

POLITECNICO DI MILANO  
UNIVERSITY CAMPUS OF COMO

School of Industrial and Information Engineering

Master of Science in  
Computer Engineering



# Recontruction of the soundfield in arbitrary locations using the Distributed Ray Space Transform

Supervisor:

DR. FABIO ANTONACCI

Assistant Supervisor:

DR. FEDERICO BORRA

Master Graduation Thesis of:

MIRCO PEZZOLI

Candidate Id n. 863382

Academic Year 2016-2017



POLITECNICO DI MILANO  
POLO TERRITORIALE OF COMO

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Laurea Specialistica in  
Ingegneria Informatica



# Ricostruzione del campo acustico in posizioni arbitrarie per mezzo della Ray Space Transform Distribuita

Relatore:

DR. FABIO ANTONACCI

Correlatore:

DR. FEDERICO BORRA

Tesi di Laurea Specialistica di:

MIRCO PEZZOLI  
Matricola n. 863382

Anno accademico 2016-2017



## ABSTRACT

---

This thesis arises from the will of exploiting the recent technologies and developments in compact representation of the acoustic field to implement one of the classic applications present in the literature, called *Virtual Miking* (VM). The goal of this thesis is to develop an efficient and flexible VM technique, analyzing the acoustic scene by means of multiple arrays arbitrary distributed in the space. The process of *Virtual Miking* consists in the acquisition of the signals from one or more *virtual microphones* which are freely placed in the analysis space. This is possible only if the acoustic field is known in the relative locations. In order to characterize the acoustic field the proposed technique uses a parametric soundfield model that is formulated in time-frequency domain. It assumes that each time-frequency instant of a microphone signal can be decomposed into three different components: the source signal, the source directivity and the propagating function. The knowledge of these three parameters allows the reconstruction of the acoustic field in every position. This gives the possibility of simulating the signal acquired by a microphone arbitrary placed in the space of analysis. A proper procedure for retrieving the parameters from the soundfield has been proposed starting from the microphone arrays signals. The use of multiple arrays allows us to capture the whole acoustic scene, but a proper representation is necessary. Among the different representation paradigms present in the literature, we adopted a geometrical acoustics inspired representation called *ray space*. In this context the soundfield is described in terms of acoustic rays which are mapped onto the ray space gaining an efficient and sturdy description: the soundfield image. Indeed, points in the ray space domain represent rays in the geometric space. Since the soundfield is analyzed by multiple microphone arrays, practically an extension of the ray space called *projective ray space* has been implemented. The parameters needed by the VM technique are extracted starting from the soundfield image and finally the VM signal is computed according to the parametric model. Through simulations and experiments we have proven that our VM technique is effective. Thanks to its flexibility the our VM procedure is suitable for being employed in a wide range of areas such as studio recording and forensic analysis.



## SOMMARIO

---

Questa tesi di laurea nasce dalla volontà di sfruttare i recenti sviluppi tecnologici e gli ultimi approcci nell'area della rappresentazione del campo acustico per implementare una delle classiche applicazioni presenti nella letteratura, denominata: *Virtual Miking*. L'obiettivo di questo lavoro è quello di ottenere una tecnica di Virtual Miking flessibile ed efficace, analizzando la scena sonora con schiere (array) di microfoni in numero e posizione arbitraria. Il processo di Virtual Miking consiste nel libero posizionamento di uno o più *microfoni virtuali* all'interno dello spazio d'analisi e nell'acquisizione dei relativi segnali. Ciò è possibile solo se si è a conoscenza del campo acustico in tali posizioni. Per descrivere il campo sonoro, la tecnica proposta sfrutta un modello parametrico del campo formulato nel dominio tempo-frequenza. Il modello assume che ad ogni istante in questo dominio, il segnale di un microfono possa essere decomposto in tre fattori differenti: il segnale sorgente, la direttività della sorgente e la funzione di propagazione. La conoscenza di questi tre parametri permette la ricostruzione del campo acustico in ogni posizione. Questo dà la possibilità di simulare il segnale acquisito da un microfono posizionato arbitrariamente nello spazio. È stata proposta un'appropriata procedura per estrarre i parametri dal campo acustico a partire dai segnali acquisiti dalle schiere di microfoni. L'utilizzo di numerosi array ci permette di catturare l'intera scena acustica, ma è necessario rappresentare l'informazione adeguatamente. Tra le diverse rappresentazioni presenti in letteratura, ne abbiamo adottata una ispirata alla acustica geometrica detta: *ray space*. In questo paradigma di rappresentazione, il campo sonoro è descritto in termini di raggi acustici che sono mappati nel *ray space* ottenendo così una descrizione efficiente e robusta: la *soundfield image*. Infatti, i punti nel dominio *ray space* rappresentano i raggi nello spazio geometrico. Nella pratica, dato che il campo sonoro è analizzato da più array di microfoni, è stata implementata una estensione della rappresentazione *ray space* chiamata *projective ray space*. I parametri richiesti dalla tecnica di VM sono estratti a partire dalla *soundfield image* e infine il segnale del microfono virtuale calcolato secondo il modello parametrico. L'efficacia della procedura di VM è stata comprovata attraverso simulazioni ed esperimenti. Grazie alla sua flessibilità, la nostra tecnica di VM si adatta ad essere impiegata in una ampia gamma di aree come lo *studio recording* e l'analisi forense.





# TABLE OF CONTENTS

---

<b>List of figures</b>	<b>ix</b>
<b>List of tables</b>	<b>xi</b>
<b>Introduction</b>	<b>xiii</b>
<b>1 Background</b>	<b>1</b>
1.1 Ray Space . . . . .	1
1.1.1 The Plenacoustic Function . . . . .	2
1.1.2 Projective Ray Space . . . . .	3
1.1.3 Acoustic Primitives in the Projective Ray Space . . . . .	4
1.1.4 Managing Multiple OW . . . . .	7
1.1.5 Soundfield Images in the Projective Ray Space . . . . .	8
1.2 Spatial Filtering . . . . .	9
1.2.1 Signal Model . . . . .	10
1.2.2 Linear Constrained Minimum Variance Beamformer . . . . .	12
1.2.3 Delay and Sum Beamformer . . . . .	13
1.2.4 Null Steering Beamformer . . . . .	14
1.2.5 Minimum Variance Distortionless Response Beamformer . . . . .	15
1.3 Summary . . . . .	16
<b>2 State of the Art</b>	<b>17</b>
2.1 Acoustic Signal Representation . . . . .	17
2.1.1 Geometric Representation . . . . .	18
2.2 Virtual Miking State of the Art . . . . .	21
2.2.1 Problem Formulation . . . . .	22
2.2.2 Parameters Estimation . . . . .	24
2.2.3 Virtual Microphone Synthesis . . . . .	27
2.3 Summary . . . . .	28

<b>3</b>	<b>Extraction of sound parameters</b>	<b>29</b>
3.1	Setting . . . . .	30
3.1.1	Problem Formulation . . . . .	31
3.2	Soundfield Image Construction . . . . .	32
3.3	Source Localization . . . . .	34
3.4	Radiance Pattern Estimation . . . . .	36
3.5	Source Signal Retrieval . . . . .	40
3.6	Summary . . . . .	42
<b>4</b>	<b>Virtual Miking</b>	<b>45</b>
4.1	Synthesis of the Virtual Microphone signal . . . . .	45
4.2	Summary of the Virtual Miking technique . . . . .	48
<b>5</b>	<b>Results</b>	<b>51</b>
5.1	Measurements Setup . . . . .	51
5.1.1	Source Localization Metrics . . . . .	53
5.1.2	Directivity Metrics . . . . .	54
5.1.3	VM Signal Metrics . . . . .	55
5.2	Simulations . . . . .	56
5.2.1	Source Localization . . . . .	57
5.2.2	Pattern Estimation . . . . .	57
5.2.3	VM Signal Synthesis . . . . .	60
5.3	Experiments . . . . .	62
5.3.1	Source localization . . . . .	65
5.3.2	VM signal synthesis . . . . .	65
5.4	Summary . . . . .	68
<b>6</b>	<b>Conclusions and future works</b>	<b>71</b>
	<b>References</b>	<b>75</b>
	<b>Appendix A Equipment specification</b>	<b>77</b>

## LIST OF FIGURES

---

1.1	The local and the global parametrization of rays. The Euclidean parametrization of rays $(\mathbf{m}, \mathbf{q})$ is defined with respect to the local reference frames. The line on the second OW $\mathbf{y}' = \mathbf{m}'\mathbf{x}' + \mathbf{q}'$ cannot be represented in the reference frame of the first OW. With the projective parametrization $(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3)$ any reference frame can be assumed as the global one, letting us to write the line incident on the second OW in the $(\mathbf{x}, \mathbf{y})$ frame. . . . .	4
1.2	A segment in the geometric space and its image in the reduced ray space . .	5
1.3	A source sensed by OW $\mathbf{p}_A \mathbf{p}_B$ and its ROI in the reduced ray space . . . . .	7
1.4	The reflector $\mathbf{p}_A \mathbf{p}_B$ and the OW $\mathbf{p}_C \mathbf{p}_D$ in the geometric space and the ROI of the reflector in the reduced ray space. Note that the ROI is obtained from the intersection of the visibility regioni of the OW and the image of the reflector. . . . .	7
1.5	The Region of Visibility of the source in the setup of Fig. 1.3 with an added reflector $\mathbf{p}_D \mathbf{p}_C$ . . . . .	8
1.6	Uniform Linear Array . . . . .	12
2.1	Relation between the wavenumber vector and the acoustic rays in a plane wave field Fig. 2.1a and spherical wave field Fig. 2.1b. The dashed lines and circles represent the eikonals while the dotted lines stand for the acoustic rays trajectories. . . . .	21
2.2	The VM processing block in the case of two microphone arrays . . . . .	27
3.1	The analysis set-up used in the <i>Virtual Miking</i> procedure. Sources (red dots) are randomly located in a circular area of ray $2[m]$ and centered in $\mathbf{o} = [2, 2]^T [m]$ . The microphone arrays are placed all around the sources area, while the virtual microphone (green dot) can be arbitrary placed in the space. . . . .	30
3.2	Most common microphone array configurations . . . . .	31
3.3	Geometry of a circular array made of $N$ microphones . . . . .	32
3.4	An example of radiance patterns with their relative formula. . . . .	33

3.5	The pseudospectrum of an array. The main peak represents the DOA of an acoustic source. The corresponding $x$ value gives the angle from which the microphone array captures the source signal, namely the Direction of Arrival (DOA). . . . .	34
4.1	The VM procedure block diagram. The signals coming from the microphones arrays are analyzed by the three main modules in order to retrieve the parameters: the sources positions $\mathbf{d}^i$ , the pattern coefficients $\mathbf{c}$ and the source signals $S$ . Then the estimated parameters and the position of the VM are used to compute the signal of the virtual microphone. . . . .	50
5.1	The setup used in the simulations. The arrays are arbitrary located all around the sources (red). The virtual microphone (green) lies almost in the center of the scene. The directivity is shown in the case of $0^\circ$ orientation for both the sources. . . . .	57
5.2	Estimated directivity in the case of one and two sources from tests 1 and 3. The blue thick line represents the reference cardioid pattern. . . . .	60
5.3	Experiment setup in the semi-anechoic chamber . . . . .	60
5.4	Uniform Circular Array structure with microphones. . . . .	63
5.5	The experiments setting. The six microphones arrays are deployed in the scene around the two sound sources in red. Four microphones (green) are placed in the scene as reference for the VMs signals . . . . .	63
5.6	. . . . .	67
5.7	The estimated radiance pattern of source 1 in experiment 1. Frequencies up to $1000\text{Hz}$ show an omnidirectional-like pattern while the highest frequencies point to $90^\circ$ . . . . .	68
A.1	Measurement microphone MM1 . . . . .	78
A.2	Measurement microphone MM1 . . . . .	79
A.3	Measurement microphone ECM 800 . . . . .	80
A.4	Measurement microphone AT 4022 . . . . .	81
A.5	Measurement microphone AT 4022 . . . . .	82
A.6	Microphones preamplifier . . . . .	83
A.7	Microphones preamplifier . . . . .	83
A.8	ADC/DAC converter . . . . .	84
A.9	Loudspeaker . . . . .	85

## LIST OF TABLES

---

5.1	General setting of tests . . . . .	52
5.2	Parameters shared by all the simulation tests. Positions are expressed in [m].	56
5.3	Source localization performance in simulated tests. For each test we report which one of the sources is active and the orientation of the pattern. The value $0^\circ$ corresponds to an orientation parallel to the $x$ axis in the positive direction while $\pm 90^\circ$ corresponds to an orientation parallel to the $y$ axis with respect to Fig. 5.1. . . . .	58
5.4	Pattern estimation performance in simulated tests. For each test we report which one of the sources is active and the orientation of the pattern. The value $0^\circ$ corresponds to an orientation parallel to the $x$ axis in the positive direction while $\pm 90^\circ$ corresponds to an orientation parallel to the $y$ axis with respect to Fig. 5.1. . . . .	59
5.5	VM synthesis performance in simulated tests. For each test we report which one of the sources is active and the orientation of the pattern. The value $0^\circ$ corresponds to an orientation parallel to the $x$ axis in the positive direction while $\pm 90^\circ$ corresponds to an orientation parallel to the $y$ axis with respect to Fig. 5.1. . . . .	61
5.6	Hardware equipment . . . . .	62
5.7	Experiment parameters . . . . .	64
5.8	Source localization performance in experimental tests. For each test we report which one of the sources is active and the orientation of the pattern. The value $0^\circ$ corresponds to an orientation parallel to the $x$ axis in the positive direction while $\pm 90^\circ$ corresponds to an orientation parallel to the $y$ axis with respect to Fig. 5.5. . . . .	66

5.9	VM signal synthesis performance in experimental tests. For each test we report which one of the sources is active and the orientation of the pattern. The value $0^\circ$ corresponds to an orientation parallel to the $x$ axis in the positive direction while $\pm 90^\circ$ corresponds to an orientation parallel to the $y$ axis with respect to Fig. 5.5. . . . .	69
-----	--	----

## INTRODUCTION

---

In the last few years an increasing interest in the microphone array signal processing techniques has been experienced. The steady decrease of the cost of acoustic transducers and the development of convenient representations of the acoustic field [10] give new possibilities of research in the field of space-time signal processing. Both researchers and companies are putting efforts into developing advanced space-time processing techniques that can be applied in many different areas such as environment monitoring, human-computer interaction and music recording. One of the most well known applications in the literature is the *Virtual Miking* (VM). The VM is a procedure which allows us to arbitrary place one or more virtual microphones and acquire their signals. In this thesis, we will introduce a novel and flexible *Virtual Miking* technique inspired by the state of art in acoustic signal representation and space-time processing literature.

Recently, the geometrical representation of the acoustic field has been exploited in the development of an efficient soundfield representation paradigm [10]. Geometrical acoustics is based on the description of the soundfield in terms of acoustic rays. Although it is a rough approximation of the acoustic field, this model has the makings of describing arbitrary complex acoustics environments, for which other representations are not practical. Starting from the geometrical representation, the concept of the ray space has been introduced as a domain whose primitives are the acoustic rays [3]. This domain allows a sparse and compact representation of the soundfield in which every point corresponds to a plane wave component of the field. In [10] the authors develop an analysis tool called soundfield imaging which aims in mapping the plenacoustic function [24](i.e. the function characterizing the acoustic radiance in the space) onto the ray space. The tool used to derive this description is the so called Observation Window, namely an array of microphone. Within the ray space context all the main acoustic primitives are mapped onto linear patterns, so, many problems such as source localization, environmental inference and radiometric analysis can be solved with pattern processing algorithms taken from the wide computer vision literature. In this thesis we are interested in managing multiple microphone arrays, so we will adopt the projective ray space [11]. In [11] the authors proposed the projective ray space as an extension of the

ray space able to overcome its main limitations, such as the possibility of managing multiple observation windows. In our thesis, we will exploit the soundfield image in the projective ray space as a tool for implementing a robust source localization [5]. Indeed, we will adopt a parametric model for describing the acoustic scene, so the geometric description will be a complementary representation useful in retrieving the parameters we need.

The *Virtual Miking* procedure developed will take advantage of a parametric description inspired by the ones available in the literature [25] [6]. Our model is simple and flexible, we are able to compute the sound pressure in every point of the space just knowing few parameters. The needed parameters are the following:

1. Sources positions.
2. Sources signals.
3. Sources radiance patterns.
4. Propagating function.
5. Virtual microphone positions.

Some parameters are known or assumed (Propagating function and Virtual microphone positions), while the others have to be extracted from the analysis of the acoustic scene. A key innovation of our VM procedure with respect to the literature [25] lies in the fact that we take into account of the source radiance pattern in the model. This envisions the possibility of developing application driven by this information. As regards the source signal and radiance pattern retrieval, we have developed two methods for extracting these informations from the signals. Both methods solves an optimization problem directly derived from the adopted model. Another important feature is represented by the acquisition devices. We can arbitrary place the microphone arrays in the space with the only limitation of using two or more devices. We choose to adopt circular array of microphones rather than the usual linear structure due to its advantages in terms of spatial filtering [26]. Moreover the circular array has been adopted also in consumer products such as: *Amazon Echo*, *Google Home* and *Apple Homepod*.

The thesis is organized as follows. In Chapter 1 we will provide an overview of the main theoretical background on which our work is based. First we will introduce the ray space and the projective ray space, then we will describe the main space-time processing techniques adopted in this thesis. In Chapter 2 we will present the state of art of acoustic signal representation and *Virtual Miking*. We will describe the main paradigm available in literature to represent the acoustic information and a *Virtual Miking* technique similar to the



one proposed in this work. In Chapter 3 we will start the description of our *Virtual Miking* procedure. First we will introduce the problem of *Virtual Miking* and the choices we made to derive a solution. We will describe the model and assumptions adopted in the development of the work. Then, we will give an in depth analysis of the source localization through the projective ray space representation. Moreover we will describe in details the developed methods for retrieving the model parameters from the microphones signals. Chapter 4 will be spent to present the synthesis of the virtual microphone signal and we will give a summary of the entire *Virtual Miking* technique. In Chapter 5 we will corroborate our VM procedure with ad hoc designed experiments and simulations. Finally in the last chapter we will draw conclusions and discuss applications and possible future developments.



## BACKGROUND

---

In this chapter we will introduce the readers to the main concepts which constitute the basis of the thesis. In particular we will describe in detail the soundfield representation used to efficiently encode the whole acoustic scene. This representation is called Projective Ray Space and has its roots in the theory of geometrical acoustics. The second important concept is spatial filtering. We present the classical spatial signal processing techniques with focus on the ones adopted in our work. We will describe in detail the state-of-the-art of beamforming methods used in the development of our *Virtual Miking* technique. First we will present the general design procedure and the optimization problems solved by beamformers. Then we will give an in depth analysis of three well-known techniques.

### 1.1 Ray Space

In the literature, we can find different soundfield representation paradigms for analyzing and characterizing the objects of an acoustic scene. Usually those techniques exploit arrays of microphones for capturing the acoustic scene. Compared with other techniques, the soundfield image [10] has the key advantage that it gathers and organizes at once and in a single representation, all the informations needed in order to develop different acoustic scene analysis applications.

We first introduce the plenacoustic function which is able to describe the acoustic radiance in every direction through every point in the scene. Spatial filtering is used to retrieve the directional informations. The plenacoustic function has been related to geometrical acoustics through the adoption of the ray space [10], as a new domain for its parameterization. The ray space is defined as the space of parameters that describe acoustic rays. In doing so the sound field gains a sparse and sturdy description. This domain is called ray space because each point in the domain corresponds to a ray in the acoustic scene. Hence every point in

the soundfield image corresponds to a ray crossing the so called Observation Window (OW), which is the microphone array that analyze the scene. As result of the change of domain, all the main acoustic primitives are mapped onto linear pattern in the soundfield image. This enables the employment of computer vision techniques to perform the most common acoustic analysis.

The Ray Space exploits an Euclidean parametrization, while Projective Ray Space [11] is based on projective geometry and it can be considered as an extension of the first one, since it is able to manage rays coming from any direction and multiple arrays. Hence, in this thesis we adopted the Projective Ray space in order to implement a robust source localization with multiple arrays distributions which is the first step of our *Virtual Miking* procedure.

### 1.1.1 The Plenacoustic Function

The Plenacoustic Function (PF) is introduced in [24] as the acoustic counterpart of the plenoptic function [2]. The PF defines the contribution of the soundfield at a given point coming from a given direction. In this thesis we limit the analysis to the 2D space, so the PF is a five-dimensional function  $f(x, y, \theta, \omega, t)$  of: position  $(x, y)$ ; direction  $(\theta)$ ; frequency  $(\omega)$  and time  $(t)$ . In [1] the PF is captured by mean of sensor arrays acting as Observation Windows of the acoustic scene. The directional information brought by the beamforming operation is used to scan the acoustic field for a discrete set of directions. A planar geometry scenario is assumed such that the soundfield  $p(\mathbf{r}, \omega)$ , at point  $\mathbf{r} = [x, y]^T$ , is described as the superposition of plane waves with direction of propagation  $\theta$ . Taking the integral over the propagating plane-wave components we can define the soundfield as follows:

$$p(r, \omega) = \frac{1}{2\pi} \int_0^{2\pi} P(\mathbf{k}(\theta)) e^{i\langle \mathbf{r}, \mathbf{k} \rangle} d\theta. \quad (1.1)$$

This expression represents the well known plane-wave decomposition of the acoustic wavefield where  $P(\mathbf{k}(\theta))$ , the Hergoltz density function, encodes the amplitude and the phase of each plane-wave component. The PF is formally defined as the integrand in 1.1:

$$f(x, y, \theta, \omega) := e^{i\langle \mathbf{r}, \mathbf{k} \rangle} P(\mathbf{k}(\theta)). \quad (1.2)$$

The estimation of  $P(\mathbf{k}(\theta))$  can be obtained by means of a beamforming operation which will be introduced in Sec. 1.2. Assuming the geometrical acoustics paradigm we can state a direct relation between plane waves and acoustic rays. Considering the *Radiance Invariance*

Law, we can characterize the acoustic ray as the oriented line orthogonal with respect to the planar wavefront of direction  $\theta$ .

The parametrization proposed in [10] and called ray space introduces a correspondence between each point in the ray space domain and each acoustic ray. More specifically, rays are identified by the parameters of the line on which they lie. These parameters are referred to each OW considered as staying on the  $y$  axis with the origin in its middle. They are: the slope  $m$  and the intercept with the  $y$  axis  $q$ . The line equation does not specify a direction, so it is conventionally assumed that the rays cross the OW only in one direction. This is one of the limitations of Euclidean ray space.

### 1.1.2 Projective Ray Space

In the previous section we have defined an acoustic ray as an oriented line perpendicular to its planar acoustic wavefront. This lets us to graphically represent it as a line. The introduced ray space definition, also known as Euclidean ray space suffers from different limitations which can be solved with the adoption of the Projective Rays Space. The Euclidean ray space represents a local parametrization of the acoustic field, namely the parameters extracted by each OW refer only to the OW it self. In fact the image coordinates (i.e. the slope  $\mathbf{m}$  and the intercept  $\mathbf{q}$ ) are related to a single window as shown in Fig. 1.1. Furthermore is not possible to specify the direction of the rays. This limitations are crucial when the goal is to jointly use multiple observation windows and we have to represent all possible rays.

On contrary, Oriented projective geometry enables the simultaneous use of multiple OW observing the acoustic scene and allows us to parametrize all possible rays, irrespective of the reference frame with the line equation  $l_1x + l_2y + l_3 = 0$ . The equation can be written in vector form as:

$$\begin{aligned}\mathbf{p}^T \mathbf{l} &= 0; \\ \mathbf{l} &= [l_1, l_2, l_3]^T; \\ \mathbf{p} &= [x, y, 1]^T.\end{aligned}\tag{1.3}$$

We can notice that this representation is homogeneous because all vectors of the form  $\mathbf{l} = k[l_1, l_2, l_3]^T, k \neq 0$  represent the same line. Therefore they form a class of equivalence. In order to distinguish between two oppositely directed rays, the scaling factor  $k$  is limited to either positive or negative values only. This oriented geometric space is defined as **Projective Ray Space  $P$**  [11]

It is always possible to reduced the Projective Ray Space to its 2D counterpart, the *reduced ray space*, by cutting it with a plane. This is particularly useful in order to visu-

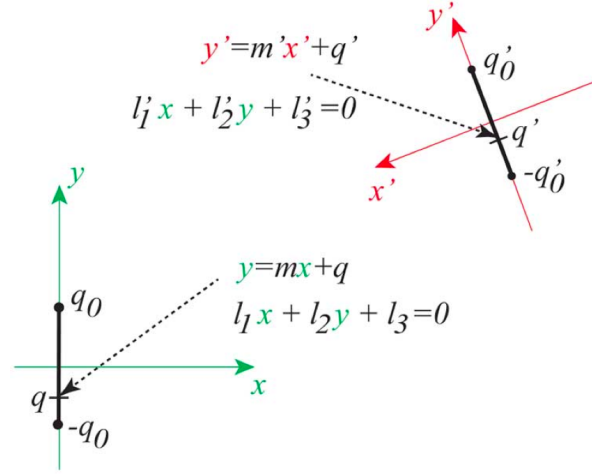


Fig. 1.1 The local and the global parametrization of rays. The Euclidean parametrization of rays  $(\mathbf{m}, \mathbf{q})$  is defined with respect to the local reference frames. The line on the second OW  $\mathbf{y}' = \mathbf{m}'\mathbf{x}' + \mathbf{q}'$  cannot be represented in the reference frame of the first OW. With the projective parametrization  $(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3)$  any reference frame can be assumed as the global one, letting us to write the line incident on the second OW in the  $(\mathbf{x}, \mathbf{y})$  frame.

alize the soundfield image. For instance, if we choose to cut with a prescribed plane, the resulting *reduced ray space* corresponds to the Euclidean (local) Ray Space described by the coordinates  $[m, q] = [-l_1/l_2, l_3/l_2]$ .

### 1.1.3 Acoustic Primitives in the Projective Ray Space

One of the main advantages of the Ray Space representation is that acoustic primitives are mapped onto linear features. In the current section we describe the main characteristics of this mapping for the most important acoustic primitives.

#### Rays

As discussed in Sec. 1.1.2 a ray in the geometric space corresponds to a point in the projective space  $P$ . Thanks to homogeneity of the representation the ray is visualized as a line, passing through the origin, of coordinates  $k[l_1, l_2, l_3]^T$ ,  $k > 0$

#### Sources

An acoustic source can be seen as the point from which all its acoustic rays originate. Therefore, the rays passing through a source of coordinates  $\mathbf{p}_A = k[x_A, y_A, 1]$ ,  $k > 0$ , clearly satisfy the following equation:

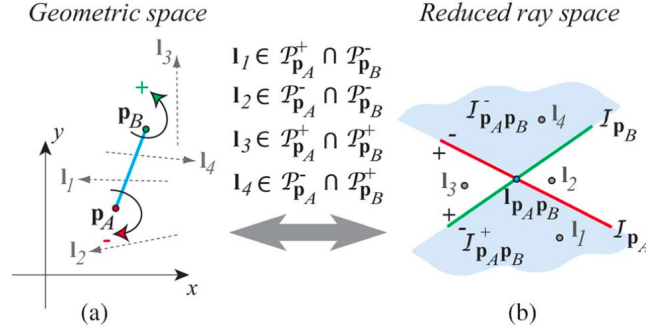


Fig. 1.2 A segment in the geometric space and its image in the reduced ray space

$$\mathbf{p}_A^T \mathbf{l} = 0$$

The identified set of rays forms, in the projective ray space, the image of  $\mathbf{p}_A$ :  $I_{\mathbf{p}_A} = \{\mathbf{l} \in P | \mathbf{p}_A^T \mathbf{l} = 0\}$ . It corresponds to a plane passing through the origin of the projective ray space which divides the space in two half-space  $P_{\mathbf{p}_A}^+ = \{\mathbf{l} \in P | \mathbf{p}_A^T \mathbf{l} > 0\}$  and  $P_{\mathbf{p}_A}^- = \{\mathbf{l} \in P | \mathbf{p}_A^T \mathbf{l} < 0\}$ . The first plane corresponds to all rays passing through  $\mathbf{p}_A$  and leaving it on their left, while the second one refers to rays leaving the source from their right.

### Segments

Segments are useful elements in the projective ray space because they are suitable to model both observation windows and acoustic reflectors. Consider two points  $\mathbf{p}_A$  and  $\mathbf{p}_B$  and the segment which connects them as shown in Fig. 1.2(a). The images of  $\mathbf{p}_A$  and  $\mathbf{p}_B$  are respectively the two plane, visualized as lines in the reduced ray space,  $I_{\mathbf{p}_A}$  and  $I_{\mathbf{p}_B}$ . We are able to determine the direction of a ray crossing the window looking at the different intersection regions of  $I_{\mathbf{p}_A}$  and  $I_{\mathbf{p}_B}$ . Referring to the Fig. 1.2 it is easy to identify the different types of rays.

The image of  $I_1$  rays is:

$$I_{\mathbf{p}_A \mathbf{p}_B}^+ = P_{\mathbf{p}_A}^+ \cap P_{\mathbf{p}_B}^-;$$

while the image of the second type is given by:

$$I_{\mathbf{p}_A \mathbf{p}_B}^- = P_{\mathbf{p}_A}^- \cap P_{\mathbf{p}_B}^+.$$

The two images are wedge-shaped regions delimited by the two planes  $I_{\mathbf{p}_A}$  and  $I_{\mathbf{p}_B}$  which meet in  $I_{\mathbf{p}_A \mathbf{p}_B}$ . This specific plane corresponds to the line which connects the endpoints in the geometric space. Rays which do not intercept the segment, named  $I_2$  and  $I_3$  in Fig. 1.2 are outside the discussed regions. Finally, we can define also the image of a non oriented segment as the union of the two images of the oriented one:

$$I_{\mathbf{p}_A \mathbf{p}_B} = I_{\mathbf{p}_A \mathbf{p}_B}^+ \cup I_{\mathbf{p}_A \mathbf{p}_B}^-.$$

### Observation Windows and Reflectors

Observation Windows are the model of sensor arrays. Their ray space representation is directly derived from the one of segments as explained above. Usually, OW are able to "sense" the rays that cross them in one of the two direction only. We are interested in managing multiple observation windows each corresponding to a different microphone arrays. For a one sided OW, we can identify the so called *Visibility Region*  $V^{(i)}$  (with superscript  $(i)$  identifying the  $i$ th OW) as one of the two wedge-shape regions  $I_{\mathbf{p}_A \mathbf{p}_B}^+$  and  $I_{\mathbf{p}_A \mathbf{p}_B}^-$ .

Acoustic reflectors are the acoustic primitives of reflective elements in the acoustic scene such as walls. Their geometric model is again the segment. The wavefront reflected by a planar wall can be thought of as originating from an image source mirrored about the wall and the source. The reflector is identified as an "illuminating window" since all rays coming from the image source are forced to pass through the reflector. It acts as an aperture that casts reflected acoustic radiance onto the scene.

### Regions of Interest (ROI)

The Region of Interest of a primitive (source or reflector) is defined as the portion of the image of that primitive visible from an observation window. We can easily obtain the ROI by intersecting  $V^{(i)}$  ( $i$ th OW visibility) with the primitive image. Let us consider a source placed in  $\mathbf{p}_s$ , its ROI can be identified as:

$$R_{\mathbf{p}_s}^{(i)} = I_{\mathbf{p}_s} \cap V^{(i)} \quad (1.4)$$

Similarly the ROI of a reflector is given by:



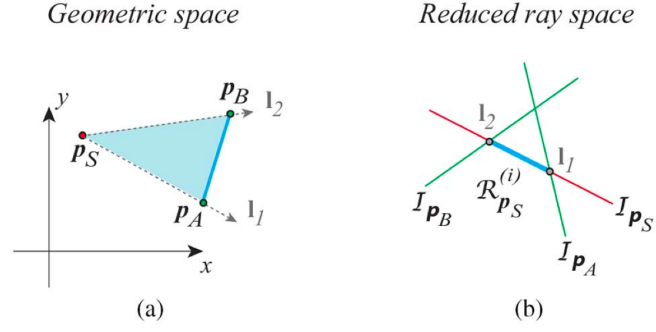


Fig. 1.3 A source sensed by OW  $p_A p_B$  and its ROI in the reduced ray space

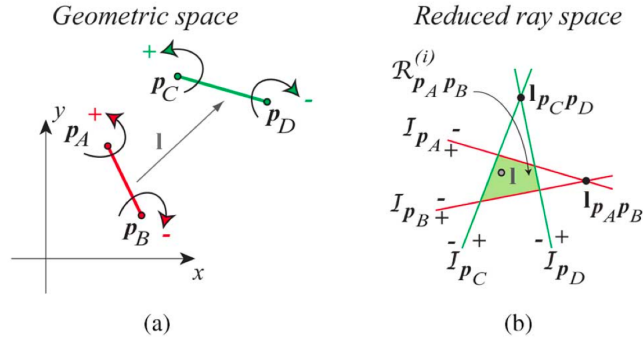


Fig. 1.4 The reflector  $p_A p_B$  and the OW  $p_C p_D$  in the geometric space and the ROI of the reflector in the reduced ray space. Note that the ROI is obtained from the intersection of the visibility regions of the OW and the image of the reflector.

$$R_{p_A p_B}^{(i)} = I_{p_A p_B} \cap V^{(i)}. \quad (1.5)$$

The ROIs are shown in figures Fig. 1.3 and Fig. 1.4.

### 1.1.4 Managing Multiple OW

For the scope of this thesis we are interested in managing an arbitrary number of observation windows. This is possible in the projective ray space. In [11] the global region of visibility is defined as the union of all ROVs relative to each OW. The region of visibility generally differs from the region of interest because it takes into account of possible acoustic shades made by the reflectors potentially placed between sources and OW. Figure Fig.1.5 illustrate the concept of ROV. With reference to Fig.1.5 we can define the ROV of the source  $p_s$  referred to the  $i$ th observation window as:

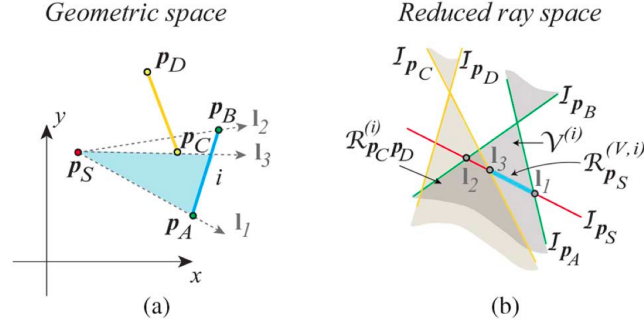


Fig. 1.5 The Region of Visibility of the source in the setup of Fig. 1.3 with an added reflector  $p_D p_C$

$$R_{p_s}^{(V,i)} = R_{p_s}^{(i)} \cap \bar{R}_{p_A p_D}^{(i)}. \quad (1.6)$$

Where  $\bar{R}_{p_A p_D}^{(i)}$  stands for the region of  $V^{(i)}$  not occupied by  $R_{p_C p_D}^{(i)}$ , i.e.

$$\bar{R}_{p_C p_D}^{(i)} = V^{(i)} \cap R_{p_C p_D}^{(i)} \quad (1.7)$$

Therefore we can define the global region of visibility  $R_{p_s}$  of a source  $p_s$  as:

$$R_{p_s}^V = \bigcup_{i=0}^N R_{p_s}^{(V,i)} \quad (1.8)$$

### 1.1.5 Soundfield Images in the Projective Ray Space

In this section we introduce the reader to the theoretical formulation of soundfield images. The goal is to map the plenacoustic function  $f(x, y, \theta)$  onto the global visibility  $V$ . In the following discussion we assume an ideal scenario, where the soundfield cameras are able to capture the acoustic radiance with no resolution losses or aliasing phenomena.

Let us consider a ray passing through a generic point  $p = [x, y]^T$  with direction  $\theta$ . The parameters of the ray are:

$$\begin{aligned} l_1 &= k \sin(\theta) \\ l_2 &= k \cos(\theta) \\ l_3 &= k[y \cos(\theta) - x \sin(\theta)], k > 0. \end{aligned} \quad (1.9)$$

We can parametrize the plenacoustic function with the above parameters obtaining the soundfield map  $p(l)$ . This function has now the projective ray space as domain. From 1.9 we can write that:

$$p(l) = \begin{cases} f(x, -\frac{l_1 x + l_3}{l_2}, -\arctan \frac{l_1}{l_2}) & l_2 \neq 0 \\ f(\frac{-l_1}{l_2}, y, \pi/2) & l_2 = 0 \end{cases} \quad (1.10)$$

Notice that the plenacoustic function is complex-valued and it carries the phase information at the considered frequency. Moreover scaling  $x$  (or  $y$  in the second case) has no effect. Thanks to the Radiance Invariant Law indeed, the value of the plenacoustic function does not change if it is picked at a different point on the same ray. Again under the hypotheses of RIL, the authors in [11] express the contribution of a source  $\mathbf{p}_s$  to the soundfield image  $\mathbf{p}_s$ :

$$p_{\mathbf{p}_s} = \begin{cases} b_{\mathbf{p}_s}(l) & R_{\mathbf{p}_s}^{(V)} \\ 0 & elsewhere \end{cases} \quad (1.11)$$

where  $b_{\mathbf{p}_s}(l)$  is the radiance pattern of a source  $\mathbf{p}_s$  which characterizes how the source radiates the sound in space along the ray  $\mathbf{l}$ .

## 1.2 Spatial Filtering

Spatial filtering is a signal processing technique that exploits spatial distribution of sensors to detect or transmit directional signals. It is an active research area which involves different fields such as radar sonar systems and acoustics.

The sensors structure is fundamental and sensors arrays are the tool to achieve spatial filtering. Indeed, the redundant information brought by multiple sensors can be exploited by properly combining the signals coming from the distributed sensors. This technique is also known as beamforming and it has several applications ranging from source extraction to source position estimation. Concerning this thesis, we are interested in the Direction of Arrival (DOA) estimation, since the DOA is the key information on which the ray space representation is devised.

In the literature, several different beamforming algorithms have been presented. In the following sections, we introduce the readers to the most used beamforming techniques. In particular, we present the theoretical analysis of:

- Linear Constrained Minimum Variance Beamformer (LCMV)
- Delay-and-sum (DAS)

- Null Steering
- Minimum Variance Distortionless Response Beamformer (MVDR)

### 1.2.1 Signal Model

Before the introduction of the spatial filtering techniques we have to clearly define the model adopted in this thesis for represent the signals. In this thesis we assume to work with a short time approach. The signal is spliced and processed frame-by-frame with the assumption that the length of the temporal frame is short enough to have a stationary signal within the frame. We start by defining the model for a single source scenario, but we will soon extend it to keep into account multiple sources. Let us consider  $I$  omnidirectional microphones. We can define the signal at the  $i$ th sensor at time  $t$  as:

$$x_n(t) = f_{i,n}(t) * s_n(t) + e_i(t), \quad (1.12)$$

where:  $s_n(t)$  is the signal emitted by the  $n$ th source at time  $t$ ,  $f_{i,n}(t)$  is the transfer function between the  $n$ th source and the  $i$ th microphone, '\*' is the convolution operator. The  $e_i(t)$  is an additive noise.  $e_i(t)$  represent the *thermal noise* of sensors electrical circuits and the *random background noise* sensed by the sensors.

Applying the Fourier Transform to 1.12 we obtain:

$$X_i(\omega) = F_{i,n}(\omega)S_n(\omega) + E_i(\omega). \quad (1.13)$$

Where  $X_i(\omega)$ ,  $F_{i,n}(\omega)$ ,  $S_n(\omega)$  and  $E_i(\omega)$  are the Fourier transform of the microphone signal, the transfer function, the source signal and the noise respectively. We can express 1.13 in a vectorial form as:

$$\mathbf{x}(\omega) = \mathbf{f}_n(\omega)S_n(\omega) + \mathbf{e}(\omega), \quad (1.14)$$

where  $\mathbf{x}(\omega) = [X_1(\omega), \dots, X_I(\omega)]^T$  is the output vector,  $\mathbf{f}_n(\omega) = [F_{1,n}(\omega), \dots, F_{I,n}(\omega)]^T$  is the so-called array transfer vector and  $\mathbf{e}(\omega) = [E_1(\omega), \dots, E_I(\omega)]^T$  is the noise vector.

Finally, we can extend 1.14 to a multi-source scenario, considering  $N$  sources. The overall model becomes:

$$\mathbf{x}(\omega) = \mathbf{F}(\omega)\mathbf{s}(\omega) + \mathbf{d}(\omega), \quad (1.15)$$

where  $\mathbf{F}(\omega) = [\mathbf{f}_1(\omega), \dots, \mathbf{f}_N(\omega)]$  and  $\mathbf{s}(\omega) = [S_1(\omega), \dots, S_N(\omega)]^T$ .

### Transfer function model

To complete our signal model we have to define the transfer function  $F_{i,n}(\omega)$ , from the  $n$ th source to the  $i$ th sensor. Since its expression strongly depends on the distance between the sources and the arrays, we can identify two main expressions. From now we are assuming to work with linear arrays. The definitions can be easily extended to circular arrays.

**Far Field** If an acoustic source is sufficiently far from the array, the wavefronts impinging on it can be considered as plane waves. This assumption let us to simplify the mathematical analysis. A source can be considered far from a circular array if the following relation on the distance between the source and the array  $\rho$  holds [26] :

$$\rho > \frac{2d^2}{\lambda} \quad (1.16)$$

Where  $d^2$  stands for the array diameter and  $\lambda$  is the wavelength. Under the far field assumption, the transfer function assumes the form of a plane wave:

$$F_{i,n}(\omega) = e^{-j\langle \mathbf{k}_n, \mathbf{r}_i \rangle} \quad (1.17)$$

Limiting our analysis to a 2D scenario, we can define the wavenumber vector of the  $n$ th source as  $\mathbf{k}_n = \frac{\omega}{c}[\cos(\theta_n), \sin(\theta_n)]^T$  and the position vector of the  $i$ th sensor as  $\mathbf{r}_i = [x_i, y_i]^T$ . The angle  $\theta_n$  is the Direction of arrival of the  $n$ th source.

**Near Field** If the far field assumption (i.e.  $\rho > \frac{2d^2}{\lambda}$ ) does not hold we are in the so called near field. In this case the assumption on the plane wavefronts is no longer valid. Hence we have to express the transfer function from source to sensor by means of a *Green's Function*.

$$F_{i,n} = \frac{e^{-j\frac{\omega}{c}\|\mathbf{r}_i - \mathbf{r}_n^\lambda\|}}{\|\mathbf{r}_i - \mathbf{r}_n^\lambda\|} \quad (1.18)$$

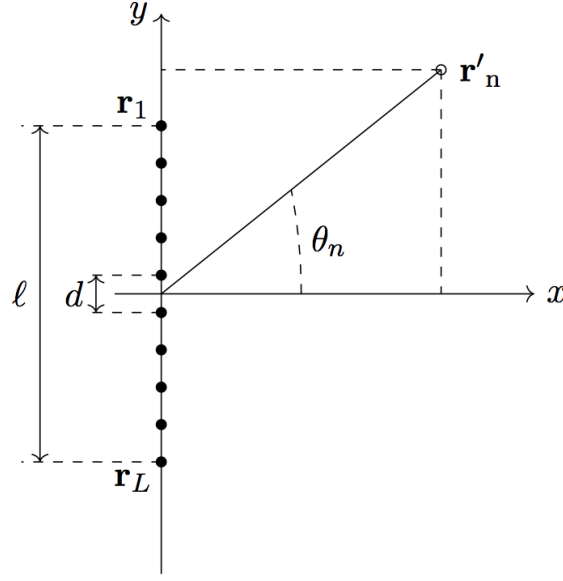


Fig. 1.6 Uniform Linear Array

where:  $\mathbf{r}'_n = [x_n, y_n]^T$  is the position of the  $k$ th source and  $\mathbf{r}_i = [x_i, y_i]^T$  is the position of the  $i$ th sensor. Again we limit the definition to a 2D case.

### 1.2.2 Linear Constrained Minimum Variance Beamformer

In this section we introduce the reader to the Linear Constrained Minimum Variance Beamformer (LCMV) [26], one of the most used beamformer in the literature. Other beamformers, presented in the next sections, can be seen as special cases of the LCMV.

Let us assume the signal model defined in 1.12, where the transfer function  $f_{i,n}$  is given either by 1.17 or 1.18. We consider the setting of a Uniform Linear Array deployed along  $y$ -axis and centered in the origin and a source placed in  $\mathbf{d}^s = \rho_s [\cos(\theta_s), \sin(\theta_s)]^T$  emitting the signal  $s(t)$  as shown in Fig. 1.6. In order to extract the signal we can apply a linear filter (FIR) to the array outputs:

$$y(\omega) = \mathbf{h}^H(\omega) \mathbf{x}(\omega), \quad (1.19)$$

where  $\mathbf{h}^H(\omega) = [h_1(\omega), \dots, h_I(\omega)]^H$  is a vector containing the filter coefficients. We can derive a solution for the filter coefficients as an optimization problem:

$$\begin{aligned} \mathbf{h}_0(\omega) = & \arg \min_{\mathbf{h}(\omega)} \quad \mathbf{h}^H(\omega) \mathbf{A}(\omega) \mathbf{h}(\omega) \\ & \text{subject to } \mathbf{h}^H(\omega) \mathbf{B}(\omega) = \mathbf{c}(\omega), \end{aligned} \quad (1.20)$$

where  $\mathbf{A} \in \mathbb{C}^{I \times I}$ ,  $\mathbf{B} \in \mathbb{C}^{I \times q}$  and  $\mathbf{c} \in \mathbb{C}^{1 \times q}$  are complex-valued matrices and  $q$  is the number of constraints. Consequently the general solution according to [23] is:

$$\mathbf{h}_0 = \mathbf{A}^{-1} \mathbf{B} (\mathbf{B}^H \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{c}^H, \quad (1.21)$$

where we have omitted the dependency on the frequency. The different beamformers described in the next sections are devised from 1.21 making different assumptions on  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{c}$ .

For the LCMV beamformer we state the following conditions:  $\mathbf{A}$  denotes the spatial covariance matrix,  $\mathbf{B}$  contains all the constraints being applied to the response  $\mathbf{c}$ . The LCMV design method is at the base of other beamformers that we present in the following sections. The optimization problem design remains the same for all these techniques, but different constraints are assumed. This results in different values assumed for matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{c}$ . Hence each beamformer is characterized by the form of its matrices, despite they all share the same solution formula. We briefly describe the assumptions of the different beamformers before describe them in details. In the design of the DAS beamformer Sec. 1.2.3 we assume that the auto-covariance matrix is equal to the identity matrix, namely that the signal  $\mathbf{x}(\omega)$  is assumed to be spatially white therefore is not correlated between one microphone and another. This implies that the matrices are:  $\mathbf{A}(\omega) = \mathbf{I}$ ,  $\mathbf{B}(\omega) = \mathbf{f}_i(\omega)$  and  $\mathbf{c} = 1$ .

The Null Steering beamformer is similar to DAS, but it adds additional constraints as specified in Sec. 1.2.4. The matrices configuration depends on the order of the constraints, but the main difference is in the form of vector  $\mathbf{c}$ . Indeed  $\mathbf{c} = [1, \varepsilon]$  contains the vector  $\varepsilon$  which represent the desired values to attenuate interferers.

Finally, the MDVR differs from the DAS in the definition of the matrix  $\mathbf{A}(\omega)$  since the assumption of having a spatially white signal is no longer valid. The auto-covariance matrix is computed from data as explained in Sec. 1.2.5.

### 1.2.3 Delay and Sum Beamformer

The Delay and Sum beamformer (DAS) is the simplest Beamformer in the literature. It follows a data-independent approach and its principle is to apply a delay to each microphones

signal to construct by summation the desired signal. We can define the the variance of 1.19 as:

$$\mathbf{E}\{|y(\omega)|^2\} = \mathbf{h}^H(\omega)\phi_{xx}(\omega)\mathbf{h}(\omega), \quad (1.22)$$

where  $\phi_{xx}(\omega) = \mathbf{E}\{\mathbf{x}(\omega)\mathbf{x}^H(\omega)\}$  is the array signal auto-covariance matrix. In the case of the DAS beamformer we assume a spatially white signal  $\mathbf{x}(\omega)$ , with no correlation among signals at different microphones. This is equal to assume the identity matrix as the auto-covariance matrix (i.e.  $\phi_{xx}(\omega) = \mathbf{I}$ ). Notice that 1.22 reminds 1.21 with  $\mathbf{A}(\omega) = \phi_{xx}(\omega)$ . We define a single constraint so that the source signal passes undistorted:

$$\mathbf{h}^H(\omega)\mathbf{f}_n(\omega) = 1 \quad (1.23)$$

Finally we can derive the DAS optimization solution substituting in 1.21  $\mathbf{A}(\omega) = \mathbf{I}$ ,  $\mathbf{B} = \mathbf{f}_n(\omega)$  and  $\mathbf{c} = 1$ :

$$\mathbf{h}_0(\omega) = \frac{\mathbf{f}_n(\omega)}{\mathbf{f}_n^H(\omega)\mathbf{f}_n(\omega)} = \frac{\mathbf{f}_n(\omega)}{\|\mathbf{f}_n(\omega)\|^2} \quad (1.24)$$

We can compute the filter power spectrum simply by inserting 1.24 in 1.22:

$$\mathbf{E}\{|y(\omega)|^2\} = \frac{\mathbf{f}_n^H(\omega)\phi_{xx}(\omega)\mathbf{f}_n(\omega)}{\|\mathbf{f}_n(\omega)\|^4} \quad (1.25)$$

### 1.2.4 Null Steering Beamformer

The Null Steering is very similar to the DAS, but it adds more constraints to attenuate the interfering signals in order to better extract the desired signal. This is achieved attenuating the signals coming from given positions. As already pointed out, also this beamformer can be expressed in terms of an optimization problem:

$$\begin{aligned} \mathbf{h}_0(\omega) = \arg \min_{\mathbf{h}(\omega)} \quad & \mathbf{h}^H(\omega)\mathbf{h}(\omega) \\ \text{subject to} \quad & \mathbf{h}^H(\omega)\mathbf{F}(\omega) = \mathbf{c}, \end{aligned} \quad (1.26)$$



where  $\mathbf{F} = [\mathbf{f}_1(\omega), \dots, \mathbf{f}_N(\omega)]$  and  $\mathbf{c} = [1, \boldsymbol{\varepsilon}]$ . With this definition of  $\mathbf{c}$  we are assuming with no loss of generality that the first source is the desired one. All the other sources are treated as interferers. We are forcing this behavior with the vector  $\boldsymbol{\varepsilon} \in \mathbb{C}^{1 \times (N-1)}$ , which contains the established values for the interferers. Doing so, we are attenuating the interfering signals and letting pass the desired signal undisturbed. This type of constraints is called *zero-order constraints* or *zero-order null*. More advanced designs are present in the literature [26]. They add a second type of constraints forcing the derivative of the beam pattern to zero. This design is strongly dependent on the transfer function model adopted, since the beam pattern depends on it.

### 1.2.5 Minimum Variance Distortionless Response Beamformer

Minimum Variance Distortionless Response Beamformer (MVDR), also known as *Capon*, can be seen as an extension of DAS, since it aims at minimizing the variance as defined in 1.22. Differently from DAS, in MVDR we do not make any assumption on the auto-covariance matrix. However, in practical situations is not possible to compute the real auto-covariance matrix  $\phi_{xx}(\omega) = \mathbf{E}\{\mathbf{x}(\omega)\mathbf{x}^H(\omega)\}$ , hence it is substituted by an estimate on  $T$  time frames:

$$\hat{\phi}_{xx}(\omega) = \sum_{t=1}^T \mathbf{x}_t(\omega)\mathbf{x}_t^H(\omega) \quad (1.27)$$

The constraint in the optimization problem is the same as 1.23, so the overall optimization problems take the form:

$$\begin{aligned} \mathbf{h}_0(\omega) = \arg \min_{\mathbf{h}(\omega)} \quad & \mathbf{h}^H(\omega) \hat{\phi}_{xx}(\omega) \mathbf{h}(\omega) \\ \text{subject to} \quad & \mathbf{h}^H(\omega) \mathbf{f}_n(\omega) = 1 \end{aligned} \quad (1.28)$$

The solution can be devised from 1.24 by properly substitute the identity matrix  $\mathbf{I}$  with the estimated auto-covariance matrix  $\hat{\phi}_{xx}$ :

$$\mathbf{h}_0 = \frac{\hat{\phi}_{xx}^{-1} \mathbf{f}_n(\omega)}{\mathbf{f}_n^H(\omega) \hat{\phi}_{xx}^{-1} \mathbf{f}_n(\omega)} \quad (1.29)$$

From 1.29 it clearly appears the data dependency of MVDR design opposed to the data independent DAS. The dependency is given by the presence of the estimated auto-covariance matrix.

## 1.3 Summary

In this chapter we have introduced the main concept involved in this thesis. First we have introduced the representation paradigm used to analyze the acoustic scene: The Ray Space. We began by describing the plenacoustic function that practically describe the acoustic radiance in the space. Then we have defined according to the literature the ray space as a new parametrization of the plenacoustic domain. We have presented the Euclidean ray space and the Projective ray space as a generalization of the first one able to overcome its limits. In fact, we adopt the projective ray space in this thesis, in order to develop a robust and flexible source localization by means of multiple microphone arrays. In Sec 1.2 we have discussed the main issues related to spatial filtering first and foremost the problem of defining a suitable signal model valid throughout the whole *Virtual Miking* procedure. Afterwards, we have presented an analysis of state of the art of beamforming methods. In particular, we have described the general design procedure of the optimization problem involved in the beamforming procedure. Then, we gave a detailed description of the beamformers used in our work spotting the differences between the different approaches.

### STATE OF THE ART

---

In this section we will present the *State of the Art* of acoustic signal representation. Three main approaches are adopted in the literature:

1. Parametric description.
2. Nonparametric description.
3. Geometric description.

In particular our focus will be on the Geometric Representation which emerged in the last years as an advantageous solution for space time soundfield analysis. In this context, the acoustic propagation is modeled on the concept of acoustic rays. After, we will describe in detail a *Virtual Miking* technique similar in the approach to the one presented in this thesis.

## 2.1 Acoustic Signal Representation

Acoustic signal processing is an active area of research that deals with the processing and extraction of information from acoustic signals. At the bottom of this discipline we find the acoustic signals representation. In the literature three main acoustic representation paradigms have emerged. What follows is a general introduction to the two main acoustic representations: **Parametric** and **Nonparametric**. Geometric representation is described in depth in section Sec. 2.1.1.

**Parametric representation** In this context the soundfield is described by means of *parameters* which characterize the scene [13]. To each sound event corresponds a sound signal with different and possibly many parameters which describe the soundfield. Examples of these parameters are: environment geometry, source locations, source radiation pattern etc.,

they depend on the a priori information available and on the desired accuracy. Microphone arrays have been traditionally used to retrieve soundfield characteristics. By means of spatial filtering Sec. 1.2 we can estimate the DOA of a sound source and localize a source given the measurements of the acoustic field. Different procedures have been developed in literature based on Time Difference of Arrival (TDoA) as [29], [8], [19], [1].

**Nonparametric representation** In this paradigm we do not assume any a priori knowledge about the acoustic scene, but it is the acoustic field itself that describes the scene. We consider the soundfield as a function of time and space, able to describe efficiently the scene. In the literature, efficient acoustic signal representations, inspired by physical models, have been developed. Indeed a point-to-point description for the acoustic field would require a big amount of data impossible to handle in practice. The most important Nonparametric representations are: **plane wave representation** introduced in [28], [27]; **spherical wave representation** [12] and **cylindrical wave representation** [15]

Nonparametric analysis aims to estimate the coefficients of a set of basis functions in which the acoustic field can be decomposed. Again the tool to sample the acoustic field is the array of microphones. The general idea is to represent the spatial dependency of a soundfield, in terms of specific set of basis functions (i.e. plane waves, spherical waves, cylindrical waves etc). In recent works such as [24] the signals from microphones are interpreted as time-space samples with their own specific temporal and spatial sampling frequencies, determined respectively by the temporal interval between adjacent time samples and by the spacing between microphones. In the literature many acoustic signal processing techniques have been proposed to decompose the acoustic field in terms of basis functions e.g. in [21] [18] [20].

Nonparametric representation is applied also in the synthesis of acoustic fields usually called *rendering*. Acoustic rendering is the process of synthesizing physically accurate acoustic fields over an extended area. This goal is achieved by means of spatially arrangements of loudspeakers. However, in this thesis we are not interested in acoustic rendering therefore acoustic rendering techniques are not presented.

### 2.1.1 Geometric Representation

Geometric representation is different from the other representations because it is devised from geometric acoustics instead of wave acoustics. In this context the propagation model is quite simple: *acoustic rays* are straight lines which carry the acoustic information. From a physical point of view this is a rough approximation of an acoustic field, in particular for near field scenarios, but the ray-based model lets us model arbitrary complex acoustic fields, for

which a physically accurate modeling is not practical. The concept of ray space, described in the next section, takes advantage of the geometric representation to define a projective domain whose primitives are acoustic rays. In the ray space domain, points correspond to plane-wave components of the acoustic field allowing a compact and sturdy soundfield representation as we have already shown in the previous chapter. Recently in [4] the author introduces the novel paradigm of beam acoustics in which the computational ease and the intuition of geometric representations are combined with the accuracy of physically-inspired representations.

Geometrical acoustics is devised from ray optics. Indeed a close analogy exists between the image formation process of an optical component and the acoustic image formation process with a microphone array. Ray acoustics is directly derived from optics and it assumes that acoustic radiance is constant along acoustic rays. Therefore this paradigm comes from the high frequency approximation of the wave equation.

### Ray Acoustics

Ray acoustics is concerned with the location and direction of acoustic rays. The trajectories can be univocally identified by the surfaces  $\psi(\mathbf{r}, \omega)$  to which they are perpendicular. The function  $\psi(\mathbf{r}, \omega)$  is usually referred as *Eikonal* [22] and depends on position  $\mathbf{r}$  and frequency  $\omega$ . Taking the direction of the gradient vector  $\nabla\psi(\mathbf{r}, \omega)$  we can identify the ray trajectories.

Let us introduce the *Homogenous Helmholtz Equation* as follows:

$$\nabla^2 P(\mathbf{r}, \omega) + \left(\frac{\omega}{c}\right)^2 P(\mathbf{r}, \omega) = 0, \quad (2.1)$$

where  $P(\mathbf{r}, \omega)$  expresses the acoustic field in a volume free of sources (i.e. Homogenous condition). Equation 2.1 is the Fourier transform of the *homogeneous wave equation*:  $\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = 0$ , which describes the field as small-amplitude variations of pressure  $p(\mathbf{r}, t)$ .

The function governing the Eikonals is derived from the high frequency approximation of the Helmholtz Equation. Let us consider the 2.1 rewritten in terms of amplitude and phase of  $P(\mathbf{r}, \omega)$ :

$$\nabla^2 \left( |P| e^{(j\psi)} \right) + \left(\frac{\omega}{c}\right)^2 |P| e^{(j\psi)} = 0. \quad (2.2)$$

Where dependency of  $P$  and  $\psi$  on space and time are omitted. Letting  $\omega \rightarrow \infty$  we obtain the *Eikonal equation* as defined in [4]:

$$\langle \nabla \psi, \nabla \psi \rangle = \left( \frac{\omega}{c} \right)^2 \quad (2.3)$$

The physical meaning of 2.3 is that acoustic rays are constrained to travel in the direction orthogonal to lines of constant phase. Hence in free-space, the acoustic rays (normals to the Eikonals  $\psi(\mathbf{r}, \omega)$ ) must be straight lines [22] and the Eikonal may be parallel planes or concentric sphere.

Consider a plane wave acoustic field propagating with wavenumber vector  $\mathbf{k}$ :

$$P(\mathbf{r}, \omega) = A(\omega) e^{j\langle \mathbf{k}, \mathbf{r} \rangle}. \quad (2.4)$$

Where  $A(\omega)$  is an amplitude factor. Also in this case we can factorize the amplitude and the phase of the acoustic field as:

$$P(\mathbf{r}, \omega) = |A(\omega)| e^{j(\langle \mathbf{k}, \mathbf{r} \rangle + \phi)}. \quad (2.5)$$

with  $\phi$  the phase contribution of  $A(\omega)$ . Therefore the Eikonal is:

$$\psi(\mathbf{r}, \omega) = \langle \mathbf{k}, \mathbf{r} \rangle + \psi \quad (2.6)$$

Substituting Sec. 2.6 in Sec. 2.3 we can find that a plane wave is a solution to the Eikonal equation as long as the  $\mathbf{k}$  satisfies the dispersion relation  $\mathbf{k} = \left( \frac{\omega}{c} \right)$ . Finally the trajectory of acoustic rays is determined by the gradient of  $\psi(\mathbf{r}, \omega)$ :

$$\nabla \psi(\mathbf{r}, \omega) = \mathbf{k}. \quad (2.7)$$

This main result establishes the strong relation between the wavenumber vector and the trajectory of acoustic rays as shown in Fig. 2.1a.

In the case of a spherical wave acoustic field we can apply the same reasoning. With no loss of generality the source is placed in the origin of the coordinate system, and the acoustic field is:

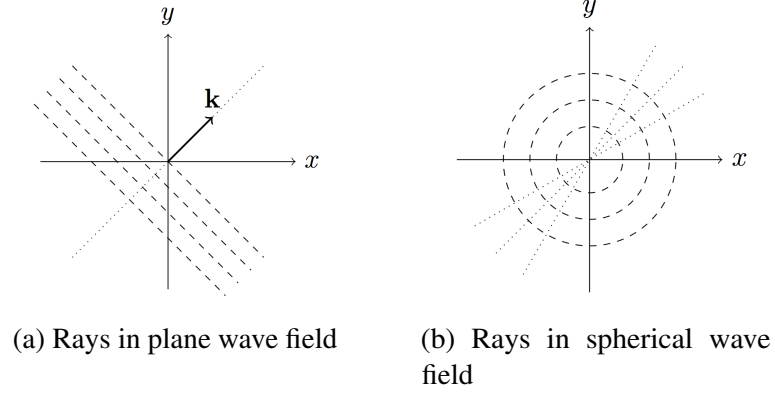


Fig. 2.1 Relation between the wavenumber vector and the acoustic rays in a plane wave field Fig. 2.1a and spherical wave field Fig. 2.1b. The dashed lines and circles represent the eikonals while the dotted lines stand for the acoustic rays trajectories.

$$P(\mathbf{r}, \omega) = \frac{e^{j\frac{\omega}{c}||\mathbf{r}||}}{4\pi||\mathbf{r}||} \quad (2.8)$$

The corresponding Eikonal is:

$$\psi(\mathbf{r}, \omega) = \frac{\omega}{c}||\mathbf{r}||, \quad (2.9)$$

and its gradient:

$$\nabla\psi(\mathbf{r}, \omega) = \frac{\omega}{c} \left( \frac{x}{||\mathbf{r}||} \hat{\mathbf{x}} + \frac{y}{||\mathbf{r}||} \hat{\mathbf{y}} + \frac{z}{||\mathbf{r}||} \hat{\mathbf{z}} \right). \quad (2.10)$$

From 2.10 it is possible to verify that spherical wavefronts are solutions to Eikonal equation. The acoustic rays belong to a set of lines crossing at the origin. Fig. 2.1b illustrates the mentioned relation.

## 2.2 Virtual Miking State of the Art

In this section we will go through the state of the art of *Virtual Miking* available in literature. In particular we focus on the work *Geometry-Based Spatial Sound Acquisition Using Distributed Microphone Arrays* [25]. In this paper, the authors present a virtual miking technique similar to the one proposed in this thesis. They follow a geometry-based analysis of the sound field

and a parametric description of the acoustic scene to compute the virtual microphone signal. In our work we have adopted a similar approach and the aim of this section is to highlight the technique adopted in [25] and to present its main characteristics.

### 2.2.1 Problem Formulation

In [25] the authors formulate a geometry-based spatial sound acquisition method capable of computing a signal corresponding to a virtual microphone with arbitrary pick-up pattern. The virtual microphone can be placed in an arbitrary position. The signal is computed in the time-frequency domain based on a parametric sound field model. The model accounts for a isotropic-point like sound source (IPLS), an homogeneous diffuse sound contribution and an arbitrary microphone pick-up pattern.

Hence, the signal model used is the following:

$$S(\mathbf{k}, n, \mathbf{d}) = S_{dir}(\mathbf{k}, n, \mathbf{d}) + S_{diff}(\mathbf{k}, n, \mathbf{d}). \quad (2.11)$$

Where  $n$  denotes the time instant,  $\mathbf{k}$  is the usual wavenumber vector and  $\mathbf{d} = [x, y, z]^T$  is an arbitrary position in cartesian coordinates. We can notice that the model assumes that the sound pressure at time  $n$  in position  $\mathbf{d}$  is given by the sum of two components: the direct one and the diffuse one.

The direct sound emitted by the IPLS located at  $\mathbf{d}_{IPLS}(\mathbf{k}, n)$  is assumed to be a monochromatic free field spherical wave. In case of more than one source we can generalized the model if the sources are sufficiently sparse. The propagation of the wave from  $\mathbf{d}_{IPLS}(\mathbf{k}, n)$  to  $\mathbf{d}$  is represented by the transfer function  $H_{dir}(\mathbf{k}, \mathbf{d}, \mathbf{d}_{IPLS})$ , i.e.:

$$S_{dir}(\mathbf{k}, n, \mathbf{d}) = H_{dir}(\mathbf{k}, \mathbf{d}, \mathbf{d}_{IPLS}) S_{dir}(\mathbf{k}, n, \mathbf{d}_{IPLS}). \quad (2.12)$$

$H_{dir}(\mathbf{k}, \mathbf{d}, \mathbf{d}_{IPLS})$  is the deterministic transfer function and can be computed for any  $\mathbf{d}$  and  $\mathbf{d}_{IPLS}$  by assuming a propagation model. Note that one can design arbitrary transfer functions to model the propagation between the source and virtual microphone, this means that any physical or non-physical behavior can be generated. In [25] authors approximate the spherical wave propagation from  $\mathbf{d}_{IPLS}$  to  $\mathbf{d}$  with:

$$H_{dir}(\mathbf{k}, \mathbf{d}, \mathbf{d}_{IPLS}) = \mathbf{r}^{-1}(\mathbf{k}) e^{j\mathbf{k}\mathbf{r}(\mathbf{k})}. \quad (2.13)$$



Where  $\mathbf{r}(\mathbf{k}) = \|\mathbf{d} - \mathbf{d}_{IPLS}(\mathbf{k})\|$  is the distance between the source and the virtual microphone.

The diffuse sound contribution at an arbitrary position  $\mathbf{d}$  is given by the diffuse sound contribution at  $\mathbf{d}_{IPLS}(\mathbf{k}, n)$  as:

$$S_{diff}(\mathbf{k}, n, \mathbf{d}) = H_{diff}(\mathbf{k}, n, \mathbf{d}) S_{diff}(\mathbf{k}, n, \mathbf{d}_{IPLS}). \quad (2.14)$$

Where  $H_{diff}(\mathbf{k}, n, \mathbf{d})$  is simply a time and frequency dependent complex value. Indeed the diffuse sound at position  $\mathbf{d}$  is typically uncorrelated if the distance between both points is sufficiently large. In the article the diffuse soundfield is assumed of being isotropic (i.e. the average sound power arriving at  $\mathbf{d}$  is uniformly distributed over all direction of incidence) and homogeneous (i.e. the average sound power does not vary with  $\mathbf{d}$ ). In [25] due to the homogeneous assumption an estimate of the power of the diffuse sound in  $\mathbf{d}$  is given by the diffuse sound in  $\mathbf{d}_{IPLS}$ . Authors have considered the phase of  $S_{diff}(\mathbf{k}, \mathbf{d}_{IPLS})$  as an estimate of  $S_{diff}(\mathbf{k}, \mathbf{d})$ , consequently the transfer function reduces to:  $H_{diff}(\mathbf{k}, \mathbf{d}, \mathbf{d}_{IPLS}) = 1$ . In practice the diffuse sound at the reference position is assumed as the sound at any  $\mathbf{d}$ .

The described model is used to compute the signal  $S(\mathbf{k}, n, \mathbf{d}_{VM})$  of a virtual microphone located at  $\mathbf{d}_{VM}$  with an arbitrary pick-up pattern given the sound pressures at the position of the IPLS. According to what previously introduced the signal is divided in its direct and diffuse components plus a propagation model. We can express the signal of a virtual microphone as follows:

$$S(\mathbf{k}, n, \mathbf{d}_{VM}) = \mathbf{w}^T(\mathbf{k}, n, \mathbf{d}_{VM}, \mathbf{d}_{IPLS}) \mathbf{s}(\mathbf{k}, n, \mathbf{d}_{IPLS}). \quad (2.15)$$

Where

$$\mathbf{s}(\cdot) = [S_{dir}(\mathbf{k}, n, \mathbf{d}_{IPLS}) \quad S_{diff}(\mathbf{k}, n, \mathbf{d}_{IPLS})]^T \quad (2.16)$$

and

$$\mathbf{w}(\cdot) = [W_{dir}(\mathbf{k}, n, \mathbf{d}_{VM}, \mathbf{d}_{IPLS}) \quad W_{diff}(\mathbf{k}, n, \mathbf{d}_{VM}, \mathbf{d}_{IPLS})]^T. \quad (2.17)$$

For example, for an omnidirectional VM the weights  $\mathbf{w}$  are:

$$w(\mathbf{k}, n, \mathbf{d}_{VM}, \mathbf{d}_{IPLS}) = [H_{dir}(\mathbf{k}, n, \mathbf{d}_{VM}, \mathbf{d}_{IPLS}) \quad H_{diff}(\mathbf{k}, n, \mathbf{d}_{VM}, \mathbf{d}_{IPLS})]^T \quad (2.18)$$

### 2.2.2 Parameters Estimation

In this section we briefly describe how the parameters to compute the VM signal are estimated in [25]. From the previous section it is clear that to compute the VM signal we need some parameters which have to be extracted from the acoustic scene. These parameters are obtained starting from the measurements of a reference omnidirectional microphone, arbitrary located in  $\mathbf{d}_{ref}$ .

**Sound Pressure at the IPLS** The direct sound  $S_{dir}(\mathbf{k}, \mathbf{d}_{IPLS})$  and the diffuse sound  $S_{diff}(\mathbf{k}, \mathbf{d}_{IPLS})$  required to compute the VM signal are estimated from the reference signal  $S(\mathbf{k}, \mathbf{d}_{ref})$ . In general we can express the direct sound at the reference point as:

$$S_{dir}(\mathbf{k}, \mathbf{d}_{IPLS}) = H_{dir}^{-1}(\mathbf{k}, \mathbf{d}_{ref}, \mathbf{d}_{IPLS}) S_{dir}(\mathbf{k}, \mathbf{d}_{ref}). \quad (2.19)$$

Where  $H_{dir}^{-1}(\mathbf{k}, \mathbf{d}_{ref}, \mathbf{d}_{IPLS})$  is the inverse of the transfer function between the source and the reference microphone. An estimate of the direct sound pressure  $S_{dir}(\mathbf{k}, \mathbf{d}_{ref})$  is obtained in [25] with the square-root Wiener filter that is given by:

$$G_{dir}(\mathbf{k}, \mathbf{d}_{ref}) = \sqrt{\frac{\Gamma(\mathbf{k}, \mathbf{d}_{ref})}{1 + \Gamma(\mathbf{k}, \mathbf{d}_{ref})}}, \quad (2.20)$$

where the signal-to-diffuse ratio (SDR)  $\Gamma(\mathbf{k}, \mathbf{d}_{ref})$  is the power ration between the direct and the diffuse sound at the reference position.

$$\Gamma(\mathbf{k}, \mathbf{d}_{ref}) = \frac{\mathbf{E}\{|S_{dir}(\mathbf{k}, \mathbf{d}_{ref})|^2\}}{\mathbf{E}\{|S_{diff}(\mathbf{k}, \mathbf{d}_{ref})|^2\}} \quad (2.21)$$

Therefore:

$$\hat{S}_{dir}(\mathbf{k}, \mathbf{d}_{ref}) = G_{dir}(\mathbf{k}, \mathbf{d}_{ref}) S(\mathbf{k}, \mathbf{d}_{ref}). \quad (2.22)$$

The same method is applied also to extract the estimation of the diffuse sound pressure:

$$\hat{S}_{diff}(\mathbf{k}, \mathbf{d}_{ref}) = G_{diff}(\mathbf{k}, \mathbf{d}_{ref}) S(\mathbf{k}, \mathbf{d}_{ref}). \quad (2.23)$$

Again the square-root Wiener filter is used:

$$G_{diff}(\mathbf{k}, \mathbf{d}_{diff}) = \sqrt{\frac{1}{1 + \Gamma(\mathbf{k}, \mathbf{d}_{ref})}} = \sqrt{1 - G_{dir}^2(\mathbf{k}, \mathbf{d}_{ref})} \quad (2.24)$$

The square-root Wiener filter is used because it has the desirable property that the estimated sound component has the correct power

$$|\hat{S}_{diff}(\mathbf{k}, \mathbf{d}_{ref})|^2 + |\hat{S}_{dir}(\mathbf{k}, \mathbf{d}_{ref})|^2 = |S_{dir}(\mathbf{k}, \mathbf{d}_{ref})|^2 \quad (2.25)$$

**Position estimation** An important parameter is the source position  $\mathbf{d}_{IPLS}(\mathbf{k})$ . The authors estimate it via triangulation based on the direction-of-arrival (DOA) of the direct sound observed at two or more different observation points. We can define the DOAs relative to  $\mathbf{M}$  microphones arrays (centered at  $\mathbf{d}_m$  with  $\mathbf{m} \in 1, 2, \dots, \mathbf{M}$ ) as the unit vectors  $\mathbf{e}(\mathbf{k}, \mathbf{d}_m)$ . Let us define  $\mathbf{c}(\mathbf{k})$  as the lines defined by the array centers and the DOA vectors:

$$\mathbf{c}(\mathbf{k}) = \mathbf{d}_m + \mathbf{e}(\mathbf{k}, \mathbf{d}_m) \xi(\mathbf{k}), \quad (2.26)$$

where  $\xi$  is an unknown real-valued scaling factor. In an ideal situation the intersection of any two lines returns the position  $\mathbf{d}_{IPLS}$ , but in a real scenario the triangulation might fail due to inaccurate DOA estimates. Hence the position  $\mathbf{d}_{IPLS}$  can be computed by minimizing the sum of the square distances of the lines in 2.26:

$$\mathbf{d}_{IPLS}(\mathbf{k}) = \underset{\mathbf{c}}{\operatorname{argmin}} \sum_{m=1}^M ||J(\mathbf{k}, \mathbf{d}_m, \mathbf{c})||^2. \quad (2.27)$$

The cost function is defined as:

$$J(\cdot) = [\mathbf{I} - \hat{\mathbf{e}}(\mathbf{k}, \mathbf{d}_m) \hat{\mathbf{e}}^T(\mathbf{k}, \mathbf{d}_m)] (\mathbf{c} - \mathbf{d}_m), \quad (2.28)$$

where  $\hat{\mathbf{e}}(\mathbf{k}, \mathbf{d}_m)$  represents the estimated DOAs. The accuracy of the position estimate is determined by factors such as the performance of the DOA estimator, the relative position of the array and the IPLS position itself.

**Signal-to-Diffuse Ratio Estimation** In [25] authors estimate the SDR  $\Gamma(\mathbf{k}, \mathbf{d}_{ref})$  at the reference point  $\mathbf{d}_{ref}$  assuming that positions  $\mathbf{d}_{1...M}$  of the  $\mathbf{M}$  microphones arrays are known. The estimation procedure is divided in two steps:

- Computing the SDRs  $\Gamma(\mathbf{k}, \mathbf{d}_m)$  for each array
- Determining the SDR at the reference point  $\mathbf{d}_{ref}$  from the estimated SDRs.

Given the active sound intensity vector as:

$$i_a(\mathbf{k}, \mathbf{d}_m) = \frac{1}{\sqrt{2}\rho_0 c} \Re\{S^*(\mathbf{k}, \mathbf{d}_m) \mathbf{v}(\mathbf{k}, \mathbf{d}_m)\}, \quad (2.29)$$

where  $\Re(\cdot)$  provides real part,  $(\cdot)^*$  represents the complex conjugate operation and  $\rho_0$  is the density of air. We can define the coefficient-of-variation (CV)  $c_v(\mathbf{k}, \mathbf{d}_m)$  of the active sound intensity vector is:

$$c_v(\mathbf{k}, \mathbf{d}_m) = \frac{\|\mathbf{E}\{i_a(\mathbf{k}, \mathbf{d}_m)\}\|}{\mathbf{E}\{\|i_a(\mathbf{k}, \mathbf{d}_m)\|\}}. \quad (2.30)$$

Authors use 2.30 to obtain an estimate of the *diffuseness* of the sound:  $\hat{\Psi}(\mathbf{k}, \mathbf{d}_m) = \sqrt{1 - c_v(\mathbf{k}, \mathbf{d}_m)}$ , which is directly related to the SDR. Finally an estimate of the SDR  $\Gamma(\mathbf{k}, \mathbf{d}_m)$  is given at the  $m$ th array as:

$$\hat{\Gamma}(\mathbf{k}, \mathbf{d}_m) = \frac{1}{\hat{\Psi}(\mathbf{k}, \mathbf{d}_m)} - 1 = \frac{1}{\sqrt{1 - c_v(\mathbf{k}, \mathbf{d}_m)}} - 1 \quad (2.31)$$

As far as the second step is concerned, the computation of the SDR at the reference position, we report in our description only the case when  $\mathbf{d}_{ref}$  is equal to one of the array positions  $\mathbf{d}_m$ . Doing so, we can directly consider the SDR estimated for the corresponding array:

$$\hat{\Gamma}(\mathbf{k}, \mathbf{d}_{ref}) = \hat{\Gamma}(\mathbf{k}, \mathbf{d}_m)|_{\mathbf{d}_m=\mathbf{d}_{ref}}. \quad (2.32)$$

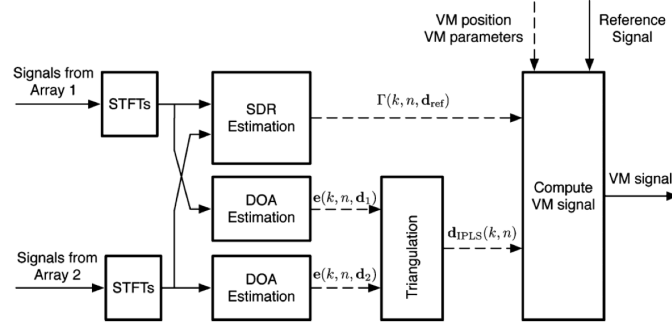


Fig. 2.2 The VM processing block in the case of two microphone arrays

### 2.2.3 Virtual Microphone Synthesis

Once all the parameters described in the previous section are available, a VM signal can be synthesized based on a the parametric representation of the sound field. we can use the knowledge of the IPLS parameters to generate a VM signal for any position  $\mathbf{d}_{VM}$  simulating a physical microphone placed in that specific position. We can express the VM signal as follows:

$$S(\mathbf{k}, \mathbf{d}_{VM}) = C(\mathbf{k}, \mathbf{l}_{VM}, \mathbf{d}_{VM}, \mathbf{d}_{IPLS})S_{dir}(\mathbf{k}, \mathbf{d}_{VM}) + Q(\mathbf{k})S_{diff}(\mathbf{k}, \mathbf{d}_{VM}), \quad (2.33)$$

where,  $C(\mathbf{k}, \mathbf{l}_{VM}, \mathbf{d}_{VM}, \mathbf{d}_{IPLS}) \in \mathbb{R}$  denotes the pick-up pattern of the VM with orientation defined by unit vector  $\mathbf{l}_{VM}$  and  $Q(\mathbf{k}) \in \mathbb{R}$  controls the sensitivity of the VM with respect to the diffuse sound and:

$$S_{dir}(\mathbf{k}, \mathbf{d}_{VM}) = H_{dir}(\mathbf{k}, \mathbf{d}_{VM}\mathbf{d}_{IPLS})S_{dir}(\mathbf{k}, \mathbf{d}_{IPLS}), \quad (2.34)$$

$$S_{diff}(\mathbf{k}, \mathbf{d}_{VM}) = H_{diff}(\mathbf{k}, \mathbf{d}_{VM}\mathbf{d}_{IPLS})S_{diff}(\mathbf{k}, \mathbf{d}_{IPLS}), \quad (2.35)$$

denote the direct and diffuse sound pressures of a virtual omnidirectional microphone. It is clear that the output signal  $S(\mathbf{k}, \mathbf{d}_{VM})$  of the VM microphone is the weighted sum of the direct and diffuse sound pressures at the position  $\mathbf{d}_{VM}$  of the VM. Therefore it is sufficient to assume a proper direct and diffuse transfer function, as explained in Sec. 2.2.1, and a pick-up pattern to compute the VM signal.

In Fig. 2.2 the entire VM procedure presented in the current section is displayed. The advantage of the parametric sound field representation is that one can assign an arbitrary pick up pattern to the VM through  $\mathbf{C}(\cdot)$ . Moreover it is possible to define a specific sensitivity of the VM with respect to the diffuse sound via the weight  $\mathbf{Q}(\mathbf{k})$ . We remark that also the transfer functions can be arbitrarily chosen. Hence every kind of behavior, physical or nonphysical, can be simulated. This flexibility makes the technique suitable for different applications.

## 2.3 Summary

In this section we have introduced the reader to the main acoustic representation paradigms and in particular to the geometric representation. We have described in details the assumption of geometric acoustics, showing how the concept of acoustic ray is fundamental in this context. Ray acoustics has been discussed in Sec. 2.1.1 as the reference paradigm of the procedure proposed in this thesis. Indeed even if it provides a rough approximation of the acoustic field it proves its importance in modeling complex acoustic environments. For this reason it has been widely adopted in literature and it can be considered the state of art of acoustic representations.

Geometric representation is at the base of the recent *Virtual Miking* technique [25] that we have discussed in Sec. 2.2. We have presented the method in details explaining its assumptions and characteristics. The work proposed in [25] presents a geometry-based spatial sound acquisition that involves the computation of one or more virtual microphone signals, which approximate the signals of physical microphones placed at the desired positions. The parametric approach assumed, let us define specific VM behavior both physical or non-physical, making this technique flexible and suitable for many different applications.

When we will present our *Virtual Miking* technique, in the following sections and we will use Sec. 2.2 as a reference highlighting similarities and differences with our approach.

## EXTRACTION OF SOUND PARAMETERS

---

In this section we will introduce our *Virtual Miking* (VM) technique. We will present how the theoretical concepts described in the previous sections have been exploited to derive a flexible and robust acoustic analysis framework. On the top of this setting we have developed the computations needed for the VM signals synthesis.

Our goal is to derive a flexible *Virtual Miking* procedure suitable for being applied in a wide range of applications. Hence, we decided to work with sensors arrays which can be arbitrary placed in the space. This choice represents an improvement with respect to already existing VM techniques in the literature, where, usually the position and the number of microphone arrays is fixed. Simultaneously we keep an high grade of flexibility which makes our procedure suitable for many different applications.

After the setting description, we will give an in depth analysis on the implemented procedure to retrieve the soundfield image. We will take advantage of the characteristics of the projective ray space, introduced in Sec. 1.1, to obtain a compact acoustic scene representation in a flexible context like the chosen one.

Later we will describe the methods that we have developed to analyze the soundfield and to retrieve the parameters useful to synthesize the virtual microphone signal. We will start from the acoustic scene representation derived in Sec. 3.2 to inspect the soundfield and to obtain the main parameters needed by our procedure. Recalling the parametric description of the soundfield proposed in Sec. 3.1 we will present the techniques adopted in our analysis procedure with their assumptions and characteristics. We will exploit the projective ray space representation of the soundfield to develop a robust and efficient source localization procedure. In order to reach this goal we will take advantage of well known computer vision algorithms. Then a parametric approach will be adopted to compute the radiance pattern of the source, more precisely we will obtain the pattern coefficients which determine the pattern by solving an optimization problem. Finally, we will present how the source signal

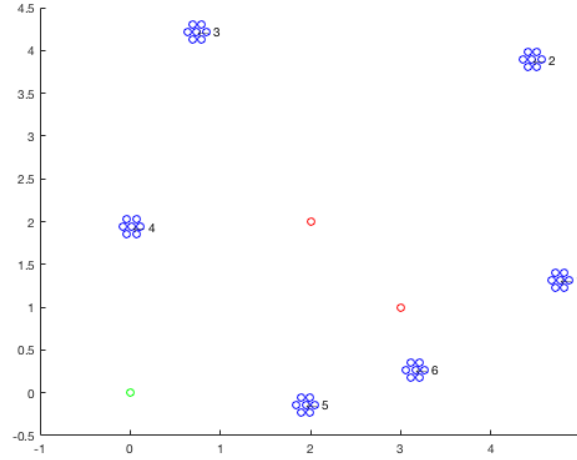


Fig. 3.1 The analysis set-up used in the *Virtual Miking* procedure. Sources (red dots) are randomly located in a circular area of ray  $2[m]$  and centered in  $\mathbf{o} = [2, 2]^T [m]$ . The microphone arrays are placed all around the sources area, while the virtual microphone (green dot) can be arbitrary placed in the space.

is retrieved from the sound pressure captured by the microphones. In this case a solution inspired by the beamforming procedure (Sec. 1.2) has been implemented.

### 3.1 Setting

In this section we present the setting of our problem. We state the assumptions we made and the general scenario of our VM technique. Following the example of [25] we exploit the geometry-based spatial sound description, as introduced in Sec. 2.1.1. This lets us to adopt the projective ray space as paradigm to represent the acoustic scene. Moreover like in [25] we use a parametric model of the acoustic scene. Our model assumes that the sound pressure at each time-frequency instant and position is given by the response of one or more sound sources. Differently from what is present in the literature we do not simply consider as source an isotropic point like sound source (IPLS), but we extend this basic model considering also the radiation pattern of the source (Sec. 3.1.1). In contrast to [25] we are not interested in taking into account for diffuse sound. In order to analyze the acoustic scene we adopt acoustic sensors organized in structure called arrays. In the literature, the use of microphone array for analyzing the acoustic field is well established. The sensors arrays are located all around the sources, with a random position. So every array observes the sources from a different perspective. The number of microphone arrays is not limited but there must be at least two devices. In Fig. 3.1 we show an example of the analysis setup just described.



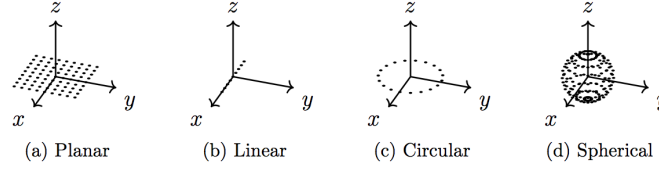


Fig. 3.2 Most common microphone array configurations

The microphone arrays can be designed almost arbitrarily, indeed, in literature many different configurations are present. In Fig. 3.2 the most common configurations are displayed. We decided to work with *Circular Arrays* of omnidirectional microphones in two dimensions. In this configuration the sensors are equally spaced on a circumference of fixed radius as shown in Fig. 3.3. We have chosen to work with circular array instead of the commonly used uniform linear array Fig. 1.6, because it has the advantage of being able to locate sources from all the directions, while the ULA suffers from the so-called *front-back ambiguity*, i.e. ULA cannot distinguish if the sound source is in front of the array or behind itself. Additionally, ULA can pinpoint only sources which are not aligned with it [26]. Furthermore the circular array has been recently deployed also in consumer products such as: *Amazon Echo*, *Google Home* and *Apple Homepod*.

### 3.1.1 Problem Formulation

Let us consider the signal model introduced in Sec. 1.2.1. Limiting our description to a single acoustic source we can extend the 1.14 to express the sound pressure impinging an acoustic sensor, taking into account the radiation pattern. The aim of the thesis is to compute the signal  $X(\omega, \mathbf{d}_{VM})$  of a virtual microphone (VM) located at  $\mathbf{d}_{VM}$  given the sound pressure at the position of the source and the radiance pattern of the source itself. More precisely,  $X(\omega, \mathbf{d}_{VM})$  is defined as follows:

$$X(\omega, \mathbf{d}_{VM}) = w(\omega, \mathbf{d}_{VM}, \mathbf{d}_S)S(\omega) + E(\omega) \quad (3.1)$$

where  $w(\cdot) = F(\omega, \mathbf{d}_{VM}, \mathbf{d}_S)P(\omega, \theta)$  are the weights given by the transfer function and the radiance pattern of the source. This two components characterize the sound pressure present at the  $\mathbf{d}_{VM}$  position.  $E(\omega)$  is an additive noise corresponding to the definition given in 1.14.  $F(\omega, \mathbf{d}, \mathbf{d}_S)$  represents the transfer function between the source at position  $\mathbf{d}_S$  and the microphone at position  $\mathbf{d}$ . Previously in 1.13 this was indicated by  $F_{n,k}(\cdot)$  as the transfer function between the  $n$ th sensor and the  $k$ th source. Note that the transfer function

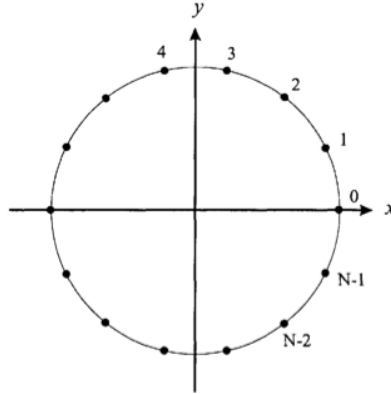


Fig. 3.3 Geometry of a circular array made of  $N$  microphones

$F(\omega, \mathbf{d}, \mathbf{d}_S)$  is deterministic and can be computed for any  $\mathbf{d}$  and  $\mathbf{d}_S$  by assuming a specific sound propagation model as the ones proposed in Sec. 1.2.1. We underline that no limits are applied to the propagation model so it can be arbitrary chosen in the VM synthesis phase.

For practical reason, we do not take into account the pick-up pattern of the virtual microphone, but it can be easily implemented in the future following the example of Sec. 2.2. Differently from [25] we focus our model on the radiation pattern of the source. This turns out to be an advanced description of real sound sources with respect to the plain IPLS model. The radiation pattern  $P(\omega, \theta)$  of a source is a real number which defines the variation of the power radiated by the source as a function of the direction  $\theta$  away from the source and the frequency  $\omega$  [7]. Fig. 3.4 shows an example of well known radiation patterns. Actually the radiation pattern can be also time dependent in the case of moving sources, but for the sake of simplicity we assume that the sources are fixed in space, so their radiation pattern does not change. However the implementation of a time dependent radiation pattern in our VM technique can be easily achieved repeating the estimation at each time frame.

## 3.2 Soundfield Image Construction

This section provides the procedure we have implemented to build the acoustic scene representation which is at the base of our VM technique. In Sec. 1.1 we have introduced the theory of soundfield imaging as a representation built from the concept of geometric acoustics (Sec. 2.1.1). What follows is a detailed description of how we build the ray space image of the acoustic scene with the analysis setup of Sec. 3.1.

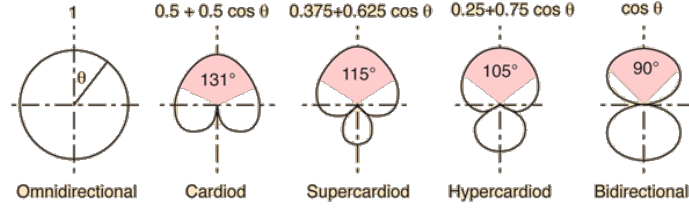


Fig. 3.4 An example of radiance patterns with their relative formula.

Let us recall the derivation of the projective ray space image as explained in Sec. 1.1. We need to compute the plenacoustic function in order to obtain the characterization of the acoustic scene in space. We report the definition of the plenacoustic function 1.2:

$$f(x, y, \theta, \omega) := e^{j\langle \mathbf{r}, \mathbf{k} \rangle} P(\mathbf{k}(\theta)).$$

The key information contained in 1.2 is represented by  $P(\mathbf{k}(\theta))$ . If the assumptions made in the previous sections are valid we can completely describe the acoustic rays if we know the direction  $\theta$ . Lets assume that only one acoustic source is present in the analysis space, we can retrieve the angular information for each microphone array by means of a beamforming procedure. Hence, we can identify an angle  $\theta$  for each array which represents the Direction of Arrival (DOA) of the ray reaching it from the source. In particular, in order to retrieve the information we exploit the Minimum Variance Distortionless Response Beamformer (MVDR) described in details in Sec. 1.2.5, which gives a better performance with respect to DAS (Sec. 1.2.3) and Null Steering (Sec. 1.2.4) beamformers.

The beamformer output is what we call *Pseudospectrum*. The pseudospectrum exhibits some peaks which represent the DOA of the acoustic rays, i.e. the angle  $\theta$  of the source position with respect to the array itself. An example of the pseudospectrum is visible in Fig. 3.5. Assuming the presence of  $K$  sources in the acoustic scene we can identify, for the  $m$ th array,  $K$  angles corresponding to the first  $K$  peaks of the pseudospectrum.

$$\mathbf{DOAs}_m = [\theta_1, \theta_2, \dots, \theta_K]^T. \quad (3.2)$$

Once the DOAs as been identified for each array we can easily build the projective ray space image which describe the acoustic scene. Recalling the definitions given in Sec. 1.1.2 we can obtain the projective ray space representation of the  $m$ th observation window with the following parameters:

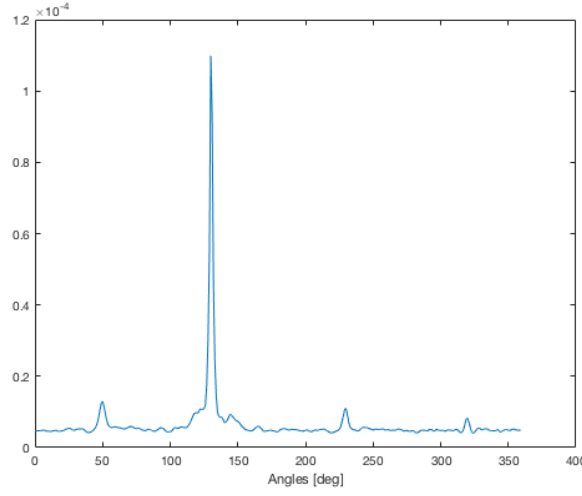


Fig. 3.5 The pseudospectrum of an array. The main peak represents the DOA of an acoustic source. The corresponding  $x$  value gives the angle from which the microphone array captures the source signal, namely the Direction of Arrival (DOA).

$$\begin{aligned} \mathbf{l}_1^{(m)} &= c \sin(\mathbf{DOAs}_k), \\ \mathbf{l}_2^{(m)} &= c \cos(\mathbf{DOAs}_k), \\ \mathbf{l}_3^{(m)} &= c[y \cos(\mathbf{DOAs}_m) - x \sin(\mathbf{DOAs}_m)], \end{aligned} \quad (3.3)$$

where  $c$  is a scaling factor, usually equal to 1,  $\mathbf{DOAs}_m \in \mathbb{R}^{K \times 1}$  is the vector of DOAs found for the  $m$ th array and  $x, y$  are the coordinates of the position of the array center. Differently from the linear OW proposed in Sec. 1.1.3, the circular array corresponds to a single plane in the projective ray space, since we identify it simply with the coordinates of its center in the geometric space. Obviously in 3.3 we compute only  $K$  points of the  $m$ th plane. The whole acoustic primitive in the projective ray space can be reconstructed with interpolation techniques.

### 3.3 Source Localization

In this section we will go through the procedure that we have implemented in order to localize the acoustic sources in the space. The problem of source localization is a well known signal processing issue in the literature. Many different solutions have been developed for achieving the ability of localize acoustic sources in the space. Recently, with the introduction of the projective ray space [11] new techniques that take advantage of computer vision algorithms has been introduced [5]. As already stated in Sec. 1.1.2 acoustic sources are mapped into

planar pattern in the projective ray space. Hence, find the position of a source results in estimating the parameters of the plane corresponding to it. Let us recall the soundfield image construction described in Sec. 3.2. After the mentioned procedure we have mapped the arrays signal in the projective ray space. As we stated in Sec. 3.1.1 the signal is frequency dependent and so by extension, also the soundfield image. For this reason, in order to deal with the localization problem, we use a wide band extension of the soundfield image. This is done adopting a geometric mean over all frequency bins of the pseudospectrum considered as in Sec. 3.2. The signal in the  $(l_1, l_2, l_3)$  domain contains both magnitude and phase information but only the first is useful for localizing the sources, so we can discard the phase taking the module of the signal. As described in Sec. 1.1.2, the equivalent of an array in the projective ray space is a plane, hence, after the wide-band analysis of the acoustic scene in the  $(l_1, l_2, l_3)$  we obtain the planes of the arrays which present peaks in magnitude. Note that for each plane the number of peaks goes up to  $K$  that is the number of sources present in the acoustic scene in analysis. That is because those points represent the intersection between the sources planes and the arrays planes. Ideally the identified peak points in the projective space cluster in linear patterns that can be detected using techniques derived from the computer vision literature. In fact there exists a wide range of robust and efficient algorithms to detect linear patterns which are flexible and able to discard outliers. Following the example of [5] we approach the localization with RANSAC (Random Sample Consensus) [14]. RANSAC is an iterative subspace clustering algorithm able to identify linear pattern with an high grade of robustness to noise and outliers. What follows is a detailed description of the application of RANSAC to the source localization problem. After the outliers have been found and discarded, we have to identify, through a linear regression the planes which represent the sources in the projective ray space. Consider a source  $S$  with position  $\mathbf{d}_S = [x_S, y_S, 1]^T$  expressed in projective coordinates. From Sec. 1.1.2 we know that equation  $\mathbf{l}^T \mathbf{d}_S = 0$  is valid for all lines passing through  $\mathbf{d}_S$ . The RANSAC algorithm finds the locations of the aligned features that we denote with  $\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_P$ . They are the rays reaching the arrays from the source, so ideally they pass through the point  $\mathbf{d}_S$  verifying the following conditions:

$$\left\{ \begin{array}{l} \mathbf{l}_1^T \mathbf{d}_S = 0 \\ \mathbf{l}_2^T \mathbf{d}_S = 0 \\ \dots \\ \mathbf{l}_P^T \mathbf{d}_S = 0 \end{array} \right. \quad (3.4)$$

Rearranging the system in matrix form we obtain:

$$\mathbf{L}\mathbf{d}_S = 0, \quad (3.5)$$

where  $L = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_p]^T$ . Therefore we are able to find the position vector  $\mathbf{d}_S$  simply by finding the null space of the matrix  $\mathbf{L}$ . Let us define:

$$\mathbf{L}_w = \text{diag}\left\{\frac{1}{e_1}, \frac{1}{e_2}, \dots, \frac{1}{e_P}\right\} \mathbf{L} = \mathbf{W}\mathbf{L}, \quad (3.6)$$

where  $\mathbf{W}$  is a weighting diagonal matrix, and  $e_p$ ,  $p = 1, \dots, P$  is the unavoidable error introduced by the peak-finding method in 3.4. The error is estimated for each peak as the width at -3dB of the lobe containing the local maximum  $l_p$ . In order to find the null space of the matrix, the next step is to compute the Singular Values Decomposition (SVD) of the matrix:

$$\mathbf{L}_w^T \mathbf{L}_w = \mathbf{U}_L \mathbf{D}_L \mathbf{V}_L^T, \quad (3.7)$$

where  $\mathbf{D}_L = \text{diag}(\sigma_{1,L}, \sigma_{2,L}, \sigma_{3,L})$  contains the singular values  $\sigma_{1,L} > \sigma_{2,L} > \sigma_{3,L}$  and  $\mathbf{U}_L$  and  $\mathbf{V}_L$  are the singular vector matrices of the decomposition. The position  $\hat{\mathbf{d}}_k$  is identified by the singular vector from  $\mathbf{V}_L$  related to the smallest singular value:

$$\hat{\mathbf{d}}_k = \mathbf{V}_{L(:,3)} \quad (3.8)$$

### 3.4 Radiance Pattern Estimation

In this section we present one of the most important block of our Virtual Miking technique. The introduction of the source radiance pattern represents a step forward in the VM literature [25]. It allows an advanced analysis of the source, envisioning the development of new pattern-based VM applications. First we describe the theoretical model assumed in order to compute the radiance pattern, then we present in details how we exploit this model in our acoustic scene description and how we are able to retrieve the pattern of a source from the signals of microphone arrays.

One of the most important property of an acoustic source is the *radiation pattern* or *directivity*, which gives the angular dependency of the sound energy radiated from the source

in the far field. The knowledge of directivity allows us to infer how the sound will propagate in the environment and to compute the VM signal as stated in Sec. 3.1.1. Let us recall 3.1 which is the model adopted to describe the acoustic scene. Highlighting the radiance pattern contribution we obtain:

$$X(\omega, \mathbf{d}_{VM}) = F(\omega, \mathbf{d}_{VM}, \mathbf{d}_S)P(\omega, \theta)S(\omega) + E(\omega), \quad (3.9)$$

where  $P(\omega, \theta)$  denotes the directivity function of the source. It depends on frequency ( $\omega$ ) and on the DOA ( $\theta$ ) of the source to the virtual microphone. Hence, we need to know the radiation pattern at all frequencies and the DOA of interest in order to reconstruct its contribution to the VM signal.  $P(\omega, \theta)$  can be thought in terms of weighting factors whose values goes from 0 (maximum attenuation) to 1 (no attenuation) based on the direction  $\theta$ . Once the source has been localized, the computation of the DOA is straightforward, but the directivity estimation needs more computations. First, we have to define a model for the radiation pattern. For the sake of simplicity, we describe the directivity function with a summation of weighted polar equation:

$$P(\omega, \theta) = \sum_{j=0}^{\frac{J}{2}-1} c_{1,j}(\omega) \cos(j\theta) + c_{2,j}(\omega) \sin(j\theta), \quad (3.10)$$

where  $J$  is the number of coefficients and  $c_{j,j}$  represent the  $j$ th coefficient related to the  $\sin(\cdot)$  or to the  $\cos(\cdot)$ . In this model the trigonometric functions are multiplied by coefficients which defines the shape of the overall pattern which values goes from 0 to 1. An example of these equations with their relative graph is visible in Fig.3.4. Consequently the directivity is completely characterized by the number and value of the pattern coefficients. For example, let us consider the polar equation of a cardioid pattern:

$$P(\omega, \theta) = \sum_{j=0}^1 0.5 \cos(j\theta) = 0.5 + 0.5 \cos(\theta). \quad (3.11)$$

In this case  $J = 2$  and  $c_{1,j} = 0.5$  while  $c_{2,j} = 0$  for all  $j$ s.

Thanks to the adopted model the radiance pattern estimation reduces to the estimation of the number and value of the pattern coefficients. Note that in the VM procedure we cannot know in advance the right number of coefficients. So, we have to assume a fixed number of

coefficients a priori, relying on the specific application and on the number of arrays. In case of extra coefficients, their contribution is negligible since they would tend to zero.

We can rewrite the 3.10 in matrix form to ease the radiation pattern estimation description:

$$\mathbf{P}(\omega) = \mathbf{B}(\theta)\mathbf{c}(\omega), \quad (3.12)$$

where  $\mathbf{c}(\omega) = [c_0, c_1, \dots, c_{J-1}]^T$  is the vector containing the  $J$  pattern coefficients and  $\mathbf{B}$  is the basis functions vector with the form:

$$\mathbf{B} = \begin{bmatrix} \cos(0\theta_1), & \dots, & \cos((J-1)\theta_1), & \sin(0\theta_1) & \dots & \sin((J-1)\theta_1) \\ \cos(0\theta_2), & \dots, & \cos((J-1)\theta_2), & \sin(0\theta_2) & \dots & \sin((J-1)\theta_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos(0\theta_R), & \dots, & \cos((J-1)\theta_R), & \sin(0\theta_R) & \dots & \sin((J-1)\theta_R) \end{bmatrix} \quad (3.13)$$

where  $\theta_r$  is the  $r$ th direction in which compute the pattern. Hence we can build the matrix  $\mathbf{B}$  with specific directions, this is useful in the analysis and synthesis phases.

The estimation procedure is directly derived from our model 3.10. Let us consider the signals captured by a sensor array made of  $N$  microphones in a one source scenario.  $X(\omega, \mathbf{d}_n)$  is the signal of the  $n$ th sensor placed in  $\mathbf{d}_n$ . In an ideal situation with the absence of errors, the directivity of the source in the direction  $\theta_n$  to the  $n$ th microphone can be expressed as:

$$P(\omega, \theta_n) = \frac{X(\omega, \mathbf{d}_n)}{F(\omega, \mathbf{d}_n, \hat{\mathbf{d}}_k)S(\omega)}, \quad (3.14)$$

where  $\hat{\mathbf{d}}_k$  is the position of the point-like source generating the sound pressure  $S(\omega)$  and  $F(\omega, \mathbf{d}_n, \hat{\mathbf{d}}_k)$  is the transfer function, in particular we adopt the near field model 1.18. The key information here, is the source position, since we assume that the microphones positions are known. Hence the performance of the source localization technique (Sec. 3.3) is fundamental. In practice, we do not know the source signal  $S(\omega)$  and a noise component is present, additionally, in general, more than one source can be emitting at the same time. As a consequence we cannot estimate the pattern as in 3.14. We can only obtain an estimation of the radiated signal, i.e. the source signal multiplied by the pattern weight, through the inversion of the transfer functions between sources and microphones. Here, we have to make an important assumption about our model. Since the microphones arrays are small compared to the distance from the source, the DOA is almost the same for each microphone in the array.



Therefore, also the pattern contribution remains almost unchanged for each microphone signal. This let us approximate the considered angles  $\theta_n$  with the one relative to the center of the array  $\theta_m$ . Doing this, we can estimate the radiated signal of the  $k$ th source from the  $N$  signals of an array as the following least square solution:

$$S_{P,k}(\omega, \theta_m) = \mathbf{F}(\omega)^\dagger \mathbf{X}(\omega), \quad (3.15)$$

where  $S_{P,k}(\omega, \theta_m) = S_k(\omega)P_k(\omega, \theta_m)$  is the the radiated source signal in the direction of the array,  $\mathbf{F}(\omega) = [F(\omega, \mathbf{d}_1, \hat{\mathbf{d}}_k), \dots, F(\omega, \mathbf{d}_N, \hat{\mathbf{d}}_k)]^T$  is the vector of the transfer functions between the source and the microphones and  $\dagger$  denotes the *pseudo inverse* operator. In general we can easily extend 3.15 to take into account more than one source at a time. In this scenario  $S_{P,k}(\omega, \theta_m)$  becomes the vector  $\mathbf{S}_P \in \mathbb{C}^{K \times 1}$ , and  $\mathbf{F}(\omega)$  becomes the matrix  $\mathbf{F} \in \mathbb{C}^{N \times K}$ . It is worth to notice that we have to perform the above estimation method for each microphone array. The result is the directed signal (source signal emitted by the point source and weighted according to the pattern)  $\mathbf{S}_P(\omega, \theta)$  sampled in the directions  $\theta_1, \theta_2 \dots \theta_M$ . Recalling the directivity function model outlined in 3.10, the final goal of the procedure is to estimate the coefficients that properly describe the pattern. For a single source, the solution of the radiance pattern estimation can be found by means of an optimization problem. To retrieve the radiance pattern we need just the magnitude of the signals  $\mathbf{S}_P(\omega, \theta) = [S_P(\omega, \theta_1), \dots, S_P(\omega, \theta_M)]^T$ , therefore from now, we will consider only its module and for the sake of simplicity we will omit the frequency dependence. The solution can be found from the following constrained optimization problem:

$$\begin{aligned} \arg \min_{\mathbf{x}} \quad & \|\mathbf{B}\mathbf{x} - \mathbf{S}_P\|^2 \\ \text{subject to} \quad & \begin{cases} \mathbf{A}_{eq}\mathbf{x} = \mathbf{b}_{eq} \\ \mathbf{A}\mathbf{x} \leq \mathbf{b} \end{cases}, \end{aligned} \quad (3.16)$$

where the vector  $\mathbf{x} \in \mathbb{R}^{1 \times J}$  contains the unknowns and matrix  $\mathbf{B} \in \mathbb{R}^{M \times J}$  is the matrix of the basis functions, as defined in 3.13, computed in the direction between microphone arrays and source. The procedure aims in finding the unknown  $\mathbf{x}$  which minimize the squared norm of the difference between the actual directed signal and the unknowns multiplied by the basis function matrix  $\mathbf{B}$ . It is clear that the solution of 3.16 does not directly retrieve the pattern coefficients. In fact the vector  $\mathbf{x}$  contains also the amplitude information of the signal. This comes evidently from the definition of  $\mathbf{S}_P$  3.15. Therefore, directly applying the solution to 3.12 we will obtain a *scaled* version of the radiance pattern. Ideally, if we compute the pattern for all the directions  $\theta$  we can easily obtain the actual radiance

pattern by normalizing the scaled pattern between 0 and 1. We can do this dividing the module of the scaled pattern by its maximum value. Doing this, the pattern will present values between 0 and 1 with the boundaries in correspondence of maximum and minimum attenuation directions. Consequently, just the weighting factor of the directivity remains and the amplification coming from the solution of 3.16 disappears. Note that from 3.12 we can compute the directivity only for  $R$  directions, so practically, we have to adopt a dense sampling on the angles  $\theta$  in 3.13 in order to obtain a good estimation of the radiance pattern with the estimated coefficients .

The constraints are important in defining the properties of the unknown  $\mathbf{x}$ . The solution of problems like 3.16 with linear cost functions and linear constraints can be found with different iterative algorithms [16] [9].

The first constraint  $\mathbf{A}_{eq}\mathbf{x} = b_{eq}$ , set the condition that the coefficients sum up to  $b_{eq}$ : the vector  $\mathbf{A}_{eq}[1, \dots, 1]$  has dimensions  $1 \times J$  and  $b_{eq} = 1$ . The second constraint  $\mathbf{Ax} \leq \mathbf{b}$  is used to obtain positive pattern values. The matrix  $\mathbf{A}$  and vector  $\mathbf{b}$  are specially made for this purpose.  $\mathbf{A}$  is built from the basis functions matrix and extended as follows:

$$\mathbf{A} = -\mathbf{D} \quad (3.17)$$

where  $\mathbf{D}$  is the matrix of 3.13 computed for all the  $R$  angles uniformly sampled between  $0^\circ$  and  $360^\circ$ . The vector  $\mathbf{b} \in \mathbb{R}^{R \times 1}$  gives the limits of the range:

$$\mathbf{b} = [0^{(1)}, \dots, 0^{(R)}]^T, \quad (3.18)$$

where the notation  $\cdot^{(r)}$  indicates the  $r$ th position in the array.

### 3.5 Source Signal Retrieval

Once the source has been localized and its pattern has been reconstruct, the last parameter we need for complete our model 3.9 is the source signal  $S(\omega)$ . This the signal emitted by the ideal point-like source without the weight of directivity. Again in an ideal situation, given the radiation pattern and proper transfer function, the procedure for retrieving the source signal from the  $n$ th microphone signal is trivial:

$$S(\omega) = \frac{X(\omega, \mathbf{d}_n)}{P(\omega, \theta_n)F(\omega, \mathbf{d}_n, \mathbf{d}_{PLS})} \quad (3.19)$$

Let us consider a multi-source scenario with  $K$  sources sensed by the  $M$  microphone arrays each one made of  $N$  microphones. If we know the source positions (Sec. 3.3) we can easily compute with the proper model (Sec. 1.2.1) the transfer function from each source to each microphones  $F(\omega, \mathbf{d}_n, \mathbf{d}_k)$  where  $\mathbf{d}_n$  and  $\mathbf{d}_k$  denote respectively the  $n$ th sensor position and the  $k$ th source location. Hence, if we add the knowledge of the directivity of each source  $P_k(\omega, \theta)$ , we can compute the weights vector of 3.1. In practice we use those informations to retrieve the source signal by means of a spatial filtering method. In particular, we take advantage of the LCMV beamformer described in details in Sec. 1.2.2. We can imagine to apply a linear filter (FIR) to the microphones signals in order to extract the source signal:

$$\hat{S}(\omega) = \mathbf{h}^H(\omega)\mathbf{X}(\omega) \quad (3.20)$$

where we have omitted the dependency on position and  $\mathbf{h}^H(\omega) = [h_1(\omega), \dots, h_{(N \times M)}(\omega)]^H$  is the vector of the filter coefficients and  $\mathbf{X}(\omega) \in \mathbb{C}^{(N \times M) \times 1}$ . From Sec. 1.2.2 we know that we can derive a solution for the filter coefficients  $\mathbf{h}$  as an optimization problem 1.20 that we report for the reader's convenience:

$$\begin{aligned} \mathbf{h}_0(\omega) = \arg \min_{\mathbf{h}(\omega)} \quad & \mathbf{h}^H(\omega)\mathbf{A}(\omega)\mathbf{h}(\omega) \\ \text{subject to} \quad & \mathbf{h}^H(\omega)\mathbf{B}(\omega) = \mathbf{c}(\omega), \end{aligned} \quad (3.21)$$

We can set the constraint in order to force the filter in returning only one source signal at a time. The knowledge of the radiance pattern (Sec. 3.4) and the source position (Sec. 3.3), let us write the proper conditions to achieve the goal. From 1.2.2 we know that matrix  $\mathbf{A} \in \mathbb{C}^{(N \times M) \times (N \times M)}$  is the covariance matrix of the signals, matrix  $\mathbf{B} \in \mathbb{C}^{(N \times M) \times K}$  contains the problem constraints and the vector  $\mathbf{c} \in \mathbb{R}^{K \times 1}$  is the desired response. According to our model 3.1, to obtain the  $k$ th source signal we can simply define the matrix  $\mathbf{B}$  with the information we already know, i.e. the weighting vector  $\mathbf{w}$  as follows:

$$\mathbf{B}(\omega) = \begin{bmatrix} w_{1,1}, & \cdots, & w_{1,K} \\ w_{2,1}, & \cdots, & w_{2,K} \\ \vdots, & \vdots, & \vdots \\ w_{(N \times M),1}, & \cdots, & w_{(N \times M),K}, \end{bmatrix} \quad (3.22)$$

where  $w_{n,k}(\omega) = P_k(\omega)F_{n,k}(\omega)$  is the weighting factor given by the directivity of the  $k$ th source and the transfer function between the  $k$ th source and the  $n$ th microphone. As regards the desired response vector  $\mathbf{c}$  we apply the same assumptions of the Null steering beamformer (Sec. 1.2.4). The vector  $\mathbf{c} = [1, 0, \dots, 0]^T$  is used to let the desired signal pass undisturbed and attenuate the interferers. The number 1 is placed in the  $k$ th location of  $\mathbf{c}$ , specifying that the desired signal is the one of the  $k$ th source. Therefore to compute the source signal of the different sources we have to properly modify vector  $\mathbf{c}$  and to repeat the procedure at each step.

### 3.6 Summary

In this section we have introduced the core of our *Virtual Miking* technique. We have described how the theoretical concepts presented in the previous sections have been implemented to build a robust analysis framework.

The set-up (Sec. 3.1) as been presented with its characteristics and assumptions. In this context we have introduced the problem formulation and our parametric description of the acoustic scene. This description let us to compute the VM signal  $X(\omega, \mathbf{d}_{VM})$  which is the ultimate goal of our work. For that reason we have derived the procedure (Sec. 3.2) to compute the projective ray space image of the acoustic scene. The projective ray space image is the fundamental description that we adopt in order to retrieve the parameters needed by our model.

The analysis of the acoustic scene is fundamental for the final result. In fact, the quality of the synthesized VM signal, presented in the next section, is strongly dependent on the estimated parameters. Its quality can decrease quickly if the parameters are unreliable. Hence we have developed the robust and efficient techniques of this section. As regards the source localization, we have played on a wide-band soundfield image of the acoustic scene to get a complete description of the soundfield. The localization has been implemented in the projective ray space by means of the efficient and robust RANSAC algorithm borrowed from the computer vision literature. Thanks to the procedure of Sec. 3.3, we have been able to implement a reliable method for retrieving the source position. With this information we can compute the transfer function present in the model 3.9. As regards the radiation pattern of the source (Sec. 3.4), first we have proposed a model 3.10 for describe the directivity in terms of weighted basis functions. Then, we have described in depth the designed method for extracting the radiation pattern. The shape of a pattern is given by the so-called pattern coefficients, so we need them to characterize the directivity of the source. In the directivity retrieval we have abandoned the soundfield image in favor of a parametric approach. In

this context we have presented the *Linear Optimization Problem* 3.16 thanks to which we are able to estimate the directivity from the sensors signals. Finally we have exploit one of the spatial filtering techniques described in (Sec. 1.2) to extract the source signal from the measurements. We have adapt the LCMV beamformer (Sec. 1.2.2) with the proper constraints, in order to obtain the source signal of one source at a time.

In this chapter we have seen how the acoustic scene is analyzed exploiting the theoretical concept introduced in the previous chapters. Moreover we have highlighted the analogies and the differences between our VM technique and state of art in the literature. The methods developed in are the core of our work because their performance is crucial in the synthesis of the VM signal that we will present in the next section.



## VIRTUAL MIKING

---

In this section we will introduce the final steps of our *Virtual Miking* technique. After the analysis of the soundfield proposed in the previous chapter we have all the informations needed by our model (Sec. 3.1.1) in order to compute the signal of a virtual microphone. In this section we will look at our soundfield parametric description and we will use the results of the acoustic scene analysis to derive the signal of the virtual microphone. We will also present the main characteristics of our method, discussing its aspects in terms of flexibility. Finally we will recap the whole *Virtual Miking* procedure to give a synthetic and methodical view of the entire process.

### 4.1 Synthesis of the Virtual Microphone signal

In this chapter we deal with the last issue of the VM procedure: the *synthesis of the Virtual microphone signal*. We will describe how to use the information derived from the analysis of the acoustic scene (Sec. 3) and the parametric soundfield representation in order to compute the signal of the desired VM. Once the parameters of the acoustic scene are estimated as described in the previous Sec. 3, a VM signal can be synthesized based on our parametric description of the soundfield (Sec.3.1.1). The knowledge of:

1. Sources position,
2. Sources directivity,
3. Sources signal,

can be used to generate the VM signal for any position  $\mathbf{d}_{VM}$  approximating a physical microphone placed at that position.

The output signal  $X(\omega, \mathbf{d}_{VM})$  of the VM in 3.1 can be computed using:

$$w(\omega, \mathbf{d}_{VM}, \hat{\mathbf{d}}_k) = F(\omega, \mathbf{d}_{VM}, \hat{\mathbf{d}}_k) \hat{P}_k(\omega, \theta_{VM}) \quad (4.1)$$

where  $\hat{\mathbf{d}}_k$  is the estimated position of the  $k$ th source.  $w(\omega, \mathbf{d}_{VM}, \hat{\mathbf{d}}_k)$  is the weighting factor determined by the assumed transfer function and the estimated source radiation pattern of source  $k$ . The estimated parameters,  $\hat{\mathbf{d}}_k$  and  $\hat{P}_k(\omega, \theta_{VM})$  are computed with the methods described in details in Sec. 3.3 and Sec. 3.4 respectively.

The overall signal is obviously determined by the superposition of the contributions given by all the  $K$  sources. Hence, we can write the output signal of the VM  $X(\omega, \mathbf{d}_{VM})$  as:

$$X(\omega, \mathbf{d}_{VM}) = \sum_{k=1}^K w(\omega, \mathbf{d}_{VM}, \hat{\mathbf{d}}_k) \hat{S}_k(\omega) = \sum_{k=1}^K F(\omega, \mathbf{d}_{VM}, \hat{\mathbf{d}}_k) \hat{P}_k(\omega, \theta_{VM}) \hat{S}_k(\omega) \quad (4.2)$$

where  $\hat{S}_k(\omega)$  denotes the source signal of the  $k$ th source estimated through the LCMV beamformer based technique that we have introduced in Sec. 3.5.

Similarly to the VM technique in [25], an high grade of flexibility is achieved by the possibility of choosing any arbitrary transfer function. Indeed the function  $F(\omega, \mathbf{d}_{VM}, \mathbf{d}_k)$  in 4.2 let us define the transfer function between the source  $k$  and the VM. This function determines the propagation of the source signal from its origin to the  $\mathbf{d}_{VM}$ . It modifies the magnitude and the phase of the source signal  $\hat{S}_k(\omega)$  according to a specific propagating model. In Sec. 1.2.1 we have describe the two main propagating model based on the physical behavior of sound. It is important to note that while we have to strictly follow the appropriate physical law during the analysis phase (Sec. 3), in the synthesis step, one can design any arbitrary transfer functions. Indeed, in order to correctly estimate the directivity and the source signal we have to reconstruct the true physical propagation of the sound from the source to the microphone arrays. On contrary, in the synthesis phase, we can generate a virtual microphone with any behavior either physical or non-physical.

The VM synthesis phase follows the opposite order of steps that we have in the analysis of the soundfield. First we compute the directivity of the source in the direction of the VM. In order to evaluate the pattern we have to compute the angle  $\theta_{VM}$  between the source and the virtual sensor. Since we know the estimate position of the source  $\hat{\mathbf{d}}_k$  and the position of the VM  $\mathbf{d}_{VM}$  is give a priori, we can easily compute  $\theta_{VM}$  as follows:

$$\theta_{VM} = \arctan \left( \frac{y_{VM} - \hat{y}_k}{x_{VM} - \hat{x}_k} \right). \quad (4.3)$$



With the direction  $\theta_{VM}$  and the estimated pattern coefficients (Sec. 3.4) we can compute the directivity according to 3.12:

$$\hat{P}_k(\omega, \theta_{VM}) = \mathbf{c}(\omega) \mathbf{B}(\theta_{VM}), \quad (4.4)$$

where  $\mathbf{B}(\theta_{VM})$  contains only the basis functions relative to the direction  $\theta_{VM}$ .

Then we compute the weighted signal according to the directivity simply by multiplying the pattern  $\hat{P}_k(\omega, \theta_{VM})$  with the estimated source signal  $\hat{S}_k(\omega)$  (Sec.3.5).

$$S_{P,k}(\omega) = \hat{P}_k(\omega, \theta_{VM}) \hat{S}_k(\omega). \quad (4.5)$$

Finally, the last step is represented by the propagation of the signal from the source to the VM with the function  $F(\omega, \mathbf{d}_{VM}, \mathbf{d}_k)$  of choice. We remark that the propagating function can simulate any physical or non-physical behavior. In general the transfer function depends on the position of the source and on the *listener* position. The model is used to describe how the information is carried from the source to an arbitrary destination, which in our case, is the position of the VM.

Since we suppose to work in the frequency domain, the propagation of the signal can be achieved by multiplying the directed signal by the transfer function:

$$X(\omega, \mathbf{d}_{VM}) = F(\omega, \mathbf{d}_{VM}, \hat{\mathbf{d}}_k) S_{P,k}(\omega). \quad (4.6)$$

From the definitions previously given, we observe that the final synthesized VM signal 4.6 is in compliance with the parametric description of the soundfield:

$$X(\omega, \mathbf{d}_{VM}) = F(\omega, \mathbf{d}_{VM}, \hat{\mathbf{d}}_k) S_{P,k}(\omega) = F(\omega, \mathbf{d}_{VM}, \hat{\mathbf{d}}_k) \hat{P}_k(\omega, \theta_{VM}) \hat{S}_k(\omega) = w(\omega, \mathbf{d}_{VM}, \hat{\mathbf{d}}_k) \hat{S}_k(\omega). \quad (4.7)$$

Clearly, the accuracy of the estimated parameters deeply influence the final VM synthesized signal. In particular, we remark the dependency on frequency and time of the soundfield description. Hence, a proper frequency and time sampling has to be considered in order to correctly retrieve the soundfield parameters. In addition, the choice of the transfer function represents a crucial point in the overall procedure: in the analysis phase, in order to obtain

good estimations of the parameters, it has to be as close as possible to the real one, while in the synthesis phase, its definition is arbitrary and strongly influences the result.

## 4.2 Summary of the Virtual Miking technique

In this section we give a compact recap of the *Virtual Miking* technique described in this thesis. We briefly express all the main steps of the procedure with their relative contribution to the final result. After the systematic description we discuss the characteristics of our VM process, its novelty compared to the literature and the differences with already existing VM methods.

---

### Algorithm 1: VM computation

---

#### Analysis

1. Compute the Short Time Fourier Transform of the array signals
2. Source localization in the projective ray space
  - (a) Construct the soundfield image in the projective ray space
    - Estimate DOAs through a beamforming operation
    - Compute the  $(l_1, l_2, l_3)$  features
  - (b) Determine the source description in  $(l_1, l_2, l_3)$  with RANSAC
3. Radiance pattern coefficients estimation
  - (a) Estimation of the directed signal (source signal weighted by the directivity)
    - Estimate the transfer function between source and microphones with a physical model
    - Retrieve the directed signal inverting the propagating functions
  - (b) Estimate the pattern coefficients solving an optimization problem
4. Source signal retrieval
  - (a) Compute the pattern and propagation weights
  - (b) Retrieve the source signal by means of the LCMV beamformer

#### Synthesis

1. Synthesis of the VM signal based on the estimated parameters
  - (a) Estimate angle between source and VM
  - (b) Compute the directivity according to the estimated coefficients and direction
  - (c) Compute the propagation contribution
  - (d) Synthesize the signal according to the model and the estimated parameters

The proposed VM technique derives from a geometry-based spatial sound acquisition integrated by a parametric model of the acoustic scene. The soundfield image is based on the geometric representation of the soundfield (Sec. 2.1.1). It is acquired by means of microphones arrays that can be placed almost arbitrary in the space. Differently from [25], we adopt the projective ray space image of the soundfield [11], which has proved its benefits in the field of acoustic analysis [5]. In this context, we have implemented a robust source localization procedure adapting the RANSAC [14] algorithm, a well known computer vision method for the identification of linear patterns. The main difference of our VM procedure lies in the parametric model of the sound scene (Sec. 3.1.1). We describe the soundfield by means of a simple model which includes also the radiation pattern, or directivity, of the sources like in [7]. Therefore, we need to estimate this key information and in order to do that we propose a model of the source directivity and a procedure to estimate it through an optimization problem. Our model represents better the real sound source with respect to what is available in literature. Moreover, it let us describe the acoustic scene with only three main parameters: the source signal, the radiation pattern and the transfer function. The knowledge of the directivity is particularly interesting because we can imagine different applications of the VM technique driven by this information. As regards the source signal, it is estimated through a beamformer as described in details in Sec. 3.5. Finally, the synthesis phase is straightforward, since it is very simple to compute the signal of the virtual microphone with the given model and the parameters as described in Sec. 4.1.

In conclusion, we sum up the whole *Virtual Miking* procedure with the following block diagram.

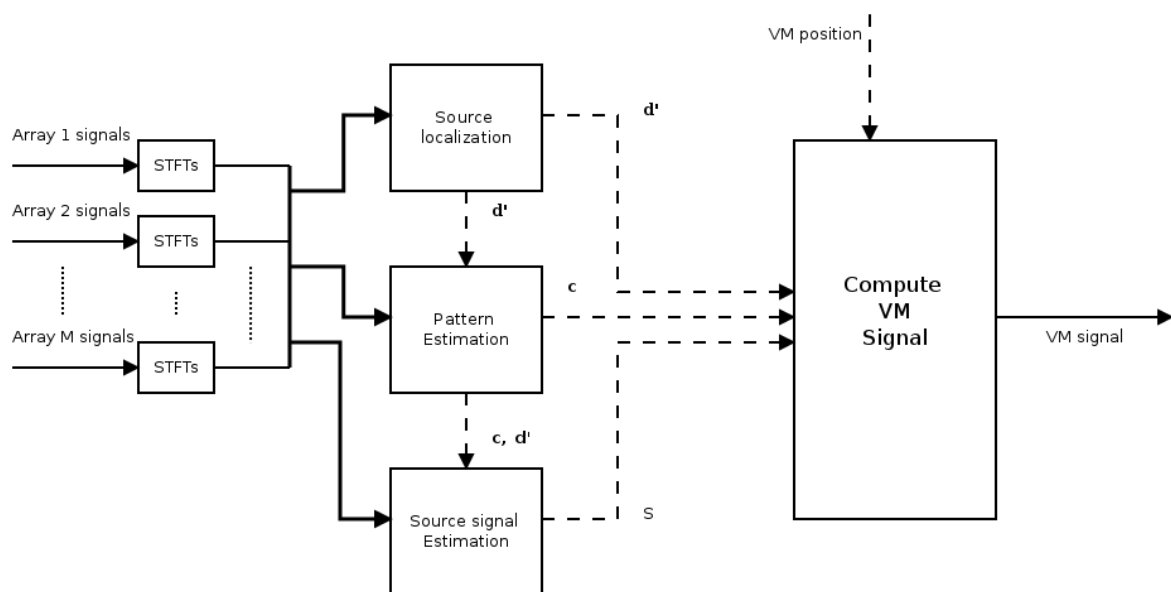


Fig. 4.1 The VM procedure block diagram. The signals coming from the microphones arrays are analyzed by the three main modules in order to retrieve the parameters: the sources positions  $\mathbf{d}'$ , the pattern coefficients  $\mathbf{c}$  and the source signals  $S$ . Then the estimated parameters and the position of the VM are used to compute the signal of the virtual microphone.

## RESULTS

---

In this chapter we will analyze the features of the *Virtual Miking* technique proposed in this thesis and we will compare its performances in different situations. We will test our procedure in two main context. The first one we will be a software simulation, through which we will analyze the performance in presence of different kind of source signals such as the white noise and speech. This analysis is important because it is close to a real application scenario and flexible in order to evaluate the behavior in different configurations. The second test context is in a real applicative scenario. In order to give a complete judgment of our work, we have set up a real scenario using microphones and loudspeakers placed as described in the theory. Also in this case, we will use both white noise and speech as source signals plus a recorded guitar.

In order to evaluate the characteristics of the technique we need appropriate metrics. To evaluate the synthesized VM signal we will compared it with a reference signal which will give the ground truth. The evaluation of the entire technique is not a trivial task. Many different distortions might affect the final result, so we have to measure the performance of the technique in the different steps which compose the VM procedure. We will test the three main tasks of the procedure (source localization, pattern estimation, VM signal synthesis) separately, reporting the values of the different metrics.

In the following sections we will introduce the evaluation metrics and we will report the results obtained in the tests.

### 5.1 Measurements Setup

The setup of the tests reflects the problem formulation described in the theory (Sec. 3.1.1). Different source signals have been used, in particular we have deployed a white gaussian noise signal and two different speech signals plus a guitar to test the system. We have

	Parameter	Value
<i>General setting</i>	Software	MATLAB R2016b
	Fs	44100 Hz
	c	340 $ms^{-1}$
	Window type	hanning
	P(window length)	882
	H(hop size)	441
	M(number arrays)	6
	J(number of pattern coefficients)	4
<i>Microphone array configuration</i>	K(number of sources)	2 or 1
	Type	UCA
	N(number of microphone in an array)	4
<i>Model parameters</i>	R(array radius)	0.1 m
	Transfer function	$F(\omega, \mathbf{d}_n, \mathbf{d}_k) = \frac{e^{-ik\ \mathbf{d}_n - \mathbf{d}_k\ }}{4\pi\ \mathbf{d}_n - \mathbf{d}_k\ }$
	J(radiance pattern coefficients)	4/2
	Frequency bands	per octave

Table 5.1 General setting of tests

developed a Matlab script which implements the whole VM procedure. It performs the analysis of the acoustic scene, as presented in Sec. 3 and synthesizes the VM signal according to Sec. 4.1. Hence our software reflects the steps derived in theory and shown in the block diagram Fig. 4.1.

In Table 5.1 we report the main parameters that we adopt both in the simulations and in the experiments with their respective values. Note that the transfer function is an assumed parameter as stated in theory (Sec. 3) and it conforms to the definition of the near field propagation given in 1.18.

As regards the position of the different elements composing the acoustic scene, we refer to the model proposed in Fig. 3.1. The microphone arrays are placed randomly around the source. They can be thought to be ideally located on a circle around the sources. The virtual microphones instead can be freely placed in the space and we will specify their positions for the tests. In general according to the source signals the tests can be divided in two groups: White noise and Speech. Tests inside the groups differ from each other by the number and orientation of the sources. We have chosen these source signals for different reasons. The white noise is useful to examine the behavior of the system in the frequency since it contains the same energy in the whole spectrum. The speech and the guitar instead, simulate a real applicative scenario of our VM technique.

As explained in Sec. 3.1.1 and in Sec. 3.4 the radiation pattern is a frequency dependent function. As regards the directivity estimation, we have to highlight some practices and assumptions made during the implementation.

It is worth to notice that portions of signal can be silent especially in speech. Hence, the pattern coefficients should be estimated in the time instants when the actual signal information is present, and then we can average the found values through a geometric mean. This operation can be implemented by means of a signal detector which detects time instants in which the source is emitting. As we explain in Sec. 3.1.1 and in Sec. 3.4 the radiation pattern is frequency dependent function. We have simplified the estimation working under the assumption that the pattern does not change considerably in the neighborhood of a frequency bin. Hence, the pattern is analyzed per-band and the signal inside a frequency band is averaged. The estimator works on a band of frequencies and not on a single specific frequency. This is especially useful with signals that present a sparse spectrum, such as speech [17]. Indeed, in these signals the energy is concentrated in the frequencies which characterize them, while almost no energy is present in the remaining spectrum. Hence, averaging on bands of frequencies helps in avoiding the error given by estimating the coefficients from a frequency with little energy contribution. We have implemented an octave frequency band estimator which gives us eleven estimation of the coefficients for the bands with the corresponding central frequencies:  $15.625Hz$ ,  $31.25Hz$ ,  $62.5Hz$ ,  $125Hz$ ,  $250Hz$ ,  $500Hz$ ,  $1000Hz$ ,  $2000Hz$ ,  $4000Hz$ ,  $8000Hz$ ,  $16000Hz$ .

### 5.1.1 Source Localization Metrics

In Sec. 3.3 we have described in details the source localization technique exploited in this thesis. We take advantage of the soundfield image in the projective ray space to obtain a robust and flexible method for localizing the sources in the acoustic scene. We have to adopt a proper measure to evaluate the performance of our method. This is not a trivial task since we rely on the soundfield image and several sources of noise can affect the final result. Indeed, as stated in Sec. 3.2, a key step of the source localization is the identification of the peaks in the pseudospectrum. This can be achieved taking the peaks greater than a predefined threshold, but clearly the results strongly depend on the choice of the threshold and it can be distorted by the presence of undesired peaks. In order to attenuate the contribution of those peaks in the localization, we have implemented a windowed version of the pseudospectrum. Since we know that the sources are surrounded by the arrays, we can simply apply a sort of window to the pseudospectrum. Practically, we attenuate the directions external to the analysis scene by putting their contribution to zero. Moreover, we have to consider that during the identification of the sources planes (Sec. 3.3), the performance of the localization is affected by the nondeterministic behavior of RANSAC. Therefore, to evaluate the performance of the localization step, we use the Mean Squared Error(MSE) of the found locations with respect to the actual positions of the sources.

**Mean Squared Error MSE** The mean squared error is a well known measure that assesses the quality of an estimator. It measures the average of the errors which is the difference between the estimated values and what has to be estimated. In our case the error is the difference between the estimated source position and the actual location. The MSE of an estimator  $\hat{Y}$  with the respect to an unknown parameter  $Y$  is defined as:

$$MSE(Y) = \mathbb{E}_Y[(\hat{Y} - Y)^2]. \quad (5.1)$$

The MSE is always non-negative and the closer it is to zero the better is the estimator performance. In practice, we cannot compute the expectation value of 5.1, so we approximate the MSE with  $N$  data as follows:

$$MSE(Y) = \frac{1}{N} \sum_{n=1}^N (\hat{Y}_n - Y_n)^2. \quad (5.2)$$

As regards the problem of evaluating the source localization method, we can apply 5.2 to the estimated positions  $\hat{\mathbf{d}}_s$  obtaining:

$$MSE(\mathbf{d}_s) = \frac{1}{K} \sum_{k=0}^K (\hat{\mathbf{d}}_k - \mathbf{d}_k)^2. \quad (5.3)$$

The MSE of the estimated position gives us a compact and intuitive metric to understand the performance of the estimator and it is useful both in the simulations and in the experiments.

### 5.1.2 Directivity Metrics

Similarly to the evaluation of the localization, derive a metric to properly evaluate the directivity estimator is not an easy task. Indeed, the estimator aims in deriving the pattern coefficients for our model, but the right number of coefficients is not known in advance. Moreover, the estimator depends on the source localization step and on the transfer function assumed. Additionally, it is worth to notice that aside simulation tests we cannot know in advance the directivity of the source. A real evaluation is not possible in these cases since we are not able to compare the estimated patterns with the real ones. Therefore, we will provide an analysis of the radiance pattern only for simulation tests, while in the experiments we will just make some observations based on the overall performance.



To measure the directivity estimation performance we have adopted a solution similar to the localization error metric inspired by the MSE. In this case, the error computed between the estimated directivity and the actual pattern is averaged not only on the different directions  $\theta$  as MSE does, but also on the frequency, in order to obtain a single compact metric for the evaluation. Therefore the metrics that we called Directivity Error (DE) is defined as follows:

$$DE_k = \frac{1}{F} \sum_{f=1}^F MSE_k(f) = \frac{1}{F} \sum_{f=1}^F \frac{1}{R} \sum_{r=1}^R (\hat{P}_k(r) - P_k(r))^2, \quad (5.4)$$

where  $k$  is the source subscript,  $f$  stands for the frequency and  $r$  is the direction for which the directivity is computed. The direction  $R$  are the sampling directions of the basis functions in 3.13.

### 5.1.3 VM Signal Metrics

Define a proper evaluation metrics for the synthesized signal is a tricky task. The idea behind the proposed evaluation metrics is the comparison of the VM signal with the real signal. For this reason, in the simulations we compute the reference signal in the position of the VM and in the experiments we place four reference microphone in order to have a signal which represent the target. Once the VM signal has been synthesized, we can compare it to the target signal called reference signal. This signal gives indeed a reference for our results. It is the ground truth to which we compare the VM. Hence, the aim of the VM procedure is to obtain a synthesized signal as close as possible to that target signal. Define a metrics to evaluate the similarity of two signal is a complex task because perception plays an important role in the evaluation. Hence it is hard to derive a complete metrics which is also compact at the same time.

As done for the other metrics we follow an approach based on the MSE to derive an objective metric called Root Mean Squared Error RMSE. The RMSE is a frequently used metrics for evaluating the difference between values. In RMSE we compute the standard deviation of the prediction error for each sample in the signal. Hence, with a single value we measures the accuracy of the estimation. The RMSE is defined as follows:

$$RMSE_{VM} = \sqrt{MSE_{VM}} = \sqrt{\frac{1}{T} \sum_{t=1}^T |x_{vm}(t) - x_{ref}(t)|^2}, \quad (5.5)$$

	Parameter	Value
<i>Sources</i>	$\mathbf{d}_{s_1}$	$[3, 3]^T$
	$\mathbf{d}_{s_2}$	$[2, 1.5]^T$
	<i>Sourcesignal<sub>1</sub></i>	WN, Female
	<i>Sourcesignal<sub>2</sub></i>	WN, Male
<i>Microphones arrays</i>	$\mathbf{d}_{OW_1}$	$[4.620, 2.208]^T$
	$\mathbf{d}_{OW_2}$	$[4.3123.661]^T$
	$\mathbf{d}_{OW_3}$	$[0.7783.981]^T$
	$\mathbf{d}_{OW_4}$	$[0.1062.827]^T$
	$\mathbf{d}_{OW_5}$	$[1.1900.204]^T$
	$\mathbf{d}_{OW_6}$	$[3.8251.097]^T$
<i>Virtual Microphone</i>	$\mathbf{d}_{VM}$	$[2.5, 2]^T$

Table 5.2 Parameters shared by all the simulation tests. Positions are expressed in [m].

where  $t$  is the time index,  $x_{vm}(t)$  is the synthesized signal and  $x_{ref}(t)$  is the reference signal. For the definition 5.5 is clear that the closer the RMSE is to zero the better is the estimator. Therefore, we can have at a first glance an evaluation of the synthesized signal.

## 5.2 Simulations

In this section, we analyze the performance of the VM procedure in software simulations. First we introduce the simulated measurements setup giving the list of the parameters used in the script. Then we evaluate the performance with the ad hoc metric (Sec. 5.1.1) and discuss the results.

The setup is shown in Fig. 5.1. It conforms to the context described in the theory where the sources are surrounded by arbitrary placed microphones arrays. We performed a total of twelve tests which we can divide based on the source signals. In the Table 5.3, Table 5.4, Table 5.5, we divide the tests based on their source signals and we specify the characteristics of each test. For each type source signal we perform six different tests. In the first and in the second tests the first and second sources are activated respectively, while in the third one the two sources emit simultaneously. Then we repeat the three simulations with a different source directivity. Some parameters do not change among all the tests, these parameters are the source and microphone positions and the VM position. Moreover, in the speech group we always assign a female voice to the first source and a male voice to the second one. In Table 5.2 we report the parameters of the setting as shown in Fig. 5.1.

In order to mimic the consumer devices available in the market, we have modified the sensor arrays adding an extra microphone in the center of the structure. In the developed software the signals are wighted by a specific directivity and propagated to the sensors in the simulated space. We adopt the well known cardioid pattern as radiance pattern introduced in Sec. 3.4 and visible in Fig. 5.1. We propagate the signals according to the near field transfer

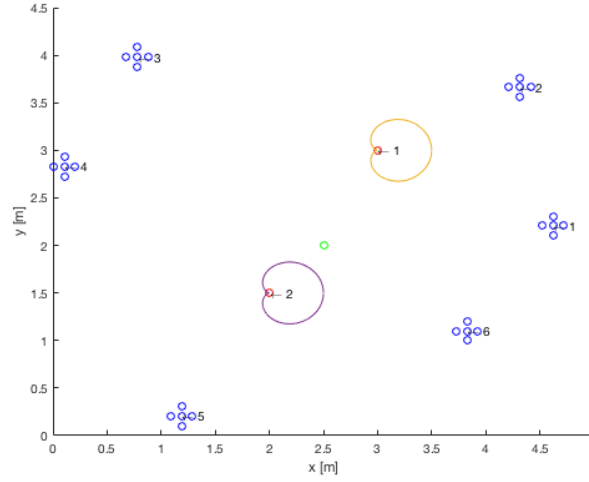


Fig. 5.1 The setup used in the simulations. The arrays are arbitrary located all around the sources (red). The virtual microphone (green) lies almost in the center of the scene. The directivity is shown in the case of  $0^\circ$  orientation for both the sources.

function 1.18. Then, the analysis of the soundfield begins with the source localization followed by the directivity estimation, the source signal retrieval and it is closed by the VM signal synthesis. The reader will notice that this structure reflects the steps introduced in theory except from the fact that we have first to simulate the signals at the microphones.

### 5.2.1 Source Localization

In the Table 5.3 we report the localization performance in the different simulations with the estimated and correct values. For each one of the twelve experiments we indicate the number of sources, their orientation and the relative MSE obtained.

The low MSE values obtained for all the tests prove the reliability of the implemented source localization. Obviously, the performance can be affected by several factors. It is mainly influenced by the setup arrangement and by the presence of multiple sources which emit simultaneously. But even with more than one source emitting at the same time we obtain a good estimation of the sources locations as we can observe from the Table 5.3.

### 5.2.2 Pattern Estimation

In a simulated scenario, we have the control over the source pattern, so it is possible to exactly evaluate the performance of our estimator since we know the directivity of the sources. In our tests, we have decided to apply the well known cardioid pattern to the sources. For each source signal category, the first three simulations present a pattern oriented towards  $0^\circ$ ,

White noise			Speech (Female-Male)		
<i>Simulation 1</i>	Source number Orientation MSE	1 0° 0.000083	<i>Simulation 7</i>	Source number Orientation MSE	1 0° 0.000106
<i>Simulation 2</i>	Source number Orientation MSE	2 0° 0.000013	<i>Simulation 8</i>	Source number Orientation MSE	2 0° 0.000023
<i>Simulation 3</i>	Source number Orientation MSE	1-2 0°, 0° 0.000046	<i>Simulation 9</i>	Source number Orientation MSE	1-2 0°, 0° 0.000146
<i>Simulation 4</i>	Source number Orientation MSE	1 −90° 0.000017	<i>Simulation 10</i>	Source number Orientation MSE	1 0° 0.000061
<i>Simulation 5</i>	Source number Orientation MSE	2 90° 0.000008	<i>Simulation 11</i>	Source number Orientation MSE	2 90° 0.000014
<i>Simulation 6</i>	Source number Orientation MSE	1-2 90°, 0° 0.000018	<i>Simulation 12</i>	Source number Orientation MSE	1-2 90°, 0° 0.000723

(a)
(b)

Table 5.3 Source localization performance in simulated tests. For each test we report which one of the sources is active and the orientation of the pattern. The value 0° corresponds to an orientation parallel to the  $x$  axis in the positive direction while  $\pm 90^\circ$  corresponds to an orientation parallel to the  $y$  axis with respect to Fig. 5.1.

White noise			Speech (Female-Male)		
<i>Simulation 1</i>	Source number Orientation DE	1 0° 0.006612	<i>Simulation 7</i>	Source number Orientation DE	1 0° 0.006238
<i>Simulation 2</i>	Source number Orientation DE	2 0° 0.036062	<i>Simulation 8</i>	Source number Orientation DE	2 0° 0.030793
<i>Simulation 3</i>	Source number Orientation $DE_1$ $DE_2$	1-2 0°, 0° 0.051792 0.041718	<i>Simulation 9</i>	Source number Orientation $DE_1$ $DE_2$	1-2 0°, 0° 0.072337 0.040498
<i>Simulation 4</i>	Source number Orientation DE	1 −90° 0.011016	<i>Simulation 10</i>	Source number Orientation DE	1 −90° 0.090905
<i>Simulation 5</i>	Source number Orientation DE	2 90° 0.008666	<i>Simulation 11</i>	Source number Orientation DE	2 90° 0.008115
<i>Simulation 6</i>	Source number Orientation $DE_1$ $DE_2$	1-2 0°, 90° 0.11717 0.05526	<i>Simulation 12</i>	Source number Orientation $DE_1$ $DE_2$	1-2 90°, 0° 0.082255 0.033562

(a)
(b)

Table 5.4 Pattern estimation performance in simulated tests. For each test we report which one of the sources is active and the orientation of the pattern. The value 0° corresponds to an orientation parallel to the  $x$  axis in the positive direction while  $\pm 90^\circ$  corresponds to an orientation parallel to the  $y$  axis with respect to Fig. 5.1.

namely this corresponds to an orientation towards the  $x$  axis in the positive direction. The other simulations present a pattern oriented both with the positive and the negative  $y$  axis (i.e. 90° and 0°). For the sake of simplicity we assume in our simulations that the pattern is not time dependent. Moreover, for the sake of simplicity we have assigned the cardioid pattern to all frequency. We remark from Sec. 5.1 that the estimation the radiance pattern coefficients is performed over bands of frequencies.

The obtained results in terms of Directivity Error (DE) 5.4, where we have sampled the direction over all the 360° with a step of 5°, are visible in Table 5.4. For each test we report the number, the directivity of the sources and the obtaining DE.

Here the results are mainly influenced by the presence of two sources that propagate signals simultaneously. Although we obtain a good DE, the performance decreases with respect to the single source tests. This is shown in Fig. 5.2.

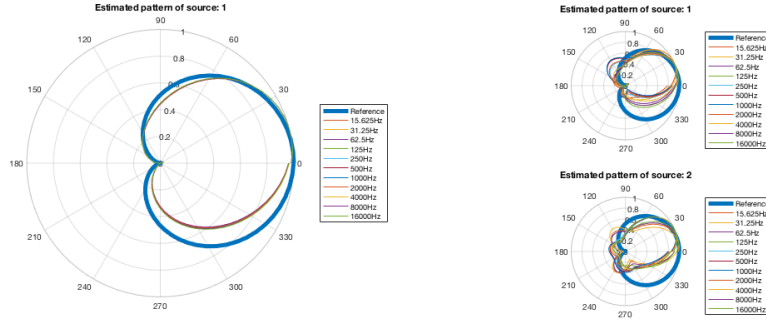


Fig. 5.2 Estimated directivity in the case of one and two sources from tests 1 and 3. The blue thick line represents the reference cardioid pattern.

### 5.2.3 VM Signal Synthesis

The synthesis of the VM signal is the final operation of the *Virtual Miking* technique. We strictly follows the steps given in Sec. 4.1. Once the parameters have been retrieved, we exploit the model of the soundfield to compute the virtual microphone signal. In our simulations we compute the signal of an omnidirectional VM placed in  $\mathbf{d}_{VM} = [2.5, 2.0]^T$ . Our code also computes the actual signal in the position of the VM, in order to have the reference needed for evaluating the goodness of the technique. In Table 5.5 the results of the synthesis phase are shown.

The data confirms the trend observed in the other metrics. The values of RMSE 5.5 are in the order of  $(10^{-2}, 10^{-3})$ , which allow us to have a good reconstruction of the VM signals. From tests number: 3, 9, 12; we can note that the estimator works better when only one source is present. This conforms to the performance of the other estimators.

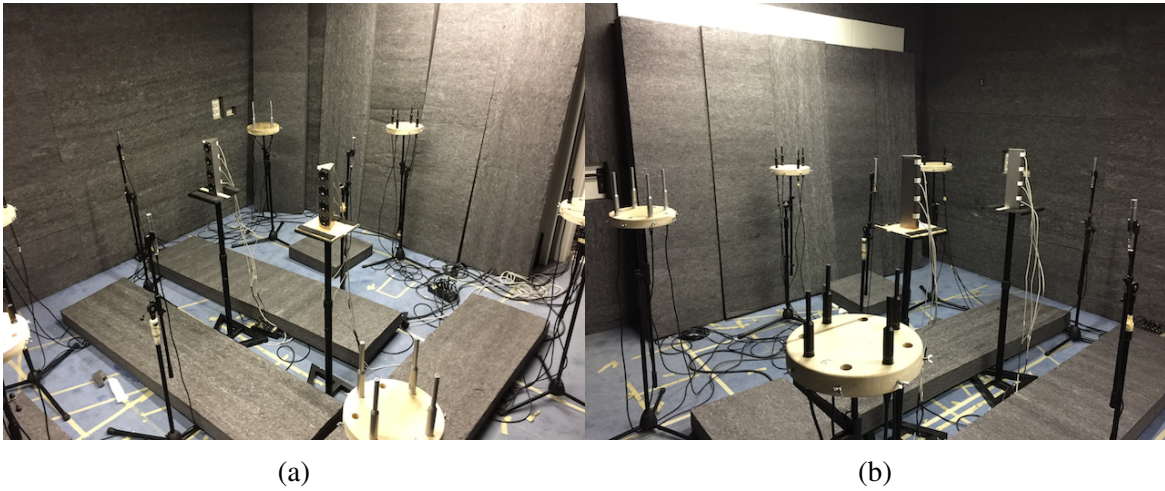


Fig. 5.3 Experiment setup in the semi-anechoic chamber

White noise			Speech (Female-Male)		
<i>Simulation 1</i>	Source number	1	<i>Simulation 7</i>	Source number	1
	Orientation	0°		Orientation	0°
	RMSE	0.008436		RMSE	0.001271
<i>Simulation 2</i>	Source number	2	<i>Simulation 8</i>	Source number	2
	Orientation	0°		Orientation	0°
	RMSE	0.017533		RMSE	0.000884
<i>Simulation 3</i>	Source number	1-2	<i>Simulation 9</i>	Source number	1-2
	Orientation	0°, 0°		Orientation	0°, 0°
	RMSE	0.010484		RMSE	0.003584
<i>Simulation 4</i>	Source number	1	<i>Simulation 10</i>	Source number	1
	Orientation	y <sup>-</sup>		Orientation	-90°
	RMSE	0.027043		RMSE	0.006044
<i>Simulation 5</i>	Source number	2	<i>Simulation 11</i>	Source number	2
	Orientation	90°		Orientation	90°
	RMSE	0.018573		RMSE	0.000595
<i>Simulation 6</i>	Source number	1-2	<i>Simulation 12</i>	Source number	1-2
	Orientation	0°, 90°		Orientation	90°, 0°
	RMSE	0.007486		RMSE	0.001333

(a)
(b)

Table 5.5 VM synthesis performance in simulated tests. For each test we report which one of the sources is active and the orientation of the pattern. The value 0° corresponds to an orientation parallel to the  $x$  axis in the positive direction while  $\pm 90^\circ$  corresponds to an orientation parallel to the  $y$  axis with respect to Fig. 5.1.

<i>Microphones</i>	Beyerdynamic MM1 Behringer ECM 8000 Audiotechnica AT4022
<i>Preamplifier</i>	Focusrite Platinum Le Octopre
<i>Loudspeakers</i>	B & C passive speakers
<i>Power amplifier</i>	B & C power amplifier
<i>ADC/DAC</i>	Lynx Aurora 16

Table 5.6 Hardware equipment

### 5.3 Experiments

In this section we will discuss the performance of the VM technique in a real applicative scenario. First we describe the measurement setting used in the experiments and then we will go through the analysis of the VM procedure in this context. Differently from the simulations cases discussed in Sec. 5.2, here we have used hardware equipment both for reproducing the signals emitted by the sources and for acquiring them. The employed hardware is summarized in Table 5.6.

The deployment of the equipment is shown in Fig. 5.5. Note that it is coherent with the one proposed in theory. The two sources are placed inside the analysis area with four extra microphones which give the references for the VM signals. In Fig. 5.3 we report two pictures of the setup for the sake of clarity. We tested the VM technique in a semi-anechoic room. Our VM procedure has been developed in a 2D space, therefore we have used a set of stands adjustable in height, to ensure that the cones of the loudspeakers and the microphones capsules were lying on the same plane. Microphones are omnidirectional and arranged in the circular array structure by means of ad hoc build supports as shown in Fig. 5.4. As regards the software, the steps are exactly the same of the simulations (Sec. 5.2) and conform with the block diagram of Fig. 4.1.

It is important to note that this kind of tests, hold in real environments with hardware equipment, are subject to many non idealities. In fact, it is practically impossible to have loudspeaker and microphones perfectly aligned at the same height or to perfectly measure the positions of the elements in the space. Moreover, even if we work in a controlled environment, some sound reflections are present due to the presence of the equipment itself and to the inability to have a perfect non-reverberant room. Additionally, the hardware equipment such as microphones, cables, preamplifiers, converters etc. are a inherent and unavoidable noise sources. Hence, the performance of our VM technique is corrupted by the those inevitable approximations. The metrics that we use to evaluate the VM procedure in the experiments are the same of the simulation tests. However, in a real context we cannot control the directivity



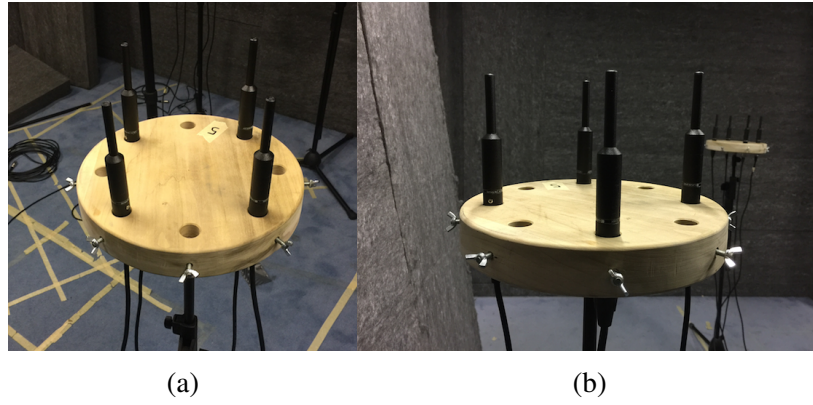


Fig. 5.4 Uniform Circular Array structure with microphones.

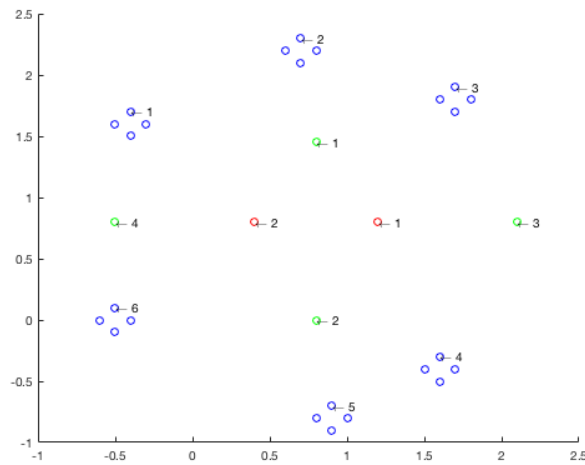


Fig. 5.5 The experiments setting. The six microphones arrays are deployed in the scene around the two sound sources in red. Four microphones (green) are placed in the scene as reference for the VMs signals

Set-up Positions	Values
Sources [m]	
$\mathbf{d}_{s_1}$	$[1.2, 0.8]^T$
$\mathbf{d}_{s_2}$	$[0.4, 0.8]^T$
$\mathbf{d}_{s_1}$ ( <i>Experiment 19</i> )	$[0.9, 0.8]^T$
Reference microphones [m]	
$\mathbf{d}_{ref_1}$	$[0.8, 1.45]^T$
$\mathbf{d}_{ref_2}$	$[0.8, 0.0]^T$
$\mathbf{d}_{ref_3}$	$[2.1, 0.8]^T$
$\mathbf{d}_{ref_4}$	$[-0.5, 0.8]^T$
Microphones Arrays [m]	
$\mathbf{d}_{OW_1}$	$[-0.4, 1.6]^T$
$\mathbf{d}_{OW_2}$	$[0.7, 2.2]^T$
$\mathbf{d}_{OW_3}$	$[1.7, 1.8]^T$
$\mathbf{d}_{OW_4}$	$[1.6, -0.4]^T$
$\mathbf{d}_{OW_5}$	$[0.9, -0.8]^T$
$\mathbf{d}_{OW_6}$	$[-0.5, 0.0]^T$

Table 5.7 Experiment parameters

of the sources, so no real evaluation of the radiation pattern is performed. We simply make some considerations about the estimated directivity, so for the sake of simplicity and in order to ease the considerations about the pattern made in Sec. 5.3.2 we adopt 2 coefficients to compute the directivity. We also report and discuss the performance of the source localization and VM signal synthesis.

We have performed nineteen tests with different source signals and setup. In particular we can divide the experiments in three categories based on the different source signals: white noise, speech, and guitar. Differently from the simulations, we have added a third source signal in order to have a better knowledge of the behavior of our technique in different situations. We use four different speech signal first we will assign the female speech to the first source and the male speech to the second one, then we swap the assigned source and change the source signals.

In Table 5.7 we report the position of the equipment in the different tests. Indeed, the setting is almost fixed for all the experiments except for the sources orientations and in the guitar test, the source position. In all the test the locations of microphone arrays and reference microphones remains the same.

### 5.3.1 Source localization

In this section we report the performance of our source localization step in a real scenario. We remark that the procedure adopted in the experiments is the same as the one implemented for simulations. Hence, we adopt also the same metrics to evaluate the results of the source localization. The MSE is used to measure the error between the estimated positions and the actual ones.

In Table 5.8 we can note that the tests differ from each others on the source signal type, the number of sources and their orientation. The measured MSE computed according to 5.3 is reported for each experiment, just after the number of sources and their orientation. We have grouped the experiment according to their source signal.

It is important to note that for all the experiments the value of the MSE is in the order of  $(10^{-3}, 10^{-4})$  or lower. These values demonstrate the reliability of the source localization method implemented. Remember that the non-deterministic behavior of the RANSAC algorithm (Sec. 3.3) affects the performance of the estimator which returns slightly different values at each execution. Obviously the performance is affected by the presence of more than one emitting source. However it is always able to find a good approximation of the sources positions as we can see from Table 5.8.

### 5.3.2 VM signal synthesis

In this section we report the results obtained by the VM technique during the experiments. The parameters extracted are used to compute the virtual signal in the desired positions  $d_{VM} = [0.8, 1.45]^T$ ,  $d_{VM} = [0.8, 0]^T$ ,  $d_{VM} = [2.1, 0.8]^T$ ,  $d_{VM} = [-0.5, 0.8]^T$ . In these points we have placed the omnidirectional microphones which give us the target signal. We evaluate the synthesized signal with respect to the reference signals. The reference positions have been chosen in order to have the most complete view of the soundfield. Indeed, we compute the VM signal in four positions all around the sources. This let us to give a rough evaluation of the directivity estimation, even if we cannot know the real radiance pattern of the source. In fact we expect that the amplitude of the VM signals vary accordingly to the orientations of the sources. This assumption seems confirmed, for example, by Fig. 5.6. We can notice from the spectrograms in Fig. 5.6 that the estimated signals and the reference ones are very similar. In particular, the amplitude of microphones in the direction of the propagation is higher with the respect to the microphones behind the source.

Fig. 5.6 refers to test number 7. In this case we have the first source emitting a female speech. The directivity is at  $90^\circ$  (i.e. aligned with the  $y$  axis in the positive direction (Fig. 5.5)). Hence, we expect to observe the highest amplitude in the  $VM_1$  signal since it is

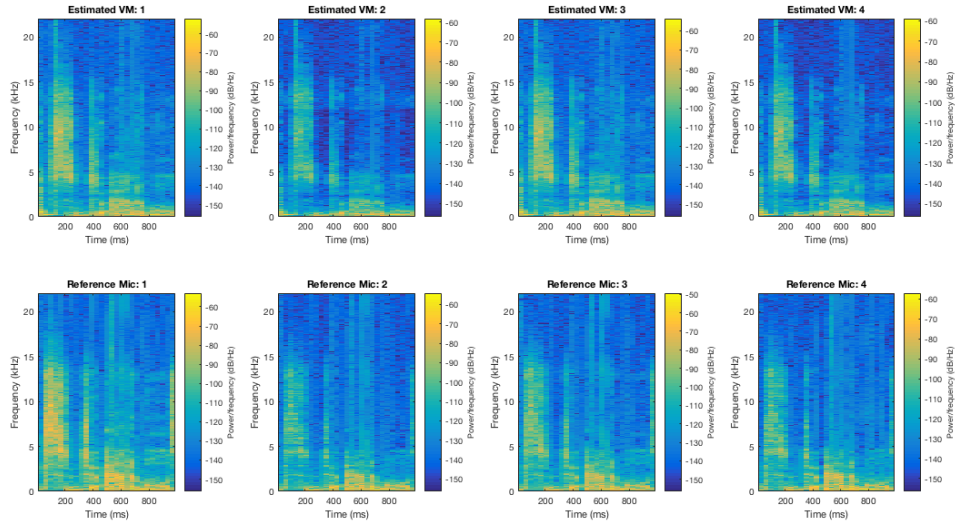
White noise			Speech (Female-Male)		
<i>Experiment 1</i>	Source number Orientation MSE	1 90° 0.000311	<i>Experiment 7</i>	Source number Orientation MSE	1 90° 0.005812
<i>Experiment 2</i>	Source number Orientation MSE	2 90° 0.001208	<i>Experiment 8</i>	Source number Orientation MSE	2 90° 0.000361
<i>Experiment 3</i>	Source number Orientation MSE	1-2 90°, 90° 0.000961	<i>Experiment 9</i>	Source number Orientation MSE	1-2 90°, 90° 0.002375
<i>Experiment 4</i>	Source number Orientation MSE	1 90° 0.000310	<i>Experiment 10</i>	Source number Orientation MSE	1 90° 0.002007
<i>Experiment 5</i>	Source number Orientation MSE	2 −90° 0.000966	<i>Experiment 11</i>	Source number Orientation MSE	2 −90° 0.001449
<i>Experiment 6</i>	Source number Orientation MSE	1-2 90°, −90° 0.002008	<i>Experiment 12</i>	Source number Orientation MSE	1-2 90°, −90° 0.000723

(a) (b)

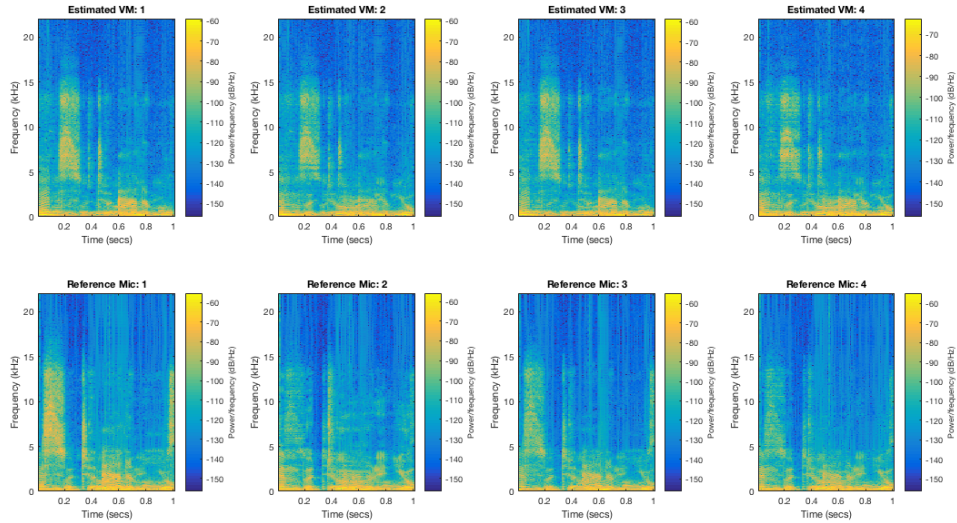
Speech (Male-Female)		
<i>Experiment 13</i>	Source number Orientation MSE	1 90° 0.001818
<i>Experiment 14</i>	Source number Orientation MSE	2 90° 0.0000875
<i>Experiment 15</i>	Source number Orientation MSE	1-2 90°, 90° 0.0009322
<i>Experiment 16</i>	Source number Orientation MSE	1 90° 0.002534
<i>Experiment 17</i>	Source number Orientation MSE	2 −90° 0.000087
<i>Experiment 18</i>	Source number Orientation MSE	1-2 90°, −90° 0.0018522
Guitar		
<i>Experiment 19</i>	Source number Orientation MSE	1 90° 0.0004334

(c)

Table 5.8 Source localization performance in experimental tests. For each test we report which one of the sources is active and the orientation of the pattern. The value 0° corresponds to an orientation parallel to the  $x$  axis in the positive direction while  $\pm 90^\circ$  corresponds to an orientation parallel to the  $y$  axis with respect to Fig. 5.5.



(a) The spectrogram of the synthesized VM with their respective reference signal from experiment 7.



(b) The spectrogram of the synthesized VM with their respective reference signal from experiment 12.

Fig. 5.6

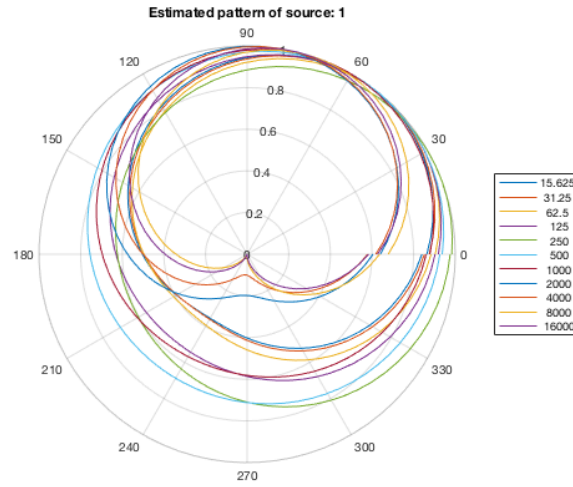


Fig. 5.7 The estimated radiance pattern of source 1 in experiment 1. Frequencies up to  $1000\text{Hz}$  show an omnidirectional-like pattern while the highest frequencies point to  $90^\circ$ .

almost in front of the source. The lowest signal is in correspondence of  $VM_2$  which is placed behind the source. Clearly from the spectrograms this conditions are verified. Moreover it is important to note that the third signal is stronger than the fourth, this behavior is coherent to the actual signals since  $VM_3$  is located closer to the source than  $VM_4$ . This trend appears in all the synthesized spectrograms according to the test setup, in fact in Fig. 5.6a the VM signals present a similar level, according to the setup of test 12. Hence, we can infer that the directivity is estimated correctly. Additionally examining the estimated radiation patterns such as the one reported in Fig. 5.7 we can roughly identify the difference in terms of directivity between high frequencies and low frequencies. In Table 5.9 we summarize the results obtained in terms of RMSE 5.5.

Inspecting the table we can note that the values of the RMSE are low and are coherent with each other. As already explained in Sec. 5.1.3 the evaluation of the similarity of two acoustic signals is a tricky task that concerns also perceptual measures. We interpret these results as a first satisfying evaluation of the VM signal.

## 5.4 Summary

In this chapter a thorough analysis of the performance of our *Virtual Miking* technique has been provided. First we have described the metrics used for evaluating the performance. Defining metrics to evaluate the entire procedure is not an easy task. We have, so analyzed the main steps of the VM technique discussing the obtained results. In this section we have

White noise					
<i>Experiment 1</i>	Source number	1	<i>Experiment 4</i>	Source number	1
	Orientation	90°		Orientation	90°
	$RMSE_1$	0.013896		$RMSE_1$	0.013862
	$RMSE_2$	0.003263		$RMSE_2$	0.002848
	$RMSE_3$	0.007562		$RMSE_3$	0.007613
	$RMSE_4$	0.004775		$RMSE_4$	0.003850
<i>Experiment 2</i>	Source number	2	<i>Experiment 5</i>	Source number	2
	Orientation	90°		Orientation	−90°
	$RMSE_1$	0.014491		$RMSE_1$	0.003444
	$RMSE_2$	0.003371		$RMSE_2$	0.011145
	$RMSE_3$	0.004550		$RMSE_3$	0.003046
	$RMSE_4$	0.006694		$RMSE_4$	0.005660
<i>Experiment 3</i>	Source number	1-2	<i>Experiment 6</i>	Source number	1-2
	Orientation	90°, 90°		Orientation	90°, −90°
	$RMSE_1$	0.017820		$RMSE_1$	0.013286
	$RMSE_2$	0.004298		$RMSE_2$	0.012908
	$RMSE_3$	0.009453		$RMSE_3$	0.007321
	$RMSE_4$	0.008016		$RMSE_4$	0.007449
Speech (Female-Male)					
<i>Experiment 7</i>	Source number	1	<i>Experiment 10</i>	Source number	1
	Orientation	90°		Orientation	90°
	$RMSE_1$	0.010801		$RMSE_1$	0.011250
	$RMSE_2$	0.006272		$RMSE_2$	0.006453
	$RMSE_3$	0.009393		$RMSE_3$	0.011638
	$RMSE_4$	0.005299		$RMSE_4$	0.005474
<i>Experiment 8</i>	Source number	2	<i>Experiment 11</i>	Source number	2
	Orientation	90°		Orientation	−90°
	$RMSE_1$	0.011542		$RMSE_1$	0.008773
	$RMSE_2$	0.008364		$RMSE_2$	0.009636
	$RMSE_3$	0.007163		$RMSE_3$	0.006443
	$RMSE_4$	0.009886		$RMSE_4$	0.010775
<i>Experiment 9</i>	Source number	1-2	<i>Experiment 12</i>	Source number	1-2
	Orientation	90°, 90°		Orientation	90°, −90°
	$RMSE_1$	0.015674		$RMSE_1$	0.014055
	$RMSE_2$	0.010860		$RMSE_2$	0.009831
	$RMSE_3$	0.012121		$RMSE_3$	0.012508
	$RMSE_4$	0.011542		$RMSE_4$	0.008416
Speech (Male-Female)					
<i>Experiment 13</i>	Source number	1	<i>Experiment 16</i>	Source number	1
	Orientation	90°		Orientation	90°
	$RMSE_1$	0.011542		$RMSE_1$	0.011289
	$RMSE_2$	0.006746		$RMSE_2$	0.007093
	$RMSE_3$	0.010380		$RMSE_3$	0.009703
	$RMSE_4$	0.006177		$RMSE_4$	0.006603
<i>Experiment 14</i>	Source number	2	<i>Experiment 17</i>	Source number	2
	Orientation	90°		Orientation	−90°
	$RMSE_1$	0.011507		$RMSE_1$	0.009388
	$RMSE_2$	0.011427		$RMSE_2$	0.011677
	$RMSE_3$	0.010374		$RMSE_3$	0.009622
	$RMSE_4$	0.012039		$RMSE_4$	0.013324
<i>Experiment 15</i>	Source number	1-2	<i>Experiment 18</i>	Source number	1-2
	Orientation	90°, 90°		Orientation	90°, −90°
	$RMSE_1$	0.01838		$RMSE_1$	0.012314
	$RMSE_2$	0.01152		$RMSE_2$	0.013726
	$RMSE_3$	0.01453		$RMSE_3$	0.011911
	$RMSE_4$	0.01347		$RMSE_4$	0.010358
Guitar					
<i>Experiment 19</i>		Source number	1		
		Orientation	90°		
		$RMSE_1$	0.003132		
		$RMSE_2$	0.002536		
		$RMSE_3$	0.002668		
		$RMSE_4$	0.002494		

Table 5.9 VM signal synthesis performance in experimental tests. For each test we report which one of the sources is active and the orientation of the pattern. The value 0° corresponds to an orientation parallel to the  $x$  axis in the positive direction while  $\pm 90^\circ$  corresponds to an orientation parallel to the  $y$  axis with respect to Fig. 5.5.

proposed three different metrics inspired by the MSE 5.1. As regards the source localization we have applied the MSE itself while for the radiance pattern we derived the *Directivity Error* 5.4 as an average on frequency of the MSE between the estimated pattern and the real one. Finally the evaluation of the VM signal has been performed through the RMSE 5.5. We have analyzed the performance of our VM procedure in two different context. The first one is a simulated scenario (Sec. 5.2) which gives us the advantage of knowing the exact values of the signals and parameters. This is fundamental for the computation of the metrics, in particular for the directivity of the source. In fact, in the second context Sec. 5.3, a real applicative scenario we have no knowledge about the radiance pattern of the source, although we were able to make some observation about source position and orientation. In the second context we have been able to test the VM technique in a semi-anechoic chamber where source and arrays of microphones have been deployed. To test the VM signals we have used four reference microphone placed in the space of analysis. In order to evaluate our work in different conditions, we adopt both in simulations and experiments various source signals. In particular we have chosen the white gaussian noise due to its frequency content and speeches and a guitar to simulate real use-cases. The results reported for all the three metrics are reported in tables to easily examine their values. In two sources scenarios the results are little lower for every metrics due to the fact that the sources are emitting simultaneously. But if we consider the context of a common conversation, the speaker barely overlaps themselves, thus the performance can be considered consistent with the one source tests. Indeed, since we work in a short-time approach, we can assume that in a conversation, no more than one source is present in each time frame. It is important to note that the evaluation of the synthesized signal is not trivial task, since it concerns an important perceptive component. We have introduced a compact and objective metrics in order to have at a first glance a measure of the synthesized signal.

Overall, the proposed *Virtual Miking* show satisfactory and promising results both in simulation and in the experiments.



## CONCLUSIONS AND FUTURE WORKS

---

In this thesis the problem of *Virtual Miking* has been addressed. The *Virtual Miking* is a well known application in the literature. It involves the computation of one or more virtual microphone (VM) signals, which approximate the signals of physical microphones placed at the VM positions. We have proposed a flexible technique inspired by the state of art of VM [25] and acoustic field representation. The designed system exploits the powerful geometrical representation of the soundfield, which gives a complete and sturdy description of the acoustic field, and a parametric model inspired by [25] and [7] to compute the VM signals. As we saw, in the recent years the geometric description of the soundfield has been deeply investigated and very interesting results have come out. Our approach starts from an efficient soundfield representation called ray space (Sec 1.1). In this domain every point represents the information of the soundfield "transported" along a ray with a given orientation. However, this transformation is based on a Euclidean parametrization of the rays that is not general enough to account for all rays in any given direction and thus is not suitable for our approach. In order to overcome this limitation, a re-parametrization of the ray space transform using a projective characterization of the rays has been done in [11]. In the projective ray space we are able to represent rays in any direction and as a consequence, to efficiently represent the signals acquired by multiple array arbitrary distributed in space. Therefore, we can analyze the acoustic scene with an higher grade of flexibility with respect to techniques where the location of the acquisition devices is strictly defined. Thanks to the soundfield parametric model introduced in Sec. 3.1.1, we are able to describe the whole acoustic scene by means of few parameters. This parameters are:

1. Source Signal.
2. Source Directivity.
3. Source Position and propagation function.

The presence of the source radiation pattern in the model represents an advance with respect to the existing VM literature. In fact, the knowledge of the source directivity let us approximate a real sound source better than the usual Isotropic Point Like Source model [25]. Obviously, the accuracy of the VM signal depends on the estimated parameters. We have designed and developed an analysis procedure able to extract the needed informations from the microphones signals. In particular, we exploit the characteristics of the soundfield image in the projective ray space to localize the source in the space by means of the RANSAC algorithm (Sec. 3.3). This information is necessary for the computation of the transfer function, which model the propagation of the acoustic rays from the source to the sensors. As regards the source directivity, in Sec. 3.4 we have derived a linear constrained optimization problem that allows us to compute the radiance pattern of the source. In Sec. 3.4 we have leveraged the LCMV beamformer in order to retrieve the source signal.

In order to validate our VM technique, both simulations and experimental sessions have been designed in which our procedure has been evaluated by means of ad hoc metrics. During the tests we adopt different source signals. In particular we have used white noise, speech and a recorded guitar. The first kind of source signal is useful to understand the system frequency behavior, while the speech and the guitar have been adopted to approximate a real applicative scenario. The metrics used to evaluate the different steps of the VM technique are inspired by the Mean Squared Error (MSE) 5.1. In particular, we have examined the performance of the system in localizing the sources positions, retrieving the source directivity and computing the VM signal. As regards the source localization, as a metrics of evaluation we have adopted the MSE of the error between the estimated position and the actual one. An averaged MSE that we have called Directivity Difference DE 5.4 has been defined in order to examine the performance of the radiance pattern estimator in the simulations. As far as the VM signal evaluation is concerned, deriving a complete and at the same time compact metric is not a trivial task, since the perceptive evaluation plays an important role here. Hence, at first, an objective metric, the RMSE 5.5, has been adopt to evaluate in a compact fashion the synthesized VM signal. The RMSE is computed with respect to a reference signal which is the signal simulated or captured by an actual microphone in the VM position. Further metrics including perceptive ones can be a matter of future works.

The proposed VM technique has shown promising results both in simulations and in experiments that let us envision future developments. For example, we can easily extend our model to take into account for the pick up pattern of the VM. This let us to simulate the behavior of any arbitrary directional microphone. Moreover we can envision the possibility of working in a reverberant environment similarly to what is done in [25]. Another possible improvement is related to the extension of the proposed VM procedure to a three-dimensional

scenario. This would open the way to a broad range of applications in music-related areas or forensic analysis. For instance, we could image the VM procedure exploited in recording studios allowing the design of novel and innovative miking technique at a very low cost. Additionally we can envision the employment of our VM technique in acoustic environment monitoring with the possibility of performing an acoustical zooming. Finally, a remarkable achievement would consist in a real time implementation of the whole *Virtual Miking* technique.



## REFERENCES

---

- [1] A. Canclini, F. Antonacci, A. S. and Tubaro, S. (2013). Acoustic source localization with distributed asynchronous microphone networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 21, no. 2:439–443.
- [2] Adelson, E. H. and Bergen, J. R. (1991). The plenoptic function and the elements of early vision. *MIT Press*, page 3–20.
- [3] Antonacci, F., Foco, M., Sarti, A., and Tubaro, S. (2008). Fast tracing of acoustic beams and paths through visibility lookup. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(4):812–824.
- [4] Bianchi, L. (2016). A unified framework for acoustic scene analysis, synthesis and processing.
- [5] Borra, F. (2016). A flexible methodology for acoustic source extraction based on the ray space transform.
- [6] Canclini, A., Mucci, L., Antonacci, F., Sarti, A., and Tubaro, S. (2015a). Estimation of the radiation pattern of a violin during the performance using plenacoustic methods.
- [7] Canclini, A., Mucci, L., Antonacci, F., Sarti, A., and Tubaro, S. (2015b). A methodology for estimating the radiation pattern of a violin during the performance. pages 1546–1550.
- [8] Carter, G. C. (1981). Time delay estimation for passive sonar signal processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-29, no. 3:463–470.
- [9] Coleman, T. F. and Li, Y. (1996). A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization*, 6(4):1040–1058.
- [10] D. Marković, F. Antonacci, A. S. and Tubaro, S. (2013). Soundfield imaging in the ray space. *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, 21(12):2493–2505.
- [11] D. Marković, F. Antonacci, A. S. and Tubaro, S. (2015). Multiview soundfield imaging in the projective ray space. *IEEE/ACM Trans. Audio, Speech, Language Process.*, 23:1054–1067.
- [12] Devaney, A. J. and Wolf, E. (1974). Multipole expansions and plane wave representations of the electromagnetic field. *J. Mathematical Physics*, 15, no. 2:234–244.
- [13] Faller, C. (2004). Parametric coding of spatial audio. *Ph.D. dissertation, École Polytechnique Fédérale de Lausanne*.

- [14] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.
- [15] G. Cincotti, F. Gori, M. S. F. F. F. F. and Schettini, G. (1993). Plane wave expansion of cylindrical functions. *Optics Communications*, 95:192–198.
- [16] Gill, P. E., Murray, W., and Wright, M. H. (1981). Practical optimization.
- [17] Kameoka, H., Nakatani, T., and Yoshioka, T. (2009). Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms. pages 45–48.
- [18] Kuntz, A. (2008). Wave field analysis using virtual circular microphone arrays. *Ph.D. dissertation, Friedrich-Alexander-Universität Erlanger-Nürnberg*.
- [19] N. Strobel, S. S. and Rabenstein, R. (2001). Joint audio-video object localization and tracking. *IEEE Signal Processing Magazine*, 18, no. 1:22–31.
- [20] P. A. Gauthier, C. Camier, Y. P. A. B. E. C. R. L. and Delalay, M. A. (2011). Beam-forming regularization matrix and inverse problems applied to sound field measurement and extrapolation using microphone array. *J. Sound and Vibration*, 330:5852–5877.
- [21] Rafaely, B. (2004). Plane-wave decomposition of the pressure on a sphere by spherical convolution. *J. Acoust. Soc. Am.*, 116, no. 4:2149–2157.
- [22] Saleh, B. E. A. and Teich, M. C. (1991). *Fundamentals of Photonics*. New York, NY, USA.
- [23] Stoica, P. and Moses, R. L. (1997). *Introduction to spectral analysis*, volume 1. Prentice hall Upper Saddle River, NJ.
- [24] T. Ajdler, L. S. and Vetterli, M. (2006). The plenacoustic function and its sampling. *IEEE Trans. Signal Process.*, 54:3790–3804.
- [25] Thiergart, O., Del Galdo, G., Taseska, M., and Habets, E. A. (2013). Geometry-based spatial sound acquisition using distributed microphone arrays. *IEEE transactions on audio, speech, and language processing*, 21(12):2583–2594.
- [26] Trees, H. L. V. (2002). Optimum array processing, part iv of detection, estimation and modulation theory.
- [27] Weyl, H. (1919). Ausbreitung elektromagnetischer wellen über einen ebenen leiter. *Ann. Phys.*, 60:481–500.
- [28] Whittaker, E. T. (1903). On the partial differential equations of mathematical physics. *Math. Ann.*, 57:333–355.
- [29] Woodward, P. M. (1964). *Probability and Information Theory with Applications to Radar*. Pergamon Press, Oxford, UK.

## APPENDIX A

### EQUIPMENT SPECIFICATION

---

## MM 1

Measurement Microphone

Order # 449.350



### FEATURES

- Linear frequency response in the diffuse field / under 90°
- Omnidirectional polar pattern
- Calibrated open circuit voltage
- Narrow tubular construction

### TECHNICAL SPECIFICATIONS

Transducer type. . . . .	Condenser (back electret)
Operating principle. . . . .	Pressure
Frequency response. . . . .	20 - 20,000 Hz (50 - 16,000 Hz $\pm$ 1.5 dB)
Polar pattern. . . . .	Omnidirectional, diffuse field calibrated
Open circuit voltage at 1 kHz. . . . .	15 mV/Pa (= -36.5 dBV) $\pm$ 1 dB
Nominal impedance. . . . .	160 $\Omega$
Nominal load impedance. . . . .	$\geq$ 2.2 k $\Omega$
Max. SPL at f = 1 kHz, k = 1%, R <sub>L</sub> = 2.2 k $\Omega$ . . . . .	122 dB <sub> SPL</sub>
S/N ratio rel. to 1 Pa. . . . .	> 57 dB
A-weighted equivalent SPL. . . . .	approx. 26 dB(A)
Power supply. . . . .	12 - 48 V phantom supply
Current consumption. . . . .	approx. 1.9 mA
Output. . . . .	electronically balanced
Connection. . . . .	3-pin XLR male
Dimensions:	
Length. . . . .	133 mm
Shaft diameter. . . . .	19/9 mm
Head diameter. . . . .	9 mm
Weight (w/out cable). . . . .	88 g

### APPLICATIONS

The MM 1 is a measurement microphone which has been designed specifically for measuring sound reinforcement and PA-systems. It is designed to work with spectrum analysers for measuring frequency response and sound pressure levels of loud speaker systems. The MM 1 is the ideal microphone for the measurement of audio signals in the research, development, for reverberation testings and other applications.

The narrow tubular construction ensures that the microphone has negligible influence on the sound field so that an increase in sound pressure is avoided with high frequencies. A natural reproduction is achieved due to the linear frequency response.

### OPTIONAL ACCESSORIES

GST 400	Microphone stand, 3/8", height 0.90 - 1.65 m, with G 400 boom. . . . .	Order # 421.294
GST 500	Microphone stand, 3/8", height 0.85 - 1.60 m, with telescopic G 500 boom. . . . .	Order # 406.252
ST 400	Microphone stand, 3/8", height 0.90 - 1.65 mm. . . . .	Order # 421.286
ST 500	Microphone stand, 3/8", height 0.85 - 1.60 mm. . . . .	Order # 406.643
WS 10	Wind shield, charcoal grey. . . . .	Order # 403.008

1 of 2

beyerdynamic GmbH & Co. KG  
Theresienstr. 8 | 74072 Heilbronn - Germany  
Tel. +49 (0) 71 31 / 617 - 0 | Fax +49 (0) 71 31 / 617 - 204  
info@beyerdynamic.de | www.beyerdynamic.com

For further distributors worldwide, please go to [www.beyerdynamic.com](http://www.beyerdynamic.com)  
Non-contractual illustrations. Contents subject to change without notice. E5 / MM 1 (05.17)

**beyerdynamic**

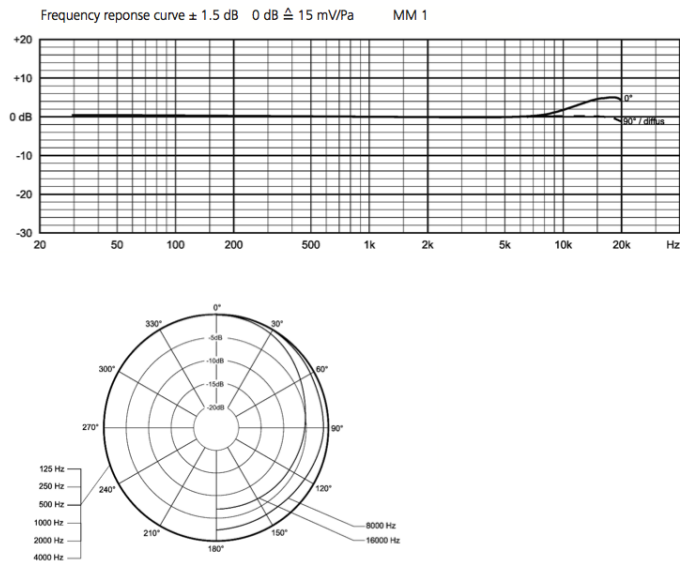
Fig. A.1 Measurement microphone MM1



## MM 1

### FREQUENCY RESPONSE & POLAR PATTERN

This polar pattern and frequency response curve (measuring tolerance  $\pm 1.5$  dB) correspond to a typical production sample for this microphone. Each microphone is supplied with an individual 0° frequency response curve. When sending a request to [proaudio@beyerdynamic.de](mailto:proaudio@beyerdynamic.de), the measured data can be sent via e-mail.



2 of 2

beyerdynamic GmbH & Co. KG  
Theresienstr. 8 | 74072 Heilbronn – Germany  
Tel. +49 (0) 71 31 / 617 - 0 | Fax +49 (0) 71 31 / 617 - 204  
[info@beyerdynamic.de](mailto:info@beyerdynamic.de) | [www.beyerdynamic.com](http://www.beyerdynamic.com)

For further distributors worldwide, please go to [www.beyerdynamic.com](http://www.beyerdynamic.com)  
Non-contractual illustrations. Contents subject to change without notice. E5 / MM 1 (05.17)

**beyerdynamic**

Fig. A.2 Measurement microphone MM1

# MEASUREMENT MICROPHONE ECM8000

## Technical Specifications

ENGLISH

Version 1.2 July 2000

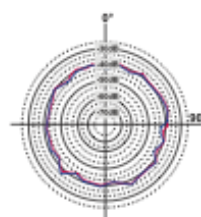
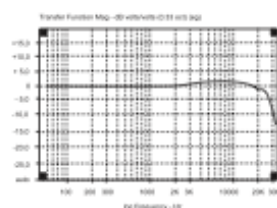
### MEASUREMENT MICROPHONE

#### FEATURES

- ▲ Precise electret condenser measurement microphone
- ▲ Ultra-linear frequency response
- ▲ Well-balanced, true omni-directional pattern
- ▲ Optimally suited for room correction applications
- ▲ Phantom powered, +15 V to +48 V
- ▲ Rugged construction and sleek, modern design
- ▲ Microphone stand adapter and windscreen for outdoor measurement included
- ▲ Perfect for use with the ULTRA-CURVE DSP8000 / ULTRA-CURVE PRO DSP8024 or any other real-time analyzer
- ▲ Manufactured under the BEHRINGER quality control

#### SPECIFICATIONS

Type	electret condenser, omni-directional
Impedance	600 Ohms
Sensitivity	-60 dB
Frequency response	15 Hz to 20 kHz
Connector	gold-plated XLR
Phantom power	+15 V to +48 V
Weight	app. 120 g



BEHRINGER is constantly striving to maintain the highest professional standards. As a result of these efforts, modifications may be made from time to time to existing products without prior notice. Specifications and appearance may differ from those listed or illustrated.

The information contained in this sheet is subject to change without notice. No part of this sheet may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording of any kind, for any purpose, without the express written permission of BEHRINGER Spezielle Studioteknik GmbH.  
BEHRINGER and ULTRA-CURVE are registered trademarks. ALL RIGHTS RESERVED.  
© 2000 BEHRINGER Spezielle Studioteknik GmbH.  
BEHRINGER Spezielle Studioteknik GmbH, Hans-Martin-Schleyer-Str. 36-38, 47877 Willich-Mönchheide II, Germany  
Tel. +49 (0) 21 54 / 92 06-0, Fax +49 (0) 21 54 / 92 06-30



Fig. A.3 Measurement microphone ECM 800

**AT4022**

## Omnidirectional Condenser End-Address Microphone

audio-technica

40 series studio microphones



To engage the high-pass filter, slide the switch toward the "bent" line.

The microphone is also equipped with a switchable 10 dB pad that lowers the microphone's sensitivity, thus providing higher SPL capability for flexible use with a wide range of users and system configurations. To engage the 10 dB pad, slide the switch toward the -10 position.

Avoid leaving the microphone in the open sun or in areas where temperatures exceed 110° F (43° C) for extended periods. Extremely high humidity should also be avoided.

**Features**

- Specially engineered to meet the most critical acoustic requirements of professional recording, broadcast and sound reinforcement
- Low self noise perfectly suited for the most sophisticated recording equipment
- Low-mass diaphragm improves transient response, increases response bandwidth and reduces handling and mechanical noise transfer
- Omnidirectional polar pattern provides maximum ambient pickup
- Rugged design and construction for reliable performance
- Integral 80 Hz high-pass filter switch and 10 dB pad switch
- State-of-the-art design and manufacturing techniques ensure compliance with A-T's stringent consistency and reliability standards

**Description**

The AT4022 is a fixed-charge condenser microphone with an omnidirectional polar pattern. It is designed to meet the demands of critical studio and live applications.

The microphone requires 48V phantom power for operation.

The output of the microphone is a 3-pin XLRM-type connector.

The microphone is equipped with a switchable 10 dB pad and a switch that permits choice of flat response or low-frequency roll-off (via integral 80 Hz high-pass filter).

The microphone is enclosed in a rugged housing. The included ATB405a stand clamp permits mounting on any microphone stand with 5/8"-27 threads. A windscreen and a protective carrying case are also included.

**Operation and Maintenance**

The AT4022 requires 48V phantom power for operation.

Output is low impedance (Lo-Z) balanced. The signal appears across Pins 2 and 3; Pin 1 is ground (shield). Output phase is "Pin 2 hot"—positive acoustic pressure produces positive voltage at Pin 2.

To avoid phase cancellation and poor sound, all mic cables must be wired consistently: Pin 1-to-Pin 1, etc.

An integral 80 Hz high-pass filter provides easy switching from a flat frequency response to a low-end roll-off. The roll-off position reduces the pickup of low-frequency ambient noise (such as traffic, airhandling systems, etc.), room reverberation and mechanically coupled vibrations.

Fig. A.4 Measurement microphone AT 4022

**Specifications**

Element	Fixed-charge back plate, permanently polarized condenser
Polar pattern	Omnidirectional
Frequency response	20-20,000 Hz
Low frequency roll-off	80 Hz, 18 dB/octave
Open circuit sensitivity	-34 dB (19.9 mV) re 1V at 1 Pa
Impedance	250 ohms
Maximum input sound level	146 dB SPL, 1 kHz at 1% T.H.D.; 156 dB SPL, with 10 dB pad (nominal)
Noise <sup>1</sup>	13 dB SPL
Dynamic range (typical)	133 dB, 1 kHz at Max SPL
Signal-to-noise ratio <sup>1</sup>	81 dB, 1 kHz at 1 Pa
Phantom power requirements	48V DC, 3.0 mA typical
Switches	Flat, roll-off; 10 dB pad (nominal)
Weight	124 g (4.4 oz)
Dimensions	144.0 mm (5.67") long, 21.0 mm (0.83") diameter
Output connector	Integral 3-pin XLRM-type
Audio-Technica case style	S1
Accessories furnished	AT8405a stand clamp for 5/8"-27 threaded stands; AT8159 windscreen; protective carrying case

In the interest of standards development, A.T.U.S. offers full details on its test methods to other industry professionals on request.

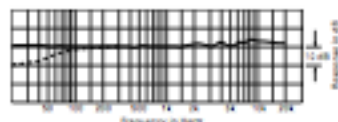
1 Pascal = 10 dynes/cm<sup>2</sup> = 10 microbars = 94 dB SPL

<sup>1</sup> Typical, A-weighted, using Audio Precision System One.

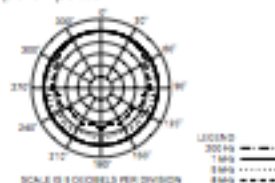
Specifications are subject to change without notice.



frequency response: 20–20,000 Hz



polar pattern



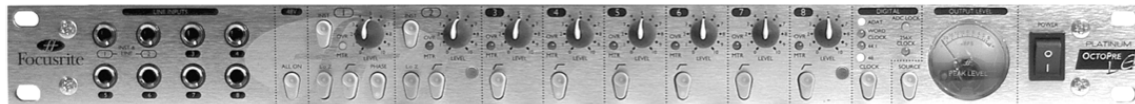
 **audio-technica**

Audio-Technica Corporation  
audio-technica.com ©2016 Audio-Technica

P52065-01

Fig. A.5 Measurement microphone AT 4022

## INTRODUCTION



The OctoPre LE features eight channels of classic Focusrite preamplification based on the popular Green Range. As well as eight analogue outputs, there is the option of a digital converter card (CODEC), which both accepts and transmits digital audio in either 44.1 or 48kHz 24-bit ADAT™ optical 'lightpipe' format. The optional CODEC means that the OctoPre LE can also be used as a monitoring control between the DAW and speaker system, for up to 7.1 surround monitoring.

The mic pres featured in the OctoPre LE have a circuit design based around the classic Focusrite Green range and feature an extremely transparent sonic character with very low distortion. They utilise the same full bandwidth philosophy of all Focusrite products, ensuring detail and clarity without obvious colouration.

Eight balanced (although unbalanced can be used) 1/4" jack inputs are located on the front panel for quick and easy access, the first two channels of which can be used for DI-free plug-in of guitars and basses. Channels 1 and 2 also allow a microphone signal (input using the rear panel XLR connections) the option of a lower impedance input, should the microphone in use have a low output impedance (around 50Ω for example).

With classic Focusrite analogue circuit design and comprehensive A/D and D/A interface options, OctoPre LE is the perfect partner for a digital audio workstation, making high quality, multi-channel recording easy. Equally, it could be utilised as part of a compact multi-channel location recording solution (with a laptop or ADAT™, for example), or as an additional set of professional-quality mic pres for any analogue or digital console or hard disk recorder.

## REAR PANEL CONNECTIONS



The OctoPre LE features eight microphone inputs on XLR connectors and eight balanced line outputs on 1/4" (TRS) jack connectors.

Fig. A.6 Microphones preamplifier

## SPECIFICATIONS

### Mic Input Response

Gain = +13dB to +60dB  
 Input Impedance = 2.5kΩ/ 150Ω on Lo Z (Ch1 & 2)  
 EIN = 124dB @ 60dB Gain with 150Ω Termination & 22Hz/22kHz Filter  
 THD+N @ Min Gain (+13dB) = 0.0006% with 0dBu input & 22Hz/22kHz Filter  
 THD+N @ Max Gain (+60dB) = 0.003% with -36dBu input & 22Hz/22kHz Filter  
 THD+N @ Max Input (+9dBu) = 0.0008% with 22Hz/22kHz Filter  
 Frequency Response:  
 Min Gain (+13dB) with -13dBu input = -0.4dB @ 10Hz & -3dB @ 122kHz  
 Max Gain (+60dB) with -60dBu input = -2.3dB @ 10Hz & -3dB @ 67kHz  
 CMRR @ Max Gain (+60dB) = 80dB

### Line Input Response

Gain = -10dB to +36dB  
 Input Impedance = 24kΩ  
 Noise @ Unity Gain (0dB) = -88dBu with 22Hz/22kHz Filter  
 S/N Ratio relative to max headroom (+36dBu) = 124dB  
 S/N Ratio relative to 0dBfs (+22dBu) = 110dB  
 THD+N @ Unity Gain (0dB) = 0.001% with 0dBfs (+22dBu) input and 22Hz/22kHz Filter  
 Frequency Response @ Unity Gain (0dB) = -0.5dB @ 10Hz & -3dB @ 110kHz

### Instrument Input Response

Gain = +13dB to +60dB  
 Input Impedance = 1MΩ  
 Noise @ Min Gain (+13dB) = -87dBu with 22Hz/22kHz Filter  
 Noise @ Max Gain (+60dB) = -42dBu with 22Hz/22kHz Filter  
 THD+N @ Min Gain (+13dB) = 0.001% with 0dBu input & 22Hz/22kHz Filter  
 Frequency Response @ Min Gain (+13dB) with -13dBu input = -0.4dB @ 10Hz & -3dB @ 122kHz

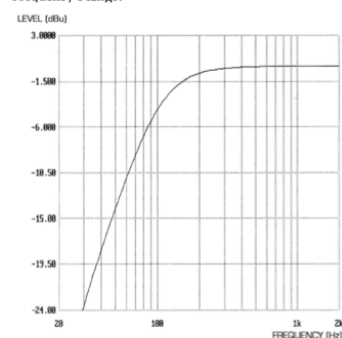
### Input Meter

Peak Level Moving Coil Meter  
 -24dBfs to +24dBfs (-2dBu to +24dBu), +22dBu = 0dBfs  
 Overload LED's triggered @ 0dBfs (+22dBu)

### High Pass Filter

Roll Off = 12dB/Octave 2 pole filter  
 Cut Off Frequency: -3dB @ 120Hz, -6dB @ 85Hz, -12dB @ 56Hz

### Frequency Range:



### OPTIONAL DIGITAL CONVERTER

#### DAC Performance

Playback Sample Frequency = 44.1kHz & 48kHz  
 Maximum Bit Depth = 24-bit  
 Maximum analogue output level = +22dBu  
 Dynamic Range = 107dB 'A' weighted

#### ADC Performance

Sample Frequency = 44.1kHz & 48kHz  
 Bit Depth = 24-bit  
 Maximum analogue input level = +22dBu (0dBfs)  
 Dynamic Range = 109.5dB 'A' weighted

#### Connections

Digital in/out: ADAT™-type optical 'lightpipe'  
 Wordclock in/out: BNC

Fig. A.7 Microphones preamplifier



## LYNX AURORA 16 AND AURORA 8 SPECIFICATIONS

### ANALOG I/O

<b>Aurora 8</b>	Eight inputs and eight outputs
<b>Aurora 16</b>	Sixteen inputs and sixteen outputs
<b>Type</b>	Electronically balanced or unbalanced,
<b>Level</b>	+4 dBu nominal / +20 dBu max. or -10 dBV nominal / +6 dBV max
<b>Input Impedance</b>	Balanced mode: 24k $\Omega$ Unbalanced mode: 12k $\Omega$
<b>Output Impedance</b>	Balanced mode: 100 $\Omega$ Unbalanced mode: 50 $\Omega$
<b>Output Drive</b>	600 $\Omega$ impedance, 0.2 $\mu$ F capacitance
<b>A/D and D/A Type</b>	24-bit multi-level, delta-sigma

### ANALOG IN PERFORMANCE

<b>Frequency Response</b>	20 Hz - 20 kHz, +0/-0.1 dB
<b>Dynamic Range</b>	117 dB, A-weighted
<b>Channel Crosstalk</b>	-120 dB maximum, 1 kHz signal, -1 dBFS
<b>THD + N</b>	-108 dB (0.0004%) @ -1 dBFS -104 dB (0.0006%) @ -6 dBFS 1 kHz signal, 22 Hz - 22 kHz BW

### ANALOG OUT PERFORMANCE

<b>Frequency Response</b>	20 Hz - 20 kHz, +0/-0.1 dB
<b>Dynamic Range</b>	117 dB, A-weighted
<b>Channel Crosstalk</b>	-120 dB max., 1 kHz signal, -1 dBFS
<b>THD + N</b>	-107 dB (0.00045%) @ -1 dBFS -106 dB (0.00050%) @ -6 dBFS 1 kHz signal, 22 Hz - 22 kHz BW

### DIGITAL I/O

<b>Number / Type</b>	Aurora 8: 8 inputs and 8 outputs Aurora 16: 16 inputs and 16 outputs 24 bit AES/EBU format, transformer coupled
<b>Channels</b>	Aurora 8: 8 in/out in single-wire mode 4 in/out in dual-wire mode Aurora 16: 16 in/out in single-wire mode 8 in/out in dual-wire mode
<b>Sample Rates</b>	All standard rates and variable rates up to 192 kHz in both single-wire and dual-wire modes

### ON-BOARD DIGITAL MIXER (VIA AES16)

<b>Type</b>	Hardware-based, low latency
<b>Routing</b>	Ability to route any input to any or multiple outputs
<b>Mixing</b>	Up to four input or playback signals mixed to any output, 40-bit precision
<b>Status</b>	Peak levels to -114 dB on all inputs and outputs

### CONNECTIONS

<b>Digital I/O Ports</b>	25-pin female D-sub connectors Port A: channels 1-8 I/O Port B: channels 9-16 I/O (Aurora 16 only) Yamaha pinout standard
<b>Analog I/O Ports</b>	25-pin female D-sub connectors. Analog In 1-8 Analog In 9-16 (Aurora 16 only) Analog Out 1-8 Analog Out 9-16 (Aurora 16 only) Tascam pinout standard
<b>External Clock</b>	75 $\Omega$ BNC word clock input and output
<b>MIDI</b>	One input and one output. Standard opto-isolated, 5-pin female DIN connectors

### REMOTE CONTROL OPTIONS

<b>Function</b>	Controls all I/O, levels, monitoring, routing and setting recall
<b>Method</b>	AES16/AES16e: With PC or Macintosh MIDI: Selected MIDI devices

### GENERAL

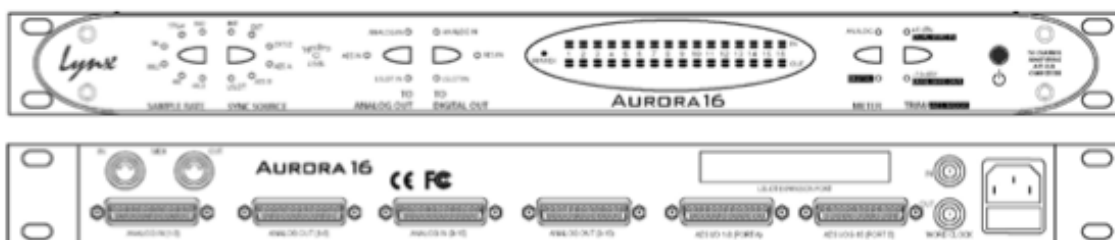
<b>AC Power</b>	100 / 115 / 230 VAC, 70 watts
<b>Size</b>	1.75" H x 19" W x 9" D
<b>Shipping Weight</b>	12 pounds
<b>Certifications</b>	CE and FCC Class B EMI, CE Product Safety

### LSLOT™ EXPANSION PORT

<b>Compatibility</b>	Supports Lynx LSlot expansion cards
<b>Channels</b>	Up to 16 input and 16 output simultaneously at up to 192 kHz sample rate

### OPTIONAL INTERFACE CARDS FOR L SLOT

<b>LT-ADAT</b>	Provides 16-channel at 48 kHz, 8-channel at 96 kHz, 4-channel at 192 kHz ADAT Optical I/O
<b>LT-HD</b>	Provides interface for Avid® ProTools   HD® systems
<b>LT-MADI</b>	Provides up to 64 channels of I/O
<b>LT-USB</b>	Provides up to 16 channels of I/O, USB 2.0
<b>LT-TB</b>	Provides up to 32 channels of I/O



Phone: 714-545-4700 Fax: 714-545-4777  
Email: [sales@lynxstudio.com](mailto:sales@lynxstudio.com)  
Website: <http://www.lynxstudio.com>

© Copyright 2014 Lynx Studio Technology, Inc. All rights reserved. Aurora 8, Aurora 16, Lynx, LS-AES, LS-ADAT, LT-ADAT, LT-HD, LT-USB, LT-TB, Stream, LSlot are trademarks of Lynx Studio Technology, Inc. All other trademarks are property of their respective holders. Designed and manufactured in the USA. All specifications are subject to change without notice. Aurora\_11\_14

Fig. A.8 ADC/DAC converter

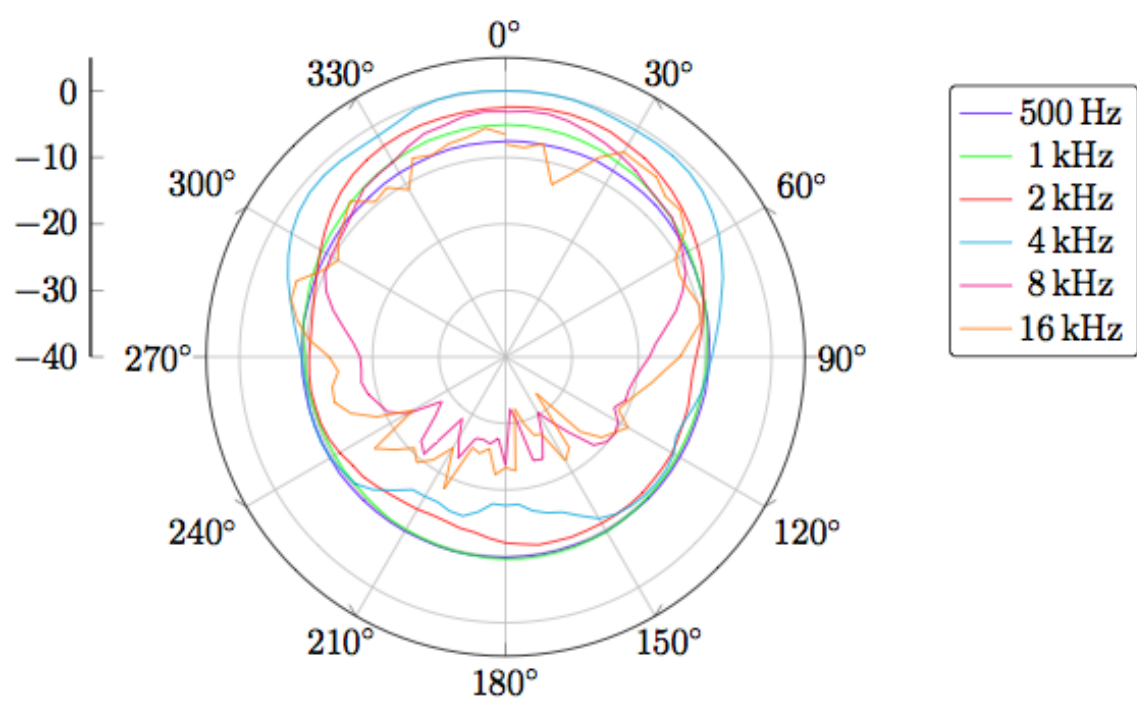


Fig. A.9 Loudspeaker

