010.

[168] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, and S. Tubaro. From direction of arrival estimates to localization of planar reflectors in a two dimensional geometry. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2620–2623, 2011.

[169] E. Mabande, K. Sun, K. Kowalzcyk, and W. Kellermann. On 2d-localization of reflectors using robust beamforming techniques. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 153–156, 2011.

[170] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann. Joint doa and tdoa estimation for 3d localization of reflective surfaces using eigenbeam mvdr and spherical microphone arrays. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 113–116, 2011.

[171] V.C. Raykar. *Position calibration of acoustic sensors and actuators on distributed general purpose computing platforms*. PhD thesis, University of Maryland, Maryland, USA, 2003.

[172] V.C. Raykar and R. Duraiswami. Automatic position calibration of multiple microphones. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume iv, pages 69–72, 2004.

[173] A. Redondi, M. Tagliasacchi, F. Antonacci, and A. Sarti. Geometric calibration of distributed microphone arrays. In *IEEE International Workshop Multimedia Signal Processing*, pages 1–5, 2009.

[174] M. Binelli, A. Venturi, A. Amendola, and A. Farina. Experimental analysis of spatial properties of the sound field inside a car employing a spherical microphone array. In *130th Audio Engineering Society Convention, London, 13-16 May 2011*, 2011. Paper number 8338.

[175] T. Lokki, H. Vertanen, A. Kuusinen, J. Pätynen, and S. Tervo. Auditorium acoustics assessment with sensory evaluation methods. In *International Symposium on Room Acoustics*, pages 29–31, 2010.

[176] T. Lokki, J. Pätynen, S. Tervo, S. Siltanen, and L. Savioja. Engaging concert hall acoustics is made up of temporal envelope preserving reflections. *The Journal of the Acoustical Society of America*, 129(6):EL223–EL228, 2011.

[177] P. Bergamo, S. Asgari, H. Wang, D. Maniezzo, L. Yip, R.E. Hudson, K. Yao, and D. Estrin. Collaborative sensor networking towards real-time acoustical beamforming in free-space and limited reverberance. *IEEE Transactions on Mobile Computing*, 3(3):211–224, 2004.

[178] W. Yan, W. Qun, B. Danping, and J. Jin. Acoustic localization in multi-path aware environments. In *International Conference on Communications, Circuits and Systems*, pages 667–670, 2007.

[179] T. Korhonen. Acoustic localization using reverberation with virtual microphones. In *International Workshop on Acoustic Echo and Noise Control*, 2008. Paper ID 9038.

[180] J. Scheuing and B. Yang. Disambiguation of tdoa estimation for multiple sources in reverberant environments. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(8):1479–1489, 2008.

[181] F. Ribeiro, C. Zhang, D.A. Florêncio, and D.E. Ba. Using reverberation to improve range and elevation discrimination for small array sound source localization. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(7):1781–1792, 2010.

[182] F. Ribeiro, D. Ba, C. Zhang, and D. Florêncio. Turning enemies into friends: using reflections to improve sound source localization. In *IEEE International Conference on Multimedia and Expo*, pages 731–736, 2010.

[183] P. Svaizer, A. Brutti, and M. Omologo. Use of reflectedwavefronts for acoustic source localization with a line array. In *Joint Workshop on Handsfree Speech Communication and Microphone Arrays*, pages 165–169, 2011.

[184] M. Omologo P. Svaizer, A. Brutti. Analysis of reflected wavefronts by means of a line microphone array. In *International Workshop on Acoustic Echo and Noise Control*, 2010. Paper ID 965.

[185] A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *108th Audio Engineering Society Convention*, 2000. Paper number 5093.

[186] A. Farina. Advancements in impulse response measurements by sine sweeps. In *122nd Convention Audio Engineering Society*, 2007. Paper number 7121.

[187] D.D. Rife and J. Vanderkooy. Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society*, 37(6):419–444, 1989.

[188] Yoiti Suzuki, Futoshi Asano, Hack-Yoon Kim, and Toshio Sone. An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses. *The Journal of the Acoustical Society of America*, 97(2):1119–1123, 1995.

[189] J. Pätynen, B.F.G. Katz, and T. Lokki. Investigations on the balloon as an impulse source. *The Journal of the Acoustical Society of America*, 129(1):EL27–EL33, 2011.

[190] A. Krokstad, S. Strom, and S. Sorsdal. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125, 1968.

[191] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J.E. West, G. Pingali, P. Min, and A. Ngan. A beam tracing method for interactive architectural acoustics. *The Journal of the Acoustical Society of America*, 115:739, 2004.

[192] P. Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[193] R. Mucci. A comparison of efficient beamforming algorithms. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 32(3):548–558, 1984.

[194] M.S. Brandstein, J.E. Adcock, and H.F. Silverman. A closed-form location estimator for use with room environment microphone arrays. *IEEE Transactions on Speech and Audio Processing*, 5(1):45–50, 1997.

[195] G.P. Yost and S. Panchapakesan. Automatic location identification using a hybrid technique. In *IEEE Vehicular Technology Conference*, volume 1, pages 264–267, 1998.

[196] R.I. Reza. *Data fusion for improved TOA/TDOA position determination in wireless systems*. PhD thesis, Faculty of the Virginia Polytechnic Institute and State University, 2000.

[197] S. Gezici and H.V. Poor. Position estimation via ultra-wide-band signals. *Proceedings of the IEEE*, 97(2):386–403, 2009.

[198] J. Yli-Hietanen, K. Kalliojärvi, and J. Astola. Low-complexity angle of arrival estimation of wideband signals using small arrays. In *IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pages 109–112, 1996.

[199] K.V. Mardia, P.E. Jupp, and KV Mardia. *Directional statistics*. Wiley, New York, NY, USA, 2000.

[200] A.D. Firoozabadi and H.R. Abutalebi. A new region search method based on DOA estimation for speech source localization by SRP-PHAT method. In *18th European Signal Processing Conference*, pages 656–660, 2010.

[201] J.P. Dmochowski, J. Benesty, and S. Affes. A generalized steered response power method for computationally viable source localization. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(8):2510–2526, 2007.

[202] L.G. da Silveira, V.P. Minotto, C.R. Jung, and B. Lee. A gpu implementation of the srp-phat sound source localization algorithm. In *International Workshop on Acoustic Echo and Noise control*, 2010. Paper ID 1062.

[203] J. Vermaak and A. Blake. Nonlinear filtering for speaker tracking in noisy and reverberant environments. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 3021–3024, 2001.

[204] D.B. Ward, E.A. Lehmann, and R.C. Williamson. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *IEEE Transactions on Speech and Audio Processing*, 11(6):826–836, 2003.

[205] F. Antonacci, D. Riva, A. Sarti, M. Tagliasacchi, and S. Tubaro. Tracking of two acoustic sources in reverberant environments using a particle swarm optimizer. In *IEEE Conference on Advanced Video and Signal Based Surveillance*, 2007.

[206] J. Kennedy and R. Eberhart. Particle swarm optimization. In *IEEE International Conference on Neural Networks*, volume 4, pages 1942–1948, 1995.

[207] R. Parisi, P. Croene, and A. Uncini. Particle swarm localization of acoustic sources in the presence of reverberation. In *IEEE International Symposium on Circuits and Systems*, pages 4739–4742, 2006.

[208] J.A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 1965.

[209] B. Alessio, O. Maurizio, and S. Piergiorgio. Multiple source localization based on acoustic map de-emphasis. *EURASIP Journal on Audio, Speech, and Music Processing*, 2010:Article ID 147495, 2010.

[210] A.V. Oppenheim and R.W. Schafer. Discrete-time signal processing (2nd ed.). *Prentice Hall Press Upper Saddle River, NJ, USA*, page 1120, 2009.

[211] S. Bellini and G. Tartara. Bounds on error in signal parameter estimation. *IEEE Transactions on Communications*, 22(3):340–342, 1974.

[212] J. Ianniello, E. Weinstein, and A. Weiss. Comparison of the ziv-zakai lower bound on time delay estimation with correlator performance. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 8, pages 875–878, 1983.

[213] B.M. Sadler and R.J. Kozick. A survey of time delay estimation performance bounds. In *IEEE Workshop on Sensor Array and Multichannel Processing*, pages 282–288, 2006.

[214] J.C. Chen, K. Yao, and R.E. Hudson. Acoustic source localization and beamforming: theory and practice. *EURASIP Journal on Applied Signal Processing*, pages 359–370, 2003.

[215] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 27(2):113–120, 1979.

[216] J. Merimaa, T. Lokki, T. Peltonen, and M. Karjalainen. Measurement, Analysis, and Visualization of Directional Room Responses. In *111th Audio Engineering Society Convention*, 2001. Paper number 5449.

[217] Norsonic. Nor848 acoustic camera. Technical report, Norsonic, Jan. 2011 (last accessed).

Appendix: Visualization examples

The visualization of early reflections is considered. Three visualization techniques are implemented and demonstrated for two reflections in a concert hall. Also, other visualization techniques for the reflections exist, such as acoustic holography [86,87]. However, it requires a line or a plane microphone array setup and differs therefore from the setup used in this example.

Overlaying the sound intensity vectors on top of a spectrogram

Possibly the first visualization of spatial room impulse responses is presented by Merimaa *et al.* [216]. The same approach is further developed and used in [1, 12]. The spatial room impulse response is divided into short time windows. For each pre-selected frequency band at the short time windows, the direction of arrival is estimated using sound intensity vectors. The vector is then plotted on top of a spectrogram consisting of these time-frequency "tiles". The azimuth and elevation of the direction of arrival are plotted separately.

An example of this visualization technique is shown in Fig. A. 1. The setup for the measurements is shown in Fig. A. 2 and the measured impulse responses in Fig. A. 3. The intensity vectors are calculated from two measurements. The first measurement with microphone array spacing of $d_{\text{spc}} = 25$ mm is used for frequencies above 1000 Hz and $d_{\text{spc}} = 100$ mm is used for frequencies below 1000 Hz.

Audio or acoustic camera

The visual and audio information is applied in several application areas [13,85,160,174]. In [13] the output of a spherical beamformer is overlayed on top of a 360 camera view of the enclosure. This is done in short time windows for an impulse response and the location of the reflections are then inspected visually. This idea is widely applied. “Acoustic cameras” (see e.g. [217]), take advantage on beamforming to enhance speech or to study, for example, noise sources.

An example of the same data as above is visualized with the acoustic camera principle in Fig. A. 3. This visualization technique lacks of frequency response information, but this information can be provided as an additional plot. The visualization technique is intuitive since the visual cues of the enclosure support the visualized reflection. One drawback is the lack of three-dimensionality in the visualization of the reflection location. It is obvious that this visualization technique requires interactive user interface to be practical.

Mapping the reflections to the geometrical model

The localized reflections can be traced back to the source via the reflective surfaces. This approach requires a priori information on the normals and the locations of the reflective surfaces, which can be extracted from the architectural models of the enclosures if available or estimated from the impulse responses. A ray-tracing approach is used inversely in Publication I to trace the reflections. The tracing is iterative. The ray is traced to the nearest surface at each iteration. Before each iteration, it is checked that the ray is long enough to reach the nearest surface. If it is not long enough, then the iteration is stopped, and ideally the ray should end in the position of the source.

An example of the same data as above is traced in Fig. A. 2. This visualization technique lacks frequency response information, but it could be easily added to the visualization.

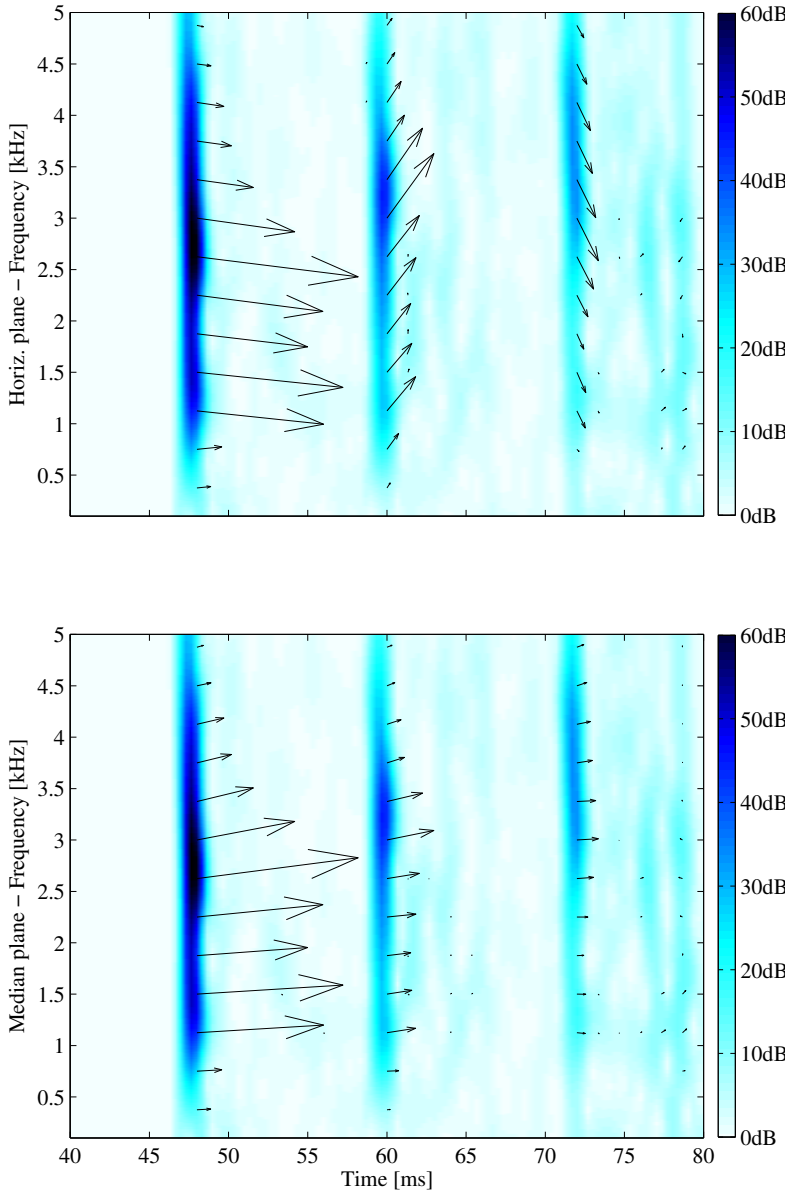


Figure A. 1. Visualization of reflections using the SIRR-framework.

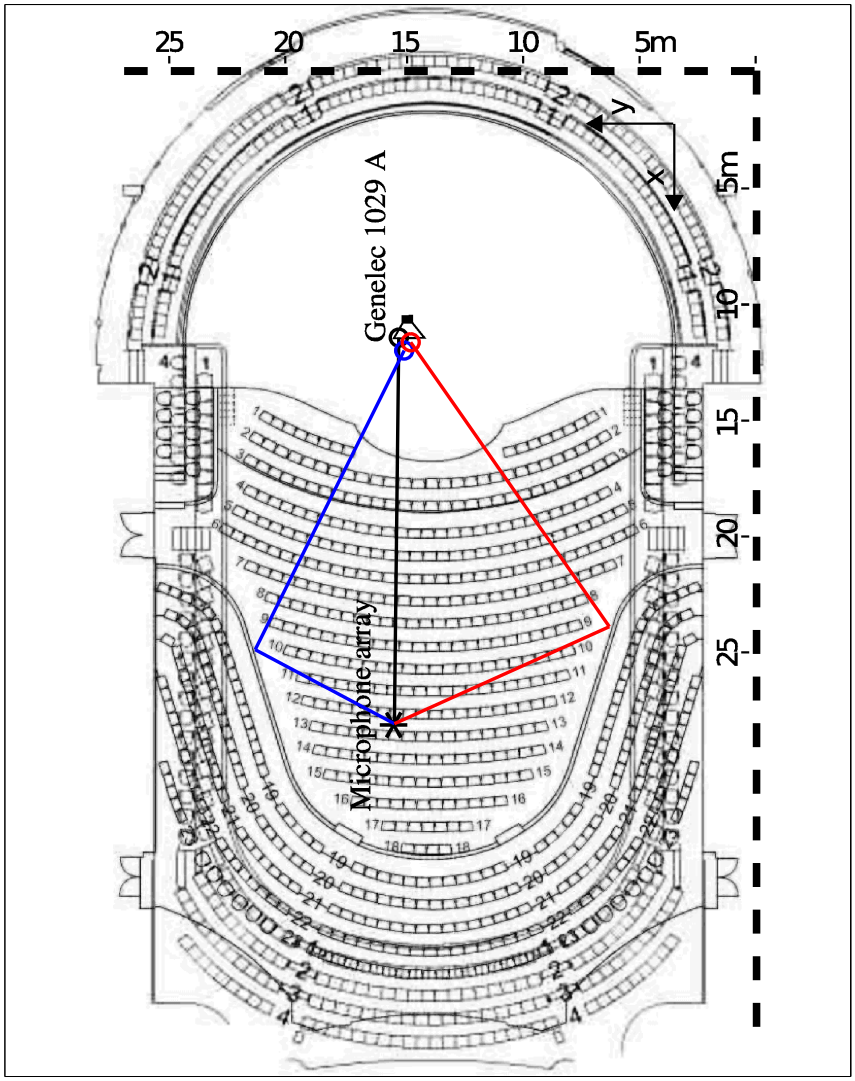
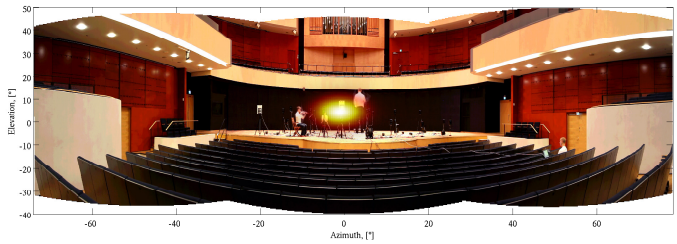
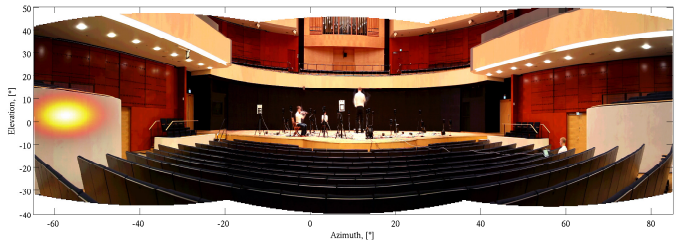


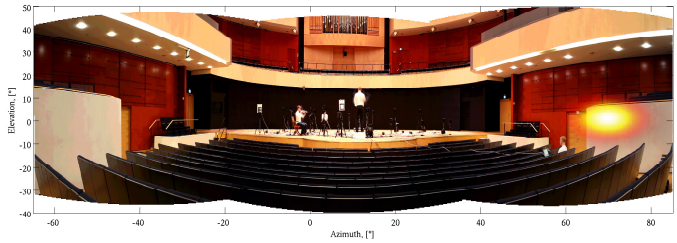
Figure A. 2. Visualization of reflections using the tracing of reflections principle.



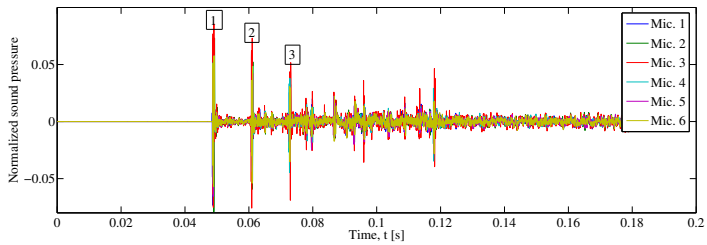
1 Direct sound, path length 16.7 m



2 Left wall reflection, path length 21.3 m



3 Right wall reflection, path length 25.2 m



Room impulse responses

Figure A. 3. Visualization using the audio camera. The steered responses are calculated using PL-TDOA. Also shown are the impulse responses for 6 microphones. The numbered boxes indicate events shown in the audio camera.

Errata

Publication IV

The character α is overloaded.

Aalto-DD 143/2011

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS

ISBN 978-952-60-4437-8
ISBN 978-952-60-4438-5 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Media Technology
www.aalto.fi