Technical University of Denmark

Filip Franek (s121303)

Localization of Incoherent Sound Sources by Three-dimensional Intensity Array

Master's Thesis, July 2014

DTU Electrical Engineering
Department of Electrical Engineering

KAIST

FILIP FRANEK (S121303)

Localization of Incoherent Sound Sources by Three-dimensional Intensity Array

Master's Thesis, July 2014

Supervisors:

Finn Agerkvist, Head of Group - Engineering Acoustics M.Sc., Associate professor

Efren Fernandez Grande, Associate Professor

Cheol-Ho Jeong, Associate Professor

DTU - Technical University of Denmark, Kgs. Lyngby - 2014

Localization of Incoherent Sound Sources by Three-dimensional Intensity Array

This report was prepared by:
Filip Franek (s121303)

Advisors:
Finn Agerkvist, Head of Group - Engineering Acoustics M.Sc., Associate professor
Efren Fernandez Grande, Associate Professor
Cheol-Ho Jeong, Associate Professor

DTU Electrical Engineering
Acoustic Technology and Hearing Systems at DTU Electrical Engineering
Technical University of Denmark
Elektrovej, Building 326
2800 Kgs. Lyngby
Denmark
Tel: +45 4525 3576

elektro@elektro.dtu.dk

Project period: February 2014- July 2014

ECTS: 30

Education: MSc

Field: Electrical Engineering

Remarks: This report is submitted as partial fulfillment of the requirements for graduation in the above education at the Technical University of Denmark.

Table of Contents

List of Figures

List of Tables

Abstract

The content of the thesis describes an approach to localize incoherent sound sources over 3D sound intensity array. The core of this work is a blind source separation combined with a 3D sound intensity measurement, what results in an angle detection of incoherent sound sources. The feasibility of the proposed method is firstly studied over simulations before the method is exampled to experimental measurements. Performance evaluation measures are suggested over statistical quantities and possible errors that arose or could have arisen are discussed.

Acknowledgement

This thesis was conducted as a part of dual-degree agreement between Technical University of Denmark (DTU) and Korean Advance Institute of Science and Technology (KAIST). Therefore, I would like to express my gratitude to Finn Agerkvist, Cheol-Ho Jeong and Olev Raven at DTU and to Jeong-Guon Ih at KAIST for creating this opportunity and allowing me to undertake the dual-degree program. The complete work was carried out at Mechanical Department at KAIST, so the main feedback was obtained from Jeong-Guon Ih at KAIST side, but online conferences and written reports were used to communicate with DTU side especially with Efren Fernando Grande and Finn Agerkvist. A great share of my acknowledgement goes to Jeong-Guon Ih for supporting my work on-site with new ideas, needed ongoing feedback and also outrageous trust which drove my progress throughout the project. Also the deepest appreciation goes to Efren Grande, who was always there for an online help and thorough feedback. I would also like to thank Finn Agerkvist who was an important person involved in many of the processes connected with my dual-degree program or the project feedback loop from DTU side.

I would not like to forget about my acoustic fellows in Denmark with who I was able to get a great deal of knowledge in acoustic field and signal processing through effective teamwork in form of endless discussions, hard and excited work on a project and very importantly, countless hours of playing table-football.

Being part of the acoustic lab at KAIST was a good opportunity to get feedback from fellow lab members to which I am grateful for introducing me laboratory instruments, Korean and Persian culture, and discussing needed topics.

Last but not least, I thank my family and close friends for keeping me fresh and giving me great deal of support and inspiration through my studies.

Filip Franek (s121303)

CHAPTER 1

Introduction

A representation of sound is believed to be fully described by the information of sound pressure, particle velocity and fluid density. On the basis of these instantaneous quantities, a complex behaviour of sound can be studied and it has a significant practical value as shown later. Sound intensity represents a flow of acoustic energy and this study takes an advantage of the flow character, which is expressed as a vector quantity and thus it gives information of its direction and strength of arriving sound. In general, sound intensity measurements give good knowledge about a directivity of a sound field. A sound field can be however composed of more sound events, which are impossible to distinguish with standard intensity methods without scanning the sound field manually or without combining an intensity computation with sound source separation methods. The idea behind this project is to tackle such sound scenario with a small fixed intensity device in order to identify directions of multiple sound sources with a high precision within a few degrees. We propose a method to solve this challenge by accommodating a combination of a 3D sound intensity computation and blind audio source separation (BASS).

1.1 Advent of sound intensity and motivation

Sound Intensity is a descriptive quantity of sound, which has been discovered relatively long time after fundamental acoustic quantities as sound pressure and particle velocity. The advent of sound intensity was approached through an investigation of sound energy, which was firstly done by Olson [22] in 1932, where an emphases were put on the energy flow of sound waves. Later in 1977, sound intensity was described in more detail and tested on laboratory experiments by Fahy [11] and Pavic [24]. This gave rise to a further development of sound intensity measurements in a few following decades. This time between 70s and 90s, as far as sound intensity is concerned, was particularly focused on analyses of sound's complex behaviour and error derivations of intensity measurements, which were predominantly done by Fahy[10] and Jacobsen[14]. In 1985, Rasmussen [27] presented a study of 3-dimensional intensity arrays, which was an important step for continuous work that has been inspired by finding a 3D sound intensity vector generated by a sound field in the 3D Euclidean space. Rasmussen introduced 3 configurations including a 6-microphone octahedron configuration, the most precise 3D array solution using microphone sensors, and 4-microphone variations as tetrahedron and perpendicular tetrahedron configurations, which pays a price for the microphone number reduction in form of a degraded intensity precision. Thereafter, more probe variations and error analyses have been presented. Usually the configurations have been inspired by Rasmussen's work and computational methods have been evolved. For example, it was work by Santos [28], Suzuki [30] and also Elko [9] who suggested mounting microphones on a rigid sphere, resulting in a reduction of a frame scattering and an increase of the intensity precision. A recent work by Ih [13] suggested a twisted configuration of array

variations what has a good impact on the intensity precision and left opportunities for further investigations.

The invention of a microflown (particle velocity sensor) in 1994 has enabled to use one vector sensor instead of 2 spaced microphones. This meant that the size of 3D intensity array was decreased to 3 microflowns and 1 omnidirectional microphone. The sensors have been even integrated on a single chip [36]. A disadvantage of such probe is the phase calibration which is more complicated compared to microphones. If such acoustic vector probe is combined with the Multiple Signal Classification algorithm (MUSIC), it was proved by Basten [1] that with n vector sensors, when one sensor is composed of 3 microflowns and 1 microphone, $(4n - 2)$ uncorrelated sources can be identified. Although this implementation lacks precision and the presented experimental results were off by up to 20°.

There has been another approach on how to measure sound intensity pioneered by Williams [34, 33]. In the study, many microphones sense the pressure around a virtual or hard sphere and spherical harmonic expansions are then used to obtain a vector intensity field. Its advantages are that it basically does not suffer from a low frequency limit and it does not obtain one intensity vector at a single point, but a vector field projection in a near-field. It is also capable of a multiple source detection. In spite of the advantages mentioned so far, this approach requires many microphones forming a spherical array with a radius of tens centimetres.

The previous work [13] has shown that precise detection of a single source can be measured by a small array that would fit into a virtual sphere with a radius of less than 2.5 cm. Other techniques for the 3D sound localization as Time Difference of Arrival (TDOA) requires high sampling frequency or large separation among microphones for the accurate source detection. Other methods as an acoustic beamforming, or an acoustic holography are used for sound source localization and these methods have a huge advantage in a visualizing of a complex sound field, but they require many microphones and also large-sized arrays for their reliable operation. A miniature acoustic vector sensor, composed of microflowns and a microphone, dramatically reduces the array dimension, but the precision for multiple sound source detection is still inaccurate. To summarize, the methods from previous work can detect multiple sound sources, but its size or angle precision is disadvantageous.

For the mentioned reasons, the development of a 3D intensity microphone array is advantageous in general for applications where a size, number of microphones, angle precision, and computational simplicity are essential factors. Also the omni-directionality of a 3D intensity array is a substantial aspect that makes 3D intensity techniques more flexible and detection of multiple sources would hence improve the performance considerably. This would result in intensity measurement without scanning a sound scape, or in a new approach for multiple sound source detection in situations where a small size, and angle precision of an acoustic array is a critical criterion. Possible application could be applied to robotic applications for human-machine interaction or for rescue operation where visual cues are not always sufficient.

1.2 Problem definition and research scope

A number of the previous works covered extensively the measurements errors, simulations, and performance analyses of the 3D intensity arrays. In most sound intensity applications, one sound intensity vector is sensed at a single position, thus if multiple sources are present, the localization of sound sources fails. Basten's approach [1] solves this problem for uncorrelated sources by using the MUSIC algorithm

with microflown technology, but only 2 sound sources can be identified with one vector array and the angle precision requires improvements. William's approach [34] is able to detect multiple sources while using tents of microphones embedded on a sphere with a radius of tens centimetres. Both aspects may be inconvenient in specific applications.

Angle detection of sound sources [13] was tested under anechoic conditions, thus detection in reverberant conditions remains to be a problem to be considered. There are numerous characters of sources to be considered. The sound sources are either dependent or independent in time or frequency domain. The most challenging case are dependent (coherent) sources in frequency and time domain. This separation has been approached only by beamforming so far. However independent sources in time and frequency domain were tackled by different approaches and it more plausible problem for sound sources separation, but the closeness of microphones and radiating sources affects the quality of sound source separation.

In this work, we propose multiple sound source detection to correctly localize dominant multiple sources using at least 4 microphones with precision within a few degrees. These sound sources can be dependent or independent on each other. This project focuses on solving scenario of the independent sound sources as it could give us a general idea on how to approach more complicated soundscapes.

1.3 Thesis structure

Chapter 3 denotes the basic definition of sound intensity and deals with approaches for the sound source separation. Chapter 3 is devoted to 1D, and 3D sound intensity, its derivations, errors, and measurement techniques. Chapter 2 focuses on general problem of sound source separation and we go deeper into the principles of BASS. chapter 4 performs simulation applied to the multiple sound source scenario and shows results under ideal conditions that are then compared in the experimental chapter 5, which applies the same techniques as has been introduced in chapter 4. Chapter 6 then discusses obtained results and Chapter 7 concludes the work and suggests future work.

CHAPTER 2

Separation of incoherent sound sources using Blind Source Separation

2.1 Overview

The substance of any sound source separation lies in the filtering of single sound sources out of mixture of sound sources. There are various techniques how to approach this task. There are 3 mostly used techniques in acoustic signal processing :

- Beamforming
- Computational Auditory Scene Analysis (CASA)
- Blind Source Separation (BSS).

Their applications vary greatly on sound event scenarios and sensor configuration requirements.The principle of beamforming for the sound visualization method can be also employed for the sound separation. Usually robust delay-and-sum beamformers (DSB) are used to find spatial information of sound pressure. By this operation, a directivity beam is initially steered in space to find sound sources positions. Subsequently, a desired signal can be obtained by adjusting the beam in order to point in the sound source direction while other sources located at different angles become partially suppressed depending on a beamformer directivity pattern. This separation technique for a small microphone array with a few microphones would cause very in accurate results in our frequency range due to the dimensions and a number of microphones, what all determines the precision of the method. Another technique CASA is based on a basic principles of the human hearing system, which are carried out over signal processing. CASA gives reasonable results in simple scenarios, but it does not surpass Blind Source Separation.

2.2 Blind source separation (BSS)

This technique is called blind, because it does not assume almost any prior knowledge about signals to be separated. In recent years, the term Blind Acoustic Source Separation (BASS) has been coined to note that an unknown environment is considered to be composed of an acoustic scene and it is based on spatial diversity of an acoustic sound field, i.e. it uses phase and amplitude differences between the sensors to recover compound signal. This separation approach always considers certain assumptions about input signals. Various techniques were developed over extensive research of BSS mostly over the past 3 decades. The basic three are:

- Principal Component Analysis (PCA)
- Independent Component Analysis (ICA)
- Independent Subspace Analysis (ISA)

The major difference between these 3 analysis lies in a statistical and temporal dependency of the sources that are to be separated. ICA and ISA assumes at most one audio source to be normally distributed and at the same time that all sources in the sound mixture are statistically independent, whereas PCA is based on the information about mutual correlation of the source signals and therefore the signals need to be dependent in time. ISA is generally a different approach of BSS which involves tasks similar to ICA together with a clustering of output components, but it does not have speed or performance advantages compared to ICA, thus it is not considered here.
An initial factors of BASS are how many sensors are employed and how many sound sources are to be separated. Thus a problem can be undetermined, determined or overdetermined. Initially, a single sensor can be assumed for a separation task, which is the most challenging example, since a character of a source which is to be analysed needs to be known and it is a strongly undetermined task [5]. Multiple sensors can be used for a smaller or equal number of sources, which can be effectively separated and it becomes the overdetermined or determined problem, respectively. Such system is also defined as Two Input Two Output (TITO) or Multiple Input Multiple Output (MIMO) depending on the number of inputs and outputs. For the sake of the sound intensity array technique which is using at least 2 sensors for 1D or 4 sensors in our case for 3D , only TITO or MIMO will be considered further on.

2.2.1 Instantaneous BSS

The basic principle of blind source separation will be explained on ICA. Let's assume, that the system comprises m microphone recordings[1], where discrete time index n spans up to N samples ($n = 1 : N$) with sampling interval T_s defined by sampling frequency f_s. The matrix of recordings is defined for $\mathbf{P} \in \mathbb{R}^{m \times N}$, as

$$\mathbf{P}_{m\times N} = \begin{bmatrix} p_1(1) & \cdots & p_1(N) \\ \vdots & \ddots & \vdots \\ p_m(1) & \cdots & p_m(N) \end{bmatrix}. \tag{2.1}$$

This is the only information we are provided with from a measurement of at least 2 microphones. Assume that sound source signals are mixed together at sensors with no delay between the sensors. This is the most simplistic case, where delays among microphones and reflections of an environment are not included and it is know as an instantaneous mixture model. A real scenario similar to these conditions in an audio field can be found e.g. in recording studios where single tracts (instruments, vocals, etc.) are processed in a mixing console. This case can be mathematically described as

$$\mathbf{P}_{m\times N} = \mathbf{A}_{m\times d}\mathbf{S}_{d\times N}. \tag{2.2}$$

Here the mixing matrix $\mathbf{A} \in \mathbb{R}^{m\times d}$ is an unknown process that mixes the source matrix $\mathbf{S}$ composed of d signals radiated from d sources, which result in the linear mixture P. The matrix of sources $\mathbf{S} \in \mathbb{R}^{d\times N}$ has

[1] In all chapters except Chap. 3, the symbols i and j describe i^{th} microphone and j^{th} source, respectively. Hence we write $i = 1, \cdots, m$ and $j = 1, \cdots, d$.

got the same structure as the mixed matrix **P** as long as the number of microphones m and sources s are equal.

$$\mathbf{S}_{d\times N} = \begin{bmatrix} s_1(n) & \cdots & s_1(N) \\ \vdots & \ddots & \vdots \\ s_d(n) & \cdots & s_d(N) \end{bmatrix} \tag{2.3}$$

Now we can obtain other relationships, supposing the basic assumption that sources are independent is fulfilled. In other words, probabilities of all sources needs to be independent, thus the following must hold: $p(s) = p(s_1, s_2, ..., s_d) = p(s_1)p(s2)...p(s_d)$. Then, the unmixing matrix **W** can be introduced as derived below. From the mixing Eq. 2.2, we can write formula for matrix **S**,

$$\mathbf{S}_{d\times N} = \mathbf{A}^{-1}_{m\times d}\mathbf{P}_{m\times N} = \mathbf{W}_{d\times m}\mathbf{P}_{m\times N}. \tag{2.4}$$

It is seen that the unmixing matrix $\mathbf{W} = \mathbf{A}^{-1}$, so if this equality is achieved in real scenario, original sources can be recovered into some extent, in ideal case as in Eq. 2.4. The separated signals are never exact copies of the original signals, but novel algorithms asymptotically approaches the original signals, when large data samples are processed, although this slows down a computational speed. In practise, we try to obtain the matrix **V** which is supposed to be a good approximate of the source matrix **S** as described in Eq. 2.5,

$$\mathbf{V}_{m\times N} = \mathbf{W}_{d\times m}\mathbf{P}_{m\times N} = \mathbf{W}_{d\times m}\mathbf{A}_{m\times d}\mathbf{S}_{m\times N} \approx \mathbf{S}_{m\times N}. \tag{2.5}$$

This procedure is depicted in Fig. 2.1. The system is a symbolic representation of a real system and it is divided into known, unknown and estimated sections. The schema can be used to represent an arbitrary number of microphones, sources, and sample lengths.

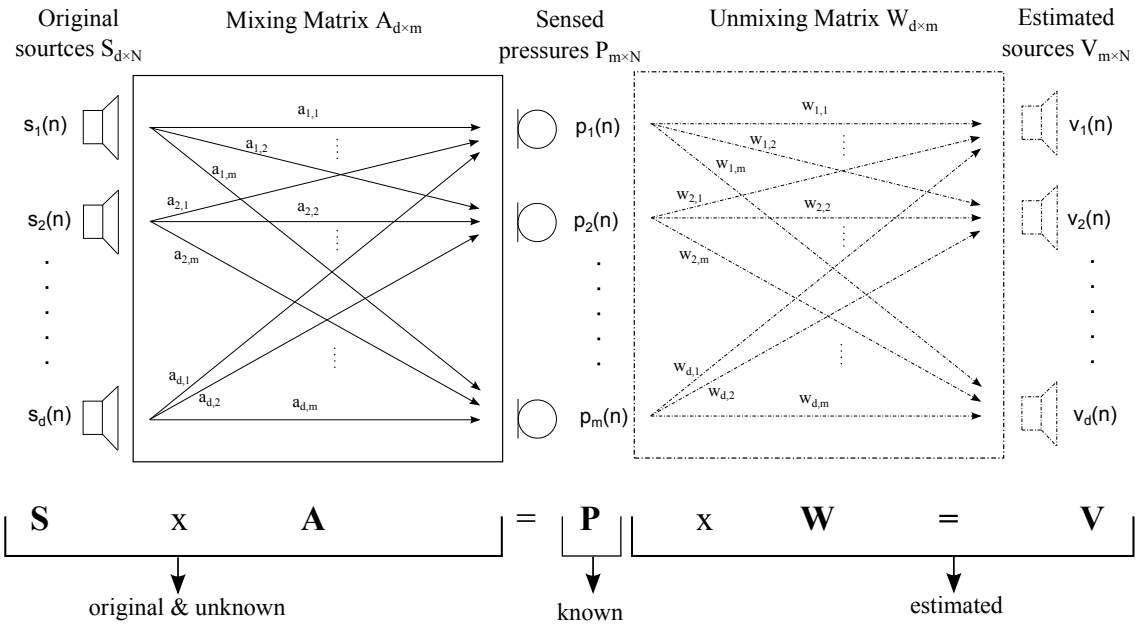

Figure 2.1: Linear mixing ICA model.

To give an example, if 4 microphones and 3 sources are present in the system, the mixing matrix **A** would consist of 3 rows and 4 columns, where each element corresponds to a scalar value $a_{i,j}$. Similarly, the unmixing matrix would have 3×4 elements.

As derived, there have to be some techniques tackling the problem of computing the unmixing matrix estimator **W**. The developed ICA methods arose from many years of intensive, and still ongoing research and they can be grouped according to different assumptions that the methods use. All these various principles are generalized into 3 classes based on non-Gaussianity, nonstationarity, or spectral diversity of original sound source signals. Some effective algorithms divided into these classes are briefly mentioned in the list below [31]:

1. non-Gaussianity
    - EFICA - Efficient Fast ICA [20]
    - JADE - Joint Approximate Diagonalization of Eigen-matrices [4]
2. nonstationarity
    - BGSEP - Block Gaussian Separation [32]
    - BGL - Block Gaussian Likelihood [25]
3. spectral diversity
    - WASOBI - Weight-Adjusted Second Order Blind Identification [8]

Other methods that are combinations of the mentioned ones are:

- Block EFICA - Block Efficient Fast ICA. Combines piecewise stationarity with non-Gaussianity [16]
- BARBI - Block-AutoRegressive Blind Identification. Combines piecewise stationarity with spectral diversity [32]
- M-COMBI - Multiple decision-driven combinations of EFICA and WASOBI. [12]

All these methods considered instantaneous mixtures as an input. Most of these algorithms are available on request from the researchers. This project utilize a few of the proposed algorithms in a final version of BASS.

2.2.2 Convolutive BSS

In a real room environment, sound waves generated by sound sources propagate through the air medium with relatively slow speed of sound $c_o = 343\,\mathrm{m/s}$ and impinges on microphones, which record a direct contribution of each sound source, as well as its delayed and filtered version due to an transfer function of a given environment between a particular source and microphone. Thus the separation problem is no longer considered the instantaneous mixture, but a convolutive mixture that yields modified problem definition rewritten in Eq. 2.6. The formula demonstrates a contribution of original sources convolved with the impulse response of an environment. The mixing matrix **A** is not any more composed of scalars, but its elements correspond to filter coefficients $a_{i,j}$, which can be interpreted as impulse responses or transfer functions between i^{th} microphone and j^{th} source. Therefore the convolutive mixing matrix **A** is added by a 3^{rd} dimension resulting in $\mathbb{R}^{m\times d\times M_{i,j}}$.

$$p_i(n) = \sum_{j=1}^{d}(a_{i,j} * s_j) = \sum_{j=1}^{d}\sum_{\tau=0}^{M_{i,j}-1}(a_{i,j}(\tau)s_j(n-\tau)) \quad , \qquad i = 1,\ldots,m;\;\; j = 1,\ldots,d \tag{2.6}$$

The convolutive problem is stated for m microphone mixtures $\mathbf{P}_{m\times N}$, from which original signals $\mathbf{S}_{d\times m}$ generated by d sources are desired to be recovered. The observed mixture $\mathbf{P}$ is formed by a convolution of an audio signal from j^{th} speaker and a transmission path between i^{th} microphone and j^{th} source. The convolution can be defined as an integral of 2 variables which form a product, where the variables are reversed and time shifted. Once we obtain the expression on the right side of Eq. 2.6, we can derive the unknown single sources s_j from estimating the impulse responses $a_{i,j}$. There have been basically 2 techniques proposed for the separation of convolutive mixtures which are:

- Time-domain BSS
- Frequency-time-domain BSS

The time-domain procedure approaches the problem directly in the time domain by a deconvolution, i.e. Blind Deconvolution. The filter coefficients can have either finite impulse response (FIR) or infinite impulse response (IIR). The convolution sum formula in Eq. 2.6 is expressed in a discrete form which is further represented by a discrete FIR filter interpretation with finite length $M_{i,j}$ of the mixing filter $\mathbf{A}$. This FIR filter solution has been chosen in this work, because IIR filters are unstable. This is because they have a feedback loop, whereas FIR filters have no feedback loop. On the other side, using IIR filters, the accuracy could be increased and there are methods of the convolutive BSS using IIR method [35]. FIR filters usually have linear-phase, meaning that there is no phase distortion, what is an important property for our ultimate purpose of deriving sound intensity. The right-sided variables of Eq. 2.6 are unknown and even though we are interested in audio signals s_j, they usually have an unknown temporal structure, thus coefficients of FIR filter $a_{i,j}$ are to be estimated since it finds impulse response of the given environment, which representation can be better anticipated. FIR impulse responses coefficients are represented in the time domain as ,

$$a_{i,j}(n) = \sum_{\tau=0}^{M_{i,j}-1} b(\tau)\cdot\delta(n-\tau) = \begin{cases} b(\tau) & {\scriptstyle 0\leq\tau\leq M_{i,j}} \\ 0 & {\scriptstyle \text{otherwise.}} \end{cases} \tag{2.7}$$

Since the FIR filter is considered (Eq. 2.6), the coefficients are non-zero only on the finite length for $\tau = 0, 1, \ldots, M_{i,j} - 1$. The impulse response is defined as a sum of impulse responses δ weighted by values $b(\tau)$ and when Eq. 2.7 is combined with a sum of original sources s_j, Eq. 2.6 is obtained, what demonstrates that the original sources are convolved with the filter response of an environment. Now, in order to estimate the original sources, we want to find the unmixing filter $\mathbf{W}$, which deconvolves the Eq. 2.6. This deconvolutive operation is mathematically described in the formula

$$s_j(n) = \sum_{i=1}^{m}\sum_{\tau=0}^{L_{i,j}-1}(w_{j,i}(\tau)x_i(n-\tau)), \tag{2.8}$$

where $w_{j,i}(\tau)$ represents the coefficients of an inverse filter which inverts the mixing process of transmission paths between sources and receivers, what eventually leads to retrieving the estimates of original signals. The computational load at time-domain techniques depends on the length of FIR filter $L_{i,j}$ and it usually increases non-linearly with L^3. Although it is possible to apply method [19] that composes matrix $\mathbf{P}^L$ using filters instead of delays and longer separating filters can be applied without the need of increasing L.

There is a way how to get around the convolution problem in time domain by transforming the recorded signal into the frequency domain where the convolution operator becomes an ordinary multiplication operator that is less computationally demanding. Hence the Fourier transform can be applied to Eq. 2.6 yielding,

$$\mathbf{P}_i(\omega) = \sum_{j=1}^{d} (\mathbf{A}_{i,j}(\omega)\mathbf{S}_j(\omega)), \tag{2.9}$$

From this assumption, it follows that the separation now reminds the instantaneous mixture problem, but the frequency transform returns complex data and therefore a complex-domain separation technique needs to be developed. However, transforming time into frequency domain requires to be processed in the way that the time information is still preserved in frequency representation, hence a Short Time Fourier Transform (STFT) is employed because it can compute spectral components of short time segments using a short time window that is handled to the Fourier transform. The problem then comply with adjusted definition

$$\mathbf{P}_i(\omega, t) = \sum_{j=1}^{d} (\mathbf{A}_{i,j}(\omega)\mathbf{S}_j(\omega, t)). \tag{2.10}$$

Such preprocessed signal can be then treated as an instantaneous mixture problem and after finding components of a required unmixing matrix, frequency-time data are transformed back to the time sequence by a Inverse Short Time Fourier Transform (ISTFT). Although, this step brings along problems as permutation and scaling ambiguity. These issues have been tackled a lot in the literature and it is possible to achieve good results with proposed solutions. This method can efficiently calculates long filters introduced by long reverberation times, although it requires a long sequence of recorded data that makes this strength less significant.

In recent years, both methods have been investigated in great detail and currently the time-domain methods perform quite well, because it eliminates some drawback of frequency domain algorithms. The scaling ambiguity is still present though, but the permutation ambiguity is inherently not a problem of the time-domain methods. It was therefore decided to focus on time-domain algorithms later on.

2.2.3 Implementation of time-domain blind audio source separation T-ABCD

The abbreviation T-ABCD stands for Time-Domain Blind Separation of Audio Sources on the Basis of a Complete ICA Decomposition of an Observation Space. It is recently developed technique for BASS by Koldovsky et al. [18]. This method is used for separating sound sources in our project. This particular technique was chosen, for the mentioned advantages compared to the frequency domain algorithms, but all of developed algorithms so far have certain pros and cons, so there is no ideal or very supreme technique available yet. The algorithm was provided in a form of MATLAB code together with permission by Koldovsky to apply the code. The procedure of the algorithm structure can be briefly described in Fig. 2.2. We consider again source matrix $\mathbf{S}_{d\times N}$ composed of d original sources. In the first step of the T-ABCD algorithm, a matrix of signal mixtures $\mathbf{P}^{L}_{mL\times N}$ is formed by introducing time-lagged copies of each microphone recording[2] as was first done in [3]. The time delay between the copies is specified by the length of separating filter L. For this reason, the longer the impulse response of a room, the separation of signals is computationally more demanding. A convenient value for the algorithm is $L_{i,j} < 40$ in order to keep computational burden low. The separation performance is good even with such short filters, but it is not suitable for very long reverberation times. Additionally, the length of a recorded signal at sampling frequency $f_s = 8\,\mathrm{kHz}$ can

[2]The superscript L is used here for every variable having its dimensions transformed by the new observation subspace $\mathbb{R}^{mL\times N}$

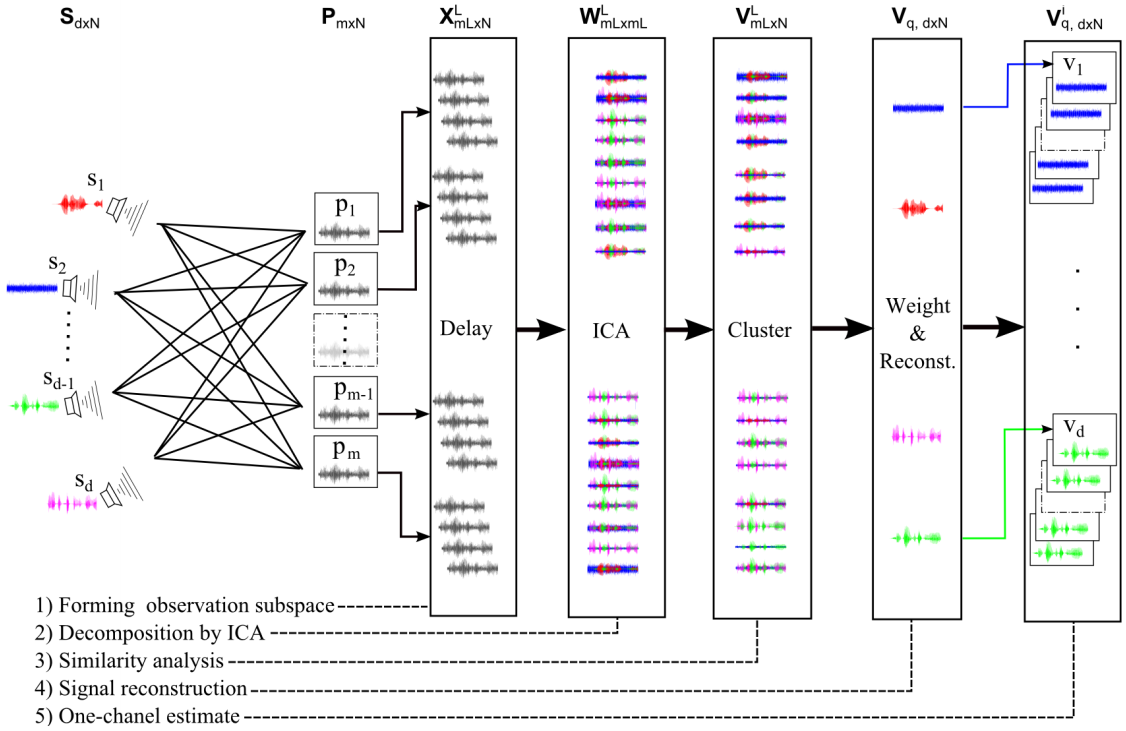

Figure 2.2: Separation procedure of T-ABCD.

be as short as 1000s samples. The matrix P^L is formed in a similar manner as demonstrated in the matrix notation below,

$$\mathbf{P}^L_{mL\times(N_2-N_1+1)} = \begin{bmatrix} p_1(N_1) & \cdots\cdots & p_1(N_2) \\ p_1(N_1-1) & \cdots\cdots & p_1(N_2-1) \\ \vdots & \ddots & \vdots \\ p_1(N_1-L+1) & \cdots\cdots & p_1(N_2-L+1) \\ p_2(N_1) & \cdots\cdots & p_2(N_2) \\ p_2(N_1-1) & \cdots\cdots & p_2(N_2-1) \\ \vdots & \ddots & \vdots \\ p_2(N_1-L+1) & \cdots\cdots & p_2(N_2-L+1) \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ p_m(N_1-L+1) & \cdots\cdots & p_m(N_2-L+1). \end{bmatrix}. \tag{2.11}$$

The time indexes N_1 and N_2 satisfy the condition $1 \leq N_1 < N_2 \leq N$. To give an example of the dimensions on a real example, consider the scenario where 4 microphones record with $f_s = 8\,\mathrm{kHz}$ for $1\,\mathrm{s}$, $N = 8000$, $N_1 = 1$, $N_2 = N$, $L_{i,j} = 20$. The dimensions of our observation space become $\mathbb{R}^{80\times 8000}$.

The heart of T-ABCD is applying a suitable ICA method. We will consider BGSEP for its fast convergence (due to dealing only with 2nd-order statistics) and capability of handling longer filters while keeping computational cost down. Employing ICA on the subspace $\mathbf{P}^L$ yields unmixing matrix $\mathbf{W}^L_{mL\times mL}$, which elements can be accessed by l^{th} and k^{th} components noted as $\mathbf{W}^L_{l,k}$, where both $l, k = (1, \ldots, mL)$. Taking ICA of the time-lagged copies is doing the troublesome deconvolution. Mutually independent outputs,

which cab be understood as MISO FIR filters of length L, are obtained by multiplying the matrix $\mathbf{W}^L$ with observation space $\mathbf{P}^L$ as already derived in Eq. 2.5.

$$\mathbf{V}^L_{mL\times N} = \mathbf{W}_{mL\times mL}\mathbf{P}_{mL\times N} \tag{2.12}$$

The row vectors of $\mathbf{W}^L$ are assumed to be independent components already and they are denoted as $\mathbf{v}^L_{L\times N_2}$. We can look at these new estimates as on arbitrary filtered versions of the original signals due to the indeterminacy of ICA. It means that the separated signals are not well separated yet, but because we have many versions of these not ideally separated sources, we can further improve the estimates.

Searching for better results, we firstly imply a clustering of a large number of separated components. They are sorted out according to similarity measures in order to compose a groups of signals most similar to each original source. We will call them clusters. The clustering in principle means to identify groups according to a certain criteria. The criterion in our case is the similarity measure. The sorting methods used in T-ABCD technique so far are hierarchical or centroid-based clusterings. Their specific subgroups here are defined as an agglomerative hierarchical and fuzzy k-means clustering and they can be further specified as hard of soft clustering, respectively. Both methods evaluate the matrix $\mathbf{V}^L$ and returns a matrix of zeroes and ones. The main difference between them is that the hard clustering do not consider any relation ship among the sorted data which are either similar (zeroes) or dissimilar (ones), whereas the soft clustering compute additionally a partition matrix with membership levels of all estimates. This is a very important point for defining on how many clusters the algorithm detects, meaning how many estimated sources will be found regardless of user's setting. It can be either defined or estimated based on well chosen clustering parameters. The resulting matrix $\mathbf{D}_{k,l}$ is computed over generalized cross-correlation with phase Transform (GCC-PHAT) as proposed in [21]. The GCC-PHAT is defined by

$$\mathbf{G}_{k,l}(\omega) = \frac{\mathcal{F}\{\mathbf{V}^L_k\}\cdot\mathcal{F}\{\mathbf{V}^L_l\}^*}{|\mathcal{F}\{\mathbf{V}^L_k\}|\cdot|\mathcal{F}\{\mathbf{V}^L_l|\}}. \tag{2.13}$$

If we take the inverse Fourier transform of $\mathbf{G}_{mL\times mL}(\omega)$, we arrive to its time domain representation $\mathbf{g}_{mL\times mL}$. The components $g_{k,l}(n)$ are equal to delayed impulse responses as long as they correspond exactly to the same source. The delay is not greater then L and thus the similarity matrix can be calculated over a sum of $g_{k,l}(n)$ as

$$\mathbf{D}_{k,l} = \sum_{n=-L}^{L}|g_{k,l}(n)|, \quad k,l = 1,\ldots,mL, \quad k\neq l. \tag{2.14}$$

The similarity between components $k = l$ is meaningless, hence the values at these components are set to zeroes. The matrix $\mathbf{D}$ is used as an input for the clustering algorithm and also for a reconstruction of the estimated signals. Once the clusters are found, we introduce new variable q which stands for cluster number and it holds that $q = 1,\ldots,b$, where b is number of found clusters. Therefore $q = d$ in an ideal case.

The reconstruction of the matrix $\mathbf{P}^L$ is conducted to transform the formed clusters into individual responses between original sound sources and microphones. Weightings Λ^q_{mL} proposed in [17] are decided according to the clustering, which can either be hard or fuzzy. Simplified version of such weighting is shown below,

$$\Lambda^q_l = \begin{cases} 1 & l\in K_q \\ 0 & \text{otherwise.}\end{cases}, \qquad \Lambda^q_l = \left(\frac{\sum_{k\in K_q, l\neq k} D_{l,k}}{\sum_{k\notin K_q l\neq k} D_{l,k}}\right)^{\alpha} \tag{2.15}$$

The left part of Eq. 2.15 shows the hard, also called binary weighting. It composes the weighting vector Λ_l^q only by filling ones to indices K_q, which denote components of q^{th} cluster. The expression on the right of Eq. 2.15 assigns a value to Λ_l^q also on the basis of similarity of each independent component calculated from Eq. 2.14, but the denominator in Eq. 2.15 takes into account similarities of l^{th} component with a component from different clusters. The α parameter correspond to so-called 'hardness' of the weighting. Increasing α leads to higher interference suppression between separating signals, but it causes a spectral distortion. α close to zero on the other side means no separation. To apply the weighting vector to our data, we write formula

$$v_q = (\mathbf{W}^L)^{-1} \operatorname{diag}[\Lambda_1^q, \ldots, \Lambda_{mL}^q]\mathbf{V}^L = (\mathbf{W}^L)^{-1} \operatorname{diag}[\Lambda_1^q, \ldots, \Lambda_{mL}^q]\mathbf{W}^L\mathbf{P}^L, \tag{2.16}$$

which reconstruct the q^{th} cluster corresponding to the j^{th} delayed original sound source playing solo. In an ideal case the cluster is composed of delayed versions of original signal on all microphones. The structure of v_q is similar to $\mathbf{P}^L$ and therefore it is possible to appropriately sum the signals to get a response of a single sound source observed by a microphone. Hereby, the estimated response of q^{th} cluster at i^{th} microphone is

$$v_q^i = \frac{1}{L}\sum_{p=1}^{L} v_{(q-1)L+p}^i(n+p-1), \tag{2.17}$$

where $p = M - d + 1$.

Finally, a beamformer method is used to find one-channel estimates of the single separated sources. This calculates correlation of microphone responses of mixed sources and then it time shifts estimated signals accordingly to each microphone position. In this order, a phase information of the separated signals is preserved, which is crucial input for sound intensity calculation.

CHAPTER 3

Sound intensity theory

3.1 1D sound intensity

Sound intensity is a vector quantity describing the instantaneous flow of sound energy per unit area $(\mathrm{W/m^2})$. In general, sound intensity I is a product of sound pressure p and particle velocity u.[1] The instantaneous or time-averaged form can be obtained by taking its corresponding product of pressure and particle velocity[2] following Euler's equation of motion,

$$p(t) = |p(t)|e^{j(\omega t - k\vec{r})}, \tag{3.1a}$$

$$\nabla p(t) = -\rho\frac{\partial \vec{u}}{\partial t} \Leftrightarrow \vec{u}(t) = -\frac{1}{\rho}\int \nabla p(t)\partial t = -\frac{1}{j\omega\rho}\nabla p(t) = \frac{j}{\omega\rho}\frac{\partial p(t)}{\partial r}, \tag{3.1b}$$

$$\vec{I}(t) = p(t)\vec{u}(t) = p(t)\frac{j}{\omega\rho}\nabla p(t). \tag{3.1c}$$

An interesting property of the active sound intensity is that its average value corresponds to the real part of pressure and conjugate velocity. As seen below, Eq. 3.1c represents the instantaneous sound intensity and Eq. 3.2 is the average intensity that is in literature often noted just as sound intensity,

$$\vec{I}(t) = \overline{p(t)\vec{u}(t)} = \lim_{T\to\infty}\frac{1}{T}\int_0^T p(t)\vec{u}(t)dt = \frac{1}{2}Re\{p(t)\vec{u}(t)^*\}. \tag{3.2}$$

This expression is already sufficient for some intensity measurement techniques in the time domain, but we can build upon the current formulas and follow Fahy's derivation in the frequency domain [10]. The time-averaged sound intensity can be defined based on statistical properties introducing a cross-correlation quantity of pressure and particle velocity $R_{pu}(\tau)$. This helps us to arrive to a frequency spectrum expression by calculating the cross-spectral density between pressure and particle velocity $S_{pu}(\omega)$, because cross-correlation and cross-spectral density functions are Fourier pairs as indicated in the following formulas:

$$R_{pu}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_0^T p(t)\vec{u}(t+\tau)dt\ , \quad S_{pu}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{pu}(\tau)e^{-j\omega\tau}d\tau. \tag{3.3}$$

[1] All the acoustic quantities in this chapter (pressure, particle velocity, intensity) are functions of position, thus the dependency on spacial variable r for 1D or $\vec{r} = (x, y, z)$ for 3D is omitted for the sake of readability unless otherwise specified. The correct notation should be e.g. $p(r, t)$ for pressure in the time domain or $P(r, \omega)$ in the frequency domain.

[2] Eq. 3.1b contains nabla operator ∇ which is a vector differential operator $\frac{\partial}{\partial r}$ and it can be approximated as the gradient of one quantity at 2 points in space $\frac{\Delta}{\Delta r}$. Other quantities follows common notation, where t is a variable of time, j in this chapter is imaginary number $(\sqrt{-1})$, ω is angular velocity, k is spacial frequency (wavenumber), λ is wavelength, and ρ is density

Supposing the time constant $\tau = 0$, Eq. 3.3 yields an important relationship between sound intensity in the time domain and the cross-spectral density function in the frequency domain,

$$\vec{\bar{I}}(t) = R_{pu}(0) = \int_{-\infty}^{\infty} S_{pu}(\omega) d\omega \tag{3.4}$$

We eager to arrive to the sound intensity expression in the frequency domain. The cross-spectral density satisfies such properties that its real part is an even function, $Re\{S_{pu}(\omega)\} = Re\{S_{pu}(-\omega)\}$, and its the imaginary part is an odd function, $Im\{S_{pu}(\omega)\} = -Im\{S_{pu}(-\omega)\}$. It is observed that the cross-spectral density has values for positive and negative frequencies, but because only positive frequencies have an appropriate meaning for us, it is useful to introduce a single-sided cross-spectral density, which is non-zero only for positive frequencies thus $G_{pu}(\omega) = 2S_{pu}(\omega)$. With this knowledge, we can rewrite Eq. 3.4 into the frequency domain and eventually end up at a simplified real part one-sided cross-spectral density G between pressure and particle velocity as derived in as follows

$$\vec{\bar{I}}(\omega) = S_{pu}(\omega) = 2Re\{S_{pu}(\omega)\} = Re\{G_{pu}(\omega)\}. \tag{3.5}$$

From the theory of random data [2] the one-sided spectral density $G_{pu}(\omega)$ can be represented as the real and the imaginary spectral part (coincident and quadrature spectral density, C and Q respectively) or by a cross product. The product in Eq. 3.6 is expressed as a limit function of the Fourier transform[3] of sound pressure p and particle velocity u,

$$G_{pu}(\omega) = C_{pu}(\omega) + jQ_{pu}(\omega) = \lim_{T\to\infty} \frac{2}{T}[P(\omega)^* \vec{U}(\omega)]. \tag{3.6}$$

In this work, we deal with microphone sensors. Therefore the pressure at the centre of one dimensional intensity probe composed of 2 microphones is estimated over an arithmetic mean and particle velocity is obtained over finite difference approximation to the sound pressure gradient,

$$p(t) = \frac{p_1(t) + p_2(t)}{2}, \tag{3.7a}$$

$$\vec{u}(t) = \frac{j}{\omega\rho}\frac{\partial p(t)}{\partial r} = \frac{j}{\omega\rho}\frac{p_2(t) - p_1(t)}{\Delta r}. \tag{3.7b}$$

Δr is a distance between the 2 microphones and recordings at microphone positions are p_1 and p_2. The Fourier transform of sound pressure and particle velocity is obtained via the estimates in set of Eq. 3.7.

$$\mathcal{F}\{p(t)\} = \int_{-\infty}^{\infty} |p(t)| e^{j\omega t} dt = \frac{P_1(\omega) + P_2(\omega)}{2} = P(\omega) \tag{3.8a}$$

$$\mathcal{F}\{\vec{u}(t)\} = \int_{-\infty}^{\infty} \frac{j}{\omega\rho}\frac{\partial p(t)}{\partial r} e^{j\omega t} dt = j\frac{P_2(\omega) - P_1(\omega)}{\rho\omega\Delta r} = \vec{U}(\omega) \tag{3.8b}$$

$$\tag{3.8c}$$

These spectral components can be inserted into Eq. 3.6. Eq. 3.9 is then derived and arranged to get one-sided cross-spectral densities between the two pressure quantities p_1 and p_2.

$$\begin{aligned} G_{pu}(\omega) &= -j\frac{1}{2\rho\omega\Delta r}\lim_{T\to\infty}\frac{2}{T}\{P_1^*(\omega)P_1(\omega) - P_2^*(\omega)P_2(\omega) + P_1(\omega)P_2^*(\omega) - P_1^*(\omega)P_2(\omega)\} \\ &= -j\frac{1}{2\rho\omega\Delta r}\{G_{p_1p_1}(\omega) - G_{p_2p_2}\omega) + 2G_{p_2p_1}(\omega)\} \end{aligned} \tag{3.9}$$

[3] Fourier transform is noted in equations as FFT and physical quantities representing Fourier transform are capitalized

Note that the last expression of Eq. 3.9 holds auto-spectral density functions $G_{p_1p_1}(\omega)$, and $G_{p_2p_2}(\omega)$, which are by its derivation always positive. It was proved in Eq. 3.5 that the averaged sound intensity in the frequency domain is equal to the real part of the one-sided cross-spectral density function. If we take the real part of the final result in Eq. 3.9, auto-spectral components are not passed through the real part, because they become imaginary,

$$\vec{\bar{I}}(\omega) = Re\{G_{pu}(\omega)\} = Im\left\{-\frac{G_{p_2p_1}(\omega)}{2\rho\omega\Delta r}\right\} \tag{3.10}$$

This is the final term of the time-averaged sound intensity expressed in the frequency domain for a 1-dimensional sound intensity probe and it is handful derivation also for more dimensional cases. We can see that the intensity expression in the frequency domain is now possible to solve by knowing only two pressure recording and taking their one-sided cross-spectral density.

3.2 3D sound intensity

Sound sources are to be estimated in 3-dimensions. In order to find the 3D sound intensity, 3 orthogonal vectors can be computed. This works optimally if an octahedron array configuration is assembled. Our interest is the tetrahedron array where single intensity vectors in between all microphone pairs do not have the same origin as they do in the octahedron configuration, a Taylor expansion is used to effectively approximate the resulting 3D intensity vector[23]. Further on, the tetrahedron configuration as depicted in Fig. 3.1 will be considered, because it employs smaller number of microphones than the octahedron configuration.

The zero and first order Taylor expansion is written in Eq. 3.11a for the 1D and then in Eq. 3.11c for the 3D sound field. The second and higher order Taylor expansion can be also derived, but the expression becomes non-linear and that could have an adverse impact on approximation results.

$$p_i(r,t) \approx p(r,t) + r\nabla p(r,t) \approx p(r,t) + x_i\frac{\partial p(r,t)}{\partial x} + y_i\frac{\partial p(r,t)}{\partial y} + z_i\frac{\partial p(r,t)}{\partial z} \tag{3.11a}$$

$$\begin{bmatrix} p_1(r,t) \\ p_2(r,t) \\ \vdots \\ p_m(r,t) \end{bmatrix} \approx \begin{bmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & y_m & z_m \end{bmatrix} \cdot \begin{bmatrix} p(t) \\ \frac{\partial p(t)}{\partial x} \\ \frac{\partial p(t)}{\partial y} \\ \frac{\partial p(t)}{\partial z} \end{bmatrix} \approx \begin{bmatrix} p(r,t) & x_1\frac{\partial p(r,t)}{\partial x} & y_1\frac{\partial p(r,t)}{\partial y} & z_1\frac{\partial p(r,t)}{\partial z} \\ p(r,t) & x_2\frac{\partial p(r,t)}{\partial x} & y_2\frac{\partial p(r,t)}{\partial y} & z_2\frac{\partial p(r,t)}{\partial z} \\ \vdots & \vdots & \vdots & \vdots \\ p(r,t) & x_m\frac{\partial p(r,t)}{\partial x} & y_m\frac{\partial p(r,t)}{\partial y} & z_m\frac{\partial p(r,t)}{\partial z} \end{bmatrix} \tag{3.11b}$$

$$\mathbf{P}_{m\times 1}(r,t) \approx \mathbf{M}_{m\times 4}(r,t) \quad \mathbf{D}_{4\times 1} \tag{3.11c}$$

The pressure $p_i(r,t)$ represents the recording of m microphones at the microphone locations $r_i = (x_i, y_i, z_i)$, where $i = 1, \ldots, m$. The measured pressures are approximated by pressure $p(r,t)$ at the centroid O_c and by pressure gradients for all Cartesian coordinates x, y, z also at the centroid of the array. The matrices $\mathbf{P}_i(r,t), \mathbf{M}(r,t)$ and $\mathbf{D}$ rewrite Eq. 3.11b into the matrix form. The matrix $\mathbf{P}$ therefore represents the measured pressures at the positions represented by the matrix $\mathbf{M}$. The matrix $\mathbf{D}$ contains estimated pressure at the centroid O_c (located ad the origin of Cartesian coordinate system) and the pressure gradients

in 3 directions of Cartesian coordinate system (x,y,z). The matrix **D** is the only unknown quantity in the formula above and it is calculated by a matrix inverse operation $\mathbf{D}_{4\times1} = \mathbf{M}^{-1}_{m\times4}\mathbf{P}_{m\times1}$.

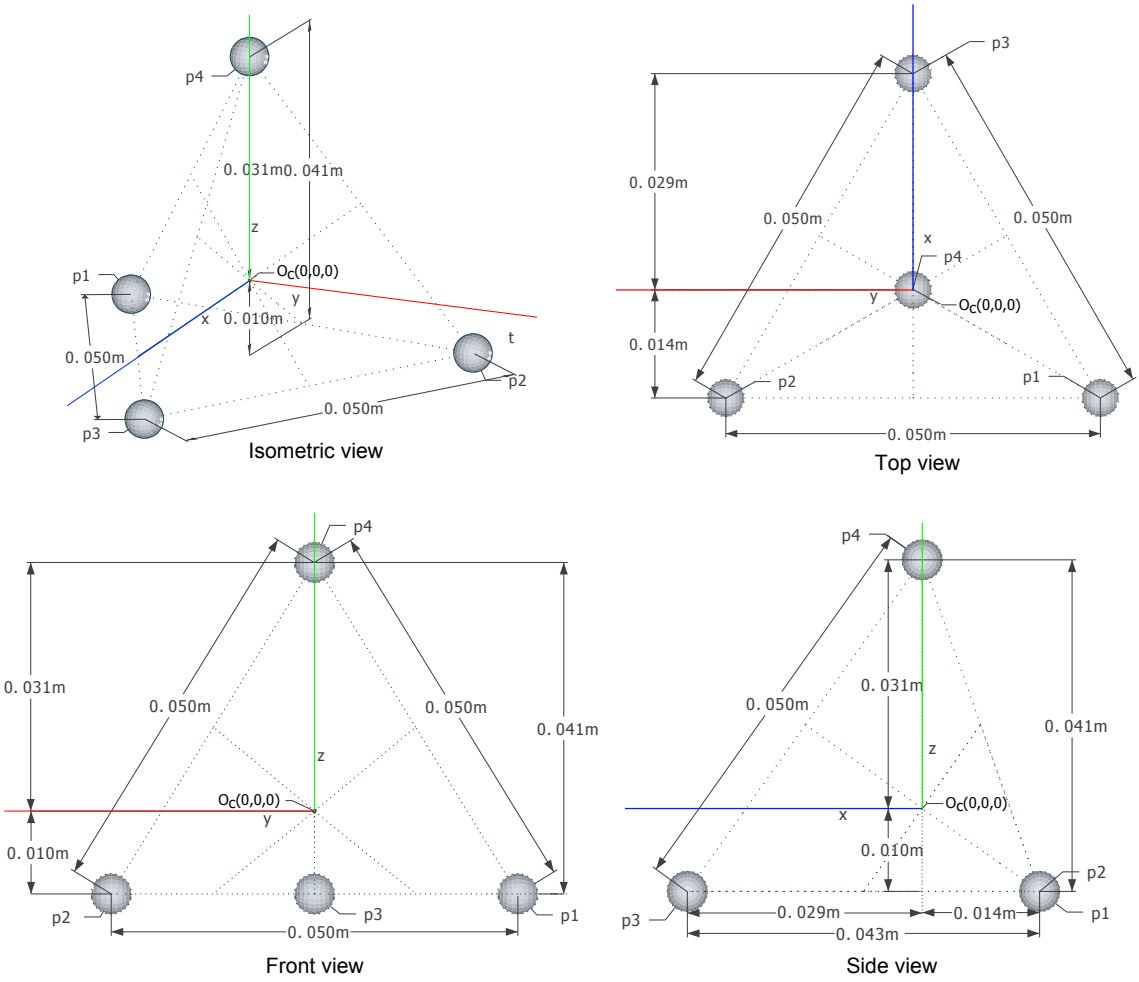

Figure 3.1: Tetrahedron configuration. Isometric view (left top), XZ plane (right top), YZ plane (left bottom), and XY plane (right bottom). Shaded spheres represent i^{th} microphone p_i and each of them is equally distanced from the centroid at the coordinate origin.

By the Taylor approximation, we arrived to the arithmetic mean of pressure, estimated at the centroid. This values is obtained from the 1^{st} element of the matrix **D**. More importantly, the pressure gradient for each Cartesian coordinate is estimated at the centroid of the array and it is contained in last the 3 elements of matrix **D**. Sound intensity in the frequency domain is derived from Eq. 3.5 and 3.6. The unknown quantities are derived in elements of the matrix **D**, which are substituted accordingly for each coordinate.

$$\vec{I}_x(\omega) = \frac{1}{2}Re\{P(w)^*\vec{U}_x(w)\} = \frac{1}{2}Re\left\{P(\omega)^*\left(\frac{j}{\rho\omega}\frac{\Delta P_x(\omega)}{\Delta r}\right)\right\} = \frac{Im\{\mathcal{F}\{D_1(\omega)^*D_2(\omega)\}\}}{\rho\omega} \quad (3.12a)$$

$$\vec{I}_y(\omega) = \frac{1}{2}Re\{P(w)^*\vec{U}_y(w)\} = \frac{1}{2}Re\left\{P(\omega)^*\left(\frac{j}{\rho\omega}\frac{\Delta P_y(\omega)}{\Delta r}\right)\right\} = \frac{Im\{\mathcal{F}\{D_1(\omega)^*D_3(\omega)\}\}}{\rho\omega} \quad (3.12b)$$

$$\vec{I}_z(\omega) = \frac{1}{2}Re\{P(w)^*\vec{U}_z(w)\} = \frac{1}{2}Re\left\{P(\omega)^*\left(\frac{j}{\rho\omega}\frac{\Delta P_z(\omega)}{\Delta r}\right)\right\} = \frac{Im\{\mathcal{F}\{D_1(\omega)^*D_4(\omega)\}\}}{\rho\omega} \quad (3.12c)$$

The resulting sound intensities are now expressed for an arbitrary configuration of an arbitrary number of microphones where pressure gradients of microphone pairs do not share the same centre. Since the procedure using the Taylor expansion produces the 3 single 1D intensity vectors, 3D intensity vector is acquired by a vector addition.

3.3 Possible errors in measuring sound intensity

It was found out that the sound intensity technique is prone to many errors mostly due to the instrumentation. In p-p techniques, both sound pressure and sound particle velocity are estimated by 2 fixed microphones what brings finite difference approximation error[29]. The basic idea of this error is depicted in Fig. 3.2, where 2 sound signals with different frequencies propagate in space along the x-axis.

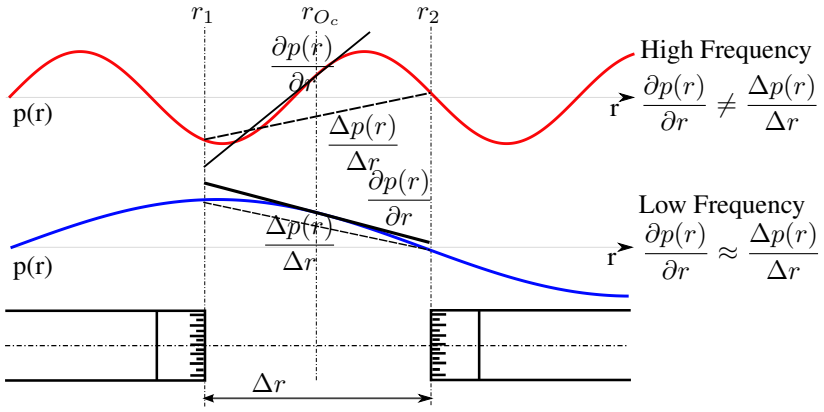

Figure 3.2: Finite difference approximation for 1D sound intensity [26].

The pressure signals have their true gradient at r_{O_c} given by $\frac{\partial p(r)}{\partial r}$, but because the 2 microphones sense the signal at 2 discrete positions r_1 and r_2, the partial derivative is not evaluated close around the point r_{O_c} and it results in finite difference estimate $\frac{\Delta p(r)}{\Delta r}$, where $\Delta r = r_2 - r_1$. This leads to an error for higher frequencies, because as seen in Fig. 3.2, low frequencies are not very affected since the rate of the pressure magnitude change over position is not fast, but with the increasing frequency, the rate of change increases, the gradient of a curve is approximated to a straight line and the error becomes more pronounced. The error can be mathematically described as ratio of the estimated and the true intensity. The Eq. 3.13 holds in ideal conditions, when a scattering of sound is neglected, what is acceptable only for lower frequencies [15],

$$\frac{I_e}{I} = \frac{\sin k\Delta r}{k\Delta r}. \tag{3.13}$$

The ratio of estimated intensity I_e and true intensity I yields an error that increases with frequency and distance between microphones. The error due to scattering and also diffraction is hardly avoidable, because a frame holding microphones is always colouring propagating waves. Although there was a suggestion by Elko[9] to mount array microphones on a rigid sphere. It has been shown, that it is not only making the microphone array frame simpler to implement, but it also has a self-correcting effect on the finite difference error. Furthermore, due to the physical distance, that a sound wave needs to travel around the sphere is increasing, the overall dimensions of the microphone array can be decreased down to $2/3^{rds}$ of the freely suspended microphone array.

A sound wave propagation can be measured on difference in pressure phase in time or in space. Sound intensity is also linked with the pressure phase change, because if there is no pressure phase change, there is no sound propagation and zero sound intensity. An example of phase change in space is demonstrated on an intensity probe in a harmonic sound field in Fig. 3.3 and the relation between sound intensity and the phase change is expressed again for a harmonic sound field in Eq. 3.14,

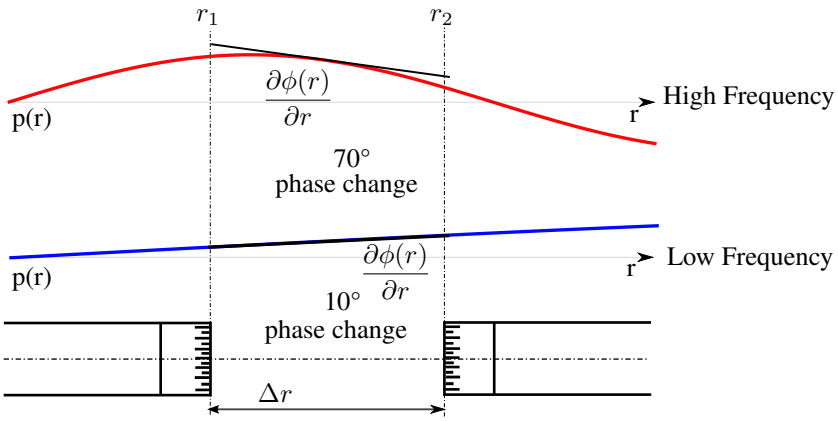

Figure 3.3: Phase change on 1D intensity probe,

$$I = -\frac{p_{rms}^2}{\omega\rho}\nabla\phi. \tag{3.14}$$

It is seen that the rate of phase change is higher for higher frequencies and very small for really low frequencies. The phase of change of $70°$ and $10°$ for microphone spacing $50\,\mathrm{mm}$ corresponds to $1.3\,\mathrm{kHz}$ and $190\,\mathrm{kHz}$ respectively.Analysing equipment always introduces a phase change between the microphone channels (typically $\pm 0.3°$), which causes an error called phase mismatch[7]. The character of the error can overestimate or underestimate the resulting intensity, since the rate of phase change is positive or negative. It can be seen that the phase mismatch has more severe effect on the low frequencies, as the sum of the physical phase change and mismatch error is influenced more by the error. The error is well approximated by Eq. 3.15,

$$\frac{I_e}{I} = 1 - \frac{\phi_e}{k\Delta r}. \tag{3.15}$$

We assume an incident angle in the x direction. The error decreases with a change of sound wave incident angle, because the phase change increases while phase mismatch remains the same for given frequencies. Despite this, the low frequency mismatch still cause large portion of sound intensity error. an example of discussed errors for $50\,\mathrm{mm}$ distanced microphone pair and for mismatch error 0.3ř is plotted in decibel scale,

It can be deduced that a trade-off needs to be found for the frequency range that is not prone to many errors. The finite difference approximation error decreases with shortening the microphone distance, so higher frequencies tend to be less erroneous compared with the larger microphone distance. On the contrary, the phase mismatch error increases with shortening the microphone gap and low frequencies with a small error then requires large separation between microphones. The compromise is then decided on the required frequency range and the typical separation distances varies from 5 millimetres up to units centimetres.

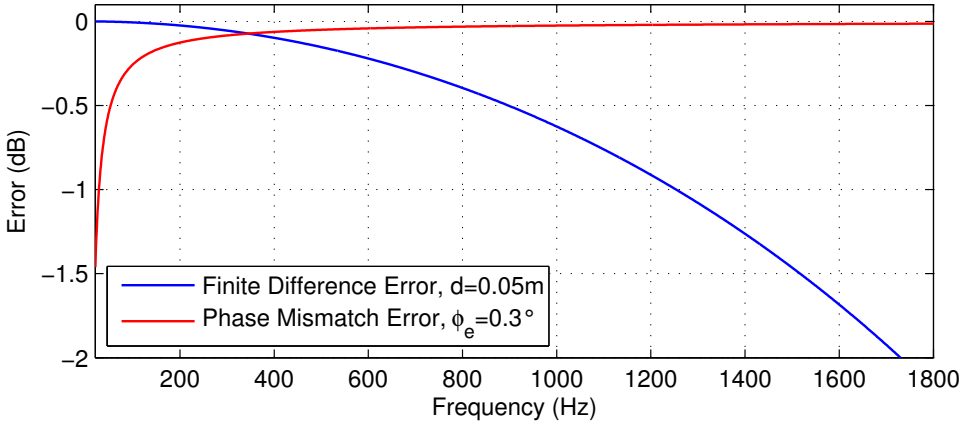

Figure 3.4: Combination of major sound intensity errors [26].

3.4 Localization of sound source by 3D sound intensity

There is no pure intensity sensor, therefore 3 possible measurement techniques were developed based on the combination of sensor types. One is a so-called velocity-velocity (u-u) method, where 3 particle velocity sensors in 3 dimension are used. Also a pressure-velocity (p-u) method is more commonly seen and it can be easily applied directly to Eq. 3.1c or Eq. 3.5. The p-u method can be implemented for a 1D, 2D or 3D design where one omni-directional microphone is always combined together with 1, 2 or 3 particle velocity sensor(s) depending on a required sound intensity dimension. In Chapter 3.1 and in this thesis we deal with a pressure-pressure (p-p) technique. It can also be implemented for arbitrary dimensions and a minimum number of microphones for 1D, 2D and 3D is 2, 3 and 4 microphones. By using the minimum number, a precision for 2D and 3D will be reduced. In order to achieve an optimal performance, 4 microphones for 2D and 6 microphones for 3D measurement should be used.

Our measurement is conducted on the tetrahedron configuration with microphone spacing r_m to be $50\,\mathrm{mm}$ depicted in Fig. 3.1 and its intensity derivation suggested in Eq. 3.12. The derivation is then processed with known microphone locations and Eq. 3.16 rises.

$$\vec{I}_x(\omega) = \frac{3Q_{31}(\omega) + 3Q_{32}(\omega) + Q_{41}(\omega) + Q_{42}(\omega) - 3Q_{43}(\omega)}{4\sqrt{3}r_m\rho\omega} \tag{3.16a}$$

$$\vec{I}_y(\omega) = \frac{2Q_{21}(\omega) + Q_{31}(\omega) + Q_{41}(\omega) - Q_{32}(\omega) - Q_{42}(\omega)}{4r_m\rho\omega} \tag{3.16b}$$

$$\vec{I}_z(\omega) = \frac{Q_{41}(\omega) + Q_{42}(\omega) + Q_{43}(\omega)}{\sqrt{6}r_m\rho\omega} \tag{3.16c}$$

By looking closely at the sound intensity derivation, we can see a certain pattern for each coordinate. This is most obvious for the z-coordinate, where we can imagine to look at the tetrahedron from the top view (see Fig. 3.1) in the direction of the negative z axis and we see a symmetrical placement of microphones where all of them are separated by the same distance from the top microphone 4. Then the approximated intensity is a weighted value of the quadrature spectral density Q between the 4^{th} microphone and all the others, so that the order of the quadrature spectral densities starts at the closest microphone towards the distanced ones. Because angles between the 4^{th} microphone and the others are the same as well as the distances, all quadrature spectral densities are weighted by the same scalar, what is not the case for the other coordinates and thus different weightings takes place as can be noticed.

The whole measurement and data acquisition for single tetrahedron configuration follows an order depicted in Fig. 3.5. This measurement procedure has been already implemented in previous years through MATLAB in our Acoustic Lab at KAIST [6].

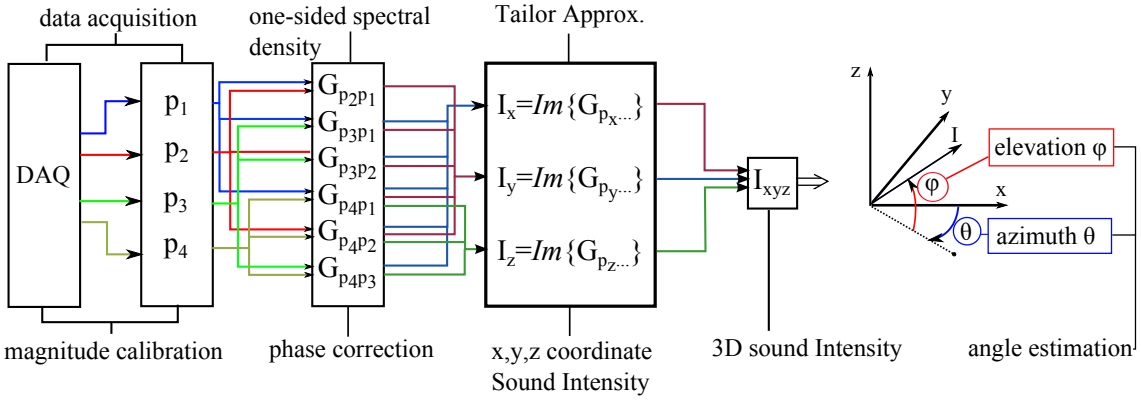

Figure 3.5: 3D Angle detection procedure by sound intensity technique

Firstly, the sound pressure is acquired through a Data Acquisition device DAQ, where pressure sensitivities are calibrated. In the subsequent step, one-sided spectral density is computed, what also allows correction of phase response of microphones and their channels. This is done by calculating transfer function between each microphone and one reference microphone. An angle information is extracted and the spectral densities are corrected for the phase mismatch. The corrected one-sided spectral densities are inserted into set of Eq. 3.16 which was derived from the Taylor approximation and it gives us 1D sound intensity vectors for each 3D Cartesian coordinate. The composition of the 3D vector is approximated over a vector summation and a bearing angle (composed of azimuth θ, and elevation ϕ) is obtained by trigonometrical operations.

CHAPTER 4

Simulation for localization of acoustic sources

The sound intensity calculation and the sound source separation was tested over a MATLAB simulation to figure out feasibility of the BSS technique for a localization of multiple sources. In this chapter, the sound intensity probe to be investigated has the tetrahedron configuration and the microphone spacing $r_m = 50\,\mathrm{mm}$. An ambient environment is assumed to have temperature $20\,\check{\mathrm{r}}\mathrm{C}$ that corresponds to speed of sound $343\,\mathrm{m/s}$. In previous work in our laboratory[13], a frequency span between $500\,\mathrm{Hz}$ and $1500\,\mathrm{Hz}$ was suggested for $r_m = 45\,\mathrm{mm}$ and it was adopted in this work as well for convenient evaluation of results. The upper limit was decided upon a finite difference approximation error $(1.2\,\mathrm{dB})$. Because different microphone spacing was used here, this error increased to $(1.5\,\mathrm{dB})$, what is still relatively low. While considering the limits of measurement, it should be taken into account that the errors presented in Sec. 3.3 applies to 1D intensity arrays, but since the intensity vector is estimated here to be off the microphone pair centres, the introduced error for 3D arrays is not exactly the same. A low frequency limit was chosen based on observations of angle error at the low frequency end. In order to evaluate a 3D sound intensity array performance, sound directivity simulation is presented in appendix A.1.

4.1 Sound intensity simulation implementation

A post-processing algorithm for obtaining a 3D intensity vector in frequency domain was implemented by [6], but a simulation of 3D intensity was desirable for a further investigation of source separation methods. The p-p sound intensity measurement is based on time and amplitude differences in sound pressure signals. In real conditions, a wavefront propagates continuously with the speed of sound. The pressure quantity is converted into a discrete form with sampling frequency as low frequency as $4\,\mathrm{kHz}$ for covering the effective frequency range. To facilitate similar continuous sound propagation conditions in the simulation, an original acoustic signal is created with a low sampling frequency and then it is resampled to a high sampling frequency in the range of MHz. That is why an accurate time shift can be now calculated from speed of sound c_o. The maximum time shift here is derived to be $\Delta t_m = \Delta r_m / c_o = 0.05/343 = 146\,\mu\mathrm{s}$. If the signal is up-sampled to $4\,\mathrm{MHz}$, the shortest time delay that can be detected is $0.25\,\mu\mathrm{s}$ which gives us a reliable precision for an angle detection and increasing up-sampling frequency does not improve precision of the simulation dramatically.

4.1.1 Sound wave propagation

A correct functionality of the simulation algorithm and an effect of the resampling was tested on 2 sound sources for clarity. The analysis in this section depicts the procedure of the resampling.

The up-sampling to a higher frequency interpolates missing non-existing values in a signal and therefore it locally modifies the shape of an audio signal. First of the tested signal is a speech. There will be 4 signals used throughout the simulation and experiment chapter, but now we are concerned only by a white noise having Gaussian distribution and speech signal 1. The character of the noise was chosen for its frequent occurrence in a noisy environment. A time and time-frequency analyses for all 4 sources are depicted in Fig. 4.1,

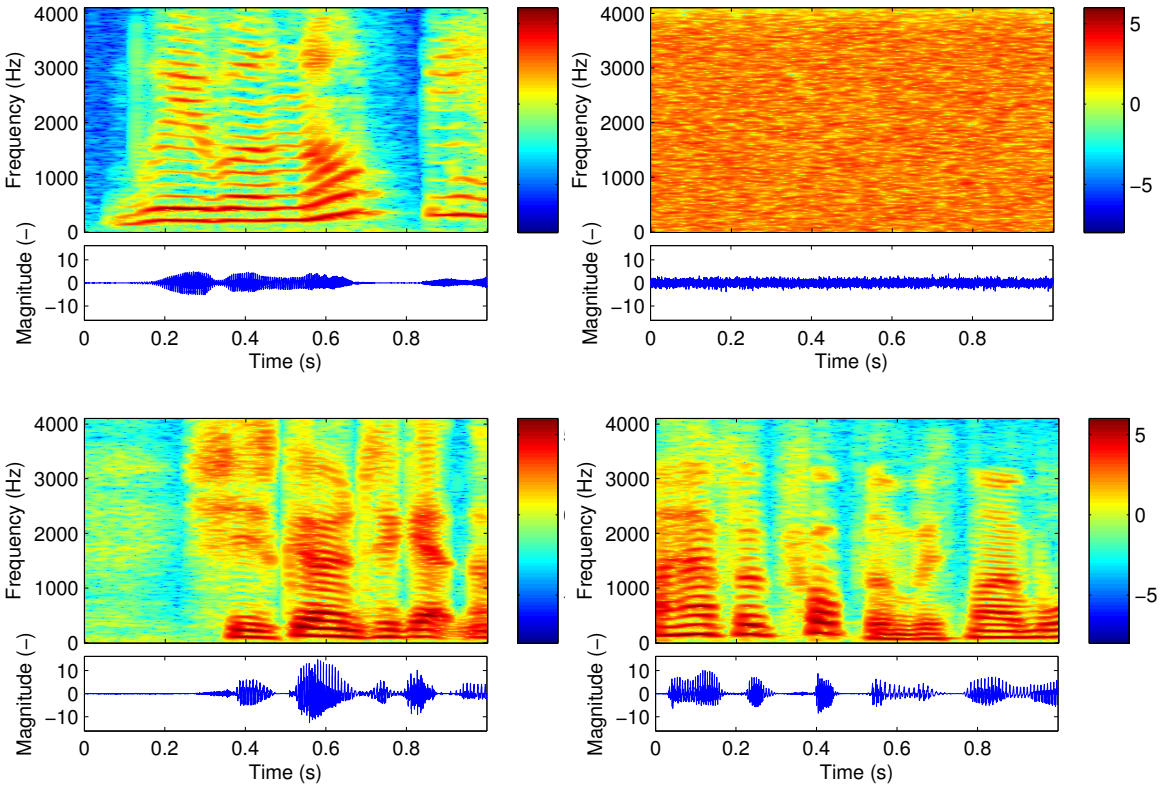

Figure 4.1: Time and spectrum analysis of single signals. Speech signal s_1 (top left), Gaussian noise s_2 (top right), speech s_3 (bottom left), speech s_4 (bottom right)

The position of these sound sources is symbolically described below.

Only the speech signal s_1 is examined in more detail. To simulate the convolutive mixture, delays between microphones for arbitrary sound source location were derived based on geometrical formulations. A demonstration of the resampling is shown on a time delay between the microphones corresponding to a half sampling interval of lower sampling frequency. The original and the delayed and resampled signals were compared in a time and a frequency domain. Fig. 4.3 shows how the waveform is changed after the resampling.

The original data sampled at the lower frequency $8\,\mathrm{kHz}$, seen as the blue line on the left part of the plot, is compared with up-sampled data at a high frequency $4\,\mathrm{MHz}$ depicted over blue line on the right. The delay between 2 microphones is not introduced represented by the red line curves. The low frequency sampled

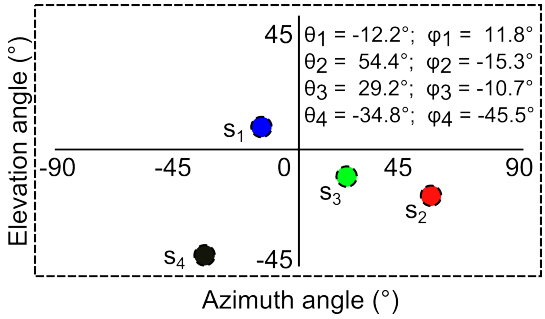

Figure 4.2: Source positions at simulation.

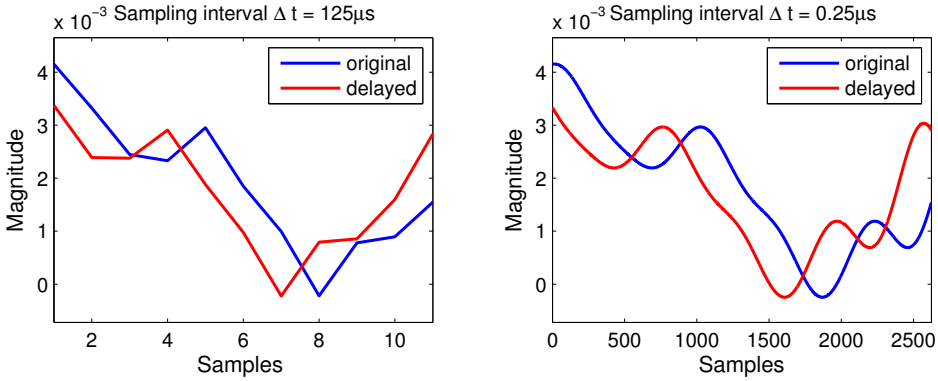

Figure 4.3: Example of original and delayed resampled data in time domain.

data change slightly in local magnitude, which introduces distorted high frequencies as shown in frequency spectrum that presents original, delayed signal, and its difference,

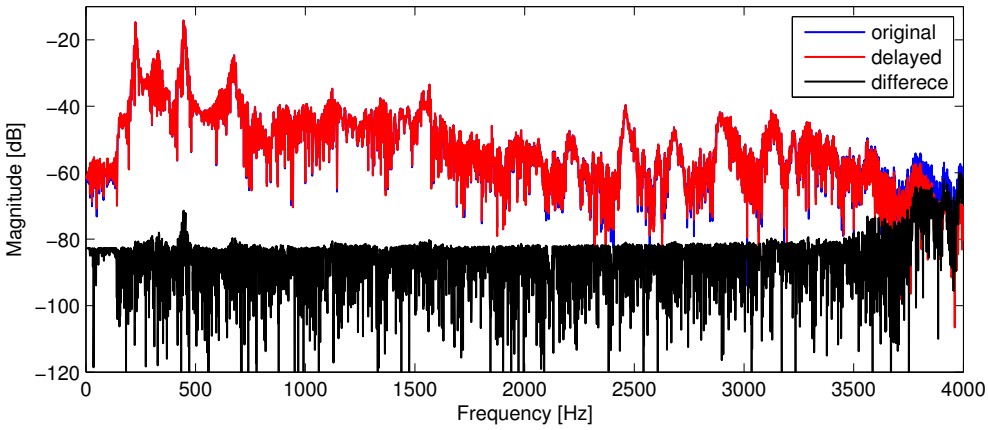

Figure 4.4: Example of original and delayed resampled data in frequency domain.

It can be observed from the Fig. 4.4 that the original and delayed signals are nearly alike in the frequency spectrum and only noticeable difference seen is below $500\,\mathrm{Hz}$ and at high end of the spectrum above $3.5\,\mathrm{kHz}$. These artefacts are not audible when listening to the original and manipulated samples.

The delay time among microphones is calculated based on a bearing angle. This is illustrated in Fig. 4.5. In our simulation, the distance of the source from microphone array do not change the final result.

The wavefront of the plane wave is represented by the blue surface, yellow dotted lines are perpendicular

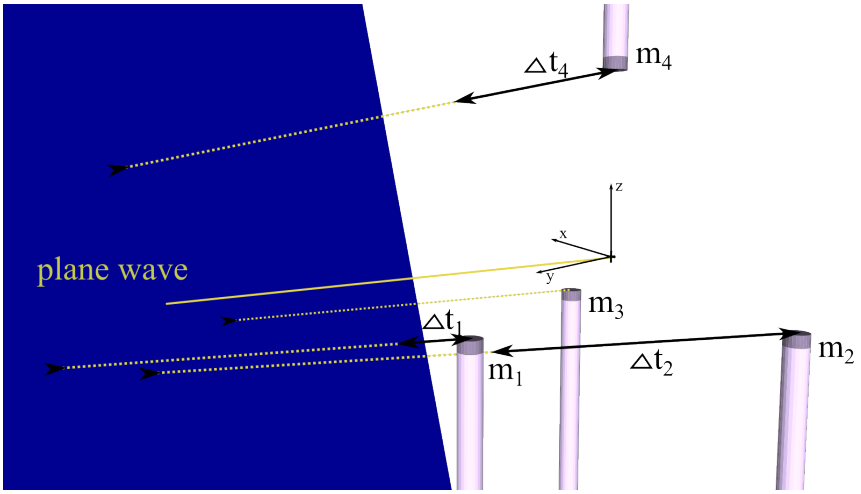

Figure 4.5: Example of delay estimation among array microphones. $\theta = 55°$, $\phi = -15°$

distances from the wavefront to an ith microphone m_i and time delays are noted as Δt_i. The microphone upon which a sound wave impinges as the first has always a zero time delay. A logical order of sound intensity simulation is demonstrated below,

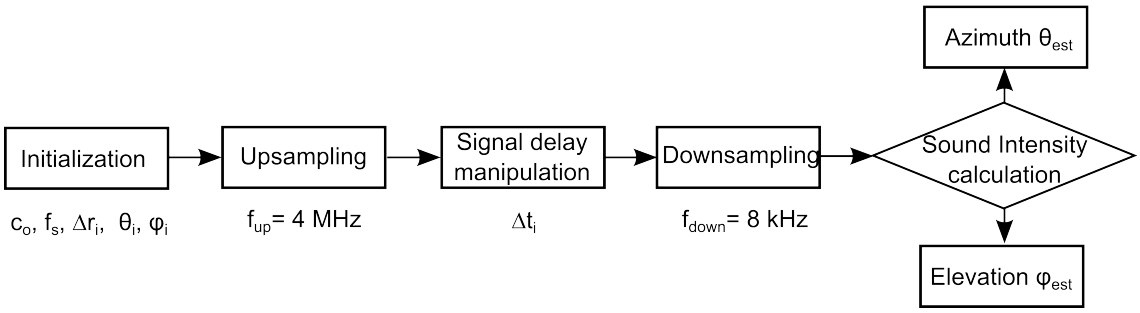

Figure 4.6: Simulation flowchart.

The bearing angle is represented through out the thesis in spherical coordinates, where azimuth corresponds to an angle θ along the horizontal line and elevation corresponds to an angle ϕ along the vertical line. The magnitude of sound intensity is not considered here.

4.2 Single source detection

The aim of our approach is to separate 2 to 4 sources from a mixture. Single sources are now considered to be processed separately in order to check a validity of the angle detection. The environment to be simulated is supposed to be anechoic and because every real measurement introduces a background noise, a Gaussian noise is added to all signals. The signal-to-noise ratio (SNR) is approximately $34\,\mathrm{dB}$ when compared to the sound source s_1. We have calculated the SNR over Eq. 4.2. This takes an average of power for frequency band between $500\,\mathrm{Hz}$ and $1500\,\mathrm{Hz}$, therefore it can be sometimes misleading to look just at the SNR value and closer look needs to be devoted to a spread of the spectrum magnitude along a considered frequency band. The comparison of all sources and the background noise is depicted as overlapping plots in Fig. 4.7. in the time domain as well in the frequency domain.

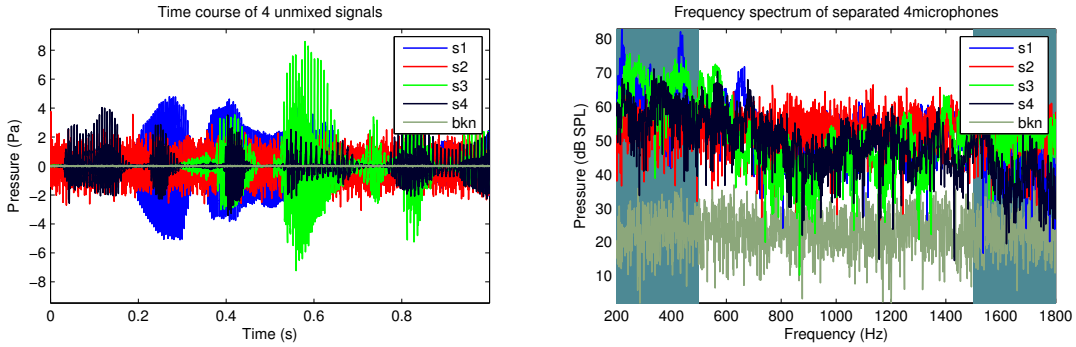

Figure 4.7: Time course and FFT of overlaying unmixed source signals

$$SNR_{band} = 20\log_{10}\frac{\sqrt{\frac{1}{N}\sum|\mathcal{F}(x_{signal})|^2}}{\sqrt{\frac{1}{N}\sum|\mathcal{F}(x_{noise})|^2}} \tag{4.1}$$

The SNR values and positions of sound sources were chosen to simulate similar conditions as were present in real anechoic experiment. For this reason, the elevation angles do not deviate a lot from $\phi = 0$ř, since it is not possible to locate sound sources to extreme angles in the vertical plane due to dimensions of our anechoic chamber. An overview of sound sources details is summarized in the Tab. 4.2. At this point, a colour of the sound source is assigned as well to indicate a notation for all angle detection plots.

Sound sources	Colour	Type	SNR (dB)	Azimuth angle $\theta(^\circ)$	Elevation angle $\phi(^\circ)$
s_1	Blue	Speech 1	34.2	-12.2	11.8
s_2	Red	Gaussian noise	33.9	54.4	-15.3
s_3	Green	Speech 2	36.6	29.2	-10.7
s_4	Black	Speech 3	32.5	-34.8	-45.5

Table 4.1: Sound sources information

The single tetrahedron array as in 3.1 has been implemented together with its twisted representation. They together form the twisted tetrahedron array for better accuracy. The angle estimation obtained over the simulation is depicted for all sources in figures below. The shaded area in figures shows frequency range out of the effective range, but it is kept here for a consistency. The dashed line corresponds to the original angles that have been assigned in an initialization step of the simulation and the color of the dashed line is shown in darker colour than the original angle line.

Each row of the figure correspond to jth source. The columns represent, from left to right, azimuth and elevation angle evaluation. In each plot, the dashed line correspond to the truth angle, the legend notes double-twisted tetrahedron configuration (twi), and single tetrahedron configurations (sg_1, sg_2).

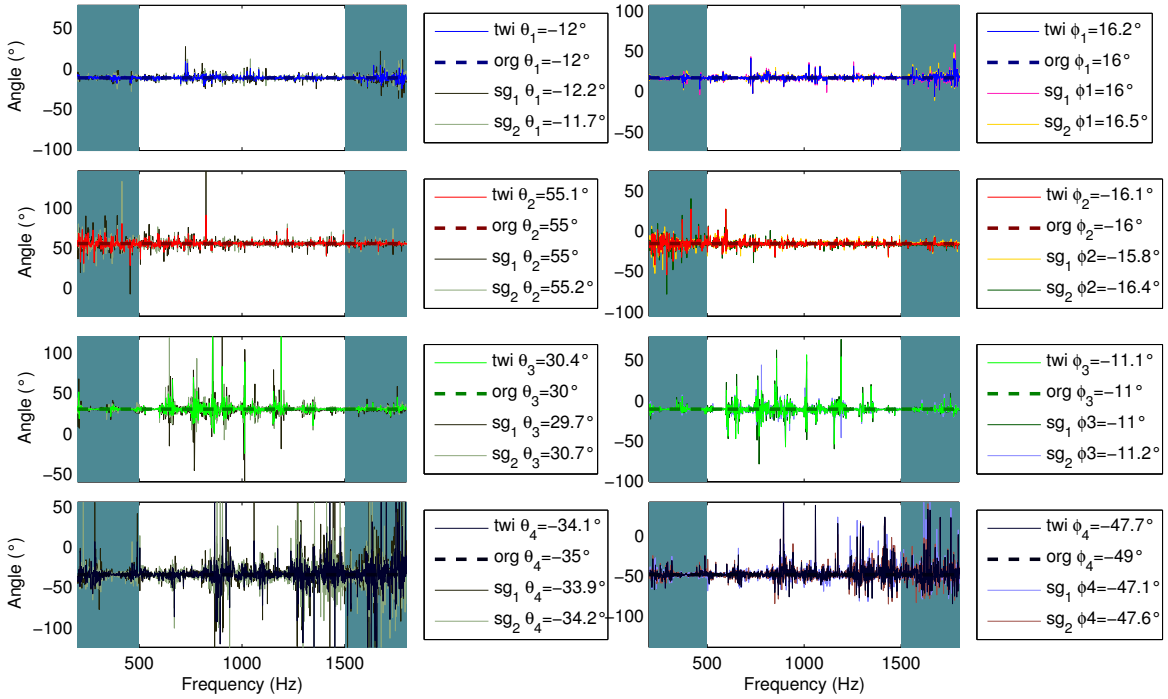

Figure 4.8: Angle detection of single sources played solo with background noise. An angle range is $90°$ about the original angles.

The sound sources are plotted from the first one at the top to the last at the bottom. The results are summarised in the Tab. 4.2, where azimuth and elevation angles of the twisted probe are analysed. The estimated angles (θ_{twi}, ϕ_{twi}) are obtained by arithmetic average taken over the effective frequency range from 500 Hz up to 1500 Hz. From the set of plot above it is seen that the single tetrahedron probes are a bit more off compared to the twisted. This is more obvious for the azimuth angle, but for the elevation angle, the angle detection is prone to a higher error due to the directivity of the probe configuration. Still the error is within 1.1 degree for the tested random angles and the algorithm can be sufficiently used for separated signals as long as they are at least a few degrees apart.

Sound source	SNR (dB)	Original $\theta_{org}(^\circ)$	Twisted $\theta_{twi}(^\circ)$	Error $\Delta\theta_{twi}(^\circ)$	Std. dev. σ°_θ	Original $\phi_{org}(^\circ)$	Twisted $\phi_{twi}(^\circ)$	Error $\Delta\phi_{twi}(^\circ)$	Std. dev. σ°_ϕ
s_1	34.2	-12.2	-12.1	0.1	0.9	11.8	12.1	0.3	1.2
s_2	33.9	54.4	54.5	0.1	1.0	-15.3	-15.4	-0.1	1.2
s_3	36.6	29.2	29.6	0.4	1.4	-10.7	-10.7	0	1.7
s_4	32.5	-34.8	-34.0	- 0.8	3.2	-45.5	-44.4	1.1	2.7

Table 4.2: Angle detection information of single unmixed sources

The SNR ratio of the sound sources predicts up to some extend the resulting error. This is although dependent on the directivity as well. Despite this fact, we can see that the signal s_4 with the lowest SNR results in biggest angle error. On the other side, the signal s_3 with the highest SNR do not have the best performance and by looking at Fig. 4.8, we see significant fluctuations compared with other signals which

have even lower SNR. This is no longer connected with the finite difference error, whereas it is probably due to an uneven spread of spectrum magnitude. By looking at Fig. 4.1, we can see that the signal s_3 has concentrated energy at low frequencies. The calculation of sound intensity is done in the frequency spectrum over power spectral densities, hence the angle of s_3 has significant fluctuations that may cause a higher error. The fluctuations are seen as oscillation around the desired angle and they are partially averaged out by each tetrahedron probe, or cancelled out by the double-twisted tetrahedron probe. It is due to these fluctuations that another measures of the sound intensity array performance need to be introduced. These are standard deviation σ_θ for the azimuth or σ_ϕ for the elevation angle. This is another statistical assessment after the arithmetic mean value and they both can be combined together to better rank a quality and preciseness of a detected source. We can see a distribution of each source on Figs. 4.9 and 4.10.

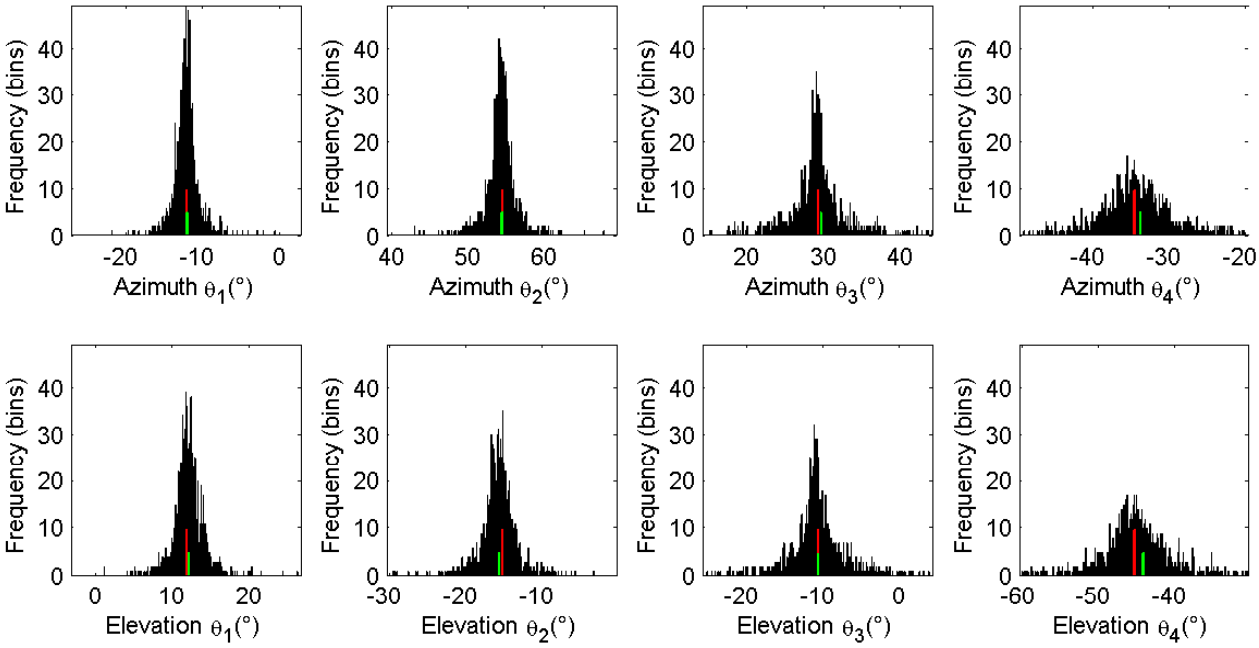

Figure 4.9: 2-dimensional histogram of azimuth angle detection at the top and elevation angle at the bottom row for single sound source playing solo. Each jth source corresponds to jth row. Red and blue vertical line marks original and estimated angles

We can see that the distributions converge to a value in the centre of the angles and at this point[1]. Now we can also match the standard deviation and estimated angle from the Tab. 4.2 to the visual representation the angle distributions. The trend of the standard deviation and the concentration of angle occurrence follows the error. For the higher standard deviation values, the error increases and also the sum of an angle occurrence in a particular bin is decreased. Therefore, based on these two measures, a range of an error can be predicted.

The Fig. 4.10 provides again histograms, but its viewed for 2 different angles in 3D to obtain better perception of the angle detection in space and a relation between the azimuth and elevation detection. The angle span is also $\pm 15^\circ$ around the original angle, thus the original angle is located exactly in the centre of the plot.

[1]The angle span is $\pm 15^\circ$ around the original angle. This scaling will be kept constant for a better comparison of all results for 2D and 3D histograms that represent the angle detection.

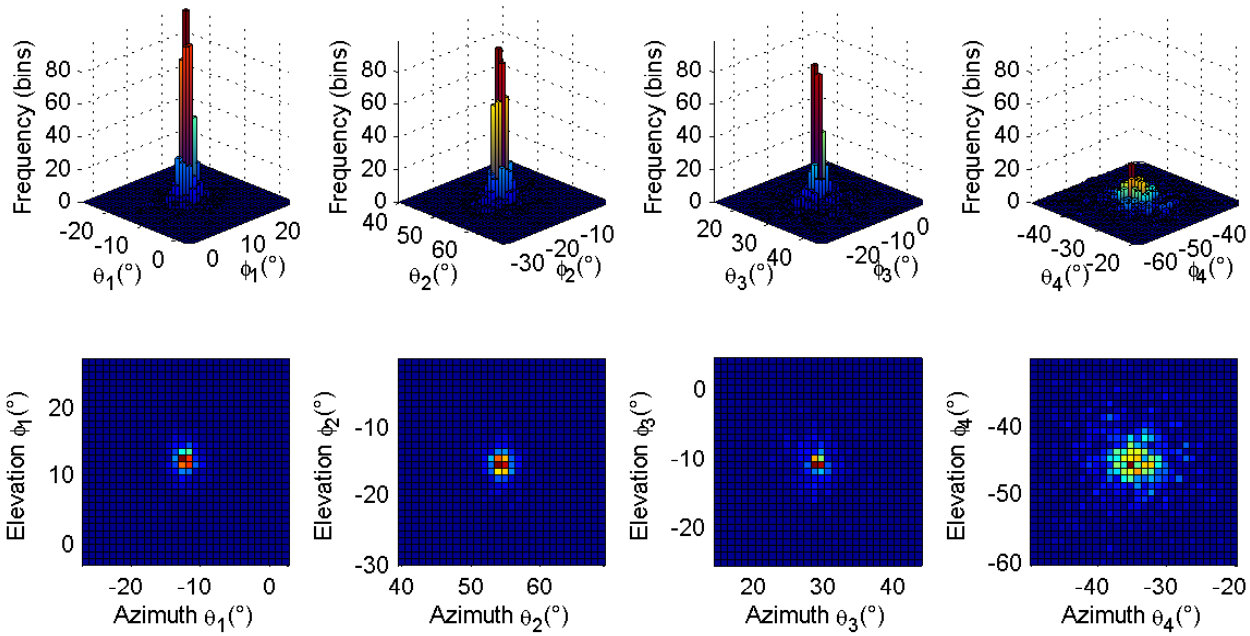

Figure 4.10: 3-dimensional histogram of azimuth and elevation angle detection for single sound sources playing solo. Top row is isometric projection, the bottom row is view from the top. Each jth source corresponds to jth row.

The 3D histogram is formed by creating a grid from the azimuth and elevation angle vectors. The size of each bin is 1° and it is composed of both, the occurrence of the azimuth and elevation angle in a particular bin. The plots reflects that the elevation and azimuth standard deviation is similar and the effect of faulty angles caused by the angle fluctuation can be eliminated to get precise detection. In case of $3^{\,\mathrm{rd}}$ and $4^{\,\mathrm{th}}$, a higher error is seen due to the larger span of the faulty bins, therefore flatter distribution resulting in higher probability of error.

4.3 Multiple Source setection

The standard intensity measurement methods estimated single 1D or 3D vector from overall contributions of present sources, so if more then 1 source is present at the same time instance, the intensity vector is evaluating a combination of sources direction. If the same sources, which were assumed in the previous Sec. 4.2, are to be excited simultaneously in the simulation, the outcome after the angle detection is evaluated with a significant dispersion among all 4 original angles as seen below,

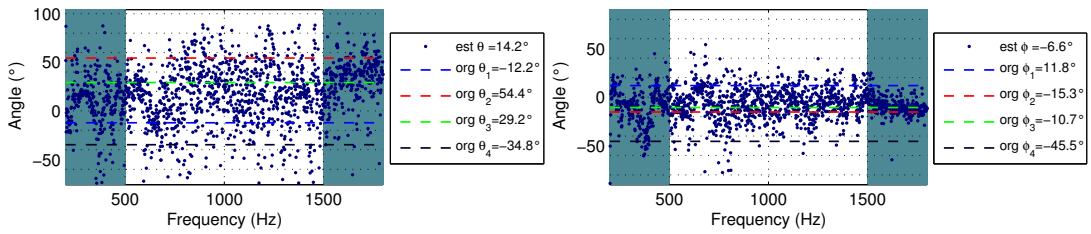

Figure 4.11: Angle detection of mixed signals.

The average of the values spanned over effective frequency spectrum is estimated to be somewhere in between, where the bias of the angle estimate depends on sound source strengths. The intensity calculation just mentioned have been used for many decades when manually scanning a source of noise, what eventually can create an intensity map. To avoid the scanning in space, BSS is attempted to be applied to the multiple source signal. The same sound sources are again used here. The signals of according sources were already plotted in Fig. 4.7, now they are mixed by adding all pressure values in pascals with an according delay to create a convolutive mixture. In order to detect multiple sources, BSS is combined with sound intensity calculation. The whole process is summarised in Fig. 4.12. Here we take advantage of the capability of T-ABCD method to separate sound sources with preserved delays between microphones for each separated source.

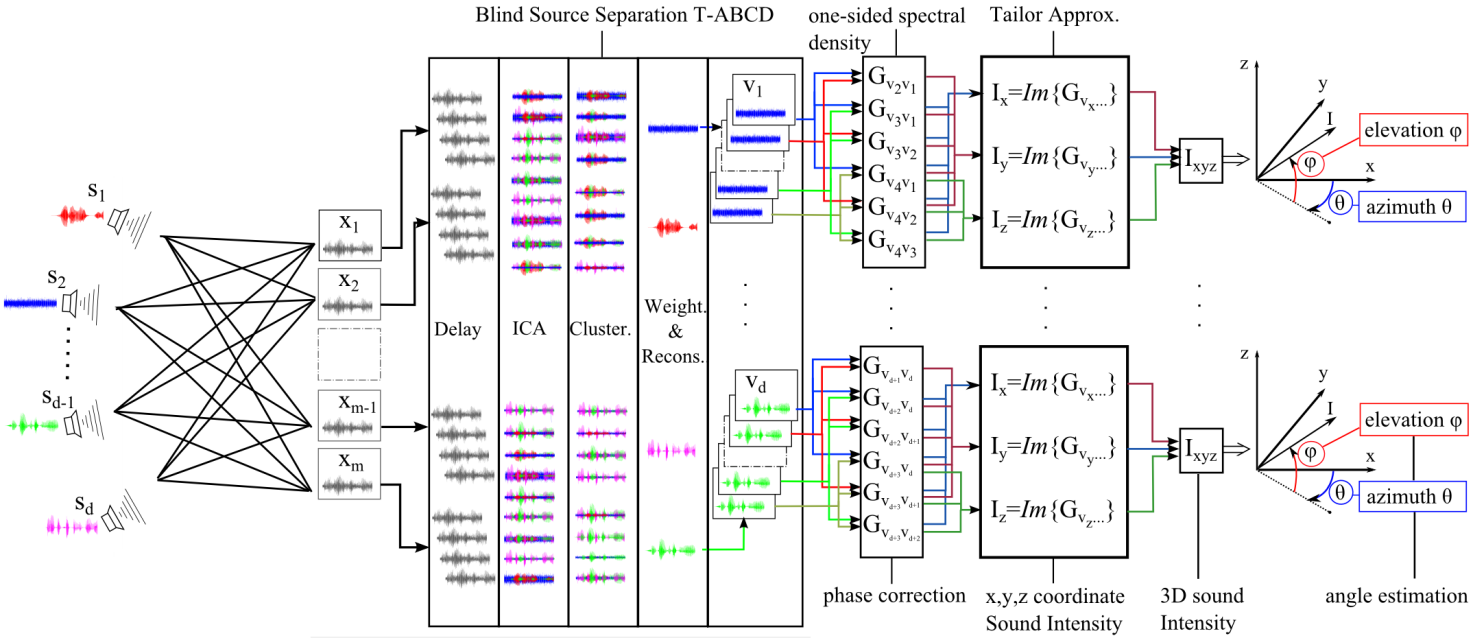

Figure 4.12: Procedure of BSS combined with sound intensity

The process basically combines the BSS method depicted in Fig. 2.2 and sound intensity method shown in Fig. 3.5. We start with d sources that are convolved and then captured by m sensors. This results in recordings x_i which frame is set to 1 second. The BASS algorithm is applied to obtained delayed estimates v_j of original sources s_j. The sound intensity computation is applied to these separated sources v_j and it follows exactly the same order as described in Chap. 3.4.

There are a few options that can be chosen for the BASS algorithm. We have used BGL ICA method has been chosen for good separation performance and low computational burden. The length of the unmixing filter $L_{i,j} = 15$ was sufficient for simulated anechoic conditions. The clustering method was chosen to be fuzzy clustering, because it shows more flexible performance. The clustering results are shown in Fig. 5.8 for both tetrahedron arrays. Even though we have array configuration of 7 microphones, the input of the BASS algorithm takes pressure data out of 4 microphones. This is firstly done because 4 sources can be sufficiently separated by determined system of 4 microphones. Secondly, the computational time increases with growing number of microphones.

We deduce from the clustering that a mixture of 4 sound sources were recognized. Each bin of the matrix corresponds to a certain independent component after the 2^{nd} step of BASS. On the basis of known cluster representation, we can also see that the 1^{st} 3 estimated sources were not separated very well so far, because

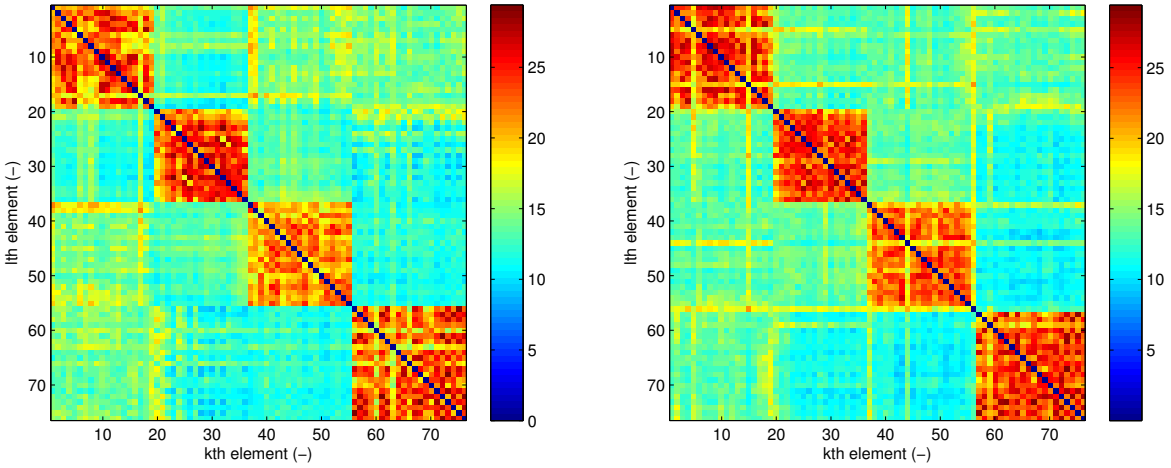

Figure 4.13: Clustering results of distance matrix with dimensions $mL \times mL$ for 1^{st} tetrahedron array on the left and 2^{nd} on the right. For both plots, 1^{st} cluster at upper left represent signals is most similar to v_3, 2^{nd} corresponds to v_4, 3^{rd} to v_1 and 4^{th} to v_2.

they are speech signals which are very similar. On the other side, the 4^{th} cluster correspond to Gaussian noise and it is seen that it is well distinguishable from speech sources v_1 and v_3, but it is correlated with v_4 what suggest worse sound separation due to the week signal v_4. This is evaluated by looking at the colours, which represent magnitude of GCC-PHAT, between each cluster. The two clusters are slightly different, but it was checked that it does not cause worse results in angle detection.

After the clustering, the weighting and sound signal reconstruction is applied. This leads to a final result of sound source separation as shown in Fig. 5.9 for 1 microphone only, because the other signals at other microphones are very similar.

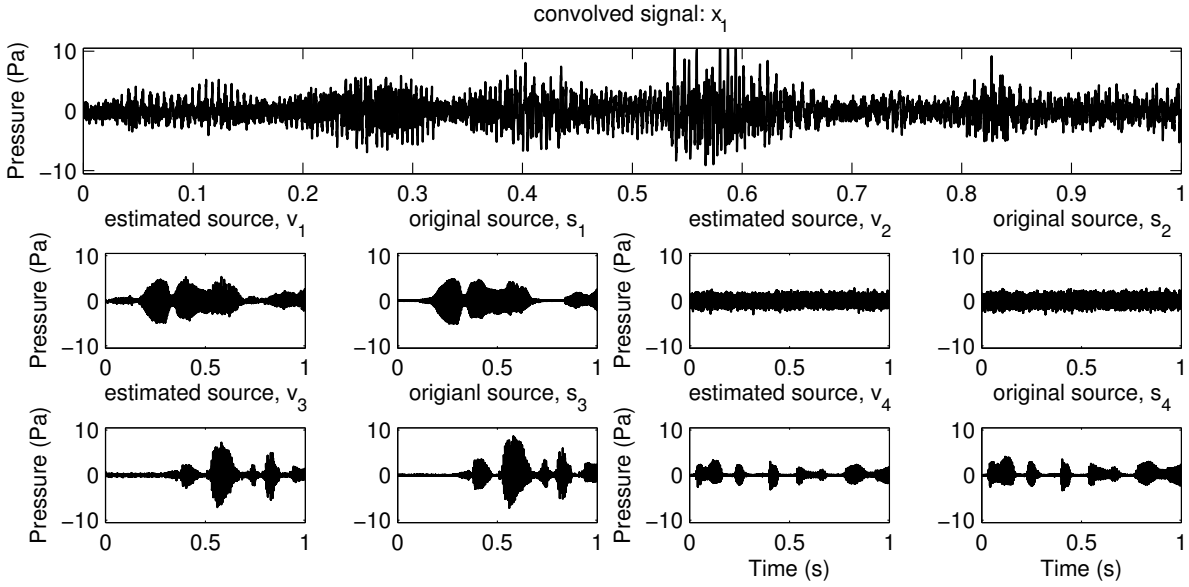

Figure 4.14: Separation results of mixed signals in time domain at 1 microphone.

The top row represents mixed 4 signals at microphone m_1. The plot in the middle row on the left depict the source v_1 estimated from the original source s_1. The other sound sources are shown in the same manner.

The estimated signal v_4 is added with a Gaussian noise as anticipated from the clustering. The source signal v_1 suffers by artifacts from the remaining speech sources, but the sound sources v_2 and v_3 sounds to be separated very well. Even though the separation of the simulated mixtures seems reasonable, the separated signals are different in amplitude when compared with original sound sources. This is dues to to the nature of BASS method, where this information about scaling is lost throughout the mixing process of 2 unknown matrices. The 4 microphones used in our set up are located very close to each other, so it is assumed that the difference in magnitude among all microphones is negligible.

Since we were able to separate the 4 independent sound signals under idealistic simulated conditions, the intensity calculation can be now conducted for the separated microphone estimates. The results for pure, not normalized data are shown in Fig. 4.15. It shows that the separated sources have been significantly separated even in angle-frequency domain, as compared to mixed sources in Fig. 4.11.

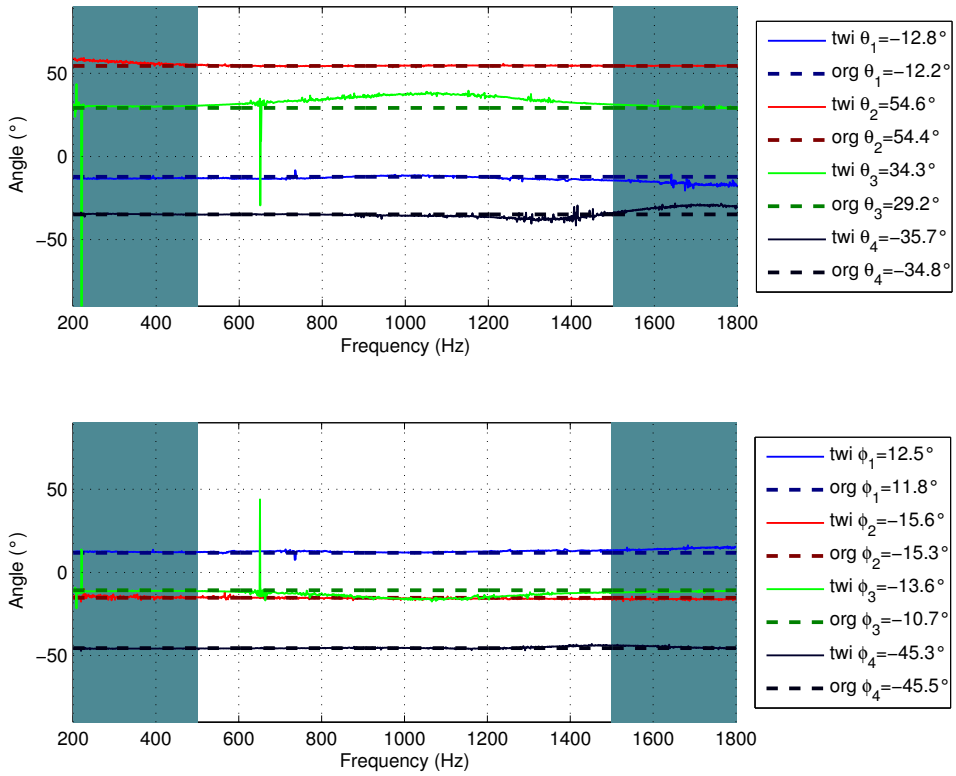

Figure 4.15: Angle estimation of multiple separated signals.

Summarised details of the angle detection are presented in Tab. 4.3. A closer look at angle detection is taken in Fig. 4.16. At this point, we should compare separately the performance of sound intensity calculation and sound source separation. We can directly notice that the error of the angle detection caused by sound intensity calculation is not correlated with error of the BASS. This can be compared between Tabs. 4.2 and 4.3. The error for the single source scenario was pronounced the most for the original source s_4 due to low SNR, but here, the most erroneous sound source is v_3 due to its imperfect separation. The origin of the separation error is firstly due to the temporal sparseness of the data. Therefore the BASS algorithm may evaluate the signal partially as a noise because when the speaker in recording s_3 is not talking a background noise is present and it has a character of Gaussian noise. The same character has the sound source s_2.

Another factor that may play a significant role is the closeness of sound sources. From the placement of sound sources in Fig. 4.2, we see that the sound sources s_2 and s_3 are the closest from all the sources and we also see the tendency of the angle detection for sound source v_3 to always incline towards the original angle of s_2.

Sound source	SNR (dB)	Original $\theta_{org}(^\circ)$	Twisted $\theta_{twi}(^\circ)$	Error $\Delta\theta_{twi}(^\circ)$	Std. dev. σ°_θ	Original $\phi(^\circ)$	Twisted $\phi_{twi}(^\circ)$	Error $\Delta\phi_{twi}(^\circ)$	Std. dev. σ°_ϕ
s_1	34	-12.2	-12.8	-0.6	0.9	11.8	12.5	0.7	0.5
s_2	33.6	54.4	54.6	0.2	0.2	-15.3	-15.6	-0.3	0.4
s_3	36.3	29.2	34.3	5.1	3.2	-10.7	-13.6	-2.9	2.4
s_4	32.3	-34.8	-35.7	-0.9	1.1	-45.5	-45.3	0.2	0.6

Table 4.3: Summary of angle detection measures for multiple sources.

The standard deviation again anticipates the error arose from the angle detection. The angle detection is more precise for the single sound sources playing solo, but we get rid of the rapid fluctuation by using our method.

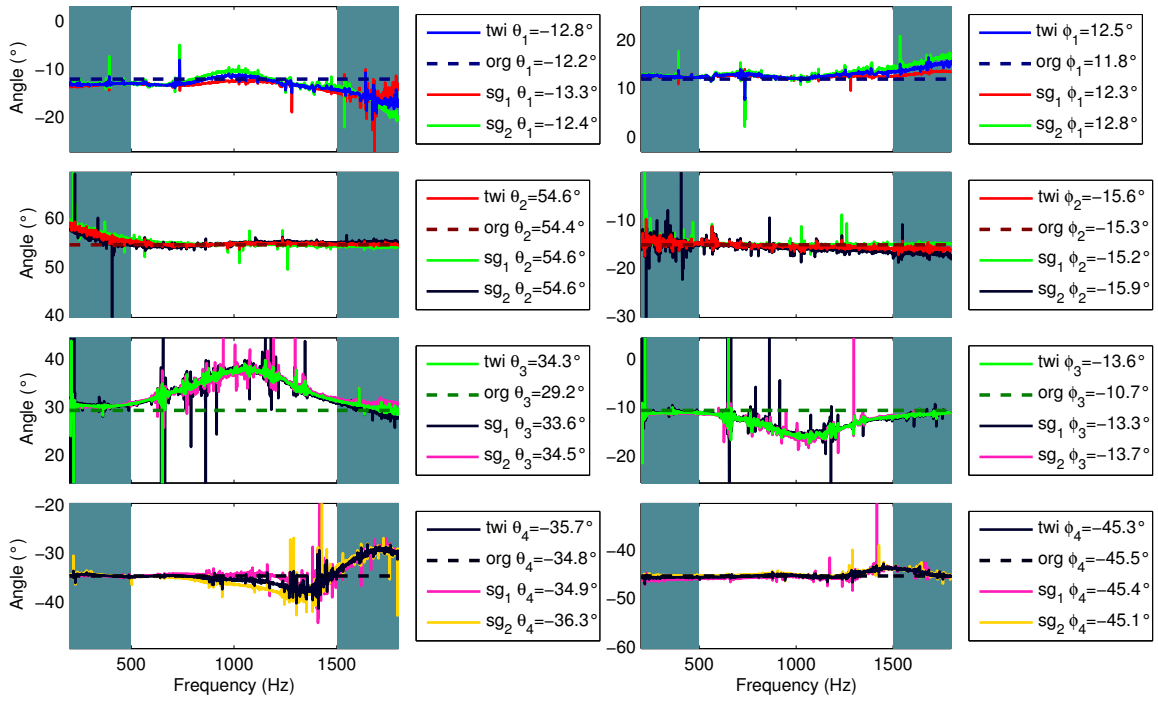

Figure 4.16: Angle detection of separated sources with background noise. An angle range is 15° about the original angles.

The detail of angle detection shows us that the double-twisted tetrahedron probe do not correct for the sound intensity error as it does for the single sources, since the error of BASS is introduced and it is much larger. Very good angle detection is achieved for Gaussian noise (v_2) because it can be very clearly separated from the mixture. Other speech signals shows error within $\pm 1^\circ$ except the discussed sound source v_3.
The histograms below shows either very concentrated or spreaded angle frequency, what is very different to the single sound source detection, since we got rid of the fluctuation, but also introduced a new error. Fig. 5.13 nicely shows the tendency of the sound source v_3 to be mixed with s_4 and thus approaching the direction the source s_4.

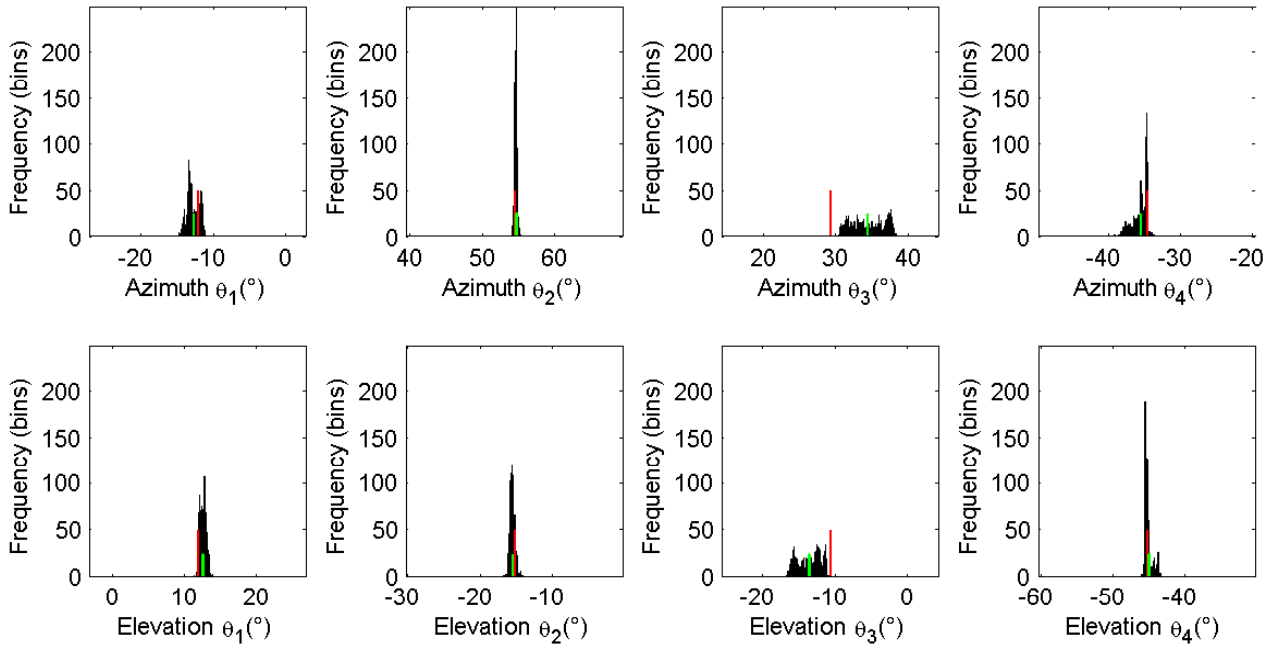

Figure 4.17: 2-dimensional histogram of azimuth angle detection at the top and elevation angle at the bottom row for separated sound sources. Each jth source corresponds to jth row. Red and blue vertical line marks original and estimated angles

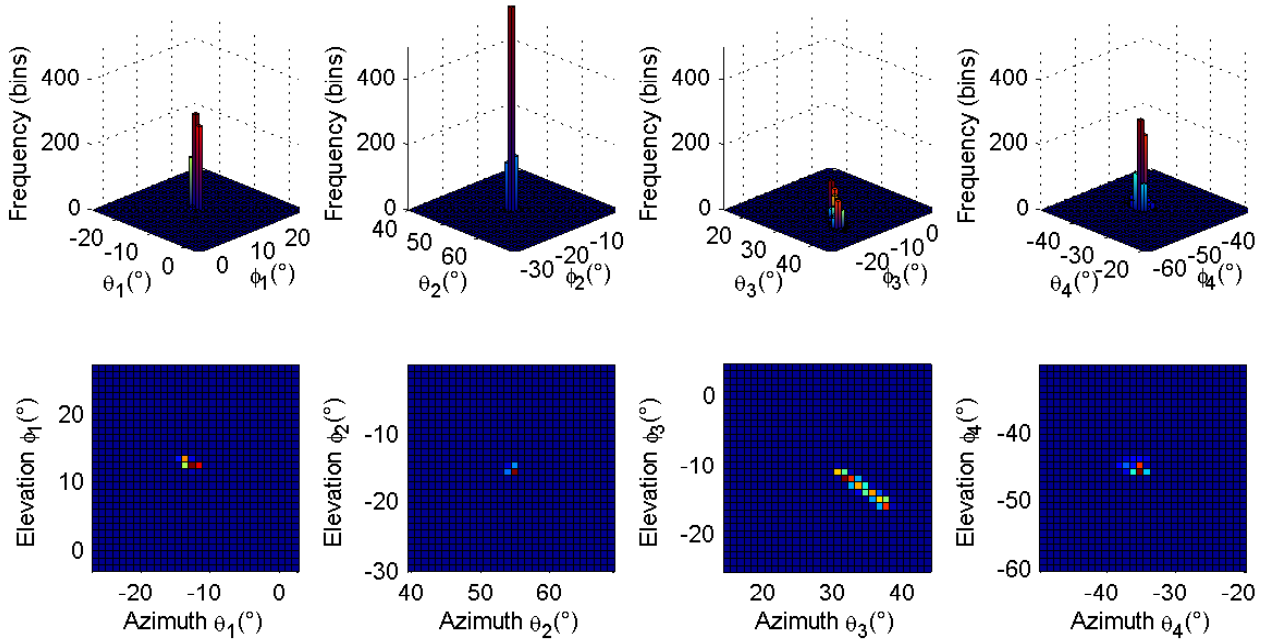

Figure 4.18: 3-dimensional histogram of azimuth and elevation angle detection for separated sound sources. Top row is isometric projection, the bottom row is view from the top. Each jth source corresponds to jth row.

The separation procedure have an adverse effect on the angle estimation for some sources more then for others, on the other side, it avoids the fluctuations caused by added background noise to the mixture.

CHAPTER 5

Experiment

The previous section showed, that under ideal, theoretical conditions including a background noise, the BASS method can effectively separate independent sources. The experiment conducted here followed similar conditions as assumed in the simulation for anechoic conditions. The probe configuration used throughout the experiment was double-twisted tetrahedron due to slightly improved results compared to the single tetrahedron.

5.1 Procedure

Array microphones were attached on a frame that was adopted from previous projects [13, 6]. The measurement arrangement has been adjusted for more convenient data manipulation using National Instrument DAQs and software (*Labview*). The recorded signal captured by the array microphones is sampled by $8\,\mathrm{kHz}$ and it is added by a DC offset as well by low frequency distortion electric noise. The used NI DAQ does not incorporates an embedded HP filter, thus the data are filtered in a post-processing algorithm with high-pass corner frequency $F_{c(hp)} = 100\,\mathrm{Hz}$. The complete overview of used devices and software is depicted in Fig. 5.1.

The sound pressure is sensed by the array microphones and transferred by a DAQ over LABVIEW into the PC unit. A wanted signal to be generated by compression drivers is send through a DAQ and multichannel amplifier. A special care had to be devoted to the grounding of the equipment, metal contraction of the microphone frame, but also the anechoic chamber floor composed of a wired grid. The electric noise was reduced this way, but still certain distortion has been evident.

The sound sources are located relatively close to the probe in distance between $0.9\,\mathrm{m}$ and $1.5\,\mathrm{m}$. We are interested in sensing the active sound intensity, which was observed by Jacobsen [14] to be further then $0.5\,\mathrm{m}$ from an acoustic source. Hereby we avoid reactive intensity generated by a source by measuring in safe distance from the source. Another constrain of the set up arrangement is also determined by the BASS method, since large distances from sources to microphones degrades the separation performance. This was not a real limitation in the anechoic chamber, but it should be bore in mind for real-world applications.

5.2 Single source detection

The positions of the sound sources and SNR were similar to the simulation configuration, so that results could be compared. The signals used in the experiments were the same as in the simulation, but all of them are are slightly time-shifted and the low frequency end is cut off. The measurement for single sources were conducted in an anechoic chamber with a cut off frequency about $150\,\mathrm{Hz}$, where an impulse response of

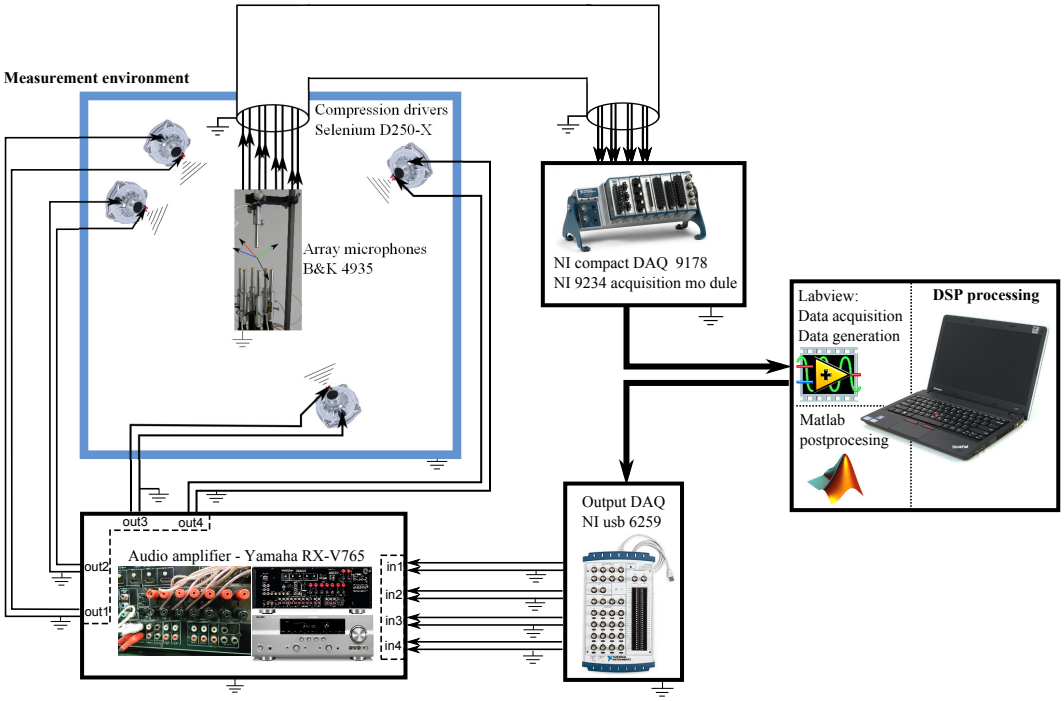

Figure 5.1: Measurement arrangement.

the chamber is supposed to be negligible. The real measurement conditions introduced a background noise into the data acquisition. The probe construction could affect recorded signals by diffraction, scattering and possible inaccuracy of the angle detection also arose from an inaccuracy of microphones placement and thus an imprecise position calibration of the probe was brought to a true source angle detection. The measured single sources playing solo in anechoic conditions are presented in Fig. 5.2.

The position of sound sources is calibrated by an angle detection and it is exactly the same as for the simulation. The measurement set-up and exact position is depicted in Fig. 5.3. Each angle is represented by a colour and also real position in anechoic chamber is depicted by the coloured squares which hold the sound source specification. The microphone array is marked in purple and it is represented by 7 microphones forming double-twisted tetrahedron configuration. The measuring equipment was placed inside the chamber for easier manipulation. This could cause an additional noise and sound reflections which should not be very noticeable though.

The comparison of the signals is shown in Fig. 5.4. The envelopes of the sources and Fourier transforms are slightly different compared to the simulation. It was caused by transferring the electrical signal into a physical quantity by transducers that do not have an ideal and flat frequency response.

The SNR was tried to be kept the same, but it was difficult due to the different frequency characteristics of the transducers. Still, the details about the sources are mentioned in Tab. 5.2, where original angles and estimated SNRs are mentioned.

The reference angles which we follow came from the sound intensity measurement for single sources plotted in Fig. 5.5. This angle detection is proved to be reasonably accurate with error of tenths or units of degree for single source detection after an averaging is applied.

We can see that the angle detection has some rapid fluctuations as well as a slow fluctuations around the

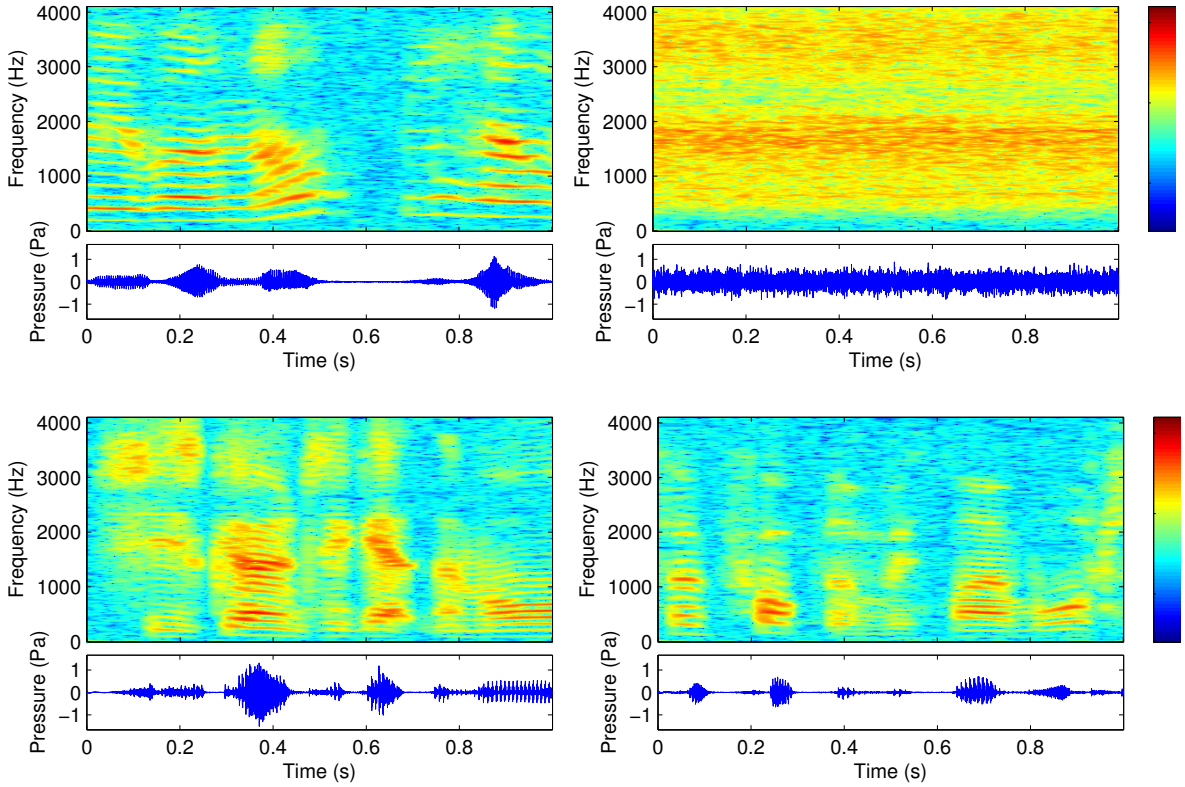

Figure 5.2: Time and spectrum analysis of single signals measured in anechoic chamber. Speech signal s_1 (top left), Gaussian noise s_2 (top right), speech s_3 (bottom left), speech s_4 (bottom right)

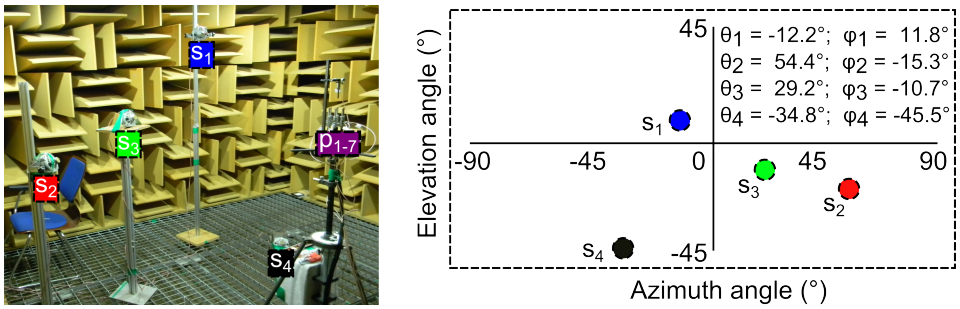

Figure 5.3: Source positions at anechoic experiment.

Sound sources	Colour	Type	SNR (dB)	Azimuth angle $\theta(^\circ)$	Elevation angle $\phi(^\circ)$
s_1	Blue	Speech 1	34.2	-12.2	11.8
s_2	Red	Gaussian noise	33.8	54.4	-15.3
s_3	Green	Speech 2	36.8	29.2	-10.7
s_4	Black	Speech 3	32.6	-34.8	-45.5

Table 5.1: Sound sources information

truth angle. The frequency of the slow fluctuations is increasing with a frequency and it has not been found out yet what is the cause of such error which is not present at the simulation results. The Tab. 5.2 presents the information of measured single sources. The most important is only the standard deviation which has increased rapidly when compared to the simulation results in Tab. 4.2.

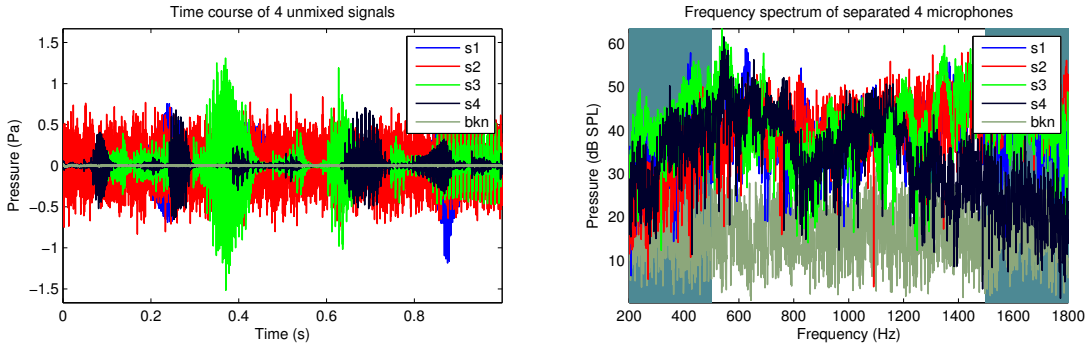

Figure 5.4: Time course and FFT of overlaying unmixed source signals measured in anechoic experiment.

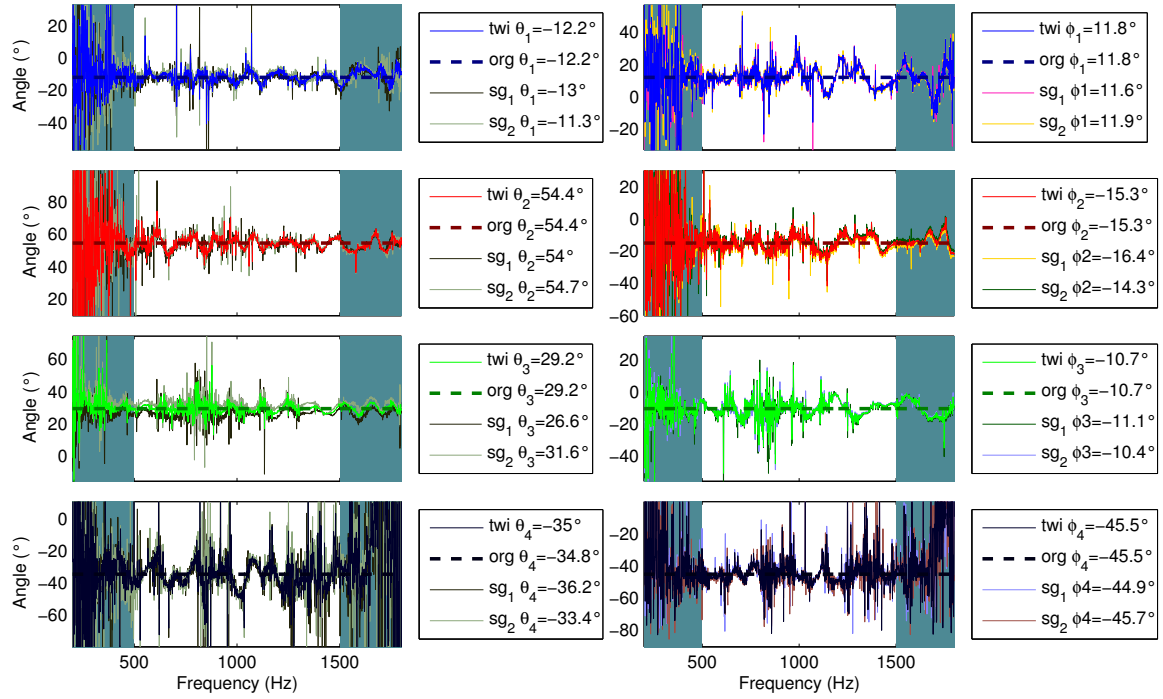

Figure 5.5: Angle detection of single sources played solo in anechoic chamber. An angle range is 45° about the original angles.

Sound source	SNR (dB)	Original $\theta_{org}(^\circ)$	Twisted $\theta_{twi}(^\circ)$	Error $\Delta\theta_{twi}(^\circ)$	Std. dev. σ°_θ	Original $\phi_{org}(^\circ)$	Twisted $\phi_{twi}(^\circ)$	Error $\Delta\phi_{twi}(^\circ)$	Std. dev. σ°_ϕ
s_1	34.2	-12.2	-12.2	0	3.8	11.8	11.8	0	6.3
s_2	33.9	54.4	54.4	0	3.7	-15.3	-15.3	0	5.3
s_3	36.6	29.2	29.2	0	3.1	-10.7	-10.7	0	5.5
s_4	32.5	-34.8	-34.8	- 0	14.1	-45.5	-45.5	0	8.5

Table 5.2: Angle detection information of single unmixed sources in anechoic conditions

The angle-frequency figures are transformed into histograms as in Figs. 5.6 and 5.7. The 2D histograms shows us that the distribution approaches normal distribution and therefore a mean is taken for detecting an angle.

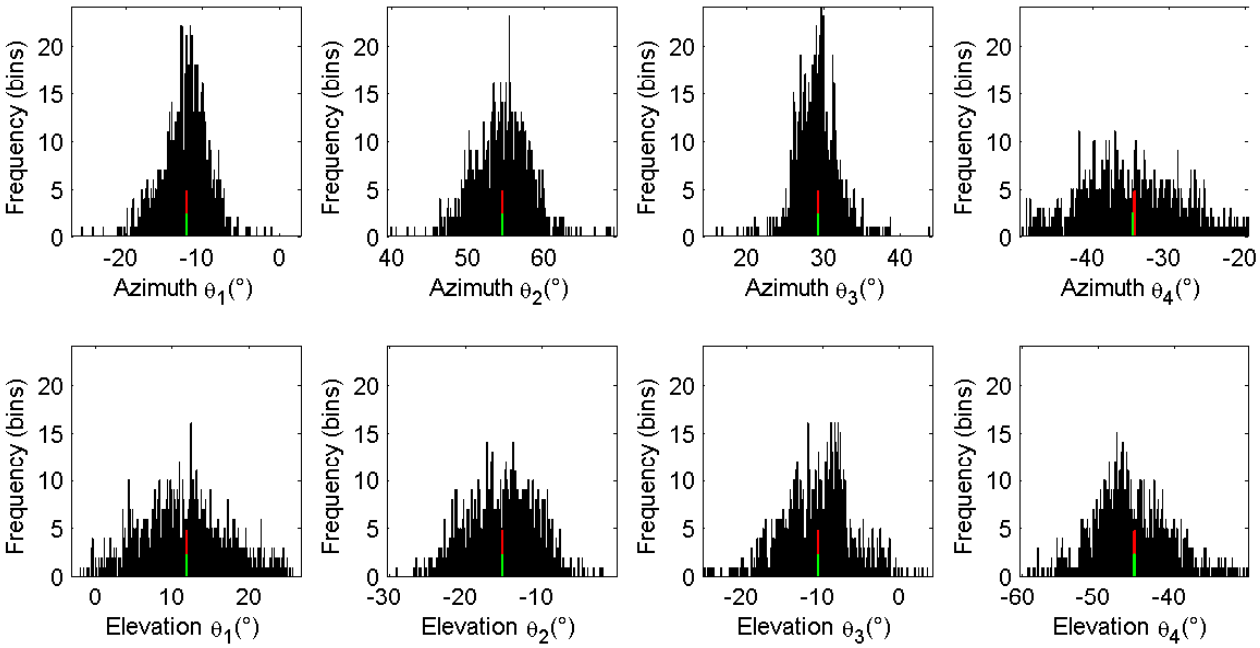

Figure 5.6: 2-dimensional histogram of azimuth angle detection at the top and elevation angle at the bottom row for single sound sources in anechoic conditions. Each jth source corresponds to jth row. Red and blue vertical line marks original and estimated angles

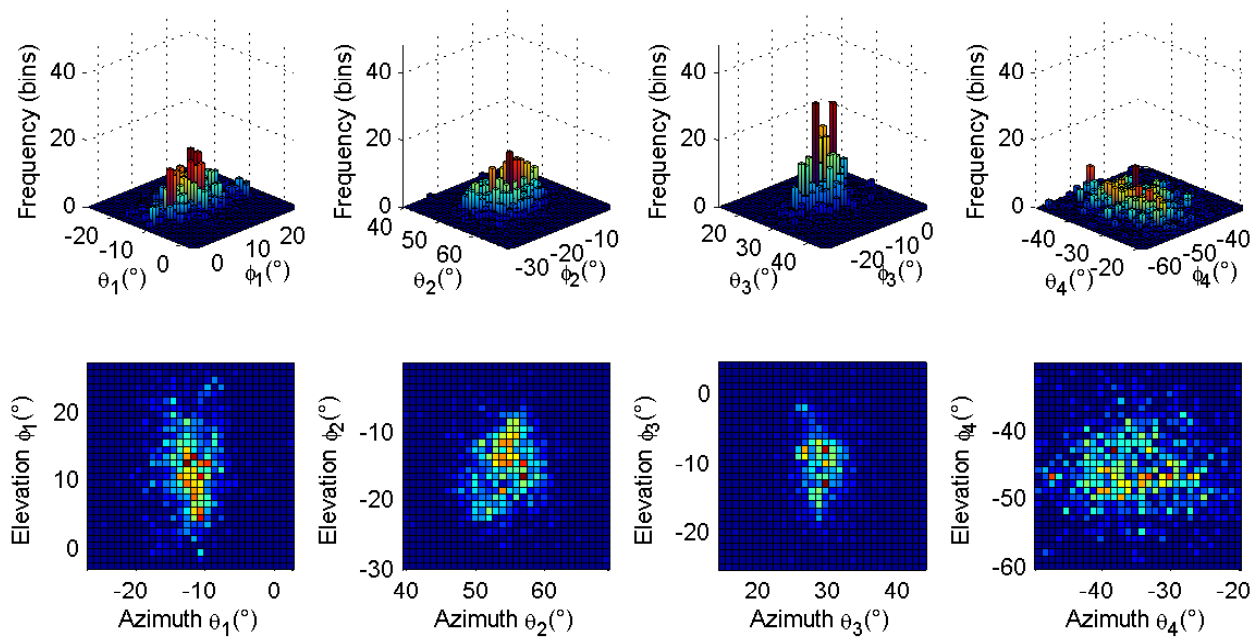

Figure 5.7: 3-dimensional histogram of azimuth and elevation angle detection for single sound sources in anechoic conditions. Top row is isometric projection, the bottom row is view from the top. Each jth source corresponds to jth row.

5.3 Multiple source detection

An anechoic chamber which was used with the same set-up as for the single sound source measurement. The errors of the probe construction were partially compensated for by measuring an averaged angles of a single source, which are taken as a reference, but this value is slightly biased by intensity measurement errors. For this reason, the accuracy is limited only to the comparison with the single sources.

The evaluation of the multiple sources follows the same order as for the simulation, so it will be presented here only briefly since the assessment approach has been discussed in Sec. 4.3. Firstly, we look at the similarity measure among clusters detected by BASS. In contrast to the simulation, the similarity here is either very high or very low what is reflected by the highest red value or the lowest blue value. This either predicts good independence of each cluster or large difference between the most and the least similar estimate. Compared to the simulation, we can subjectively say that the separation is slightly worse then for the simulation example.

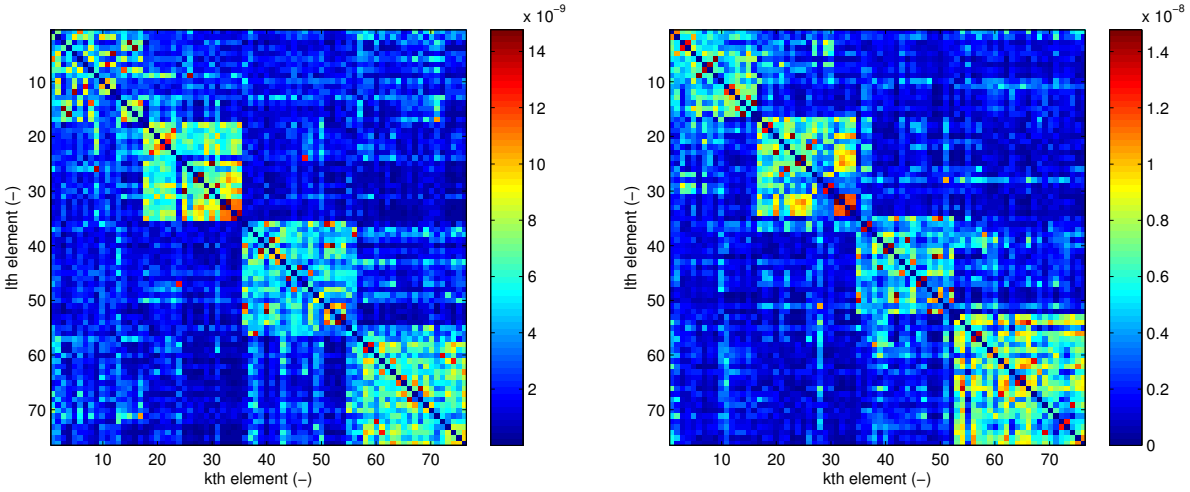

Figure 5.8: Clustering results of BASS distance matrix in anechoic conditions with dimensions $mL \times mL$ for 1^{st} tetrahedron array on the left and 2^{nd} on the right. For both plots, 1^{st} cluster at upper left represent signals is most similar to v_3, 2^{nd} corresponds to v_4, 3^{rd} to v_1 and 4^{th} to v_2.

The clusters later results in separated sources which are quite well separated, but a moderate contribution of a noise is detected in all separated sources. The magnitude scale of the separated sources is again changed.

The angle detection of separated sources plotted in Fig. 5.10 shows similar error trend as in the simulation, but the error is more enhanced for the anechoic conditions.

The source v_3 is again problematic since it converges towards the Gaussian source v_2. Also the source v_4 is prone to the error due to its low SNR and therefore inaccurate sound source separation as well as distorted angle detection over sound intensity. The Tab. 5.3 sums up the angle detection measures.

Sound source	SNR (dB)	Original $\theta_{org}(^\circ)$	Twisted $\theta_{twi}(^\circ)$	Error $\Delta\theta_{twi}(^\circ)$	Std. dev. σ°_θ	Original $\phi(^\circ)$	Twisted $\phi_{twi}(^\circ)$	Error $\Delta\phi_{twi}(^\circ)$	Std. dev. σ°_ϕ
s_1	34.2	-12.2	-14.2	-2	5.8	11.8	13.6	1.8	3.6
s_2	33.8	54.4	55.4	1	4	-15.3	-15.3	0	1.5
s_3	36.8	29.2	31.7	2.5	6.3	-10.7	-11.7	-1	2.4
s_4	32.6	-34.8	-26.1	8.7	16.9	-45.5	-46.2	-0.7	5.3

Table 5.3: Summary of angle detection measures for multiple sources.

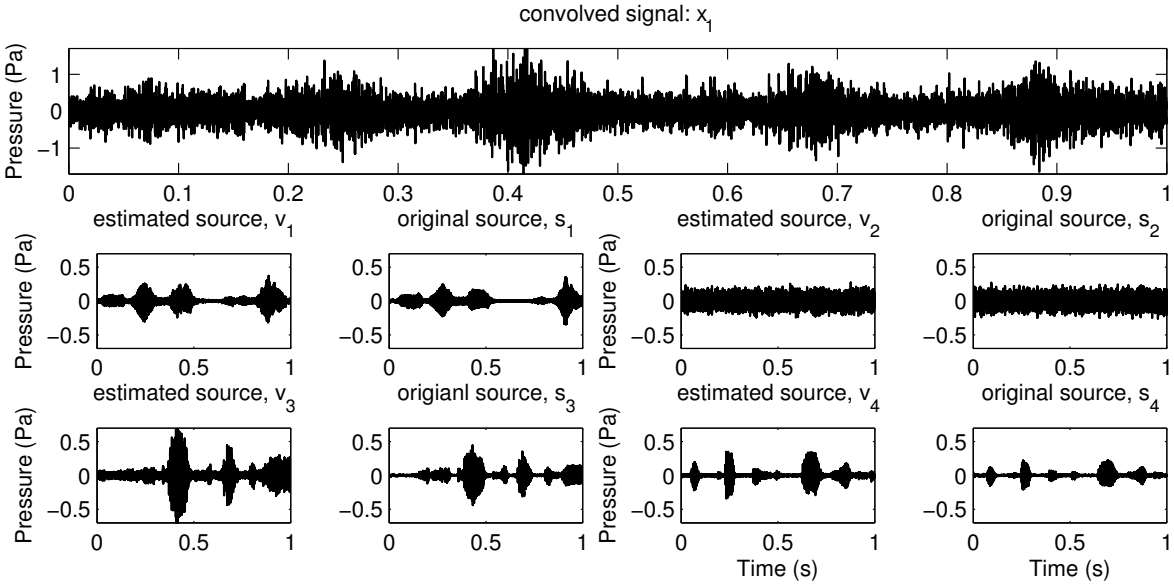

Figure 5.9: Separation results of mixed signals in time domain at 1 microphone.

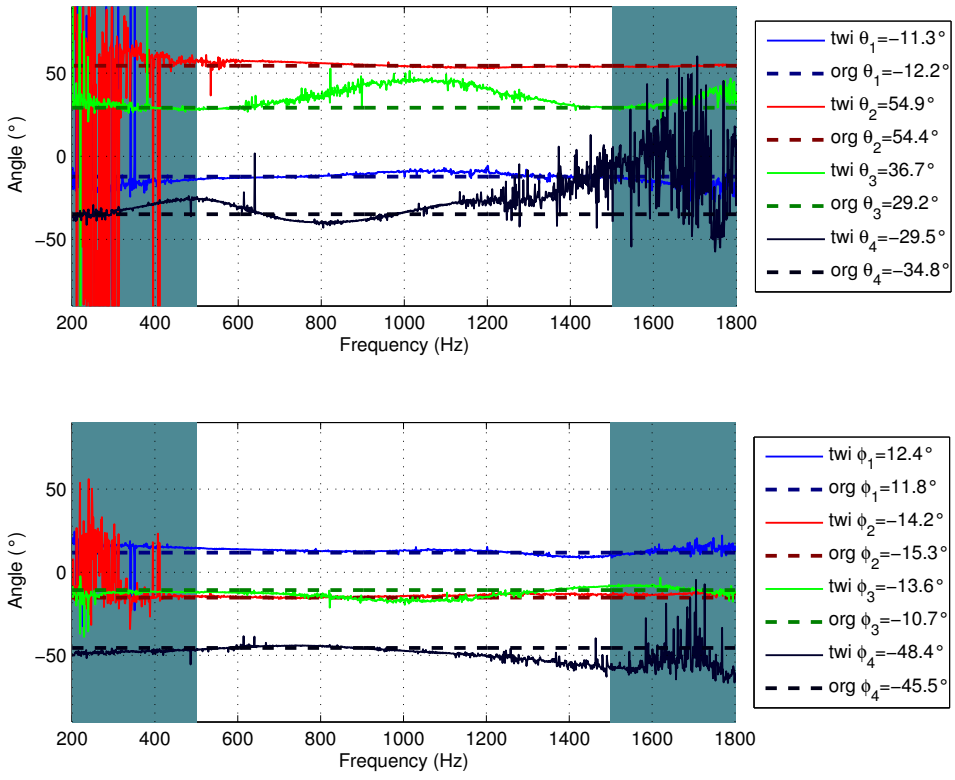

Figure 5.10: Angle estimation of multiple separated signals.

It is shown that the error gets high in experimental conditions supposing the SNR is low. This result in error about 9 degree. This error can be anticipated by the high standard deviation. The other sources have got reasonable error within a few degrees. The detail of the sources is presented in Fig. 5.11 and the influence of double-twisted tetrahedron array can be evaluated. The combination of the single tetrahedron

probes sometimes correct for the error, but it is not a rule.

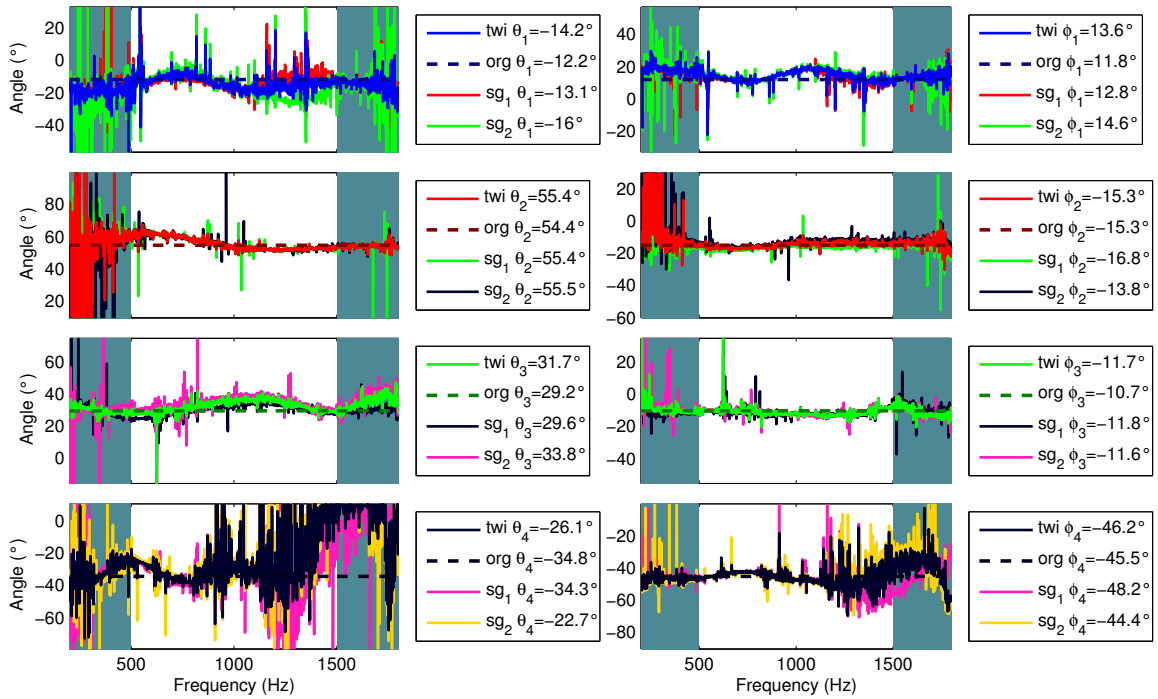

Figure 5.11: Angle detection of separated sources in anechoic conditions. An angle range is 45° about the original angles.

The histograms shows that a compromise has to be made in angle evaluation, but a precision is decreased. The distribution usually have two peaks along the angle axis and usually non of the peaks correspond exactly to the true angle.

We can see how the estimated sound source v_3 always incline to the estimate v_2 and how the signal v_4 is dispersed. The signal v_1 converts towards no source, what is an interesting note.

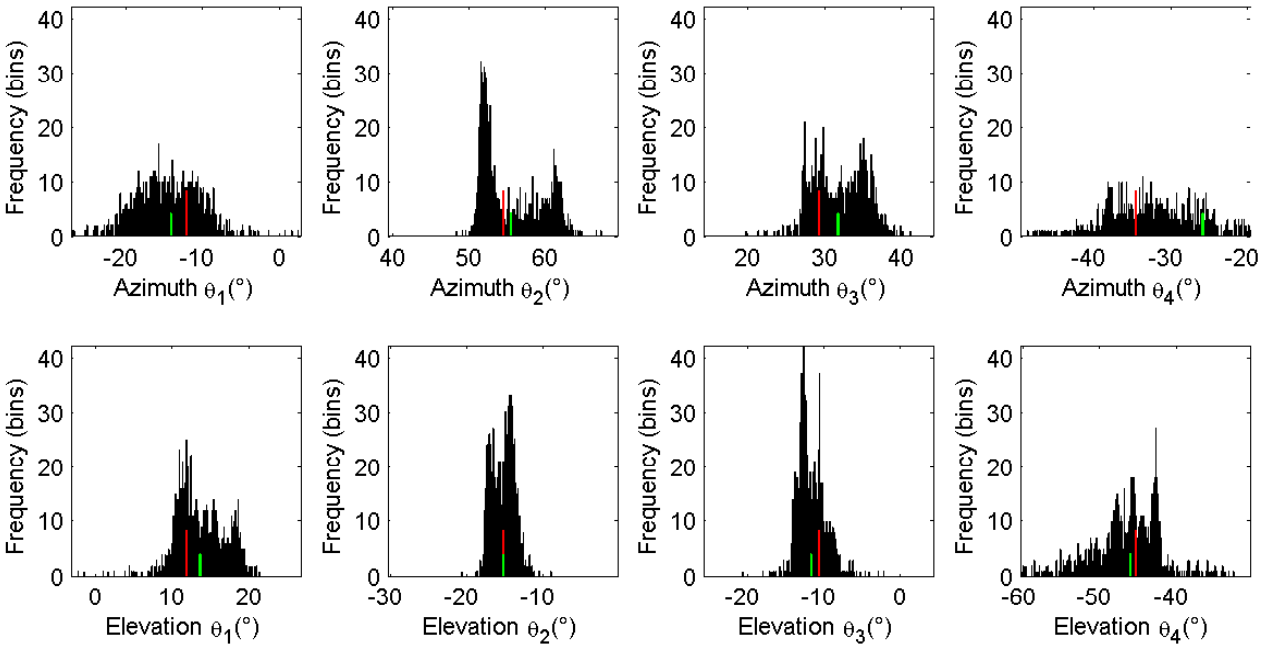

Figure 5.12: 2-dimensional histogram of azimuth angle detection at the top and elevation angle at the bottom row for separated sound sources. Each jth source corresponds to jth row. Red and blue vertical line marks original and estimated angles

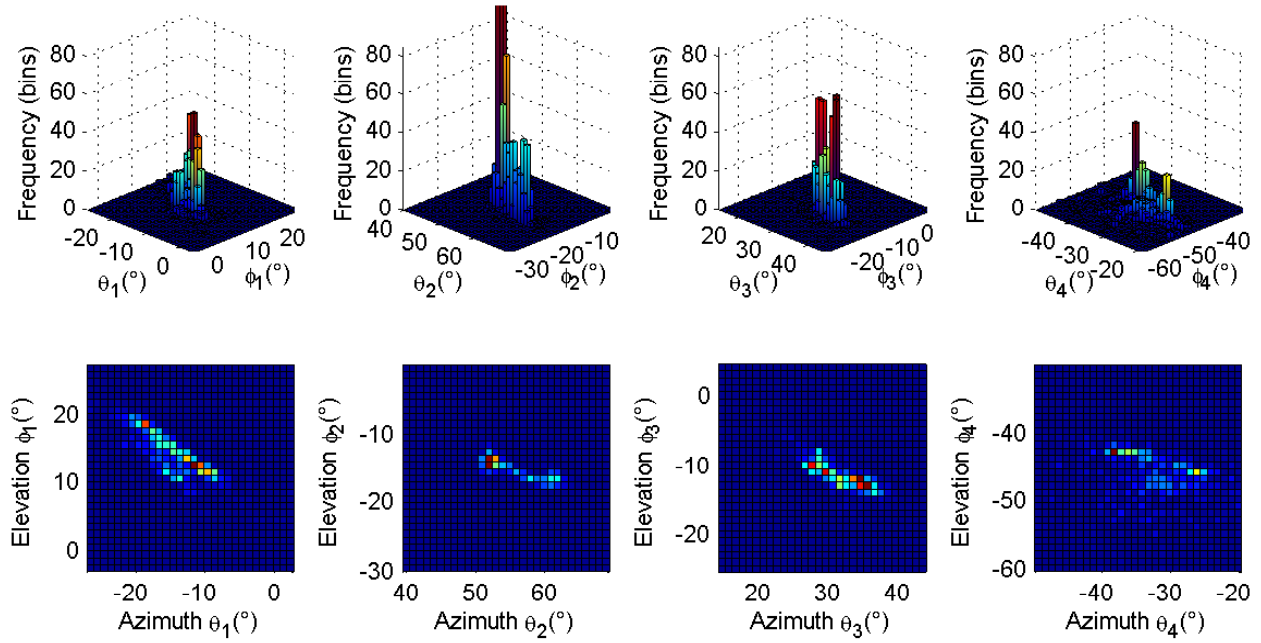

Figure 5.13: 3-dimensional histogram of azimuth and elevation angle detection for separated sound sources. Top row is isometric projection, the bottom row is view from the top. Each jth source corresponds to jth row.

CHAPTER 6

Conclusion

The problem of the master thesis was stated to be focused on 4 independent source. So far, 4 sources has been separated over the simulation and then over the experiment, where additional inaccuracy has occurred. It was although proven that it is possible to distinguish an angle of incoherent sound sources by the time-domain BSS. The comparison between the simulation and experimental results suggests us that the increase of error can arise from factors as scattering, diffraction, microphone positioning and background noise, because all these can worsen the accuracy. Several statistical measures were suggested for a performance evaluation of detected sources, so we can rank the separated sources according to their SPL magnitude and quality of separation.

CHAPTER 7

Further Work

The experiment was done in an anechoic chamber so far, so real conditions with reverberant room will be evaluated. Also more incoherent sources can be considered to detect. For the double-twisted array it could be 7 sources, but with adding the microphones, the computational load will be increasing, so it should be taken into consideration and evaluate whether the detection of a few most significant source is to be separated within shorter time segment, or whether a maximum sources are desired to be separated for the price of a higher computational load. The performance ranking is dependent mostly on 2 factors, a spread of SPL magnitude along a time frame and a quality of separated sources. Therefore it needs to be distinguished in order to prevent ambiguity in evaluation.

Appendices

APPENDIX A

Appendix

A.1 Directivity characteristics

The directivity simulation was implemented by Cho[6] and it is approached by assuming a plane wave and calculating phase difference between array microphones. The angle information is calculated from Eq. 3.16 and resulting directivity is plotted over a solid angle. 2D polar plots for 3 cross-sections of double-twisted tetrahedron array are plotted in the A.1. The directivity for d= 50 mm is very omnidirectional, therefore the radial coordinate of the plots were adjusted to get greater detail.

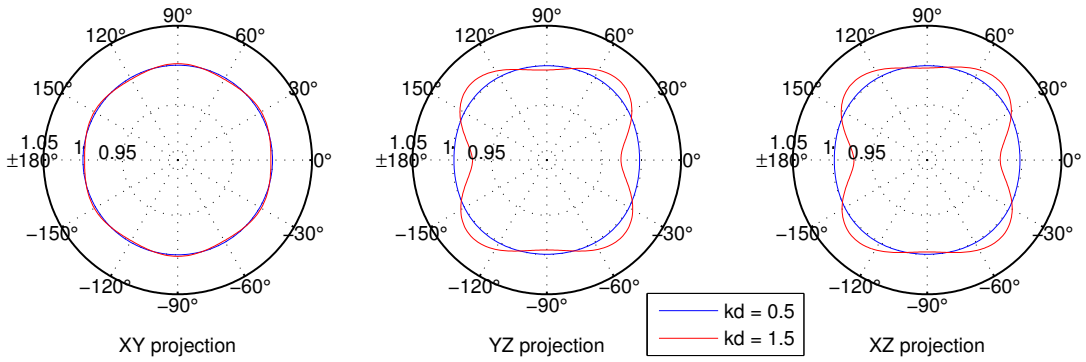

Figure A.1: 2D polar plots of double-twisted tetrahedron array

The values kd were chosen to depict limits of our effective frequency bandwidth. $kd = 0.5$ corresponds to 545 Hz and $kd = 1.5$ shows directivity for 1637 Hz. It can be said that the configuration is omnidirectional for our interest. If we zoom, lower frequencies are still omnidirectional, but with increasing frequency, directivity lobes starts to form and increase with increasing frequency. Notice also that the directivity in XY projection is very uniform and thus it also shows the best accuracy in angle detection. As the elevation angle is increasing to positive or negative values up to $90°$, the bias in the intensity calculation becomes to be stronger and it is reaching its minima ($0°$ and $\pm 180°$) and maxima ($\pm 45°$ and $\pm 135°$).

To get the whole picture of the directivity pattern, 3D plots are shown in Figs. A.2 and A.3 for $kd = 0.5$ and $kd = 1.5$, respectively.

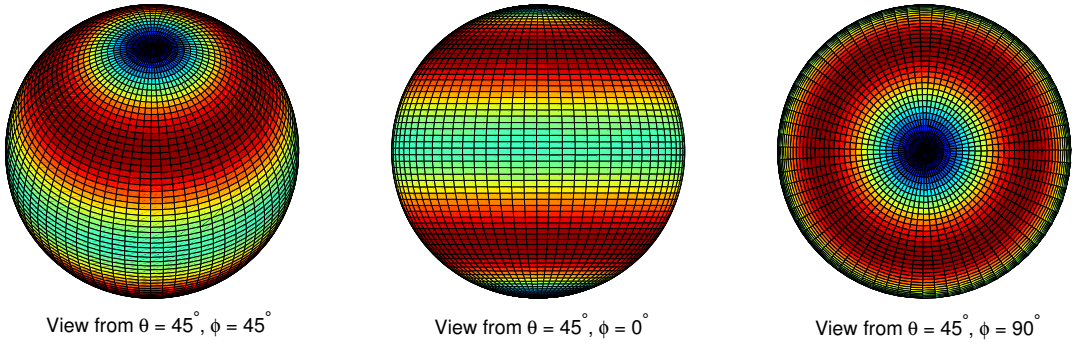

Figure A.2: 2D polar plots of double-twisted tetrahedron array for $kd = 0.5$

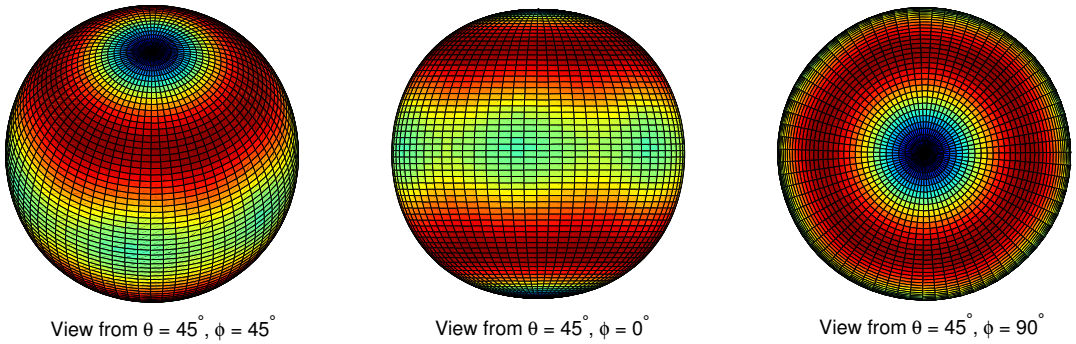

Figure A.3: 2D polar plots of double-twisted tetrahedron array for $kd = 1.5$

A.2 Angle detection of single sources

The simulation of angle detection without background noise present is depicted in Fig. A.4. Each row of the figure correspond to jth source. The columns represent, from left to right, azimuth and elevation angle evaluation. In each plot, the dashed line correspond to the truth angle, the legend notes double-twisted tetrahedron configuration (twi), and single tetrahedron configurations (sg_1, sg_2).

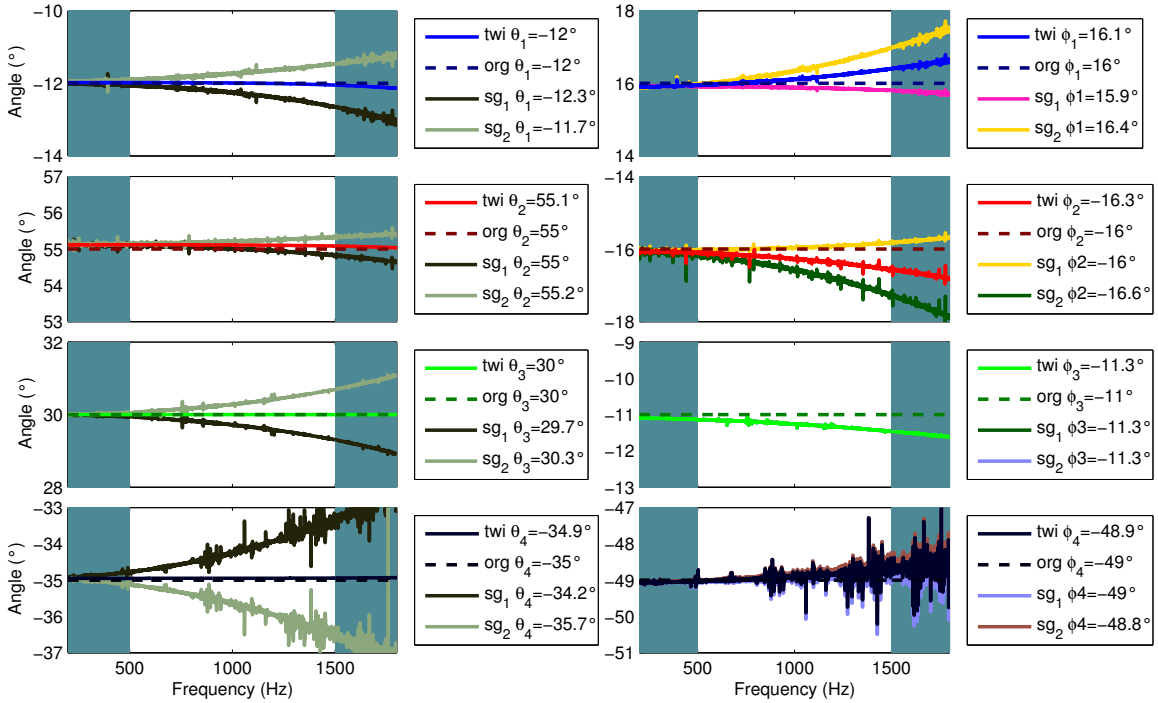

Figure A.4: Angle detection of single sources played solo with background noise. An angle range is 2° about the original angles.

The error is more pronounced as frequency increases due to finite difference error. It is seen that some angles are corrected by applying double-twisted tetrahedron configuration and some angles do not show almost any or no improvement. This can be explained by directivity patterns, since some angles has got the same directivity for single and double-twisted configurations.

A.3 Magnitude normalization of separated signals

A magnitude normalization was tried to be applied to the signals, so that the lost scaling information from the BASS can be partially recovered. The normalization is performed in the way, that each data point of the estimated signal is divided by a maximum absolute value of the same signal. It was observed that the temporal envelopes of the estimated signals do have similar shape and therefore this normalization is not affected by random peaks that would biased the normalization.

$$v_{j,norm}(n) = \frac{v_j(n)}{max(|v_j|)} \tag{A.1}$$

In the case sharp peaks would occur, different normalization which ignores outliers would have to be applied to obtain unbiased results.
The detail of the time signal through out the process of resampling and separation is depicted in Fig. A.5. The Figure is similar to the Fig. 4.5 in section 4.1, but now it is plotted for 4 microphones at once and additionally, the recovered signals after separation are analysed as well.

On the left top is the resampled and shifted signals at 4 microphone positions. The sound wave of the source 1 firstly arrives to the 3^{rd} microphone, it is delayed by 52 samples at microphone m_2, by 67 samples at microphone m_3 and ultimately it the sound wave impinges on microphone m_1 after 536 samples. It is seen that down-sampled data hold similar shape of the data and after the resampling, the shape still remains similar, although a distortion in the local waveform shape can be noticeable. For this reason, the normalization is applied to correct for this imperfection. An improvement due to this step is not very visible locally, but the normalization evens out the differences in amplitudes.

When source angle is detected by the normalized signals, we obtain the Figs. A.6, A.7 and A.8. The details of the figures are summarized in Tab. A.1.

The normalization has almost no effect on the azimuth angle detection, but it strongly effect the elevation angle detection. From the Tab. A.1, we can note that the standard deviation σ_ϕ° has increased rapidly and it could bring uncertainty into the evaluation. Although by looking at the elevation angle error, the values remains similar when compared data with and without normalization. There can be seen slight improvement in elevation angle detection of source v_2, but also degradation of angle precision for source v_1.

Sound source	SNR (dB)	Original $\theta_{org}(^\circ)$	Twisted $\theta_{twi}(^\circ)$	Error $\Delta\theta_{twi}(^\circ)$	Std. dev. σ_θ°	Original $\phi(^\circ)$	Twisted $\phi_{twi}(^\circ)$	Error $\Delta\phi_{twi}(^\circ)$	Std. dev. σ_ϕ°
s_1	34	-12.2	-12.4	-0.2	0.8	11.8	12.9	1.1	4.6
s_2	34	54.4	55.1	0.7	0.2	-15.3	-15.4	-0.1	9.5
s_3	36.3	29.2	34.3	5.1	2.5	-10.7	-10.1	0.6	17.3
s_4	32.3	-34.8	-35.4	-0.6	1.1	-45.5	-45.1	0.4	5.3

Table A.1: Summary of angle detection measures for multiple sources.

This method is not very transparent and the results may have slightly better angle detection, but it comes with a price for higher standard deviation. For these reasons, we do not consider this procedure for angle evaluation.

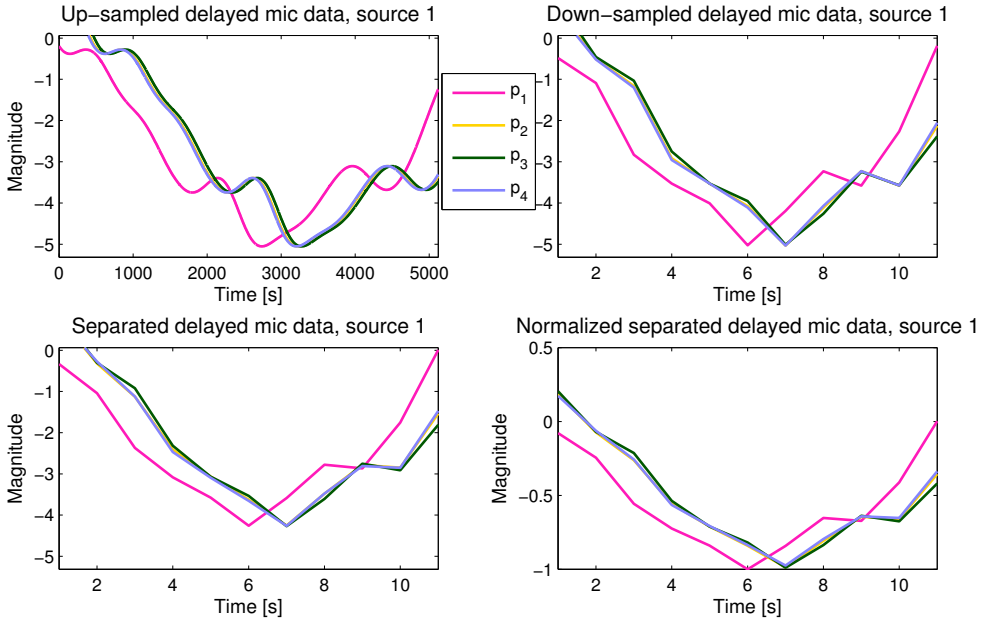

Figure A.5: Detail of resampling, separation and normalization process for source s_1.

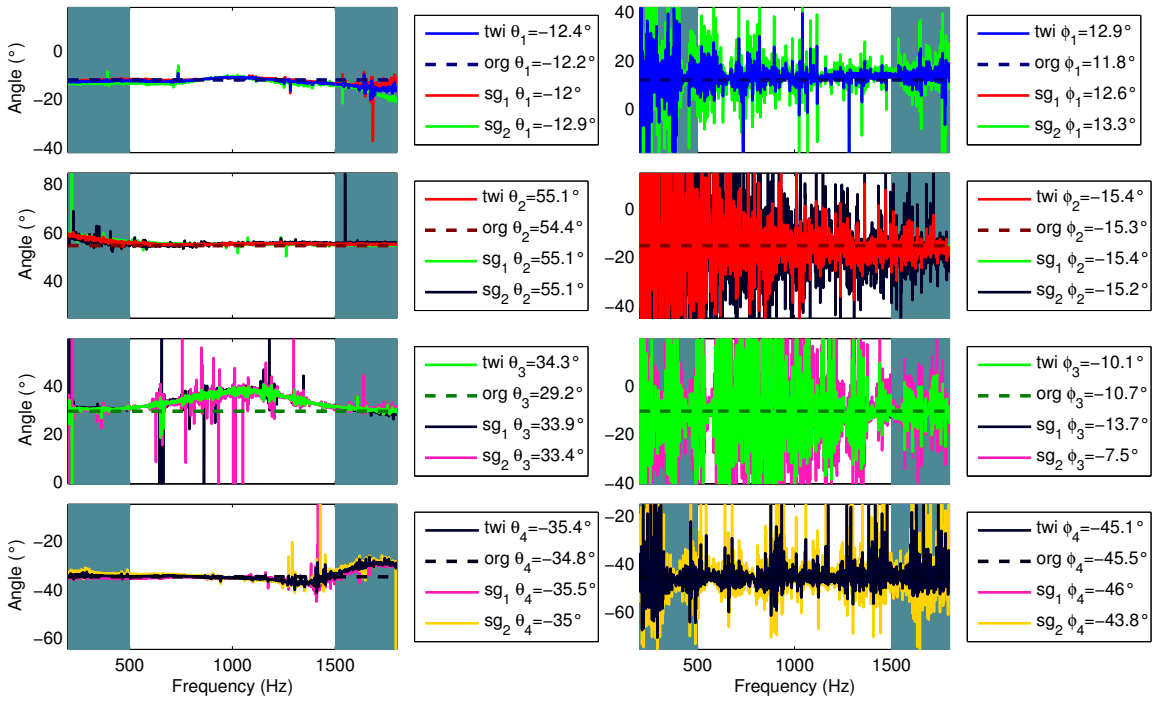

Figure A.6: Angle detection of normalized single sources played solo with background noise. An angle range is 90° about the original angles.

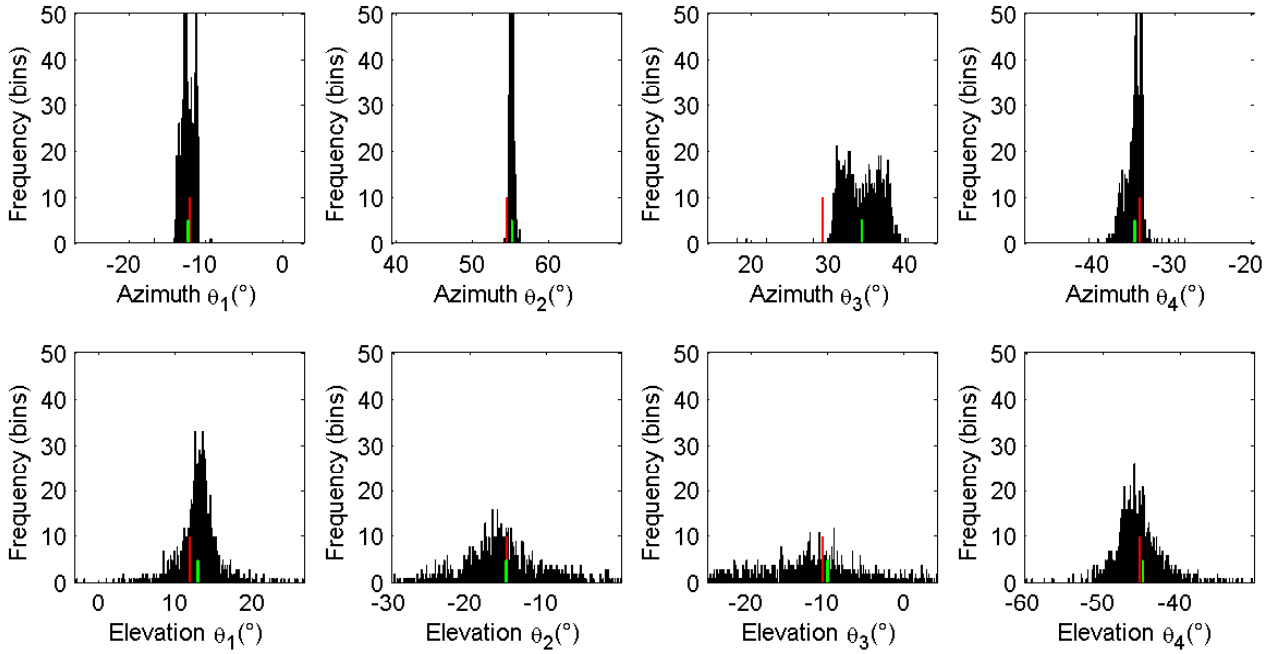

Figure A.7: 2-dimensional histogram of azimuth angle at the top and elevation angle at the bottom row for normalized single source playing solo. Each jth source corresponds to jth row. Red and blue vertical line marks original and estimated angles

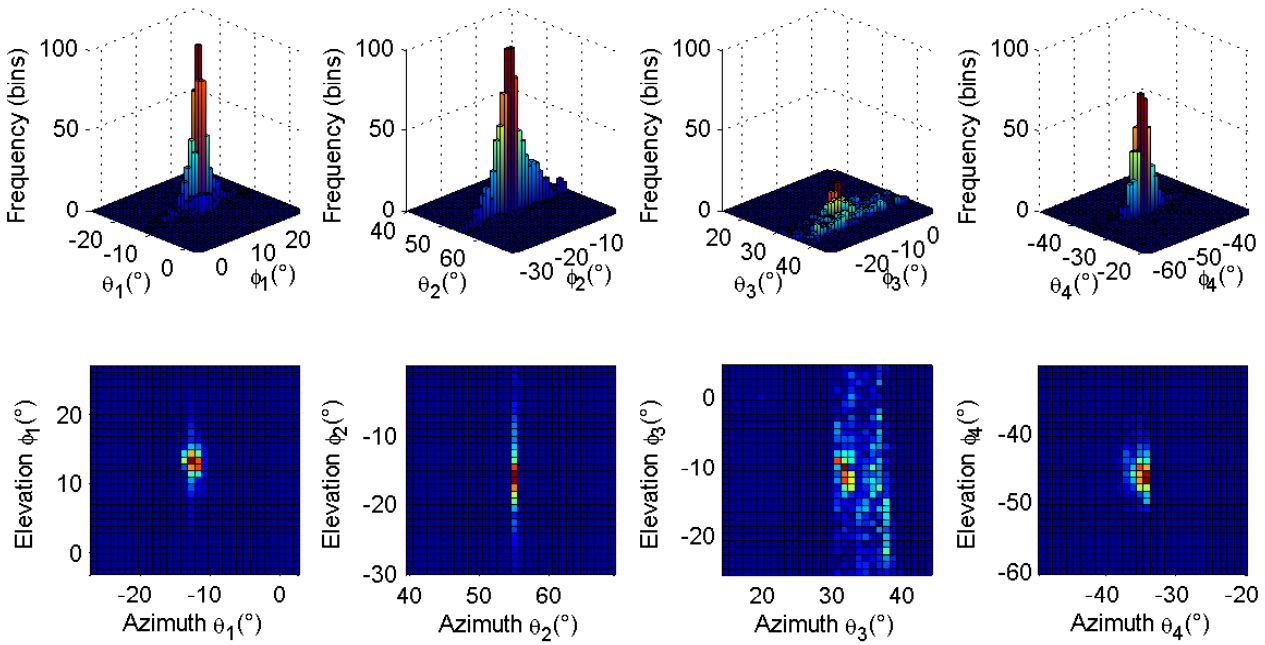

Figure A.8: 3-dimensional histogram of azimuth and elevation angle for normalized single sources detection playing solo. Top row is isometric projection, the bottom row is view from the top. Each jth source corresponds to jth row.

Bibliography

[1] Basten, T., de Bree, H., Druyvesteyn, W. and Wind, J.: 2009, Multiple incoherent sound source localization using a single vector sensor, *ICSV16, Krakow, Poland* .

[2] Bendat, J. S. and Piersol, A. G.: 2011, *Random data: analysis and measurement procedures*, Vol. 729, John Wiley & Sons.

[3] Buchner, H., Aichner, R. and Kellermann, W.: 2005, A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics, *Speech and Audio Processing, IEEE Transactions on* **1**3(1), 120–134.

[4] Cardoso, J.-F. and Souloumiac, A.: 1993, Blind beamforming for non-gaussian signals, *IEE Proceedings F (Radar and Signal Processing)*, Vol. 140, IET, pp. 362–370.

[5] Cazzolato, B. S. and Ghan, J.: 2005, Frequency domain expressions for the estimation of time-averaged acoustic energy density, *The Journal of the Acoustical Society of America* **1**17, 3750.

[6] Cho, S.-K.: 2012, *Acoustic source localization by using double-module intensity array*, Master's thesis, School of Mechanical, Aerospace & System Engineering, Division of Mechanical Engineering, KAIST.

[7] Chung, J.: 1978, Cross-spectral method of measuring acoustic intensity without error caused by instrument phase mismatch, *The Journal of the Acoustical Society of America* **6**4, 1613.

[8] Doron, E., Yeredor, A. and Jan, N.: 2005, Computationally feasible implementation of asymptotically optimal blind separation algorithm sobi (wasobi) for ar source, *ÚTIA AV ČR,(Praha 2005) Internal Publication* **2**9.

[9] Elko, G. W.: 1991, An acoustic vector-field probe with calculable obstacle bias, *Proceedings of Noise-Con*, Vol. 91, pp. 525–532.

[10] Fahy, F.: 2002, *Sound intensity*, Taylor & Francis.

[11] Fahy, F. J.: 1977, Measurement of acoustic intensity using the cross-spectral density of two microphone signals, *The Journal of the Acoustical Society of America* **6**2, 1057.

[12] Gómez-Herrero, G., Koldovskỳ, Z., Tichavskỳ, P. and Egiazarian, K.: 2007, A fast algorithm for blind separation of non-gaussian and time-correlated signals, *Proccedings of the 15th European Signal Processing Conference. EUSIPCO*, Citeseer, pp. 2007–07.

[13] Ih, J.-G., Woo, J.-H. and Cho, S.-K.: 2013, Acoustic source localization by using twisted double-module 3d intensity array, *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Vol. 247, Institute of Noise Control Engineering, pp. 2779–2784.

[14] Jacobsen, F.: 2003, Sound intensity and its measurement and applications, *Current Topics in Acoustical Research* **3**, 87–91.

[15] Jacobsen, F., Cutanda, V. and Juhl, P. M.: 1998, A numerical and experimental investigation of the performance of sound intensity probes at high frequencies, *The Journal of the Acoustical Society of America* **1**03(2), 953–961.

[16] Koldovskỳ, Z., Málek, J., Tichavskỳ, P., Deville, Y. and Hosseini, S.: 2009, Blind separation of piecewise stationary non-gaussian sources, *Signal Processing* **8**9(12), 2570–2584.

[17] Koldovsky, Z. and Tichavsky, P.: 2008, Time-domain blind audio source separation using advanced component clustering and reconstruction, *Hands-Free Speech Communication and Microphone Arrays, 2008. HSCMA 2008*, IEEE, pp. 216–219.

[18] Koldovsky, Z. and Tichavsky, P.: 2011, Time-domain blind separation of audio sources on the basis of a complete ica decomposition of an observation space, *Audio, Speech, and Language Processing, IEEE Transactions on* **1**9(2), 406–416.

[19] Koldovskỳ, Z., Tichavskỳ, P. and Málek, J.: 2010, Time-domain blind audio source separation method producing separating filters of generalized feedforward structure, *Latent Variable Analysis and Signal Separation*, Springer, pp. 17–24.

[20] Koldovsky, Z., Tichavsky, P. and Oja, E.: 2006, Efficient variant of algorithm fastica for independent component analysis attaining the cram&# 201; r-rao lower bound, *Neural Networks, IEEE Transactions on* **1**7(5), 1265–1277.

[21] Málek, J., Koldovskỳ, Z. and Tichavskỳ, P.: 2010, Adaptive time-domain blind separation of speech signals, *Latent Variable Analysis and Signal Separation*, Springer, pp. 9–16.

[22] Olson, H. F.: 1932, System responsive to the energy plow op sound waves. US Patent 1,892,644.

[23] Pascal, J.-C. and Li, J.-F.: 2008, A systematic method to obtain 3d finite-difference formulations for acoustic intensity and other energy quantities, *Journal of Sound and Vibration* **3**10(4), 1093–1111.

[24] Pavić, G.: 1977, Measurement of sound intensity, *Journal of Sound and Vibration* **5**1(4), 533–545.

[25] Pham, D.-T. and Cardoso, J.-F.: 2001, Blind separation of instantaneous mixtures of nonstationary sources, *Signal Processing, IEEE Transactions on* **4**9(9), 1837–1848.

[26] Primer, S. I.: 1993, Brüel & kjaer application notes.

[27] Rasmussen, G.: 1985, Measurement of vector fields, *Proceedings of the 2nd International Congress on Acoustic Intensity*, pp. 53–58.

[28] Santos, L., Rodrigues, C. and Bento Coelho, J.: 1989, Measuring the three-dimensional acoustic intensity vector with a four-microphone probe.

[29] Shirahatti, U. and Crocker, M. J.: 1992, Two-microphone finite difference approximation errors in the interference fields of point dipole sources, *The Journal of the Acoustical Society of America* **9**2(1), 258–267.

[30] Suzuki, H., Oguro, S., Anzai, M. and Ono, T.: 1995, Performance evaluation of a three dimensional intensity probe, *Journal of the Acoustical Society of Japan (E)* **1**6(4), 233–238.

[31] Tichavskỳ, P. and Koldovskỳ, Z.: 2011, Fast and accurate methods of independent component analysis: A survey, *Kybernetika* **4**7(3), 426–438.

[32] Tichavsky, P., Yeredor, A. and Koldovsky, Z.: 2009, A fast asymptotically efficient algorithm for blind separation of a linear mixture of block-wise stationary autoregressive processes, *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, IEEE, pp. 3133–3136.

[33] Williams, E. G. and Takashima, K.: 2010, Vector intensity reconstructions in a volume surrounding a rigid spherical microphone arraya), *The Journal of the Acoustical Society of America* **1**27(2), 773–783.

[34] Williams, E. G., Valdivia, N., Herdic, P. C. and Klos, J.: 2006, Volumetric acoustic vector intensity imager, *The Journal of the Acoustical Society of America* **1**20(4), 1887–1897.

[35] won Lee, T., Bell, A. J. and Lambert, R. H.: 1997, Blind separation of delayed and convolved sources.

[36] Yntema, D., Wiegerink, R., Van Honschoten, J. and Elwenspoek, M.: 2007, Fully integrated three dimensional sound intensity sensor, *Micro Electro Mechanical Systems, 2007. MEMS. IEEE 20th International Conference on*, IEEE, pp. 51–54.