ENTROPY-BASED DIRECTION-OF-ARRIVAL ESTIMATION METHODS FOR RIGID SPHERICAL MICROPHONE ARRAYS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ORHUN OLGUN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MODELLING AND SIMULATION

JULY 2019

Approval of the thesis:

ENTROPY-BASED DIRECTION-OF-ARRIVAL ESTIMATION METHODS FOR RIGID SPHERICAL MICROPHONE ARRAYS

submitted by **ORHUN OLGUN** in partial fulfillment of the requirements for the degree of **Master of Science in Modelling and Simulation Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, Graduate School of **Informatics** _____

Assist. Prof. Dr. Elif Sürer
Head of Department, **Modelling and Simulation, METU** _____

Assoc. Prof. Dr. Hüseyin Hacıhabiboğlu
Supervisor, **Modelling and Simulation, METU** _____

Examining Committee Members:

Prof. Dr. Aydın Alatan
Electrical & Electronics Engineering, METU _____

Assoc. Prof. Dr. Hüseyin Hacıhabiboğlu
Modelling and Simulation Department, METU _____

Assoc. Prof. Dr. Berke Gür
Mechatronics Engineering, BAU _____

Assist. Prof. Dr. Elif Sürer
Modelling and Simulation Department, METU _____

Prof. Dr. Alptekin Temizel
Modelling and Simulation Department, METU _____

Date: 26.07.2019

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Orhun Olgun

Signature :

ABSTRACT

ENTROPY-BASED DIRECTION-OF-ARRIVAL ESTIMATION METHODS FOR RIGID SPHERICAL MICROPHONE ARRAYS

Olgun, Orhun

M.S., Department of Modelling and Simulation

Supervisor: Assoc. Prof. Dr. Hüseyin Hacıhabiboğlu

July 2019, 35 pages

Direction-of-arrival (DOA) estimation of sound sources is a popular research topic and has several different applications including spatial audio. Recent advances in microphone arrays made more accurate sound field analysis possible. Spherical microphone arrays afford a trivial calculation of spherical harmonic decomposition of sound fields and can be employed in different DOA estimation methods in spherical harmonics domain.

This thesis proposes two extensions to the a novel DOA estimation method called Hierarchical Grid Refinement (HiGRID) for rigid spherical microphone arrays (RSMA). HiGRID is based on the calculation of the sector averaged directional response power of a steered beam over a sparse set of directions on the unit sphere. The selection of the direction for which response power is to be calculated is determined using spatial entropy as a criterion. This is followed by clustering of the resulting DOA map using a method based on connected components labelling is also proposed for counting sources and estimating their DOAs.

This thesis also investigates the extensions of several state-of-the-art DOA estimation techniques. These include the improvement of DOA estimation performance or computational efficiency of Eigenbeam Multiple Signal Classification (EB-MUSIC) and Direct Path Dominance (DPD) test. HiGRID is first used as source counting method prior to EB-MUSIC to decrease the computational cost of DOA estimation. HiGRID is then used as a DOA estimation method following the DPD test which increases the DOA estimation accuracy while reducing the total computational cost. A new data-driven statistical method for DPD test threshold selection is also proposed.

This allows the an informed selection of DPD test threshold based on effective rank statistics of spatial correlation matrices obtained from RSMAs.

Comparison of HiGRID with previous DOA estimation methods with real and simulated recordings are presented. Evaluations of proposed algorithms for EB-MUSIC and DPD test are also presented in terms of DOA estimation errors using simulated recordings. HiGRID and its combinations with EB-MUSIC and DPD test performed favourably in comparison with other state-of-the-art DOA estimation methods indicating the utility of the proposed methods in multiple source DOA estimation.

Keywords: direction-of-arrival estimation, spherical harmonics, spherical microphone arrays, source localisation

ÖZ

MİKROFON DİZİNLERİ İÇİN ENTROPİ TEMELLİ VARIŞ YÖNÜ KESTİRME YÖNTEMLERİ

Olgun, Orhun

Yüksek Lisans, Modelleme ve Simülasyon Anabilim Dalı Bölümü

Tez Yöneticisi: Doç. Dr. Hüseyin Hacıhabiboğlu

Temmuz 2019 , 35 sayfa

Ses kaynaklarının varış yönünü kestirme popüler bir araştırma konusudur ve farklı uygulamalarda önemli rolü vardır. Küresel mikrofon dizinlerinin gelişmesi ile daha doğru varış yönü kestirme ve ses ortamı analizi mümkün olmuştur. Küresel mikrofon dizinleri küresel harmoniklerin hesaplamasını kolaylaştırmış ve bu küresel harmonik alanında ses kaynağı tespiti yapan yöntemlerin geliştirilmesini sağlamıştır.

Bu tez çalışmasında özgün bir varış kestirme yöntemi sunulmuştur. Önerilen yöntem küresel mikrofon dizinleri için geliştirilmiş olup entropi temelli hiyerarşik sistem düzenlemesi (HiGRID) olarak tanımlanabilir. HiGRID, birim kürede için yönsel güç yanıtlarının sektörel bölgelerde ortalamaların hesaplanmasıyla gerçekleşmektedir.

HiGRID yöntemine ek olarak bu yöntemin eigen sinyal bazlı çoklu sinyal sınıflandırması (EB-MUSIC) ve baskın doğrusal yol (DPD) testi ile kombinasyonları da sunulmuştur. EB-MUSIC ile kullanılan HiGRID ses kaynağı sayma yöntemi olarak kullanılmıştır. Diğer yöntemde ise HiGRID, DPD testinden sonra uygulanmış ve toplam işlem yükü azaltılmıştır. Ayrıca daha doğru DPD testi için istatiksel bir eşik belirleme yöntemi önerilmiştir.

HiGRID yönteminin güncel yöntemlerle karşılaştırması sunulmuştur. Bu karşılaştırma simüle edilen ses sahmelerindeki ses kaynaklarının tespitindeki hataların hesaplanması şekilinde gerçekleştirilmiştir. HiGRID'in EB-MUSIC and DPD testi ile kombinasyonlarının değerlendirmesi yinesimule edilmiş farklı senaryolarda ses kaynaklarını tespitindeki hataların hesaplanması ile yapılmıştır. Varış yönü kestirme performansı açısından bakıldığında ve güncel yöntemlerle karşılaştırıldığında HiGRID ve kombinasyonları için olumlu sonuçlar elde edilmiştir.

Anahtar Kelimeler: varış yönü kestirme, küresel harmonikler, ses kaynağı tespiti, küresel mikrofon dizinleri

This thesis is dedicated to my family, my friends and the people that care about me.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor Assoc. Prof. Dr. Hüseyin Hacıhabiboğlu for the continuous support during my thesis and my studies for master's degree. His knowledge and guidance helped me through during this three years long journey.
Thanks to my puppy, Jeffrey, whom I adopted while writing this thesis. Thanks a bunch to all of my friends who made this chapter of my life somewhat more enjoyable. Finally, I would like to thank my mom and dad for being fully supportive of my academic career since the beginning, without their support I wouldn't be able persevere throughout my journey.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

2D	2 Dimensional
3D	3 Dimensional
AIV	Augmented Intensity Vector
CCL	Connected Components Labelling
DOA	Direction of Arrival
DPD	Direct Path Dominance
D/R	Direct-to-Reverberant
EB	Eigenbeam
GPD	Generalized Pareto Distribution
HiGRID	Hierarchical Grid Refinement
MUSIC	Multiple Signal Classification
NNL	Neighbouring Nodes Labelling
PIV	Pseudo-Intensity Vector
RSMA	Rigid Spherical Microphone Array
SHD	Spherical Harmonic Decomposition
SH	Spherical Harmonic
SRP	Steered Response Power
SRPD	Steered Response Power Density
SS-PIV	Subspace Pseudo-Intensity Vector
STFT	Short-time Fourier Transform
SVD	Singular Value Decomposition
TF	Time-Frequency

CHAPTER 1

INTRODUCTION

1.1 Motivation and Problem Definition

Direction of arrival estimation is a important research topic for acoustic environment analysis, it can be applied for robotics field, augmented reality applications and many additional areas. Motivation of this thesis comes from developing a novel DOA estimation method that performs robustly in different acoustics condition using RSMAs. This is then improved by combining it with other methods.

Estimating direction of arrival of the sources in sound field has become really convenient using RSMAs. Working with a rigid microphone array enables harmonic decomposition of sound field which gives a different perspective of the sound field analysis. Comprehending spherical harmonics in terms of signal processing is an insightful experience. This thesis also helped me to learn about state-of-the-art methods and realize these methods have some setbacks and limitations and the work reported in this thesis tries offer solutions with a novel approach for DOA estimation.

1.2 Proposed Methods and Models

Entropy-based hierarchical grid refinement (HiGRID) is proposed with modified steered response functional called steered response power map (SRPD). Combinations of HiGRID with EB-MUSIC and DPD is proposed with a probability based threshold selection method.

1.3 Contributions and Novelties

The main contributions reported in this thesis are as follows:

- A novel method, entropy based steered response map based HiGRID is proposed for DOA estimation.
- Limitations of EB-MUSIC are alleviated by using HiGRID as an efficient source counting method.
- Threshold selection for DPD-test is a with probability based methodology.

The work reported in this thesis made the following publications possible:

- M. B. Çöteli, O. Olgun, and H. Hacıhabiboglu, “Multiple sound source localization with steered response power density and hierarchical grid refinement,” IEEE/ACM Trans. on Audio, Speech and Lang. Process., vol. 26, pp. 2215 – 2229, November 2018.
- O. Olgun and H. Hacihabiboglu, "Localization of Multiple Sources in the Spherical Harmonic Domain with Hierarchical Grid Refinement and Eb-Music," 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC), Tokyo, 2018, pp. 101-105.
- O. Olgun and H. Hacihabiboglu, "Data-driven Threshold Selection for Direct Path Dominance Test," 2019 23th 23rd International Congress on Acoustics (ICA), Aachen. Sept 2019. (Accepted)

1.4 The Outline of the Thesis

This thesis is structured as follows. A two part background section is presented. The first part of background includes an introduction to spherical harmonic decomposition and the theory of rigid spherical microphone arrays (RSMA). The first part of background section builds the theoretical basis for the second part which includes explanations of earlier DOA estimation methods using RSMA. After background, the work done is presented in Chapter 3 where SRPD and HiGRID are introduced which are components of our proposed algorithm for DOA estimation. Then, HiGRID is combined with state-of-the-art methods EB-MUSIC and DPD-test separately and the chapter concludes with a recently developed threshold selection method DPD-test. Following the presented work evaluation chapter presents an analysis of DOA estimation errors of the proposed methods. Comparison of HiGRID with other state-of-the-art methods is also presented. Conclusion chapter ends this thesis work with discussion of results for proposed methods.

CHAPTER 2

BACKGROUND

In this section, first of all the definitions of spherical harmonics and spherical harmonics decomposition are presented. Then rigid microphone arrays and plane-wave composition of a sound field are explained. After technical background previous work on DOA estimation methods are reviewed in detail. SRP, EB-MUSIC and DPD-test presented in this section are directly associated with work presented in the following parts of this thesis.

2.1 Definitions

2.1.1 Physical Spherical Coordinate System

Consider a point defined in Cartesian coordinates as $\mathbf{x} = (x, y, z)$, this point can be transformed to spherical coordinates using equations in 2.1. Spherical coordinate of the point is defined as $\mathbf{r} = (r, \theta, \phi)$ where radial distance, azimuth and elevation are defined as r, ϕ, θ respectively The relation of spherical coordinates with Cartesian coordinates is visualized in Fig. 2.1.

$$\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned} \tag{2.1}$$

2.1.2 Spherical Harmonic Functions

Spherical harmonics are set of special functions defined on surface of unit sphere which are defined in spherical coordinate system. Spherical harmonics are central for this thesis since the methods presented include functions related to rigid spherical microphone arrays which are directly connected to spherical harmonics representation of sound field.

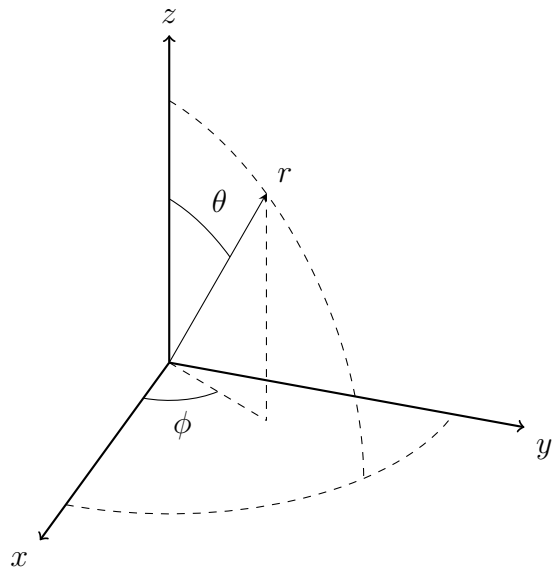


Figure 2.1: Spherical coordinate system variables (r, ϕ, θ) correspondents in Cartesian coordinate system.

Basis function for spherical harmonics defined as follows [1]:

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{\mathrm{i}m\phi} \tag{2.2}$$

where $n \in \mathbb{N}$ and $m \in \mathbb{Z}$ with, $-n \leq m \leq n$, where $P_n^m(\cdot)$ are the associated Legendre functions, θ and ϕ are inclination and azimuth angles. m and n defined as function degree and order spherical harmonics respectively. Fig. 2.2 depicts real parts of the first five orders of spherical harmonic function where Mollwiede projection for visual representation.

2.2 Spherical Harmonic Decomposition

The spherical harmonic functions of order $n \in \mathbb{N}$ and degree $m \in \mathbb{Z}$ are defined in 2.2, for the unit sphere corresponding spherical harmonic coefficients are defined as:

$$f_{nm} = \int_0^{2\pi} \int_0^{\pi} f(\theta,\phi)[Y_n^m(\theta,\phi)]^* \sin\theta d\theta d\phi \tag{2.3}$$

$f(\theta,\phi)$ function projected onto the spherical harmonic basis is called the spherical harmonic decomposition (SHD). It should be noted that amount of spherical harmonics is infinite and finite approximation will be introduced in the following sections. Band-limited functions and distributions on sphere can be represented using SHD.

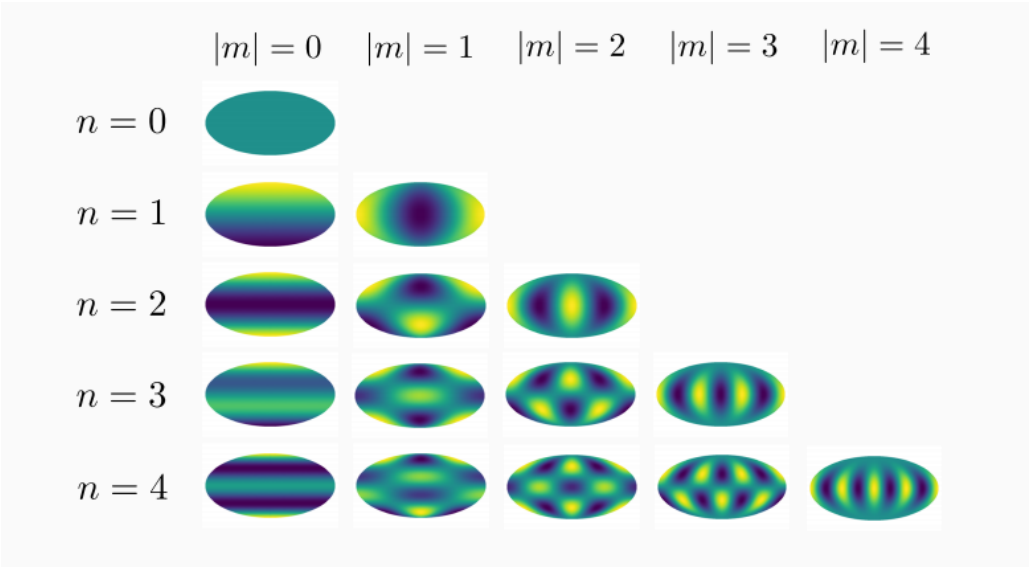


Figure 2.2: Real parts of the first five orders of the spherical harmonic function, $\mathrm{Re}[Y_n^m(\theta, \phi)]$.

2.3 Plane-wave Composition of Sound Field

A sound field can be represented by a superposition of an infinite number of plane waves. In order to obtain a general representation of such a sound field must be expressed in terms of spherical harmonic functions. Sound pressure at point $\mathbf{r} = (r, \theta, \phi)$ due to plane wave, incident from (θ_l, ϕ_l) direction can be represented as:

$$p(k, r, \theta, \phi) = e^{j\mathbf{k}\cdot\mathbf{r}} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi i^n j_n(kr) Y_n^m(\theta, \phi)[Y_n^m(\theta_l, \phi_l)]^*. \tag{2.4}$$

where $j_n(kr)$ is spherical Bessel function, r is radius at the surface of sphere and $k = 2\pi f/c$ denotes the wavenumber where f is frequency and c is the speed of sound. Notice that we omitted time dime dependency for simplicity. Since in a real life scenario the computation of infinite ordered summation for spherical harmonics wouldn't be possible, sound pressure from a single wave is typically approximated as finite summation as:

$$p(k, r, \theta, \phi) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} 4\pi i^n j_n(kr) Y_n^m(\theta, \phi)[Y_n^m(\theta_l, \phi_l)]^*. \tag{2.5}$$

for smaller values of N the sinusoidal behavior is distorted when plane wave is represented using spherical harmonics [1].
A sound field composed of multiple plane waves with directional amplitude $a(k, \theta_l, \phi_l)$,

the sound pressure is defined as follows:

$$\begin{aligned} p(k,r,\theta,\phi) &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi i^n j_n(kr) Y_n^m(\theta,\phi) \\ &\times \int_0^{2\pi}\int_0^{\pi} a(k,\theta_l,\phi_l)[Y_n^m(\theta_l,\phi_l)]^* \sin\theta_l d\theta_l d\phi_l. \qquad (2.6) \\ &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi i^n a_{nm}(k) j_n(kr) Y_n^m(\theta,\phi) \qquad (2.7) \end{aligned}$$

where $a_{nm}(k)$ is spherical harmonic decomposition of $a(k,\theta_l,\phi_l)$, the pressure distribution on the sphere from Eq. 2.5, the following relation for single, unit amplitude plane wave can be derived:

$$a_{nm}(k) = [Y_n^m(\theta_l,\phi_l)]^* \qquad (2.8)$$

Similar to single-wave case, now Eq. 2.7 can be written as finite summation to approximate sound field employing multiple plane waves,

$$p(k,r,\theta,\phi) = \sum_{q=1}^{Q}\sum_{n=0}^{N}\sum_{m=-n}^{n} a_q 4\pi i^n a_{nm}(k) j_n(kr) Y_n^m(\theta,\phi) \qquad (2.9)$$

where $a_q \in \mathbb{C}$ is the amplitude of the k-th plane wave. This representation allows further analysis of general sound fields and will be employed later as a starting point.

2.4 Rigid Spherical Microphone Arrays

Spherical microphone arrays can be broadly classified into two groups: open and closed arrays. Open arrays comprise microphones positioned on the surface of an open sphere. Closed arrays comprise microphones positioned on a rigid spherical baffle. For the latter, the effect of the spherical scatterer changes the expression given above for a plane-wave.
The spherical rigid body imposes a boundary condition on its surface of zero radial particle velocity. Pressure sensitive microphones are located at the surface rigid body for the retrieval of desired spherical harmonics. In this sense, this section divided into two parts; sampling the sphere and sound pressure on surface rigid sphere.

2.4.1 Sound Pressure on the Surface of Rigid Sphere

Defining sound pressure on the surface of rigid sphere due to simple sources such as a plane wave or a point source is for using the recordings made using such arrays in sound field analysis. The sound field around a rigid sphere consists of a combination of the incident field and the scattered field from the sphere. Consider a sphere with radius r_s at the surface of sphere $r = r_s$, the incident sound pressure on sphere

in spherical harmonics domain by a single complex monochromatic plane-wave is defined as same as in Eq. 2.7:

$$p_i(k, r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_{nm}(k) 4\pi i^n j_n(kr) Y_n^m(\theta, \phi). \tag{2.10}$$

where a_{nm} is defined as in Eq. 2.8. In addition to incident sound pressure p_i, scattered sound pressure p_s from rigid body is defined as [1]:

$$p_s(k, r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} c_{nm}(k) {h_n}^{(2)}(kr) Y_n^m(\theta, \phi). \tag{2.11}$$

where ${h_n}^{(2)}(\cdot)$ is the spherical Hankel function of the second kind. As mentioned earlier rigid spherical employs zero radial particle velocity at the surface and by conversation of momentum, derivative of incident pressure on sphere is equal to derivative of total scattered pressure outwards from the rigid body. This relation results with following equation:

$$c_n m(k) = -a_n m(k) 4\pi i^n \frac{j_n^{'}(kr_s)}{h_n^{(2)'}(kr_s)}. \tag{2.12}$$

since $p = p_i + p_s$ the following holds true for total pressure:

$$p(k, r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi i^n a_{nm}(k) \left[j_n(kr) - \frac{j_n^{'}(kr_s)}{h_n^{(2)'}(kr_s)} h_n^{(2)}(kr) \right] Y_n^m(\theta, \phi) \tag{2.13}$$

the structure of pressure function in spherical harmonic is similar to Eq. 2.9 and by defining a new functional $b_n(kr)$ as:

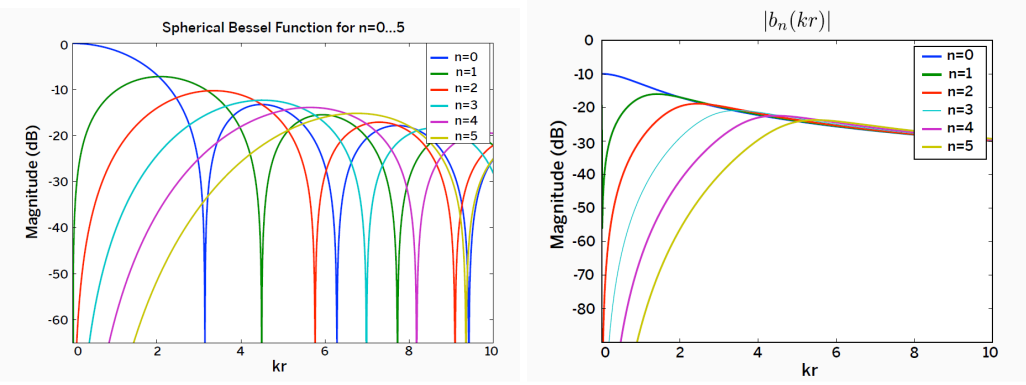
$$b_n(kr) = j_n(kr) - \frac{j_n^{'}(kr_s)}{h_n^{(2)'}(kr_s)} h_n^{(2)}(kr). \tag{2.14}$$

where j_n spherical Bessel function, ${h_n}^{(2)}$ is spherical Henkel function, finally $(\cdot)'$ and $(\cdot)^{(2)'}$ denote first derivatives of these functions.

2.4.2 Sampling the Sphere

The design of rigid microphone array using defines the hardware complexity and accuracy of system. The number of microphone on a spherical constellation determines accuracy of reconstruction of the sound pressure function. Increasing number of microphones increases the accuracy and decreasing number of microphones decreases complexity. If N is maximum harmonic order that can be attained then $p(k, r, \theta, \phi)$ has $(N + 1)^2$ spherical harmonics, then the number of microphones on array, Q, should satisfy the following condition:

$$Q \geq (N + 1)^2 \tag{2.15}$$



(a) (b)

Figure 2.3: $j_n(kr)$ (a) and $b_n(kr)$ (b) functions with different order and $r_s = r$ (at the surface) grid.

The locations of microphones as well as number of microphones are important since the sampling of sphere should enable computation of spherical Fourier transform for order-limited functions.

The microphone array that was used in this thesis was em32 Eigenmike32® which is a microphone array with multiple electret microphones embedded on the surface of rigid spherical baffle (see Fig.2.4). Early prototype of Eigenmike32 is introduced in [2] by mh acoustics. As mentioned earlier a standard RSMA consists multiple pressure sensitive microphones on the surface, in the case of em32 Eigenmike32®, 32 microphones are located around 4.2 cm rigid sphere which enables computation of spherical harmonic order up to $N = 4$, since it holds true for Eq. 2.15. Microphones located at the faces of a truncated icosahedron and at the center of each face [3] .

Spherical harmonic decomposition (SHD) of a sound field can be obtained using recordings obtained via RSMAs. Spatial sampling of a sound field using spherical array involves transfer functions in spherical harmonics domain mentioned Section 2.4.1. We know the pressure distribution on spherical surface is known and with $b_n(kr)$ defined in Eq.2.14, sound pressure on the surface of rigid sphere using spherical harmonics can be expressed as:

$$p_{nm}(k, r_s) = 4\pi i^n a_{nm}(k) b_n(kr) [Y_n^m(\theta_k, \phi_k)]^* \tag{2.16}$$

for a plane wave incident from (θ_k, ϕ_k), p_{nm} can also be named as SHD coefficients or eigenbeams and for Q points sampled on spherical microphone array, total p_{nm} can be approximated as:

$$p_{nm}(k, r) = \sum_{q=1}^{Q} w_q p(\theta_q, \phi_q, k) [Y_n^m(\theta_q, \phi_q)]^* \tag{2.17}$$

where w_q is quadrature weights on spherical array and (θ_q, ϕ_q) are locations of microphones on RSMA. One of the important considerations for selection of the sampling discrete orthonormality condition [4] such that $p_{nm}(k, r)$ converge to the real SHD

Figure 2.4: em32 Eigenmike spherical microphone array

coefficients. For a band-limited pressure distribution $p(k, r, \theta, \phi)$ defined on a spherical surface:

$$p(k, r, \theta, \phi) = \sum_{n=0}^{N} \sum_{m=-n}^{n} p_{nm}(k, r) Y_n^m(\theta, \phi) \tag{2.18}$$

where N is the maximum order of the decomposition. SHD coefficients can be equalised to eliminate the effect of the scattered field. For a monochromatic plane wave with the complex amplitude $\alpha_l(k) \in \mathbb{C}$, the normalized SHD coefficients can be defined as:

$$\tilde{p}_{nm}(k) = \frac{p_{nm}(k, r)}{4\pi i^n b_n(kr)} = \alpha_l(k) \left[Y_n^m(\theta_l, \phi_l)\right]^* . \tag{2.19}$$

The SHD is a linear operation and a linear combination of plane waves can be represented using linear combination of spherical harmonics coefficients [5]. Using this linearity property and going back to Section 2.3 a sound field consisting multiple plane waves can be represented linear combination of the SHD coefficients. Consider a sound field comprising L plane waves, then the frequency-equalized SHD coefficients of that sound field is:

$$\tilde{p}_{nm}(k) = \sum_{l=1}^{L} \alpha_l(k) \left[Y_n^m(\theta_l, \phi_l)\right]^* \tag{2.20}$$

for which matrix notation of this decomposition can be expressed as:

$$\mathbf{p}_{nm}(k) = \mathbf{Y}_s^H \mathbf{a}(k) \tag{2.21}$$

where $(\cdot)^H$ is conjugate transpose and $\mathbf{a}(k) = [a_1(k), a_2(k), ..., a_L(k)]^T$ is the $L \times 1$ complex amplitude vector. $\mathbf{Y}_s$ is $L \times (N+1)^2$ beamspace manifold matrix with the

l-th column given by:

$$\mathbf{y}(\theta_l, \phi_l) = [Y_0^0(\theta_l, \phi_l), Y_1^{-1}(\theta_l, \phi_l), Y_1^0(\theta_l, \phi_l), ..., Y_N^N(\theta_l, \phi_l)] \tag{2.22}$$

2.5 Previous Work

This section explains state-of-the-art methods for DOA estimation using RSMAs and is divided into 3 main parts: beamforming-based, vector-based and subspace-based methods. In addition, the direct path dominance (DPD) test is reviewed.

2.5.1 Beamforming-based Approach

In this section, steered-beamformer approach for source localization is presented. Steered Response Power (SRP) beamformer and Minimum Variance Distortionless Response (MVDR) are examples of beamformers. SRP is explained in detail since it is related to work reported in this thesis. MVDR [6] is an optimal spatial-filtering method which operates on cross-power spectral matrix and beamformer is designed to minimize variance of output array in spherical harmonics domain however not covered in this section since it is not entirely relevant in the context of the presented work.

2.5.1.1 Steered Response Power (SRP)

Steered Response Power (SRP) [7] involves a directive beam pattern steered in the direction of source that maximizes steered beam response in a sound field. SRP map includes DOA estimates maximum output response from direction (θ, ϕ). The resolution of SRP map is related to directivity pattern of beam which is related to the maximum SHD order, N_{max}. Beam pattern for SRP map is usually chosen as *regular beam pattern* [8] to get a maximally directional response. Regular beam pattern is obtained by selecting beamformer coefficients as selecting beamformer coefficients as $Y_n^m(\theta_b, \phi_b)$ in the steering direction (θ_b, ϕ_b).
SHD coefficients, p_{nm}, of sound field with multiple plane-waves can be rewritten using Eq. 2.16 :

$$p_{nm}(k) = 4\pi i^n b_n(kr_a) \sum_{l=1}^{L} \alpha_l(k)[Y_n^m(\theta_l, \phi_l)]^* \tag{2.23}$$

where $\alpha_l(k)$ is amplitude of single-wave component. Let us define array output functional $y_N(\theta, \phi, k)$ using plane-wave decomposition (PWD) with the approximation of SHD to a maximum order of N (see Eq. 2.18).:

$$y_N(\theta, \phi, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{p_{nm}(k)}{4\pi i^n b_n(kr_a)} Y_n^m(\theta, \phi) \tag{2.24}$$

When SHD is not order-limited the SRP becomes a combination of Dirac delta functions defined on the unit sphere such that:

$$\lim_{N\to\infty} y_N(\theta,\phi,k) = \sum_{l=1}^{L} \alpha_l(k)\delta(\cos\theta - \cos\theta_l)\delta(\phi-\phi_l) \tag{2.25}$$

where $\delta(\cdot)$ denotes Dirac delta function.
Now let us call array output using N SHD coefficients as $y_N(\theta,\phi,k)$. Using the spherical harmonics addition theorem [8] leads to *regular beam pattern* which has the maximum directivity:

$$y_N(\theta,\phi,k) = \frac{N+1}{4\pi(\cos\Theta_l - 1)} \sum_{l=1}^{L} \alpha_s(k)\left[P_{N+1}(\cos\Theta_l) - P_N(\cos\Theta_l)\right] \tag{2.26}$$

where Θ_l is defined as the angle between the source direction, (θ_l,ϕ_l), and steering direction (θ,ϕ). $y_N(\theta,\phi,k)$ is named as *steered response power* (SRP) and SRP map can be defined over sphere as finding (θ_s,ϕ_s) pairs that maximize this functional. DOA estimation using SRP involves finding the global maximum of the SRP map such that:

$$(\hat{\theta_s},\hat{\phi_s}) = \underset{\theta_s,\phi_s}{\operatorname{argmax}} |y_N(\theta,\phi,k)|^2. \tag{2.27}$$

SRP is one of the simplest methods for DOA estimation. However steering beams in every possible direction is not an efficient method for higher resolution DOA estimation. SRP usually combined with more sparse DOA estimation methods to establish more precise estimation over an area of interest to optimize computational complexity. An important consideration about SRP is that when interpreted as distribution on the unit sphere spatial resolution is approximately π/N, defined by *Rayleigh condition* [9]. The main lobe of beam pattern defines the resolution for spatial separations of sources. Sound sources must have π/N separation between each other for accurate DOA estimation.

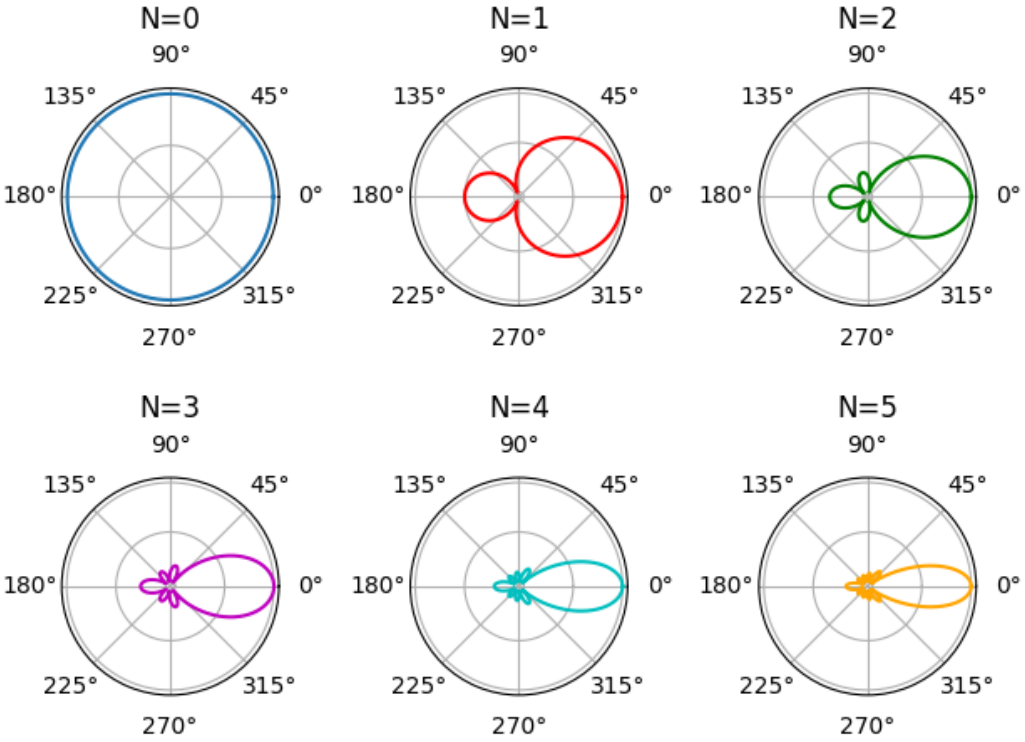


Figure 2.5: Hyper-cardioid beam patterns for orders N = 0,1,2,3,4,5

2.5.2 Vector-based Approach

In this section vector-based approaches are explained mainly centered around PIV since further methods like AIV and SS-PIV are extended version of PIV. PIV and AIV are subsectioned as Vector-based approaches since they are inspired by sound intensity vectors [10].

2.5.2.1 Pseudo-Intensity Vectors (PIV)

Pseudo-intensity Vector (PIV) [11] is proposed for 3D DOA estimation of single source where pseudo-intensity vector is directed at source. Pseudo-intensity vector is calculated by using eigenbeams.

As mentioned earlier sound intensity vector is defined as:

$$\mathbf{I} = \frac{1}{2}\mathrm{Re}\{p(k)^* \cdot \mathbf{v}(k)\}. \tag{2.28}$$

where p is sound pressure and $\mathbf{v} = [v_x v_y v_z]^T$ is particle velocity in Cartesian coordinates system and $\mathrm{Re}\{\cdot\}$ is real part of a complex number. Particle velocity is related to direction of arrival so (θ, ϕ) and can be transformed into spherical coordinates (using Eq. 2.1) as:

$$\mathbf{v} = -\frac{p}{\rho_0 c}\begin{bmatrix}\cos\theta\sin\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{bmatrix}$$

Pseudo-intensity vector is defined as an approximation of intensity vector using zeroth and first order spherical eigenbeams which can be easily obtained using a spherical microphone array. To obtain zeroth and first order eigenbeams Eq. 2.16 can be used where for $p_{nm}(n = 0, 1)$ so that pseudo-intensity vector $\mathbf{I}(k)$ can be defined as:

$$\mathbf{I}(k) = \frac{1}{2}\text{Re}\left\{ p_{00}(k)^* \begin{bmatrix} p_x(k)\\ p_y(k)\\ p_z(k)\end{bmatrix}\right\} \tag{2.29}$$

where

$$p_D(k) = \sum_{m=1}^{-1} Y_1^m(\Omega_D) p_{1(m)}(k), \quad D \in \{x, y, z\} \tag{2.30}$$

since DOA have negative direction to RSMA coordinates, the appropriate directions or rotated eigenbeams are obtained, by using following equations:

$$\begin{aligned}\Omega_x &= (\pi/2, \pi)\\ \Omega_y &= (\pi/2, -\pi/2)\\ \Omega_z &= (\pi, 0)\end{aligned} \tag{2.31}$$

Estimated direction of pseudo-intensity vector as an unit vector can be defined as:

$$\mathbf{u}(k) = \frac{\mathbf{I}(k)}{\|\mathbf{I}(k)\|} \tag{2.32}$$

where $\|\cdot\|$ is L2-norm of the vector.

2.5.2.2 Augmented Intensity Vectors (AIV)

Augmented Intensity Vector (AIV) [12] is an enhanced version of PIV which employs higher order spherical harmonics. Recall in Sec. 2.5.2.1 PIV uses only zero and first order harmonics however higher order harmonics also includes spatial information which can refine DOA estimation.

Consider a plane wave $S(\tau, \kappa)$ with amplitude $\alpha(k)$ impinging at angle $\Omega_u = (\theta_u, \phi_u)$ to RSMA with , SHD of this plane wave is defined as:

$$p_{nm}(\tau, \kappa) = S(\tau, \kappa)\left[Y_n^m(\Omega_u)\right]^* + n_m^n(\tau, \kappa) \tag{2.33}$$

where n_m^n is noise and reverberation component. For a noise free scene $S(\tau, \kappa)$ is approximated as:

$$S(\tau, \kappa) \approx \sqrt{4\pi} p_{00}(\tau, \kappa) \tag{2.34}$$

Now, using Eqs. 2.33 and 2.34 a direction dependent error function is defined as:

$$E_{nm}(\tau, \kappa, \Omega) = p_{nm}(\tau, \kappa) - \sqrt{4\pi} p_{00}(\tau, \kappa) Y_n^m(\Omega) \tag{2.35}$$

where cost function is:

$$C(\tau, \kappa, \Omega) = \sum_{n=0}^{L} \sum_{m=1}^{1} |E_{nm}(\tau, \kappa, \Omega)|^2, \quad D \in \{x, y, z\} \tag{2.36}$$

where,

$$\Omega_{aiv} = \underset{\Omega}{\operatorname{argmin}}\, C(\tau, \kappa, \Omega) \tag{2.37}$$

minimising the cost function will give optimized vector direction Ω_{aiv}. Original norm PIV, $\mathbf{I_{piv}}$, vector is combined with new direction $\mathbf{u}_{\Omega_{aiv}}$ unit vector to estimate DOA:

$$\mathbf{I_{aiv}}(\tau, \kappa) = -\mathbf{u}_{aiv}(\tau, \kappa) \|\mathbf{I_{piv}}(\tau, \kappa)\|. \tag{2.38}$$

2.5.3 Subspace-based Approach

In this section, subpsace-based approaches for DOA estimation are investigated. EB-MUSIC (Eigenbeam Multiple Signal Classification) [13], EB-ESPRIT (Eigenbeam Estimation of Signal Parameters via rotational invariance techniques) [14] and recently proposed subspace pseudo-intensity vector(SS-PIV) [15], which extends PIV, are popular subspace-based methods. These exploit signal and noise subspaces to estimate DOA. EB-ESPRIT formulation is based on recurrence relation for associated Legendre function for spherical harmonic domain mentioned in Sec. 2.1.2. For the solution of EB-ESPRIT, eigenvalues are computed where DOA estimation is obtained by phase and amplitudes of eigenvalues. EB-ESPRIT is not detailed but this section is rather focused on EB-MUSIC since the work presented later is related to EB-MUSIC algorithm.

2.5.3.1 Eigenbeam Multiple Signal Classification (EB-MUSIC)

MUltiple **SI**gnal **C**lassification (MUSIC) [16] is an algorithm originally implemented for antennas receiving narrowband signals, e.g. radio, which determines DOA by using noise and signal subspace, more specifically the eigenspace. **E**igen**b**eam **MU**ltiple **SI**gnal **C**lassification (EB-MUSIC) is a version of original MUSIC algorithm adopted for broadband signals such as music and speech. It is a subspace method since it employs the properties of the signal subpspace and noise subspace to estimate DOA. For the ideal case the SHD coefficients defined in Eq. 2.21 are free from noise. However in a realistic scenario, SHD coefficients are defined as:

$$\mathbf{p}_{nm}(k) = \mathbf{Y}_s^H \mathbf{a}(k) + \mathbf{n}(k) \tag{2.39}$$

where $\mathbf{n}(k)$ is a $(N+1)^2 \times 1$ vector that includes noise components. EB-MUSIC is performed in time-frequency (TF) domain which can be achieved with the short-time

Fourier transform. SHD coefficients of a single TF-bin can be defined as $\mathbf{p}_{nm}(\tau,\kappa)$ where (τ,κ) pair is time and frequency respectively. The vector form of SHD coefficients, $\mathbf{p}_{nm}$, in TF-domain is used for the calculation of the spatial correlation matrix:

$$\begin{aligned}\mathbf{R}_p(\tau,\kappa) &= E\left[\mathbf{p}_{nm}(\tau,\kappa)\mathbf{p}_{nm}^H(\tau,\kappa)\right] \\ &= \mathbf{Y}^H\mathbf{R}_s(\tau,\kappa)\mathbf{Y}+\mathbf{R}_n(\tau,\kappa)\end{aligned} \tag{2.40}$$

where $E[\cdot]$ is statistical expectation and $(\cdot)^H$ is Hermitian transpose. $\mathbf{R}_s$ is signal correlation matrix and $\mathbf{R}_n$ denotes noise correlation matrix.

In practice, expectation calculation is carried out by averaging time, j_τ, and frequency, j_κ frames [17]:

$$\mathbf{R}_p(\tau,\kappa)=\frac{1}{(J_\tau+1)(J_\kappa+1)}\sum_{j_\tau=-\frac{J_\tau}{2}}^{\frac{J_\tau}{2}}\sum_{j_\kappa=-\frac{J_\kappa}{2}}^{\frac{J_\kappa}{2}}\mathbf{p}_{nm}(\tau+j_\tau,\kappa+j_\kappa)\mathbf{p}_{nm}^H(\tau+j_\tau,\kappa+j_\kappa). \tag{2.41}$$

It should be noted that $\mathbf{R}_p$ is Hermitian symmetric. Signal and noise subspaces can be obtained by the eigendecomposition of spatial correlation matrix such that:

$$\mathbf{R}_p=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H=[\mathbf{U}_s\ \mathbf{U}_n]\begin{bmatrix}\mathbf{\Lambda}_s & \mathbf{0}\\ \mathbf{0} & \mathbf{\Lambda}_n\end{bmatrix}\begin{bmatrix}\mathbf{U}_s^H\\ \mathbf{U}_n^H\end{bmatrix}, \tag{2.42}$$

where the diagonal matrices $\mathbf{\Lambda}_s$ and $\mathbf{\Lambda}_n$ contain the eigenvalues, $\mathbf{U_s}$ and $\mathbf{U_n}$ are signal and noise subspaces including the corresponding eigenvectors to define EB-MUSIC spectrum. Now, MUSIC spectrum at a given direction where $\Omega=(\theta_l,\phi_l)$ can interpreted as:

$$S_{MUSIC}(\Omega)=\frac{1}{\|\mathbf{U}_n^H\mathbf{y}^H(\Omega)\|^2}. \tag{2.43}$$

DOA estimation is carried out finding peaks in MUSIC spectrum obtained at the directions $\{\Omega_l\}$ where $S_{MUSIC}(\Omega)$ has its local maxima.

2.5.3.2 Subspace Pseudo-Intensity Vectors (SS-PIV)

PIV uses lower order spherical harmonics can be used to improve DOA estimation. Subspace PIV uses higher order SH components by analyzing subspace model of sound field and frequency smoothing in TF domain.

In the TF domain for the case of a single plane wave, spatial correlation matrix $\mathbf{R}_p$ is defined in Eq. 2.41 and decomposed using SVD:

$$\mathbf{R}(\tau,\kappa)=\mathbf{U_s}\Lambda_s{\mathbf{U_s}}^H+\mathbf{U_n}\Lambda_n{\mathbf{U_n}}^H \tag{2.44}$$

where $\mathbf{U_s}=[p_{00},p_{1(-1)},p_{1(0)},p_{1(1)},\cdots,p_{NN}]$ is $1\times N$ signal subspace matrix, $\mathbf{U_v}$ is noise subpsace, (τ,κ) is omitted for simpler notation. SSPIV uses similar method to Eq. 2.29 but with instead of plain eigenbeams, it uses decomposed SH signal subspace matrix:

$$\mathbf{I_{ss}}(\tau,\kappa)=\frac{4\pi\sqrt{4\pi}}{3}\mathrm{Re}\left\{p_{00}(\tau,\kappa)^*\begin{bmatrix}p_x(\tau,\kappa,\mathbf{U_s})\\ p_y(\tau,\kappa,\mathbf{U_s})\\ p_z(\tau,\kappa,\mathbf{U_s})\end{bmatrix}\right\} \tag{2.45}$$

where p_x, p_y, p_z are the components with largest 3 eigenvalues after eigenvalue decomposition. Using this formulation vector is directed to source DOA and although $n = 0, 1$ order harmonics are used, value of $\mathbf{U_s}$ depends on higher order SH.

2.5.4 Direct Path Dominance (DPD) Test

Direct Path Dominance (DPD) is a DOA estimation technique which involves identification of TF bins dominated by direct-path from sources to microphone array [18]. DPD test carried out in subspace and singular value decomposition (SVD). Let us apply short-time Fourier transform (STFT) on plane wave decomposition signal of a sound field as presented in Sec. 2.2, and let us define the signal:

$$\mathbf{p}_{nm}(k) = \mathbf{Y}_s^H \mathbf{s}(\tau, \kappa) + \mathbf{n}(\tau, \kappa) \tag{2.46}$$

where $\mathbf{s}$ signal amplitude vector and frequency response from each plane-wave or source. For each TF-bin spatial correlation matrix is calculated which is averaged over neighbouring time, $J\tau$ and frequency, J_κ bins yielding following notation:

$$\tilde{\mathbf{R}}_{\mathbf{p}}(\tau, \kappa) = \frac{1}{J_{tot}} \sum_{j_\tau=-J_\tau}^{J_\tau} \sum_{j_\kappa=-J_\kappa}^{J_\kappa} \mathbf{p}_{nm}(\tau + j_\tau, \kappa + j_\kappa) \mathbf{p}_{nm}^H(\tau + j_\tau, \kappa + j_\kappa) \tag{2.47}$$

where $J_{tot} = (2J_\tau + 1)(2J_\kappa + 1)$. TF smoothing approximates expectation for spatial correlation matrix. DPD test determines whether a time-frequency bin is dominated by a single source. This is done by examining SVD of each averaged TF-bin spatial correlation matrix. Eigen-value decomposition enables the computation of effective rank (erank) and DPD test is defined as follows:

$$\mathcal{R}_{DPDtest} = \Big\{ (\tau, \kappa) : \text{erank}\Big(\tilde{\mathbf{R}}_{\mathbf{p}}(\tau, \kappa)\Big) = 1 \Big\} \tag{2.48}$$

where,

$$\mathcal{E} = \text{erank}\Big(\tilde{\mathbf{R}}_{\mathbf{p}}(\tau, \kappa)\Big) = 1 \quad \text{if} \quad \frac{\sigma_1(\tau, \kappa)}{\sigma_2(\tau, \kappa)} \geq T_{\text{DPD}}, \tag{2.49}$$

where $\sigma_1(\tau, \kappa)$ and $\sigma_2(\tau, \kappa)$ are the largest and second-largest singular values of $\tilde{\mathbf{R}}_{\mathbf{p}}$ obtained by employing SVD and T_{DPD} is a threshold value which is chosen larger than 1 to guarantee $\tilde{R}_p$ has a dominant singular vector. The ratio of the singular values given in (2.49) can never be less than unity.

CHAPTER 3

DIRECTION-OF-ARRIVAL ESTIMATION WITH HIERARCHICAL GRID REFINEMENT

In this chapter, a new steered response functional called the Steered Response Power Density (SRPD) and recently developed Hierarchical Grid (HiGRID) are introduced. Following sections present combinations of HiGRID with EB-MUSIC and DPD-test separately. A threshold selection method for DPD-test is also proposed at the end of chapter.

3.1 Steered Response Power Density (SRPD)

Steered response power described in 2.5.1.1 is used to estimate DOA of sources by steering a directive beam that maximises the output power however this creates high computational cost. A new, improved version SRP, called Steered Response Power Density (SRPD) is proposed to resolve this heavy loaded source direction seeking method. SRPD aims to reduce the weight of the process by calculating power density of an area instead of specific direction.

A spatially band-limited approximation of plane wave-decomposition can be obtained with practical RSMAs due to the order limitation that results from sampling the pressure at a finite number of points. The order-limited *steered response functional* (SRF) is given as in Eq. 2.24:

$$y_N(\theta, \phi, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{p_{nm}(k)}{4\pi i^n b_n(kr_a)} Y_n^m(\theta, \phi) \tag{3.1}$$

Steered response power (SRP) maps are obtained by steering a maximally directive beam in all possible directions and seeking the directions that provide the maximum output power. For example, an SRP map for a single source will contain a single maximum corresponding to the DOA of the sound source, such that:

$$(\theta_s, \phi_s) = \underset{\theta,\phi}{\operatorname{argmax}} \, |y_N(\theta, \phi)|^2 \tag{3.2}$$

where (θ_s, ϕ_s) is estimated DOA of sound source is spherical coordinate system.

One of the main disadvantages of DOA estimation using SRP technique is that the steered response is calculated for discrete directions only and finding the local maxima requires search on a fine resolution grid. In order to allow using coarser grids

steered response power density (SRPD) was proposed [19]. SRPD is defined as:

$$\mathcal{P}_i(k) = \frac{1}{A_i} \int_{\mathcal{S}_i} |\mathbf{y}(\theta_s, \phi_s)^T \tilde{\mathbf{p}}_{nm}|^2 d\mathcal{S}_i \tag{3.3}$$

where S_i is a surface element on the unit sphere with its centre positioned at (θ_i, ϕ_i), and A_i is its area.

Computation of SRPD can be substantially simplified by dimensionality reduction techniques. SRPD can be expressed using (3.1) such that:

$$\mathcal{P}_i(k) = \sum_{n,m,n',m'} \frac{p_{nm}(k) p^*_{n'm'}(k)}{b_n(kr_a) b^*_{n'}(kr_a)} Q^{m,m'}_{n,n'}(\mathcal{S}_i), \tag{3.4}$$

where

$$Q^{n',m'}_{n,m}(\mathcal{S}_i) = \frac{1}{(4\pi)^2 A_i} \int_{\mathcal{S}_i} Y^m_n(\theta, \phi) \left[Y^{m'}_{n'}(\theta, \phi) \right]^* d\mathcal{S}_i. \tag{3.5}$$

which can be calculated via a suitable numerical quadrature technique.

The summation in (3.4) can be expressed as the grand sum of the matrix:

$$\mathbf{H}_i = \mathbf{P} \circ \mathbf{Q}_i \tag{3.6}$$

where $\circ$ is the Hadamard (i.e. element-wise) product, $\mathbf{P} = \mathbf{p}\mathbf{p}^H$ is a matrix with dimensions $(N+1)^2 \times (N+1)^2$ and the cross spatial density matrix $\mathbf{Q}_i$ is also an $(N+1)^2 \times (N+1)^2$ matrix.

An identity of the Hadamard product can be used for simplifying the grand sum of $\mathbf{H}_i$, such that [20]:

$$\mathcal{P}_i = \mathbf{e}^T (\mathbf{P} \circ \mathbf{Q}_i) \mathbf{e} = \mathrm{tr}(\mathbf{V}_i^H \mathbf{P} \mathbf{V}_i \mathbf{D}_i) \tag{3.7}$$

where $\mathbf{e}$ is a column vector of ones and $\mathrm{tr}(\cdot)$ is the trace operator. Here, the columns of $\mathbf{V}_i$ are the eigenvectors, and the diagonal matrix $\mathbf{D}_i$ contains the eigenvalues $\lambda_{i,m}$ of $\mathbf{Q}_i^T$. The calculation can be simplified by selecting the largest eigenvalues and eigenvectors of $\mathbf{Q}_i^T$. SRPD was shown to be equivalent to SRP at medium and high grid resolutions in terms of its computational cost [19].

3.2 Hierarchical Grid Refinement (HiGRID)

In this section HiGRID [19] which employs a hierarchical multi-resolution search grid based on information gain to estimate DOA is reviewed. Information gain is obtained from SRDP map (see Sec. 3.1) however before generating SRPD map and calculating information gain, there are preprocessing stages.
First of all, STFT from each microphone signal are calculated to get time-frequency domain representation. Second stage of prepocessing is bin selection using spectrum based onset detection. Selecting bins with onsets will improve DOA accuracy because TF-bins with onsets assumed to have one or more sources. Superflux [21] is used for detecting onsets in original HiGRID algorithm however there are others method for

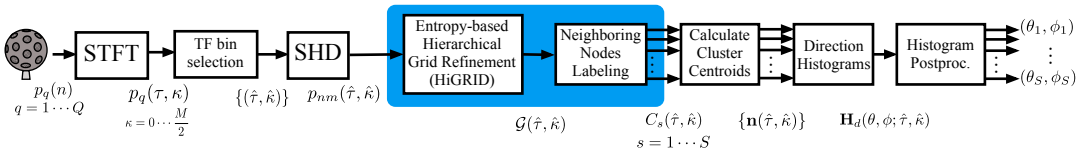


Figure 3.1: The flow diagram of the proposed algorithm. The core part is highlighted with a blue box. © 2018, IEEE

selection TF-bins. After, selecting TF-bins to be processed, SHD coefficients of these bins are calculated to obtain SRPD map.

SRPD values can be calculated on a HEALPix grid [22] which is a hierarchical tessellation of the unit sphere consisting of grid elements which are quadrilateral and non-overlapping pixels. Each pixel can be subdivided into 4 higher-resolution elements which gives quadtree representation of the underlying steered response map. The total number of pixels in a HEALPix grid at the resolution level is, $h \in \mathbb{N}$ is 12×2^{2h}, the angular resolution is,

$$\Theta_\Delta = \sqrt{\frac{3}{\pi}} \frac{\pi}{3 \cdot 2^h}. \tag{3.8}$$

and the area is equal to,

$$A_l = \frac{4\pi R^2}{12 \cdot 2^{2l}}. \tag{3.9}$$

Note that magnitude of area is dependent to resolution level. In lowest resolution where $l = 0$ has 12 pixel elements where increasing resolution decreases area of each pixel equally. HEALPix can have any resolution level where $l \in [0, 10]$. Resolution levels from $l = 1$ to $l = 4$ are visualized in Fig. 3.2 where angular resolution or the distance between centers of two nodes are $29.32°$, $14.66°$, $7.33°$, $3.66°$ respectively. Let us define grid elements as $\mathcal{S}_{l,m}$ at l resolution level where m is index of element on HEALPix grid where the probability of a source being present in a given pixel as:

$$\gamma(\mathcal{S}_{l,m}) = \frac{\mathcal{P}_{l,m}}{\sum_{\forall \mathcal{S}_{l,m} \in \mathcal{G}} \mathcal{P}_{l,m}} \tag{3.10}$$

where $\mathcal{G}$ is a multi-resolution tessellation of the unit sphere where $\mathcal{P}_{l,m}$ is defined in Eq. 3.4. The total spatial entropy of the SRPD representation is given as:

$$H(\mathcal{G}) = -\sum_{\forall \mathcal{S}_{l,m} \in \mathcal{G}} \gamma(\mathcal{S}_{l,m}) \log \frac{\gamma(\mathcal{S}_{l,m})}{A(\mathcal{S}_{l,m})}. \tag{3.11}$$

HiGRID involves a multi-resolution and adaptive refinement of the HEALPix grid by minimizing the total spatial entropy [23]. This is equivalent to maximizing the information gain of the representation for each time-frequency bin resulting in concentrated high-resolution regions around local maxima. The candidate grid, $\mathcal{G}'_t$ is

obtained by refining existing grid, $\mathcal{G}_t$, so each element is refined based on a subdivision would whether or not reduce the spatial entropy where the decision criterion to refine the representation can be defined as:

$$\mathcal{I}(l,m) = H(\mathcal{G}_t) - H(\mathcal{G}'_t). \tag{3.12}$$

where $\mathcal{I}(l,m)$ is defined as *information gain* for and $\mathcal{I}(l,m) > 0$ means refinement of that leaf node $S_{l,m}$. HiGRID starts at the lowest resolution while forming an SRPD map and gradually refines it based on information gain. The result is a multiresolution power map which contracts around local maxima which representing probable source directions.

Neigbouring nodes labeling (NNL) [19] is used for clustering source directions using association between neighbouring nodes on quadtree representation of sound field. NNL is similar to *connected components labeling* [24] (CCL) that is used on uniformly sampled pixels (i.e. images) where NNL can operate on grid pixels with different resolution levels to identify regions containing a sound source. The centroid of these regions are stored as DOA estimations in a 2D histogram with 1° bin size to represent DOA in spherical coordinates bounded by $\phi = [0, 2\pi]$, and $\theta = [0, \pi]$.

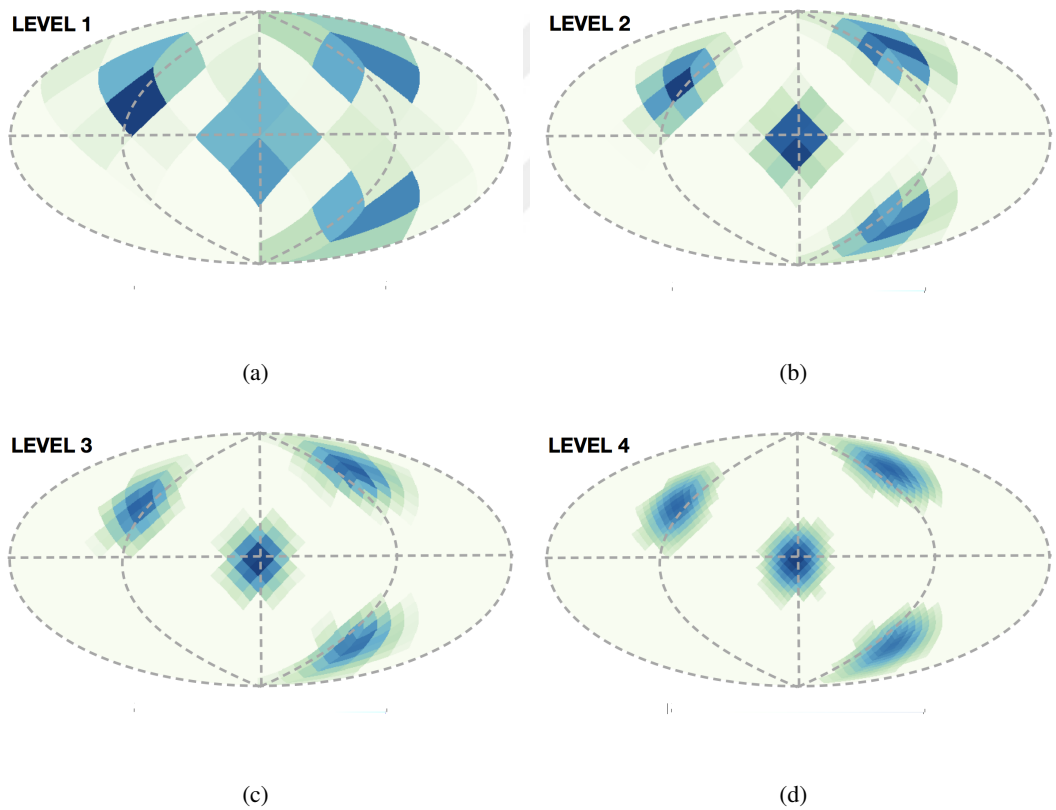


Figure 3.2: The progression of the HiGRID method for a simulated case with four unit amplitude monochromatic plane waves from $\Omega_1 = (\pi/2, 0)$, $\Omega_2 = (\pi/3, \pi/2)$, $\Omega_3 = (5\pi/6, -\pi/2)$ and $\Omega_4 = (\pi/6, -\pi/2)$. Mollweide projection is used in the figures. Resolution levels are shown $l = 1, 2, 3, 4$ for (a)-(d) respectively. © 2018, IEEE

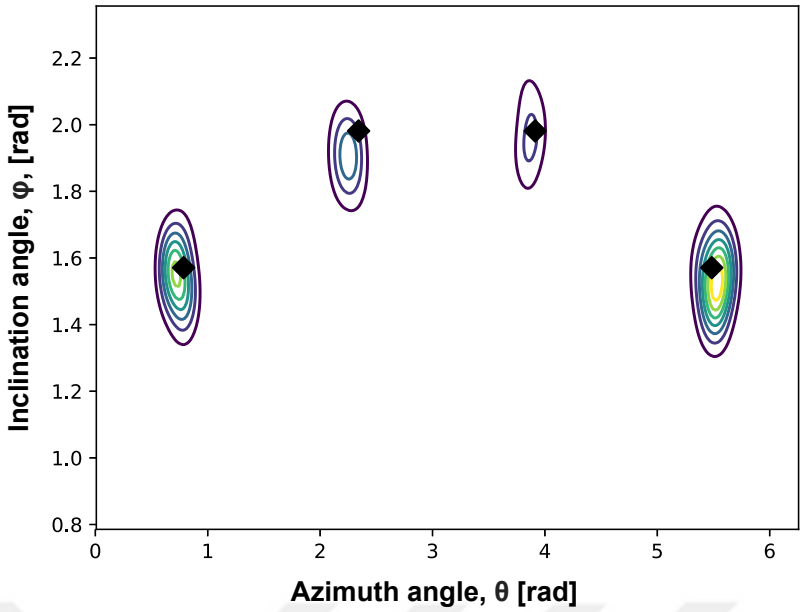


Figure 3.3: HiGRID-MUSIC map showing the peaks of the post-processed DOA histogram and the true DOAs. © 2018, IEEE

3.3 HiGRID-MUSIC

A combination of HiGRID and EB-MUSIC allows estimating DOAs of multiple, coherent and incoherent, sources. Prior information about number of sources in sound field is needed to employ signal and noise subspaces, this is a setback if there are multiple sources concurrently in the same TF-bin. This disadvantage of EB-MUSIC is solved using HiGRID as a source counting for a TF bin.

For applying HiGRID, standard preprocessing stage described in Sec.3.2 is carried out. For selected TF bins HiGRID is applied so that set of low-resolution grid elements with a single source $\mathcal{K}_i^{(l)}$ at resolution level, l, are detected. The global grid can contain K sources for a TF bin such that:

$$\mathcal{K}^{(l)} = \bigcup_{k=1}^{K} \mathcal{K}_k^{(l)} = \bigcup_{l=1}^{K} \{\mathcal{S}_i^{(l)}\}_k. \tag{3.13}$$

where $\mathcal{S}_i$ is i^{th} low-resolution grid element at resolution level, l, defined in Sec. 3.2. The number of sources K is used to defined signal and noise subspaces defined in Eq. 2.42. For a higher-resolution DOA estimation, EB-MUSIC is applied to SRPD map regions which are obtained using low-resolution HiGRID explained above. For the global grid higher resolution EB-MUSIC can be in defined as:

$$\mathcal{K}_k^{\mathbb{U}} = \bigcup_{q=0\cdots Q} \left\{ \bigcup_{i|\mathcal{S}_i \in \mathcal{K}_k^{(l)}} \mathcal{S}_j^{(l+q)} \;\middle|\; j = 4^q i, \cdots, 4^q(i+1) - 1 \right\}$$

where Q is number of refinements. With each refinement angular resolution increases by 2^{Q+1} times. EB-MUSIC spectrum computational cost is substantially reduced by calculating EB-MUSIC only at the centre positions (θ_j, ϕ_j) of elements in $\bigcup_l \mathcal{L}_l^{\mathbb{U}}$ and not for all for the sphere.

For estimates of DOA as 2D histogram, DOA estimates for each TF bin are denoised using a median filter then smoothed using Gaussian window (see. Fig. 3.3). Spherical k-means is employed to cluster final DOA pairs (θ, ϕ) in spherical coordinate system.

3.4 DPD-HiGRID

In this section combination of DPD with HiGRID is presented. DPD, explained in section 2.5.4, originally used as a preprocessing stage for EB-MUSIC [18]. Identified TF-bins are processed using HiGRID for DOA estimation. Note that, in its original formulation, HiGRID was executed only on time-frequency bins that are close to the onsets in the recorded audio. It was shown that the computational cost of HiGRID increases with the number of source components in the time-frequency bin [19]. Therefore, combining HiGRID with DPD serves not only to select time-frequency bins, but also to reduce the computational cost of DOA estimation.

Since each bin selected by the DPD test will contain a single dominant source, the global maximum of the SRPD map can be identified as the DOA estimate for the analyzed bin. DOA estimations obtained from all selected bins are then clustered using the spherical k-means algorithm [25]. The cluster centers are identified as the source DOAs.

Fig. 3.4 shows the azimuth/inclination histogram of the DOA estimates using proposed DPD-HiGRID method in a Mollwiede projection. The histograms show distribution of TF bins for four sources incident from different azimuth angles but the same inclination angle. The estimates are obtained from TF bins selected via DPD test by processing with the HiGRID algorithm. Decomposition level of the HEALPix grid is $h = 3$ where number of pixels are equal to N_{pix}=768. The true DOAs were $(90°, 0°)$, $(90°, 90°)$, $(90°, 180°)$, and $(90°, 270°)$. The estimated DOAs were $(88.56°, 357.77°)$, $(90.92°, 89.26°)$, $(92.84°, 179.51°)$, and $(91.27°, 268.54°)$.

3.4.1 Threshold selection for DPD-test

Threshold for DPD-test which is defined in Sec. 2.5.4 is usually chosen in an *ad hoc* manner [26, 27]. A more convenient threshold selection is proposed to provide better TF-bins for improve accuracy DOA estimation and since number of TF-bins selected has a profound effect on computational cost which can be reduced.

The histograms of effective ranks of the spatial correlation matrices can be observed to come from right-tailed probability distributions (see. Fig 3.5). The distribution depends number of sources existing in a sound field, the reverberation time of closed space as well as direct-to-reverberant (D/R) ratio.

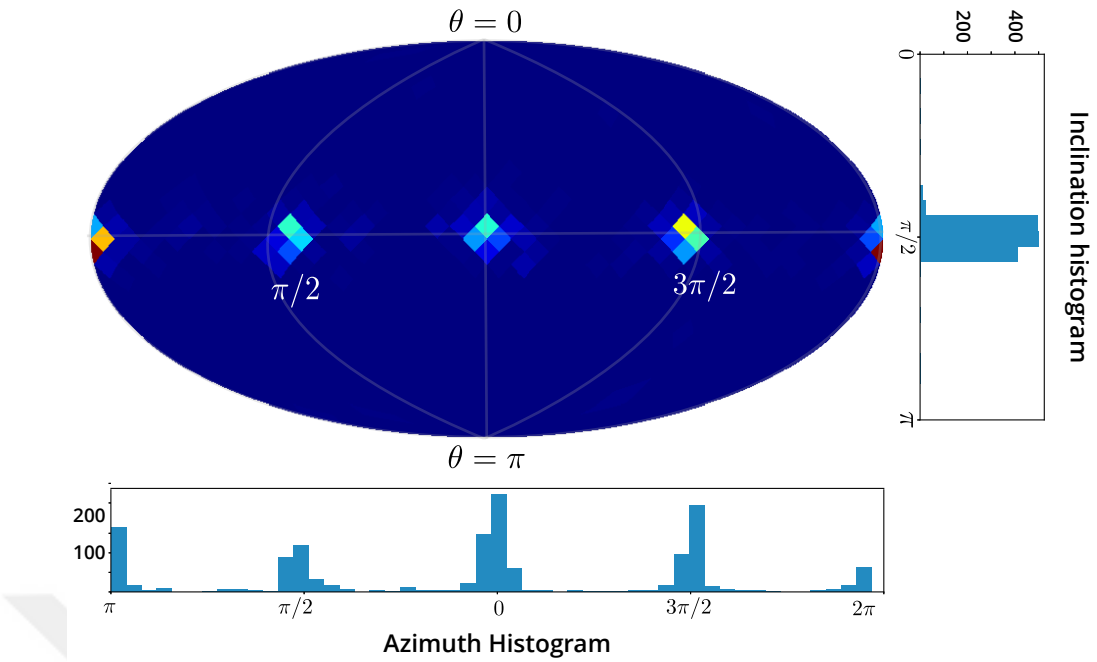


Figure 3.4: Mollweide projection of DPD-HiGRID result showing the peaks of source locations on pixel tessellation (N_{pix}=768).

Proposed threshold selection method assumes that the effective ranks calculated for each TF bin comes from a generalized Pareto distribution (GPD) [28] whose probability density function is given as:

$$f_{\zeta,\sigma,\mu}(x) = \frac{1}{\sigma}\left[1 - \frac{\zeta(x-\mu)}{\sigma}\right]^{(1-\zeta)/\zeta} \tag{3.14}$$

where $\zeta \leq 0$ is the shape parameter, $\sigma > 0$ is the scale parameter, and $\mu > 0$ is the location parameter.
In reverberant environments, the probability of observing a high ratio of singular values is low where low ratio of singular has a high probability of occurrence. Accordingly, instead of using ratio threshold the selection of threshold can be made based on a probability threshold so the probability of observing a ratio greater than a desired threshold, T_{DPD} is defined as:

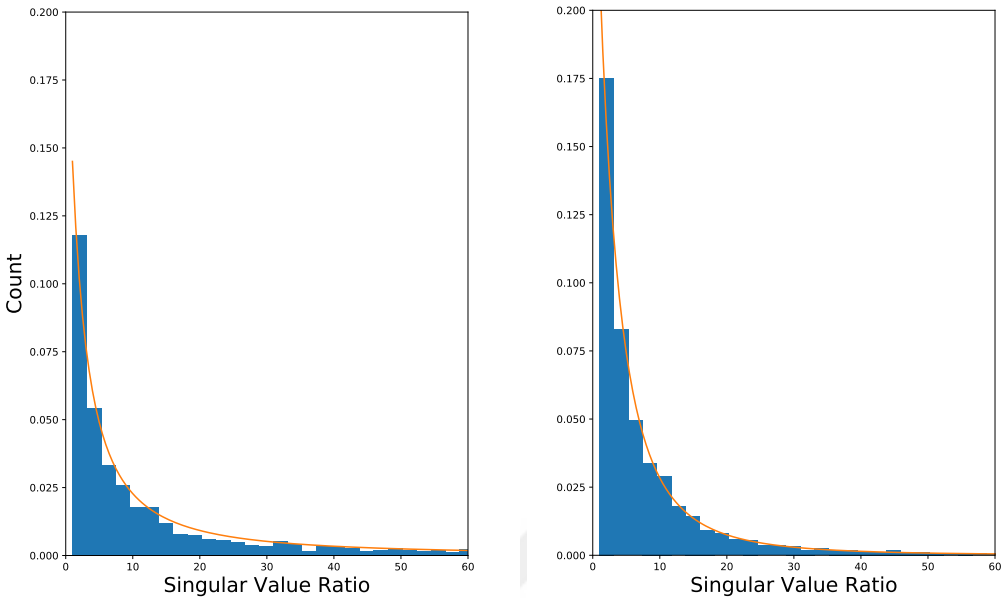
$$P[\mathcal{R} > T_{\mathrm{DPD}}] = 1 - \int_{1}^{T_{\mathrm{DPD}}} f_{\zeta,\sigma,1}(x)dx \tag{3.15}$$

$$= 1 - F_{\zeta,\sigma,1}(T_{\mathrm{DPD}}) + F_{\zeta,\sigma,1}(1) \tag{3.16}$$

where $P[\cdot]$ represents the probability, and,

$$F_{\zeta,\sigma,\mu}(x) = 1 - [1 - \zeta(x-\mu)/\sigma]^{1/\zeta} \tag{3.17}$$

is the cumulative distribution function (CDF) of the generalized Pareto distribution. Note that $F_{\zeta,\sigma,1}(1) = 0$. DPD-test threshold can be selected by estimating the scale



(a) (b)

Figure 3.5: Ratio histograms of the nearest 0.5 m (a) and the furthest 2.6 m (b) sources in measurement grid.

and shape parameters of GPD. Selected threshold is the minimum value which satisfies:

$$P[\mathcal{R} > T_{\mathrm{DPD}}] = 1 - F_{\zeta,\sigma,1}(T_{\mathrm{DPD}}) < \widehat{P} \tag{3.18}$$

The threshold to be used in the DPD-test in order to guarantee that the ratio for the selected bins have the probability of occurrence, $\widehat{P}$ should then be selected to satisfy:

$$T_{\mathrm{DPD}} > \sigma(1 - \widehat{P}^{\zeta})/\zeta \tag{3.19}$$

The scale and shape parameters of the underlying distribution can be obtained via maximum likelihood estimation [29, 30].

CHAPTER 4

EVALUATION AND DISCUSSION

Proposed methods are evaluated in terms of DOA estimation errors for scenes that involve one or more coherent and incoherent sources. In this section first general evaluation setup is introduced then detail of evaluations with results are presented for HiGRID, HiGRID-MUSIC and DPD-HiGRID in that order. Finally, results are discussed.

4.1 Evaluation Setup and Recordings

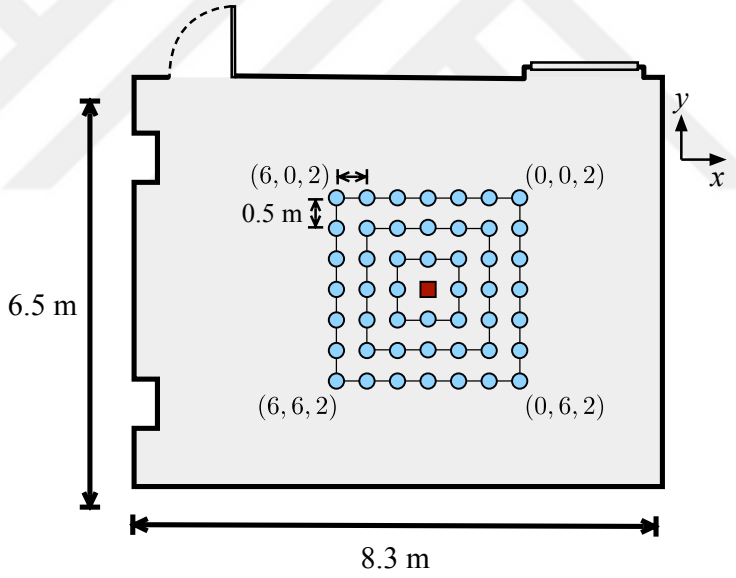


Figure 4.1: Top view of the classroom with the measurement positions. Red square in the center denotes is Eigenmike em32. © 2018, IEEE

Acoustic impulse response (AIR) is defined as the response of a room to an impulse stimulus as measured by a microphone. For evaluations, Eigenmike em32 spherical microphone array (see Fig. 2.4) were used to measure multi-channel AIRs in an empty classroom at METU Graduate School of Informatics. The classroom was emptied to avoid interference and has a high reverberation time ($T_{60} \approx 1.12$). Room

dimensions were $6.5 \times 8.3 \times 2.9$ m and em32 Eigenmike array was positioned at 1.5 m height. Logarithmic Sine sweep method [31] was used and the sound source was a Genelec 6010A loudspeaker. A total of 240 AIR measurements were made on a $7 \times 7 \times 5$ measurement grid (see Fig. 4.1). The grid's horizontal resolution was 0.5 m and vertical resolution was 0.3 m. Using the same locations on grid, room impulse measurements were also made using an Alctron M6 omnidirectional microphone to calculate the D/R ratios.
For evaluation scenarios different recordings were used where all the recordings were convolved with the measured AIRs to simulate acoustic scenes with multiple sources. All the recordings used in this work have the sampling rate of 48 kHz.

4.2 Evaluation of HiGRID

Evaluation of HiGRID is based on comparison of state-of-the-art DOA estimation methods with proposed method.These methods are PIV [11], DPD-MUSIC [18] and SS-PIV [15].

For the evaluation of HiGRID, 4 seconds (01:00-01:04) of anechoic recordings of the fourth movement of Mahler's Symphony Nr. 1 [32] and the first 4 seconds of anechoic speech recordings from B&O Music for Archimedes CD [33] were used. 4 seconds sample of Mahler's Symphony includes four violins playing the same phrase in unison and speech recordings are in English and Danish from two female and two male speakers.
The results of DOA estimation error for scenarios with different number of sources are presented in Table 4.1 for speech and Table 4.2 for violins. Source positions for the evaluation scenario are $(90°, 45°)$, $(90°, 135°)$, $(90°, 225°)$ and $(90°, 315°)$ with all sources equally distant from Eigenmike em32 with 1.41 m. The D/R ratios for these source positions were, 3.04, 3.29, 3.56, and 2.39 dB, respectively.

It may be observed from the reported results that HiGRID showed a performance similar to SRP where in both scenarios small estimation errors are obtained for multiple sources where computational cost significantly lowered when compared to SRP. In Table 4.1, PIV performed worst for speech signal where other methods performed acceptable level of accuracy and in Table 4.2 coherent sources caused failure for PIV where DPD-MUSIC and SS-PIV less effected however DOA estimation errors are much higher than speech case.

Real recordings were also used to estimate DOA using HiGRID without any simulated sources however this case not included for comparison. The recordings are done in Erimtan Concert Hall with reverberation time $T_{60} \approx 1.19$. Eigenmike em32 is placed in the center of Nemeth Quartet (consisting of two violins, a viola and a cello) and recorded pre-concert rehearsal (see Fig 4.2). For the evaluation 5 seconds of excerpt Beethoven's String Quartet Nr. 11, Op. 85 was selected where instruments playing individual parts not unison, unlike the case in comparison evaluation. The results are presented for two different frequency ranges ($2608 - 5216$ Hz and $1304 - 5216$ Hz). It may be observed that even the reference DOAs are not static and precise since depend on the musicians' movements during performance, the results are presented to show HiGRID works well in a real life scenario for DOA estimation.

Table 4.1: DOA estimation errors for four concurrent speech sources using different methods. © 2018, IEEE

Source	SRP	HiGRID	PIV	SSPIV	DPD-MUSIC
1	1.54°	1.15°	13.12°	0.87°	0.69°
2	1.87°	1.34°	5.76°	1.21°	2.04°
3	1.39°	0.32°	6.86°	1.56°	1.17°
4	0.80°	0.36°	3.51°	1.28°	0.56°
Average	1.40°	0.79°	7.31°	1.23°	1.12°

Table 4.2: DOA estimation errors for four concurrent violin sources using different methods. © 2018, IEEE

Source	SRP	HiGRID	PIV	SSPIV	DPD-MUSIC
1	1.21°	1.15°	4.89°	6.60°	3.45°
2	1.05°	1.34°	10.07°	4.11°	4.21°
3	1.15°	0.70°	20.41°	2.31°	2.28°
4	1.05°	1.18°	11.92°	7.20°	1.12°
Average	1.12°	1.09°	11.82°	5.06°	2.77°

Figure 4.2: Setup for the recording of the classical quartet. © 2018, IEEE

Table 4.3: Reference and estimated DOAs for the microphone array recording of the classical quartet. © 2018, IEEE

(θ, ϕ)	**Violin 1**	**Violin 2**	**Viola**	**Cello**
Reference	$(116.1°, 92.4°)$	$(114.9°, 32.3°)$	$(124.4°, 327.8°)$	$(132.4°, 268.4°)$
$2608 - 5216$ Hz	$(109.7°, 89.4°)$	$(109.7°, 28.1°)$	$(112.3°, 340.4°)$	-
$1304 - 5216$ Hz	$(118.1°, 88.5°)$	$(107.1°, 30.9°)$	$(115.0°, 337.7°)$	$(138.1°, 276.4°)$

4.3 Evaluation of HiGRID-MUSIC

Table 4.4: Average DOA estimation errors for HiGRID-MUSIC. © 2018, IEEE

D/R ratio [dB]	$L = 1$	$L = 2$	$L = 3$	$L = 4$
-1.48	$2.88°$	$3.74°$	$4.74°$	$6.25°$
0.25	$2.57°$	$3.35°$	$3.35°$	$5.55°$
1.98	$2.75°$	$2.98°$	$4.09°$	$4.30°$
3.71	$2.36°$	$3.09°$	$3.86°$	$3.94°$
5.44	$4.10°$	$3.95°$	$3.00°$	$4.57°$

For the evaluation of HiGRID-MUSIC, only 4 seconds (01:00-01:04) of anechoic recordings of the fourth movement of Mahler's Symphony Nr. 1 [32] was used.
A different approach was selected for the evaluation of HiGRID-MUSIC. The simulated source positions (i.e. AIRs) were grouped into five clusters according to the D/R ratios at the recording positions. The D/R ratios of cluster centroids were -1.48, 0.25, 1.98, 3.71, 5.44 dB, respectively. The other condition for selecting sources was that the separation between any two AIRs must be greater than $\pi/4$ in each simulated scenario to satisfy the Rayleigh condition [9]. Four cases where $L = 1, 2, 3, 4$ were evaluated with 8 randomly generated scenarios for each D/R ratio cluster, where L is the number of sources. This evaluation resulted in a total of 160 randomly generated test scenarios.
Windowed Fourier transform with 1024-point FFT and a Hamming window with 25% overlap is used. For EB-MUSIC spectrum $J\tau = 4$ and $J_\kappa = 15$ were selected as smoothing parameters. All simulated scenarios were analyzed in frequency range between 2608 and 5314 Hz which allows decomposition order $N = 4$.
Extreme value for DOA estimation is defined as error value between the true and estimated DOAs paired using Kuhn-Munkres algorithm [34], if the error value was found to be larger than $\pi/4$ this excluded from the mapping. HiGRID analysis was at resolution level $l = 2$ (defined in Sec. 3.2) to complete the source count with $14.66°$ angular resolution. EB-MUSIC DOA estimation was carried out at $7.33°$ angular resolution.

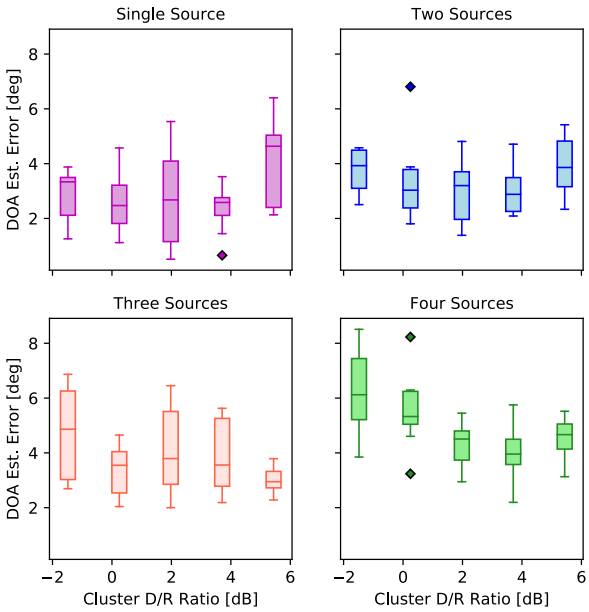


Figure 4.3: DOA estimation errors for 1, 2, 3 and 4 source cases. © 2018, IEEE

Table 4.4 presents DOA errors where Fig. 4.3 shows the distribution of DOA estimation errors for the tested cases using HiGRID-MUSIC. The results show that despite the challenging scenarios all of the DOA estimations are under angular resolution however for multiple source case performance of HiGRID-MUSIC is dependable to D/R ratio of the source location.

4.4 Evaluation of DPD-HiGRID with data-driven threshold selection

For the evaluations of DPD-HiGRID, sound scene emulations include two types of sources which were music and speech. Music signals were 4 seconds (01:00-01:04) of anechoic recordings of the fourth movement of Mahler's Symphony Nr. 1 and speech signals were energy normalized dry speech signals of 2 male and 2 female speakers recorded in METU SPARG audio lab. Four sources are positioned at distances 0.5 m, 1 m, 1.5 m far from RSMA with directions $(90°, 0°)$, $(90°, 90°)$, $(90°, 180°)$ and $(90°, 270°)$. All scenarios include four sources with fixed directions but changing distances correspond to average direct-to-reverberant (D/R) ratios of 10.72, 5.69, and 2.12 dB, respective to distance. Scenario 1 is the closest evaluation setup where Scenario 3 has the most remote source locations.
$J\tau = 4$ and $J_\kappa = 15$ were selected as smoothing parameters for DPD-test. All simulated scenarios are analyzed in frequency range between 2608 and 5216 Hz which allows decomposition order $N = 4$ and SHD coefficients order is selected $N = 3$ for the calculation of spatial correlation matrices. Windowed Fourier transform with 1024-point FFT and a Hamming window with 75% overlap is used for DPD-HiGRID evaluation.
After obtaining effective ranks for each bin and fitting them into GPD, selection of bins using DPD-test with proposed threshold selection were done for five different

probability values, $\widehat{P} = 0.00625, 0.0125, 0.025, 0.05$, and 0.1.
Table 4.5 shows the threshold values calculated for different scenarios at different probability levels for DPD-test where values with bold typeface indicate cases where less than four sources were localised. Secondly, table 4.6 shows the number of bins selected from a total of 42552 time-frequency bins (corresponding 4 seconds signal duration) in each tested case. Finally, table 4.7 demonstrate DOA estimation errors for each scenario. Results showed that proposed threshold selection decreases number of TF-bins selected significantly while maintaining DOA estimation errors at an acceptable level.

Table 4.5: DPD test thresholds for different probabilities, $\widehat{P}$.

$\widehat{P}$	0.00625	0.0125	0.025	0.05	0.1
Scenario 1 (Violins)	79.6	53.7	35.6	23.0	14.2
Scenario 2 (Violins)	46.7	34.9	25.5	18.0	12.2
Scenario 3 (Violins)	**34.7**	26.6	20.0	14.5	10.0
Scenario 1 (Speech)	**459.5**	**226.5**	110.9	53.6	25.2
Scenario 2 (Speech)	**94.0**	**58.7**	36.2	21.8	12.6
Scenario 3 (Speech)	**31.6**	**22.7**	16.0	11.0	7.2

Table 4.6: Number of bins selected for different probabilities, $\widehat{P}$.

$\widehat{P}$	0.00625	0.0125	0.025	0.05	0.1
Scenario 1 (Violins)	379	675	1414	2536	4426
Scenario 2 (Violins)	386	727	1344	2433	4733
Scenario 3 (Violins)	**502**	828	1328	2335	4570
Scenario 1 (Speech)	**55**	**191**	797	2389	5102
Scenario 2 (Speech)	**67**	**287**	1013	2638	5317
Scenario 3 (Speech)	**164**	**519**	1364	2825	5494

Table 4.7: DOA estimation errors in degrees for DPD-HiGRID.

$\widehat{P}$	0.00625	0.0125	0.025	0.05	0.1
Scenario 1 (Violins)	3.27°	2.77°	2.24°	1.91°	1.52°
Scenario 2 (Violins)	4.36°	2.92°	2.06°	1.10°	1.38°
Scenario 3 (Violins)	**4.01°**	3.34°	1.56°	0.99°	1.21°
Scenario 1 (Speech)	**8.31°**	**3.31°**	3.53°	3.14°	2.30°
Scenario 2 (Speech)	**0.02°**	**2.41°**	2.61°	2.74°	1.92°
Scenario 3 (Speech)	**1.12°**	**3.32°**	2.87°	1.56°	1.05°

CHAPTER 5

CONCLUSION

5.1 Discussion

In this section, results of evaluations for proposed methods are discussed in the order of evaluations. First, HiGRID results showed that HiGRID performed favourably when compared to state-of-the-art methods for DOA estimation. Estimation of multiple sources in highly reverberant environment is a challenge and HiGRID can accurately estimate coherent sources as well as incoherent one (i.e. speech). Another proof of adequacy for HiGRID as an DOA estimation method is that it operated well real recordings (see Table 4.3). Usually prior DOA estimation methods (see Section 2.5) fail under reverberation and however HiGRID is robust to reverberation and additive noise. For example, PIV performed poorly since it uses only zeroth and first order spherical harmonics and designed for single source detection however even in single source case it suffers from high reverberation of room. SRP, SSPIV and DPD-MUSIC are state-of-the-art methods however all of them took longer time to process when compared to HiGRID.

On the other hand, SRPD map is based SRP so it should be noted that an important limitation of HiGRID, along with other methods based on steered response in that two sources further than $\Theta_{\Delta} = \pi/N$ cannot be discriminated where N is the maximum decomposition order, this is called Rayleigh condition mentioned in Section 2.5.1.1. In the case of HiGRID evaluation, in which em32 microphone array is used, maximum decomposition order is $N = 4$ so if two sources are closer than 45° then both DOA estimation for sources are affected from this condition.

In the case HiGRID-MUSIC all DOA estimations were below angular resolution of search grid except the lowest D/R ratio with four sources which can be considered as an extreme case. Applying HiGRID reduced computational cost of EB-MUSIC by decreasing regions to be processed. HiGRID also solved a major constraint of EB-MUSIC spectrum which is the prior information of source count to define signal and noise subspaces. The DOA error results showed that proposed method performed accurate DOA estimations of multiple coherent sources as long as sources have similar D/R ratio in a reverberant environment.

Finally, results of DPD-HiGRID with new threshold selection method showed that it can improve both DPD test and HiGRID. Looking at Table 4.5 and Table 4.6 it can be observed that decreasing probability $\widehat{P}$ corresponds to a higher DPD test threshold which results number of selected bins approximately coincide with $\hat{P} \times 100\%$ of the total number of tested bins. It is possible to use high DPD thresholds without degrad-

ing DOA estimation performance however it should be noted selecting threshold too high will result with information loss and less accurate DOA estimation. Using DPD-test as a TF-bin selection method reduced the computational cost of HiGRID since only single source dominant bins are selected and computation of HiGRID increase with the number of source existing in that TF bin. The proposed method mutually improved DPD-test and HiGRID.

5.2 Conclusion

In this final chapter of thesis, the work reported is summarized with reflections on proposed methods and possible future work. The work presented in this thesis investigates various types of DOA estimation methods with evaluations including multiple sound sources in a reverberant room conditions. HiGRID is a novel proposed DOA estimation method and the center of this thesis where other methods are centered around HiGRID. Derivations of method HiGRID-MUSIC and DPD-HiGRID have perform well in acoustically adverse conditions.

HiGRID, HiGRID-MUSIC and DPD-HiGRID show promising results in terms of DOA estimation errors. Combinations of HiGRID with EB-MUSIC and DPD-test are indicators of versatile nature of proposed algorithm. As possible future work HiGRID can be manipulated to work with other RSMA based methods for achieving faster and accurate DOA estimation. Real time applications needs faster processing performance, all proposed methods are implemented using Python 3 and for real time application they can be implemented machine language like C++.

To summarize, the proposed methods are suitable for many spatial audio applications including surveillance systems, teleconference systems and user-oriented applications like hand-held devices and augmented reality applications.

To conclude my thesis, during the development of the reported work I have learned a lot about spherical harmonics and DOA estimation methods using RSMAs. Being a part of developing a novel method and extending it with other methods informed me and improved my personal knowledge about the area, last but not least, prepared me for the my future academia path.

REFERENCES

[1] B. Rafaely, *Fundamentals of Spherical Array Processing*. Springer-Verlag, 2015.

[2] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. II–1781–II–1784, May 2002.

[3] mh acoustics, "em32 eigenmike® microphone array release notes (v17.0)," 2013.

[4] Z. Li, R. Duraiswami, E. Grassi, and L. S. Davis, "Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays," *Proc. IEEE Int. Conf. on Acoust. Speech and Signal Process. (ICASSP 2004)*, pp. IV–41 – IV–44, 2004.

[5] E. G. Williams, *Fourier Acoustics*. London: UK:Academic Press, 1999.

[6] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, pp. 1408–1418, Aug 1969.

[7] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*, vol. 9 of *Springer Topics in Signal Processing*. Springer, Aug. 2016.

[8] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, "Spherical Microphone Array Beamforming," in *Speech Processing in Modern Communication* (G. S. Cohen I, Benesty J, ed.), pp. 281–305, Heidelberg, Germany: Springer-Verlag, 2010.

[9] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J. Acoust. Soc. Am.*, vol. 116, pp. 2149–2157, Oct. 2004.

[10] F. Jacobsen, *Sound Intensity*, pp. 1053–1075. New York, NY: Springer New York, 2007.

[11] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," *Proc. 18th European Signal Process. Conf. (EUSIPCO 2010)*, pp. 442–446, August 2010.

[12] S. Hafezi, A. H. Moore, and P. A. Naylor, "Multiple source localization in the spherical harmonic domain using augmented intensity vectors based on grid search," in *Proc. 24th European Signal Process. Conf. (EUSIPCO 2016)*, (Budapest, Hungary), pp. 602–606, August 2016.

[13] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing," *J. Acoust. Soc. Am.*, vol. 131, pp. 2828–2840, April 2012.

[14] H. Teutsch and W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," in *Proc. IEEE Int. Conf. on Acoust. Speech and Signal Process. (ICASSP-08)*, (Las Vegas, NV, USA), pp. 5276–5279, Mar. 31 - Apr. 4 2008.

[15] A. H. Moore, C. Evers, and P. A. Naylor, "Direction of arrival estimation in the spherical harmonic domain using subspace pseudointensity vectors," *IEEE/ACM Trans. on Audio, Speech and Language Process.*, vol. 25, pp. 178–192, January 2017.

[16] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, pp. 276–280, 1986.

[17] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," vol. 10, (New Paltz, NY, USA), pp. 221–224, October 2009.

[18] O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," *IEEE/ACM Trans. on Audio, Speech and Language Process*, vol. 10, pp. 1494–1505, October 2013.

[19] M. B. Çöteli, O. Olgun, and H. Hacıhabiboğlu, "Multiple sound source localization with steered response power density and hierarchical grid refinement," *IEEE/ACM Trans. on Audio, Speech and Lang. Process.*, vol. 26, pp. 2215 – 2229, November 2018.

[20] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, Feb. 1990.

[21] S. Böck and G. Widmer, "Maximum filter vibrato suppression for onset detection," in *Proc. 16th Int. Conf. on Digital Audio Effects (DAFx-13)*, (Maynooth, Ireland), Sept. 2-5 2013.

[22] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, "HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere," *Astrophys. J.*, vol. 622, pp. 759–771, 2005.

[23] M. Batty, "Spatial entropy," *Geographical Analysis*, vol. 6, pp. 1–31, 1974.

[24] P. Soille, *Morphological image analysis: principles and applications,*. Heidelberg, Germany: Springer-Verlag, 2004.

[25] C. Buchta, M. Kober, I. Feinerer, and K. Hornik, "Spherical k-means clustering," *J. Stat. Software*, vol. 50, pp. 1–22, October 2012.

[26] A. Moore, C. Evers, P. A. Naylor, D. L. Alon, and B. Rafaely, "Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test," in *Proc. 23rd European Signal Process. Conf. (EUSIPCO-15)*, (Nice, France), pp. 2296–3000, August 2015.

[27] B. Rafaely and D. Kolossa, "Speaker localization in reverberant rooms based on direct path dominance test statistics," in *2017 IEEE Int. Conf. on Acoust. Speech and Signal Process. (ICASSP'17)*, (New Orleans, USA), pp. 6120–6124, March 2017.

[28] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*, vol. 208. Springer, 2001.

[29] J. R. M. Hosking and J. R. Wallis, "Parameter and Quantile Estimation for the Generalized Pareto Distribution," *Technometrics*, vol. 29, pp. 339–349, March 1987.

[30] V. Choulakian and M. A. Stephens, "Goodness-of-fit tests for the generalized pareto distribution," *Technometrics*, vol. 43, pp. 478–484, November 2001.

[31] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. 108th Audio Eng. Soc. Convention*, no. Preprint #5093, (Paris, France), Feb. 1 2000.

[32] J. Pätynen, V. Pulkki, and T. Lokki, "Anechoic recording system for symphony orchestra," *Acta Acust. united with Acust.*, vol. 94, pp. 856–865, November 2008.

[33] Bang and Olufsen, "Music for Archimedes." Audio CD, 1992.

[34] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logist. (NRL)*, vol. 2, pp. 83–97, January 1955.