Soundfield Imaging in the Ray Space

D. Marković, F. Antonacci, A. Sarti, S. Tubaro

Abstract

In this work we propose a general approach to acoustic scene analysis based on a novel data structure (ray-space image) that encodes the directional plenacoustic function over a line segment (Observation Window, OW). We define and describe a system for acquiring a ray-space image using a microphone array and refer to it as ray-space (or "soundfield") camera. The method consists of acquiring the pseudo-spectra corresponding to a grid of sampling points over the OW, and remapping them onto the ray space, which parameterizes acoustic paths crossing the OW. The resulting ray-space image displays the information gathered by the sensors in such a way that the elements of the acoustic scene (sources and reflectors) will be easy to discern, recognize and extract. The key advantage of this method is that ray-space images, irrespective of the application, are generated by a common (and highly parallelizable) processing layer, and can be processed using methods coming from the extensive literature of pattern analysis. After defining the ideal ray-space image in terms of the directional plenacoustic function, we show how to acquire it using a microphone array. We also discuss resolution and aliasing issues and show two simple examples of applications of ray-space imaging.

I. Introduction

The interest in space-time audio processing algorithms has considerably grown in the past decade. Numerous products, in fact, have appeared in the market, which take advantage of multiple sensors (microphones) to localize, track and extract acoustic sources in space with the purpose of improving their SNR [1] or Signal-to-Reverberation ratio [2]; or of enabling new human-machine interaction mechanisms [3]. These solutions are today widely employed in applications to telecommunications, gaming and entertainment [4]. As the expectations on space-time audio processing algorithms increase, so does the interest in acoustic scene analysis, intended as the process of acquiring geometric and "radiative"

The authors are with the Dipartimento di Elettronica, Informazione e Bioingegneria at Politecnico di Milano, Milano, Italy `dmarkovic/antonacc/sarti/tubaro@elet.polimi.it`

information on acoustic sources (e.g. [5]) and reflectors. A number of recently published environment-aware sound processing algorithms, in fact, exploit the available information on the acoustic scene to boost the performance of space-time processing methods [6], [7], [8]. These solutions rely on acoustic scene analysis for collecting the required information. This is generally done by gathering measurements and combining the related constraints, through a process that is specifically developed for the problem at hand. In this manuscript we follow a different route, which consists of collecting the information that is available on the acoustic scene all at once; organizing it into a data structure that displays it in a ready-to-interpret fashion; and performing the analysis of the collected data afterwards, using various methodologies, typically from pattern analysis.

The method that we propose is inspired by the concept of *plenoptic function* [9], [10], which describes the optical wave field intensity as a function of position and direction (plus time and frequency). Its acoustic counterpart was first introduced in [11], [12] in two different forms: *directional* and *omnidirectional* plenacoustic function. The former mirrored the definition of plenoptic function, whereas the latter dropped the dependency on direction. Optical wavelengths, in fact, are much smaller than sensors and imaged objects, therefore they enable extreme directional selectivity. In acoustics, on the other hand, directivity is always an issue. This is why [11] and [12] decided to work on the omnidirectional definition of the plenacoustic function. In this manuscript we go back to the *directional* definition of plenacoustic function [13], assuming that the directional information will be recovered through space-time processing. We will be working under the hypotheses of geometrical acoustics, as this will allow us to use acoustic rays to describe listening points and *look* directions in a compact and effective representation framework.

If we want to measure the plenacoustic function in a single point, we can do so by centering a microphone array in that location, and estimating (through beamforming) the acoustic radiance along all look directions. A device of this sort is commonly known as “acoustic camera”, and is often based on the computation of a pseudo-spectrum [14] as a function of the direction of arrival. A natural extension of this concept would be that of a “plenacoustic camera”, intended as a theoretical device that acquires the plenacoustic function over a spatially extended “Observation Window” (OW) facing the acoustic scene, as done in the literature of computer vision [15], [16]. In the case of 2D geometry, the OW is a line segment through which the acoustic scene is being “observed”. If the OW became infinite (a whole line), then knowing the plenacoustic function on it would mean knowing it everywhere in space. This is indeed true because of Huygens’ Principle, but it is also true because of the Radiance Invariance Law, which is a fundamental principle of geometrical acoustics that states that the acoustic radiance remains constant along rays. By limiting our knowledge of the plenacoustic function to an OW of finite extension, the plenacoustic function will only be approximately known in space, the approximation depending on visibility and occlusion conditions. Nonetheless, knowing the directional plenacoustic function over a whole OW means gathering a great deal of information on the global acoustic wave field. In this manuscript we elaborate this idea by exploring how this information can be acquired, organized, analyzed and exploited. In order to approach the problem with the necessary progression, as done for example in [11], we address here the 2D case only, as it allows us to define a parameter space (ray space) that can be readily visualized and understood. Although based on the same principles and approach, in fact, the 3D case requires a different set of geometric tools, therefore it deserves to be discussed in a separate manuscript. The 2D case, however, is also relevant for a variety of applications, ranging from source localization and separation to wave field rendering, and is valid for a wide range of enclosures [17].

We are interested in implementing a device that captures the plenacoustic function over an OW based on an array of microphones. One rather straightforward way of doing so is to think of this device as an array of acoustic cameras that sample the OW. The unavoidable compactness of these cameras, however, causes one such device to exhibit severe resolution limitations. This means that this system cannot represent the direct acoustic counterpart of a plenoptic camera. We will, however, introduce a novel parameterization for the domain of the plenacoustic function (ray space), which conveniently displays (as an image) all the elements of the acoustic scene in such a way to facilitate its analysis despite this loss of resolution. The resulting image will be here referred to as “ray-space” image, and the device for capturing it, we will call “ray-space” or “soundfield” camera. With this new parameterization, acoustic primitives such as sources and reflectors, are mapped onto rectilinear features/regions of the ray-space image, which greatly simplifies acoustic scene analysis algorithms. In fact, this allows us to approach space-time processing problems with pattern analysis tools, which are readily available in the rich literature of computer vision and multidimensional signal processing. One other key aspect of this approach is that we are defining a single space-time processing layer (that transforms acoustic acquisitions into ray-space images), which can be shared “as is” by a wide variety of applications. In order to exemplify this aspect, in this manuscript we describe two simple examples of applications: multiple near-field source localization and reflector localization. These problems have been addressed numerous times in the literature. For example a near-field beamforming method for the localization of acoustic sources is proposed in [18]. Reflector localization methods were proposed in [19] and [20]. These, however, were effective ad-hoc solutions devised for the specific problem at hand. We will address such problems with the sole purpose of showing how they can be successfully turned into problems of pattern analysis.

The manuscript is organized as follows: in Section II we define the domain of the ray-space images and show how geometric primitives of interest and acoustic measurements are mapped onto it. Section III describes more in detail the acquisition process of ray-space images. Here we also discuss the impact of spatial sampling and the related issues of resolution and aliasing phenomena. Section IV discusses the two examples of application related to source and reflector estimation. We also show some simulative and experimental results to prove the feasibility of the proposed technique.

II. The Plenacoustic Function and its Parameterization

In this Section we derive a suitable parameterization for sound fields, which serves as a basis for defining the soundfield camera and understanding the structure of the pattern that it captures.

A. The Plenacoustic Function

Quite symmetrically to its optical counterpart, the plenacoustic function can be thought of as a parameterization of the sound field, which is a function that describes the acoustic radiance in every direction through every point in space. This means that, in the case of a 2D geometric domain, it can be written as a function $f(x, y, \theta, \omega, t)$ of position (x, y); direction θ; frequency ω; and time t [11]. In particular, we are interested in the dependency on space (x, y) and direction θ, therefore we simplify the notation by dropping both ω and t. We will specify later in the manuscript whether the dependency from time and frequency is to be considered. Under the hypothesis of validity of geometrical acoustics, expressing the soundfield as a function of the spatial/directional parameters x, y and θ, corresponds to adopting a representation based on acoustic rays.

We recall that (in a homogeneous medium) an acoustic ray is an oriented line that identifies a planar acoustic wavefront component and is inherently perpendicular to it (i.e. it is collinear with the wave vector). A beam of acoustic rays originating from an acoustic source, therefore, identifies an infinite combination of infinitesimal planar wavefront contributions, each identified by a ray that will be locally orthogonal to the wavefront. In geometrical acoustics (just like in optical radiometry) we can rely on the principle of Radiance Invariance Law, which states that the acoustic radiance remains constant along the acoustic path. In fact, the reduction of sound intensity with distance is explained by the fact that the density of acoustic rays per unit area decreases as the receiver moves farther from the source [21]. This is, of course, true in the absence of propagation losses due to absorption, etc. This invariance, in fact, tells us that $f(x, y, \theta)$ has only two degrees of freedom instead of three, which suggests us that we should look for an alternate and more compact parameterization for the soundfield, as done in the optical domain [22],[23]. In the acoustic domain one such parameterizations was introduced in [17],

[24] for defining visibility diagrams and combining them into a data structure that could be iteratively looked up for readily tracing beams of acoustic rays in enclosures. The parameterization that we adopt in this manuscript is designed after that one, as it has already proven effective not just for applications of acoustic modeling, but also for acoustic scene analysis [25] and rendering [26].

B. Parametrization: the ray space

We want to define a compact and simple parameterization for the rays on an Observation Window (OW). As we are interested in defining a soundfield camera, our parameterization will be "one-sided", as it will cover only the rays that cross the OW in just one of the two possible directions. The invariance of the acoustic radiance along the direction of rays, allows us to establish an equivalence between rays and oriented lines that cross that window in the same direction. We therefore need a rule for implicitly and uniquely specifying the orientation of a line given the line parameters.

Let us consider a reference frame positioned in such a way that the OW lies on the y axis between $y = -q_0$ and $y = q_0$. The equation

$$y = mx + q \ , \tag{1}$$

of parameters (m, q), describes any line that is not parallel to the y axis ($|m| < \infty$). This line has two possible directions: one pointing towards the y axis from the "positive" half-space $x > 0$, and one against. As we are interested in defining a soundfield *camera* whose OW lies on the y axis, we conventionally assign the line the orientation towards the y axis from the positive half-space $x > 0$. This allows us to establish an equivalence between rays and lines, which is why we refer to the (m, q) space as the "ray space". From now on, therefore, we will be able to interchangeably talk about rays and lines.

If the space $\mathcal{P}$ of all possible parameters (m, q) covers the rays that point towards the y axis from the positive half-space, the subset of such rays that only "illuminate" the OW lies within the region $\mathcal{V} = \{(m, q) \in \mathcal{P} : -q_0 \leq q \leq q_0\}$, which we call "visibility region" of the OW, as done in [17]. Given an acoustic primitive (a source, a reflector, etc.), we are interested in finding which of the "visible" rays (those in $\mathcal{V}$) are coming from points of that primitive, in order to assess "what" of the radiance produced by that primitive could be picked up by the soundfield camera. This region of the ray space, referred to as the Region Of Interest (ROI) of the primitive, is closely related to the concept of visibility region introduced in [17]. In order to have a better idea of what a ray-space image is expected to look like, let us begin with characterizing the ROIs of some acoustic primitives.

1) Points: A point $\boldsymbol{p} = [\overline{x}, \overline{y}]^T$, $\overline{x} > 0$, can be equivalently thought of as the set of all the lines $\mathbf{r}$ that pass through it. These lines, in fact, identify only those rays that depart from the source and point towards the y axis. The region of the ray space describing the parameters of such lines is called the *dual* [17] $\mathcal{I}_{\boldsymbol{p}}$ of the point $\boldsymbol{p}$ and is represented by the line $q = -\overline{x}m + \overline{y}$. The ROI of $\boldsymbol{p}$ is the set of lines that pass through both $\boldsymbol{p}$ and the OW:

$$\mathcal{R}_{\boldsymbol{p}} = \mathcal{V} \cap \mathcal{I}_{\boldsymbol{p}} = \left\{ \mathbf{r} = [m, q]^T \in \mathcal{V} \;:\; q = -\overline{x}m + \overline{y} \ \right\} \ . \tag{2}$$

As shown in Fig. 1 $\mathcal{R}_{\boldsymbol{p}}$ divides $\mathcal{V}$ in the two regions $\mathcal{V}^+_{\boldsymbol{p}}$ and $\mathcal{V}^-_{\boldsymbol{p}}$. Rays in $\mathcal{V}^+_{\boldsymbol{p}}$ reach the OW after going around $\boldsymbol{p}$ in a clockwise fashion (i.e. while keeping $\boldsymbol{p}$ on their right); while rays in $\mathcal{V}^-_{\boldsymbol{p}}$ fall on the OW after going around $\boldsymbol{p}$ counterclockwise (i.e. while keeping $\boldsymbol{p}$ on their left). A similar definition can be given for the two half-spaces $\mathcal{P}^+_{\boldsymbol{p}}$ and $\mathcal{P}^-_{\boldsymbol{p}}$ that $\mathcal{I}_{\boldsymbol{p}}$ divides the parameter space into. These definitions will come at handy later.

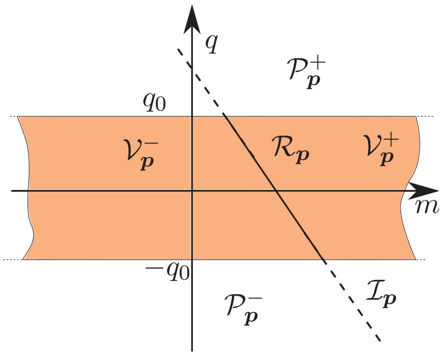

q
$\mathcal{P}^+_{\boldsymbol{p}}$
q_0
$\mathcal{V}^-_{\boldsymbol{p}}$
$\mathcal{R}_{\boldsymbol{p}}$
$\mathcal{V}^+_{\boldsymbol{p}}$
m
$-q_0$
$\mathcal{I}_{\boldsymbol{p}}$
$\mathcal{P}^-_{\boldsymbol{p}}$

Fig. 1. ROI $\mathcal{R}_{\overline{\boldsymbol{x}}}$ of the point $\overline{\boldsymbol{x}}$, and related regions of visibility that this ROI defines on $\mathcal{V}$.

2) Segments: As done for points, the dual $\mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B}$ of a segment $\boldsymbol{p}_A\boldsymbol{p}_B$ is the region of the plane $\mathcal{P}$ corresponding to the set of all lines passing through $\boldsymbol{p}_A\boldsymbol{p}_B$. The ROI $\mathcal{R}_{\boldsymbol{p}_A\boldsymbol{p}_B}$ of the segment $\boldsymbol{p}_A\boldsymbol{p}_B$ is the set of rays that pass through the segment and are, at the same time, visible from the OW:

$$\mathcal{R}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B} \cap \mathcal{V} \ .$$

With reference to Fig. 2, $\mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B}$ can be found by first determining the duals $\mathcal{I}_{\boldsymbol{p}_A}$ and $\mathcal{I}_{\boldsymbol{p}_B}$ of the endpoints, and then by identifying the related half-spaces $\mathcal{P}^+_{\boldsymbol{p}_A}$, $\mathcal{P}^-_{\boldsymbol{p}_A}$, $\mathcal{P}^+_{\boldsymbol{p}_B}$, and $\mathcal{P}^-_{\boldsymbol{p}_B}$ as done above. Such half-spaces allow us to identify the set of rays that cross the segment from one side

$$\mathcal{I}^{(1)}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{P}^-_{\boldsymbol{p}_A} \cap \mathcal{P}^+_{\boldsymbol{p}_B} \ ,$$

which is a wedge-shaped region in the ray space $\mathcal{P}$; or those that cross the segment from the other side

$$\mathcal{I}^{(2)}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{P}^{+}_{\boldsymbol{p}_A} \cap \mathcal{P}^{-}_{\boldsymbol{p}_B} ,$$

which is the opposite wedge to the previous one. All rays that cross the segment are therefore given by the double wedge

$$\mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{I}^{(1)}_{\boldsymbol{p}_A\boldsymbol{p}_B} \cup \mathcal{I}^{(2)}_{\boldsymbol{p}_A\boldsymbol{p}_B} .$$

Correspondingly, the rays that pass through the OW after crossing the segment from one side only are $\mathcal{R}^{(1)}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{I}^{(1)}_{\boldsymbol{p}_A\boldsymbol{p}_B} \cup \mathcal{V}$ and $\mathcal{R}^{(2)}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{I}^{(2)}_{\boldsymbol{p}_A\boldsymbol{p}_B} \cup \mathcal{V}$, respectively. The duals $\mathcal{I}_{\boldsymbol{p}_A}$ and $\mathcal{I}_{\boldsymbol{p}_B}$ of the endpoints of the segment are the lines that delimit the double wedge $\mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B}$ in the ray space. Such lines meet in the point $\mathbf{r}_{AB} \in \mathcal{P}$ of coordinates

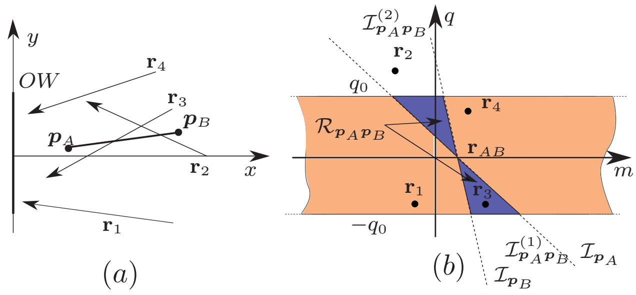

y
OW
$\mathbf{r}_4$
$\mathbf{r}_3$
$\boldsymbol{p}_B$
$\boldsymbol{p}_A$
$\mathbf{r}_2$
x
$\mathbf{r}_1$
(a)
$\mathcal{I}^{(2)}_{\boldsymbol{p}_A\boldsymbol{p}_B}$
q
$\mathbf{r}_2$
q_0
$\mathbf{r}_4$
$\mathcal{R}_{\boldsymbol{p}_A\boldsymbol{p}_B}$
$\mathbf{r}_{AB}$
m
$\mathbf{r}_1$
$\mathbf{r}_3$
$-q_0$
$\mathcal{I}^{(1)}_{\boldsymbol{p}_A\boldsymbol{p}_B}$
$\mathcal{I}_{\boldsymbol{p}_A}$
(b)
$\mathcal{I}_{\boldsymbol{p}_B}$

Fig. 2. A segment in the geometric domain (a) and the corresponding ROI (b). Examples of rays and corresponding points in the ray space.

$$m = \frac{y_A - y_B}{x_A - x_B} , \quad q = -\frac{y_B x_A - y_A x_B}{y_B - y_A} ,$$

which are the parameters of the line that $\boldsymbol{p}_A\boldsymbol{p}_B$ lies upon, corresponding to the side view of the segment.

3) Managing multiple primitives: Let us now consider two acoustic sources $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$ (points) lying on a line that passes through the OW, as shown in Fig. 3(a). In the ray space $\mathcal{P}$ this line corresponds to the point $\bar{\mathbf{r}} = [\overline{m}, \overline{q}]^T$ of intersection between the ROIs $\mathcal{R}_{\boldsymbol{p}_A}$ and $\mathcal{R}_{\boldsymbol{p}_B}$, which exists because the ray $\bar{\mathbf{r}}$ points to the OW. In fact, the ray $\bar{\mathbf{r}}$ is the only direction of observation of the many covered by the OW that sees the sources $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$ aligned.

The situation becomes more complex when we need to account for occlusions. The ROI defined above, in fact, does not do so. For example, let us consider the two acoustic reflectors (segments) of Fig. 4(a). Here the reflector $\boldsymbol{p}_A\boldsymbol{p}_B$ occludes a portion of the rays that depart from $\boldsymbol{p}_C\boldsymbol{p}_D$ and point to the OW. This occlusion results in two overlapping ROIs in the ray space. As $\boldsymbol{p}_A\boldsymbol{p}_B$ occludes $\boldsymbol{p}_C\boldsymbol{p}_D$, $\mathcal{R}_{\boldsymbol{p}_A\boldsymbol{p}_B}$ replaces

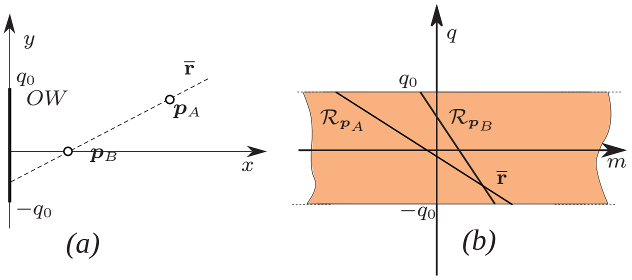


Fig. 3. The sources $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$ in the geometric domain (a) and the corresponding ROIs (b), which generate an overlap.

$\mathcal{R}_{\boldsymbol{p}_C\boldsymbol{p}_D}$ in the overlap. The Region Of Visibility (ROV) of the reflector $\boldsymbol{p}_C\boldsymbol{p}_D$ is a subset of its ROI, after visibility culling, i.e. after removing the portion of ROI occluded by $\mathcal{R}_{\boldsymbol{p}_A\boldsymbol{p}_B}$:

$$\mathcal{R}^{(V)}_{\boldsymbol{p}_C\boldsymbol{p}_D} = \mathcal{R}_{\boldsymbol{p}_C\boldsymbol{p}_D} - \left(\mathcal{R}_{\boldsymbol{p}_A\boldsymbol{p}_B} \cap \mathcal{R}_{\boldsymbol{p}_C\boldsymbol{p}_D}\right) .$$

The reduction of the ROI into a ROV can be similarly defined for reflectors occluding sources or other configurations of the acoustic scene.

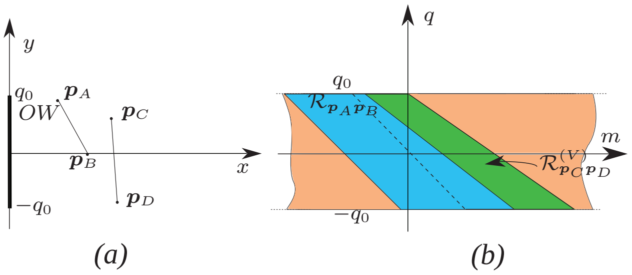


Fig. 4. Reflectors $\boldsymbol{p}_A\boldsymbol{p}_B$ and $\boldsymbol{p}_C\boldsymbol{p}_D$ in the geometric domain (a) and the corresponding ROVs (b). $\mathcal{R}_{\boldsymbol{p}_C\boldsymbol{p}_D}$ is partially occluded by $\mathcal{R}_{\boldsymbol{p}_A\boldsymbol{p}_B}$, therefore the corresponding ROV $\mathcal{R}^{(V)}_{\boldsymbol{p}_C\boldsymbol{p}_D}$ is smaller.

III. Soundfield Images in the Ray Space

We now introduce the concept of ray-space image as the ray-space parameterization of the sound field.

A. The Ideal Soundfield Camera

In order to define a ray-space image, we start from the classical parameterization $f(x, y, \theta)$ of the plenacoustic function discussed in Section I, and map it onto the ray space $\mathcal{P}$ defined in Section II. This mapping is defined by $x = 0$ (the OW is on the y axis); $\theta = \arctan(m)$, $-\pi/2 < \theta < \pi/2$; and $q = y$.

The resulting ray-space image is therefore $p(m,q) = f(0,q,\arctan(m))$. This image carries information on both magnitude and phase of the acoustic radiance, therefore it is generally complex-valued. For reasons that will be clearer later, however, the images that we will work with in this manuscript are power images such as $P(m,q) = |p(m,q)|^2$. Depending on the application, however, phase information can be used as well.

Consider the acoustic scene of Fig. 5(a), where a ray-space camera (OW) acquires direct acoustic paths from a source $\boldsymbol{p}_S$ in direct visibility, as well as acoustic paths that bounce off reflector $\boldsymbol{p}_A\boldsymbol{p}_B$. These reflective paths can be thought of as generated by the source $\boldsymbol{p}_{S'}$, image of $\boldsymbol{p}_S$ on the acoustic "mirror" $\boldsymbol{p}_A\boldsymbol{p}_B$. We immediately notice that the source $\boldsymbol{p}_S$ is "visible" from only some of the points of the OW, due to a partial occlusion on the part of the reflector. Also the image source $\boldsymbol{p}_{S'}$ is only "visible" by a portion of the OW, this time because visibility must be guaranteed "*through* the mirror". It is important to underline that reflectors always act as occluders except for the image sources that they generate, in which case they act as a "window of visibility". The two acoustic beams (i.e. wedges delimited by dashed lines) of Fig. 5(a), one originating from $\boldsymbol{p}_S$ and one from $\boldsymbol{p}_{S'}$, delimit the rays that actually end up on the OW. Those originating from $\boldsymbol{p}_S$ work their way around the reflector while those originating from $\boldsymbol{p}_{S'}$ must pass through the reflector.

Fig. 5(b) illustrates the same situation in the ray space, where the above beams of ray are now visualized as segments. As the points of these segments correspond to the only rays that illuminate the OW, these are the only points where the ideal ray-space image takes on non-zero values. The ROV $\mathcal{R}^{(V)}_{\boldsymbol{p}_S}$ of the source $\boldsymbol{p}_S$ can be readily obtained as

$$\mathcal{R}^{(V)}_{\boldsymbol{p}_S} = \mathcal{I}_{\boldsymbol{p}_S} \cap \overline{\mathcal{I}}_{\boldsymbol{p}_A\boldsymbol{p}_B} \cap \mathcal{V} \ ,$$

where $\overline{\mathcal{I}}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{P} - \mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B}$ is the complementary region of the ROI of the reflector; while the ROV $\mathcal{R}^{(V)}_{\boldsymbol{p}_{S'}}$ of the source $\boldsymbol{p}_{S'}$ is given by

$$\mathcal{R}^{(V)}_{\boldsymbol{p}_{S'}} = \mathcal{I}_{\boldsymbol{p}_{S'}} \cap \mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B} \cap \mathcal{V} \ .$$

The plenacoustic function in these ROVs can be determined using the radiance beampattern $b_{\boldsymbol{p}_S}(\theta)$ of the source, which is the distribution of acoustic radiance produced by the source, as a function of the angle $\theta = \arctan(m)$. The invariance of the acoustic radiance along the ray allows us to write

$$p_{\boldsymbol{p}_S}(m,q) = \begin{cases} b_{\boldsymbol{p}_S}(\arctan(m)) & , \quad (m,q) \in \mathcal{R}^{(V)}_{\boldsymbol{p}_S} \\ 0 & , \quad \text{elsewhere} \end{cases} \ . \tag{3}$$

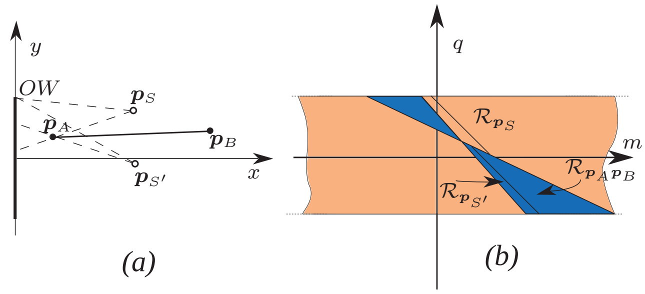

y
OW
$\boldsymbol{p}_S$
$\boldsymbol{p}_A$
$\boldsymbol{p}_B$
$\boldsymbol{p}_{S'}$
x
(a)
q
$\mathcal{R}_{\boldsymbol{p}_S}$
m
$\mathcal{R}_{\boldsymbol{p}_A\boldsymbol{p}_B}$
$\mathcal{R}_{\boldsymbol{p}_{S'}}$
(b)

Fig. 5. A simple acoustic scene “observed” by an ideal soundfield camera (a) and the corresponding representation in the ray space (b).

The contribution of the image source $\boldsymbol{p}_{S'}$ is

$$p_{\boldsymbol{p}_{S'}}(m,q) = \begin{cases} k_{\boldsymbol{p}_A\boldsymbol{p}_B} b_{\boldsymbol{p}_{S'}}(\theta) & , \quad (m,q) \in \mathcal{R}^{(V)}_{\boldsymbol{p}_{S'}} \\ 0 & , \quad \text{elsewhere} \end{cases} , \tag{4}$$

where $b_{\boldsymbol{p}_{S'}}(\theta)$ is the radiance beampattern of $\boldsymbol{p}_{S'}$, which is a specularly reflected version of $b_{\boldsymbol{p}_S}(\theta)$; and $k_{\boldsymbol{p}_A\boldsymbol{p}_B}$ is a function that accounts wall reflection attenuation, a function that generally depends on the parameters (m,q) of the incident ray as well as the frequency. The ray-space image, shown in Fig. 5(b), will be the sum of the two complex contributions (3) and (4). This image is a simplification of reality for a twofold reason: the camera is idealized (no issues of limited resolution or aliasing phenomena); and the scene is idealized (no diffraction or diffusion phenomena). The issues related to the camera will be discussed in the next Section.

When multiple reflectors are present in the acoustic scene, the ray-space image collects numerous contributions, each coming from either a real source or an image source. The computation of the individual contributions of such sources follows a similar approach to that described above. Given an image source of order i, i.e. resulting from i consecutive wall reflections, we first compute its ROV through the intersection of the ROIs of the intermediate reflecting walls and the visibility region $\mathcal{V}$, and then we compute the value of the ray-space image within the ROV as the product between the beam pattern of the image source and the reflection functions of the intermediate reflectors. If we wanted to predict the ray-space image from an object-based description of the acoustic scene, we would need to keep track of the image sources and their visibility, which can be done by using the beam tracing algorithm introduced in [17]. In normal conditions, however, second-order or higher-order reflective paths do not produce relevant contributions to the ray-space image. We can also neglect the impact of diffraction and diffusion, as we assessed from preliminary measurements that these phenomena generate features in the ray-space image whose magnitude is much smaller with respect to echoes associated to direct and first-order reflective paths.

Notice that, as the spatial extension of the OW increases, so does the thickness of the strip $\mathcal{V}$. An infinitely wide OW in fact, could ideally capture the plenacoustic function $p(m, q)$ over the whole ray space $\mathcal{P}$, as discussed in the Introduction.

B. The Real Soundfield Camera

So far we have discussed the ideal soundfield (ray-space) camera and the structure of the images that it captures. We now discuss how to acquire a ray-space image using a microphone array. In principle, just like in the optical domain, the soundfield camera can be thought of as an array of acoustic cameras, placed on a grid that samples the OW. Different setups are possible for this measurement procedure. If the acoustic scene is static and the signal emitted by the sources is stationary, the soundfield could be measured by simply moving an acoustic camera along the OW. If the acoustic scene is not static, then we need to resort to a one-shot acquisition procedure based on a spatially extended microphone array. In order to do so, we can adopt different geometric configurations of microphones. The simplest is a Uniform Linear Array (ULA), partitioned into smaller compact sub-arrays. An alternate configuration that we define in this manuscript is obtained by organizing the microphones in three parallel and staggered linear arrays, which allows us to define a linear and uniform distribution of small hexagonal sub-arrays. Whatever the configuration, we apply beamforming to each sub-array and map the output onto the ray space to form one row of the ray-space image.

The resulting image is in the complex domain. If the application does not require phase information (as in the two examples discussed in this manuscript), the image formation process simplifies. In this case it is convenient to construct the power ray-space image $P(m, q)$: for each location of the array the angular distribution of acoustic power is estimated through the computation of a pseudospectrum [14].

We remark that the use of geometrical acoustics is consistent with the near-field assumption (spatially extended array). In fact, near-field refers to the fact that acoustic sources produce wavefronts that cannot be considered as planar over the whole extension of the array, while they can be confused with planar wavefronts if observed on the (smaller) sub-arrays. Under this condition each sub-array is able to consistently determine the directions of arrival of the sources, as if they were in the far-field. Different sub-arrays, on the other hand, observe the sources under different angles (i.e. from different positions), due to the spherical shape of wavefronts.

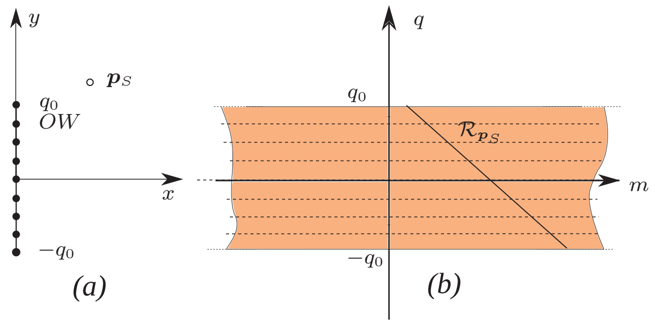


Fig. 6. Implementation of a soundfield camera using a ULA.

1) ULA soundfield camera: Consider the simple setup in Fig. 6. The acoustic source is located in $\boldsymbol{p}_S$ and the microphone array is placed on the y axis, between $y = q_0$ and $y = -q_0$. The ith microphone, in particular, is in $\boldsymbol{m}_i = [0, q_0 - 2q_0(i-1)/(M-1)]^T,\ i = 1, \dots, M$. Let us consider a sub-array centered in $\boldsymbol{m}_i$ (the microphone in $\boldsymbol{m}_i$ is the reference sensor of the sub-array). The sensors in the sub-array are located at $\boldsymbol{m}_j, j = i - \frac{W-1}{2}, \dots, i + \frac{W-1}{2}$, where W is the (odd) number of microphones of the sub-array. The signals acquired by the sensors in the sub-array are $s_j(t), j = i - \frac{W-1}{2}, \dots, i + \frac{W-1}{2}$. In this example of implementation, we chose to process $s_j(t)$ through a wideband Minimum Variance Distortionless Response (MVDR) beamformer [27], although various alternatives could be employed. The first step is to process $s_j(t)$ with a filter bank to obtain $s_j(t, \omega_k), k = 1, \dots, K$, ω_k being the central frequency of the kth sub-band. The signals produced by the filterbank are then stacked into the vector $\mathbf{s}_i(t, \omega_k) = [s_{i-(W-1)/2}(t, \omega_k), \dots, s_{i+(W-1)/2}(t, \omega_k)]^T$, which allows us to compute the autocorrelation matrix

$$\mathbf{R}_{i,k} = \sum_{t=1}^{T} \mathbf{s}_i(t, \omega_k)\mathbf{s}_i(t, \omega_k)^H \ . \tag{5}$$

The MVDR pseudospectrum [27] of the kth sub-band, relative to the sub-array centered on the ith microphone is

$$h_{i,k}(\theta) = \frac{1}{\mathbf{a}^H(\theta, \omega_k)\mathbf{R}_{i,k}^{-1}\mathbf{a}(\theta, \omega_k)} \ , \tag{6}$$

where $\mathbf{a}(\theta, \omega_k)$ is the propagation vector for frequency ω_k and direction θ [28]. The wideband version of the pseudospectrum is obtained as

$$H_i(\theta) = \prod_{k=1}^{K} h_{i,k}(\theta) \ ,\ i = \frac{W+1}{2}, \dots, M - \frac{W+1}{2} \ , \tag{7}$$

i being the index of the subarray. In general, we can choose whether we want to work with separate sub-bands (e.g. in the presence of frequency-dependent reflectors), or with wideband MVDR pseudospectra.

Whatever the choice, pseudospectra must be mapped onto the ray space. We recall that the pseudospectrum $H_i(\theta)$ measures the power distribution of rays passing through the location $\boldsymbol{m}_i$ of the ith microphone. An acoustic ray passing through such point at an angle θ has parameters

$$\begin{aligned} m &= \tan(\theta) \\ q_i &= q_0 - 2q_0 \tfrac{i-1}{M-1} \ , \end{aligned} \tag{8}$$

therefore we can write

$$\widetilde{P}(m, q_i) = H_i(\arctan(m)) \ , \tag{9}$$

where $i = (W+1)/2, \ldots, M-(W+1)/2$. The scanlines $q = q_i$ are the dashed ones in Fig. 6(b). The real ray-space image $\widetilde{P}(m, q_i)$ that we obtain will differ from what we would obtain with an ideal soundfield camera (see Subsection III-A) for a twofold reason: it is sampled along q (due to the limited number of subarrays); and it is blurred along m (due to the limited number of sensors in each subarray). We will see that, given a total number of microphones, increasing the sampling density along q and increasing the resolution along m are two contrasting needs.

Fig. 8(a) shows an acoustic scene that includes a ULA of 15 microphones spaced of 0.11 m. An acoustic source placed in $\boldsymbol{p}_S = [1, 1]^T$ produces a pass-band signal whose spectrum lies between 300 Hz and 10 kHz. The corresponding simulated ray-space image is shown in Fig. 8(b). For clarity of visualization, the resulting image is displayed after order-zero interpolation (piecewise constant) with respect to q (along m the number of samples is very large). The dashed line of Fig. 8(b) is the dual $\mathcal{I}_{\boldsymbol{p}_S}$ of the source, i.e. the representation of the source in the ray space. As we can see, the ray-space image $\widetilde{P}(m, q)$ exhibits a ridge in the same location as $\mathcal{I}_{\boldsymbol{p}_S}$. This ridge is, in fact, a blurred version of the visible portion of the dual of the source and the magnitude of the blurring varies with both q and m. This is due to the fact that a ULA (subarray) does not exhibit a uniform resolution over θ [28]. As we can see in Fig. 8(b), in particular, the farther the point from the source, the larger the incidence angle, the greater the blurring. This loss of resolution could prevent us from being able to tell multiple acoustic objects apart when they lie too close to each other.

In the lower left area of the ray-space image in Fig. 8(b) we also notice a rather large bright area, caused by aliasing. The signal emitted by the source, in fact, has frequency content that goes beyond the spatial Nyquist frequency. This phenomenon and its impact will be better characterized later on in the manuscript.

2) Quincunx soundfield camera: In order to address, at least in part, the problems of resolution seen with ULA cameras, we consider here a different configuration of microphones. The Quincunx

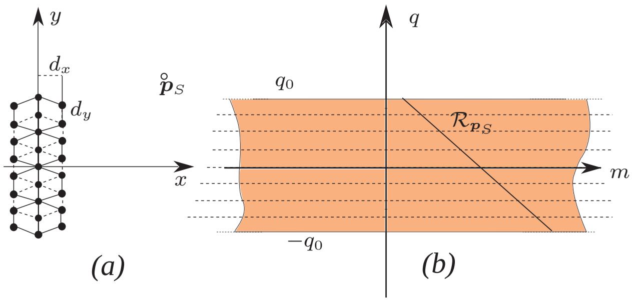
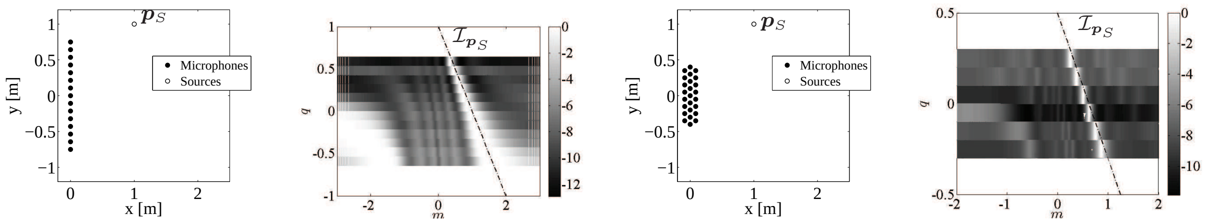


Fig. 7. Implementation of a soundfield camera based on three staggered parallel ULAs. Single sub-arrays are marked with alternate solid and dashed strokes.

(or hexagonal) array of Fig. 7(a) is formed by three parallel and staggered ULAs of N, $N+1$ and N microphones, respectively. The former lies on the line $x = -d_x$; the second in $x = 0$; and the latter in $x = d_x$. Distances between adjacent sensors are d_x and d_y. This setup allows us to group microphones into $N-1$ hexagonal sub-arrays, as shown in Fig. 7. If $d_x = \sqrt{3}/2d_y$, then the hexagons become regular, and the subarrays become Uniform Circular Arrays (UCA). Pseudospectra $H_i(\theta)$, $i = 2, \ldots, N-1$ are computed for each sub-array and the ray-space image is obtained as in eq. (9). UCAs are known for offering a more uniform resolution than ULAs [28], we can therefore expect this camera to introduce more of a uniform blurring throughout the image. This improvement, however, comes at a cost. The number of pseudospectra $H_i(\theta)$ contributing to the ray-space image, in fact, is $N-1$, whereas the number of sensors is $3N+1$. For example, with 23 microphones we can only extract seven pseudospectra, therefore the ray-space image $\widetilde{P}(m, q_i)$, is made of 7 rows. If we want to build a ray-space image of 7 rows with a ULA camera, we need a minimum of 9 microphones (7 lapped subarrays of $W = 3$ microphones each). Fig. 8(c) shows an acoustic scene acquired with a quincunx camera. The source, as in the previous example, is in $\boldsymbol{p}_S = [1, 1]^T$ and produces a signal with a pass-band spectrum ranging from 300 Hz to 10 kHz Hz. Fig. 8(d) shows the corresponding ray-space image. Notice that the resolution is now quite uniform throughout the ray space and aliasing issues are much more under control with respect to the case of ULA cameras.

C. Angular Aliasing

Aliasing is a well known phenomenon in space-time processing, which causes an error in the localization of the acoustic source. An aliased pseudospectrum exhibits multiple lobes of comparable magnitude, known as grating lobes [28], which are replicas of the main lobe. In order to prevent aliasing, the distance



(a) Geometric space (b) $\widetilde{P}(m,q)$[dB] (c) Geometric space (d) $\widetilde{P}(m,q)$[dB]

Fig. 8. An acoustic scene with a ULA camera (a) and with a quincunx camera (c) for a source at coordinates $\boldsymbol{p}_S = [1,1]^T$ and the corresponding ray-space images $\widetilde{P}(m,q)$ (b) and (d), respectively. The amplitude has been normalized and expressed in a dB scale.

d between adjacent sensors needs to be kept sufficiently small.

As far as ULAs are concerned, the no-alias condition is $d < \lambda/2$, where λ is the wavelength corresponding to the maximum frequency contained in the signal and d is the distance between adjacent sensors. We focus here on the impact of spatial aliasing on ray-space images. The presence of aliasing depends only on the deployment of sensors. As a closed-form characterization of aliasing for MVDR is not possible, we present analytical results for the case of delay-and-sum beamformer [28], which can be applied with some approximation to the MVDR.

Let us consider the ith subarray, whose central microphone is in $\boldsymbol{m}_i = [0, q_0 - 2q_0(i-1)/(M-1)]^T$. The angle under which this sub-array sees the acoustic source in $\boldsymbol{p}_S = [x_S, y_S]^T$ is

$$\theta_i = \arctan\left(\frac{q_0 - 2q_0(i-1)/(M-1) - y_S}{-x_S}\right) .$$

The acoustic source produces a single tone of wavelength λ. For the delay-and-sum beamformer, the contribution of the sub-array to the ray-space image is

$$H_i(\theta) = C\frac{\sin\left[\frac{\pi W d}{\lambda}(\sin\theta - \sin\theta_i)\right]^2}{\sin\left[\frac{\pi d}{\lambda}(\sin\theta - \sin\theta_i)\right]^2} , \tag{10}$$

where C is a positive constant. Spatial aliasing occurs when the denominator is zero, i.e. when

$$\frac{\pi d}{\lambda}(\sin\theta - \sin\theta_i) = l\pi,\ l \in \mathbb{Z} ,$$

which gives

$$\theta = \arcsin\left(\frac{l\lambda}{d} + \sin\theta_i\right) \ , -\pi/2 \leq \theta < pi/2 \ , l \in \mathbb{Z} \tag{11}$$

Fig. (9) shows the same ray-space image of Fig. 8(b). Small crosses mark the location of aliasing peaks, as detected with a peak-picking algorithm. In this figure a solid curve marks the location of the grating

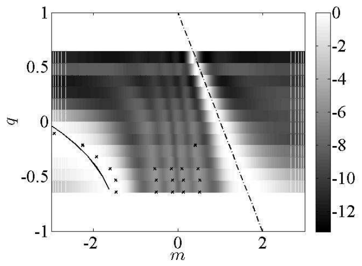


Fig. 9. Example of ray-space image with aliasing. The geometric setup is the same as in Fig. 8(a). The dashed line is the dual of the source; crosses mark the detected aliasing peaks; the continuous line, computed from (11) approximately predicts the location of aliasing peaks.

lobes as predicted with eq. (11), with $l = -1$. As we can see, although derived for the delay-and-sum beamformer, the curve well approximates the location of the grating lobes also in the case of MVDR beamforming. We also notice that the grating lobes form on the image a nonlinear pattern. This means that any line detection tool such as the Hough transform would allow us to easily discriminate between peaks related to real sources and peaks due to grating lobes.

In the case of the quincunx camera the derivation of the alias contribution of each subarray to the ray-space image is more complex [28] and goes beyond the scope of this manuscript. It is worth mentioning, however, that the no-aliasing condition in this case becomes

$$d_y \leq \frac{5\lambda}{4\pi} \; . \tag{12}$$

D. Resolution

The resolution is defined as the minimum angular distance between two sources that makes the related pseudospectrum peaks discernible. The discriminating ability depends on the adopted peak-picking algorithm, therefore it is more of an operative definition. In this manuscript the resolution is evaluated by sizing the width of the lobe of the pseudospectrum corresponding to the direction of arrival of a point-like source, i.e. we characterize it as a point-spread function (on scan lines of the ray-space image). Let m_{max} be the value of m corresponding to the Direction Of Arrival (DOA) of the source and $\delta m^+ > 0$ be the interval on the m axis such that

$$\widetilde{P}(m_{\text{max}} + \delta m^+, q_i)|_{\text{dB}} = \widetilde{P}(m_{\text{max}}, q_i)|_{\text{dB}} - \Delta \; ,$$

Δ being a given threshold. Similarly we define $\delta m^- < 0$ as the value of m such that

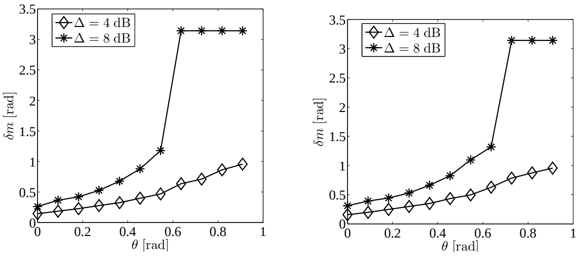
$$\widetilde{P}(m_{\text{max}} + \delta m^-, q_i)|_{\text{dB}} = \widetilde{P}(m_{\text{max}}, q_i)|_{\text{dB}} - \Delta \; .$$

Finally, we define the width of the lobe as

$$\delta m = \delta m^{+} - \delta m^{-} \ , \tag{13}$$

which clearly depends on Δ. For the applications presented in this manuscript (see next Section) we preliminarily verified that the peak picking algorithm adopted in this manuscript requires $\Delta \geq 4$ dB in order to discriminate between peaks in the pseudospectra associated to multiple sources.

A closed-form expression of δm can only be found for the delay-and-sum beamformer, and not for MVDR beamformer. This is why we performed simulations. Fig. 10 plots δm for $W = 3$ and $W = 5$ for various source positions. More specifically, the source is placed at a distance of 1.5 m, and the angle θ formed by the source and the x axis varies between $0°$ and $45°$. The microphone array has the same configuration of Fig. 8(a). For visualization convenience, the value of δm has been converted in angles, as the range of variability of m is too large for a clear representation. Notice that for $\Delta = 8$ dB there are angles for which $\delta m|_{\text{rad}} = \pi$. In this situation δm^{+} is not defined, as the lobe of $\widetilde{P}(m, q_i)|_{\text{dB}}$ does not decrease to $\widetilde{P}(m_{\text{max}}, q_i)|_{\text{dB}} - \Delta$ for $m > m_{\text{max}}$. Notice that there is no significant improvement on the resolution for $W = 3$ and $W = 5$ when $\Delta = 4$ dB. As for the quincunx camera, the resolution turns out to be almost constant as the angle θ formed by the source and the x axis varies. When the angle of the source ranges from $0°$ to $90°$ the width of the lobe for $\Delta = 4$ dB oscillates between $3°$ and $13°$.



(a) $W = 3$ (b) $W = 5$

Fig. 10. δm as a function of the angle of the source for the ULA camera.

To summarize, resolution is a critical factor for deciding which array configuration to use. The quincunx camera has relevant advantages over the ULA camera when the source is viewed from a particularly disadvantaged angle (e.g. $\theta > \pi/4$).

E. Computational complexity

In this subsection we aim at providing an upper-bound estimation of the cost, in terms of multiplication and memory accesses, for the generation of a ray-space image. We identify five steps performed to generate the image. First the signals acquired by each microphone $s_j(t), j = 1, \ldots, M$ are divided in T frames made by N samples. FFT is applied to each frame to obtain $s_j(t, \omega_k), t = 1, \ldots, T, k = 1, \ldots, K$ (step 1). The autocorrelation matrix $\mathbf{R}_{i,k}$ of eq. (5) is computed (step 2) and inverted (step 3) for each subarray and each sub-band. The total number of subarrays is $S = M - W + 1$, where M is the total number of microphones and W is the number of microphones per subarray. The MVDR pseudospectra $h_{i,k}(\theta_l)$ are computed for directions $\theta_l, l = 1, \ldots, L$, as in eq. (6) (step 4). Finally, the wideband pseudospectra $H_i(\theta_l)$ are obtained as in eq. (7) (step 5). For each of these steps we estimate the number of multiplications with accumulation (MACs) and of the accesses to the memory (AMs). The results are shown in Table I. Notice, however, that the image generation is a highly-parallelizable application-independent pre-processing step to optimize which several solutions can be devised [29], [30].

TABLE I

ESTIMATED COST FOR THE GENERATION OF A RAY-SPACE IMAGE.

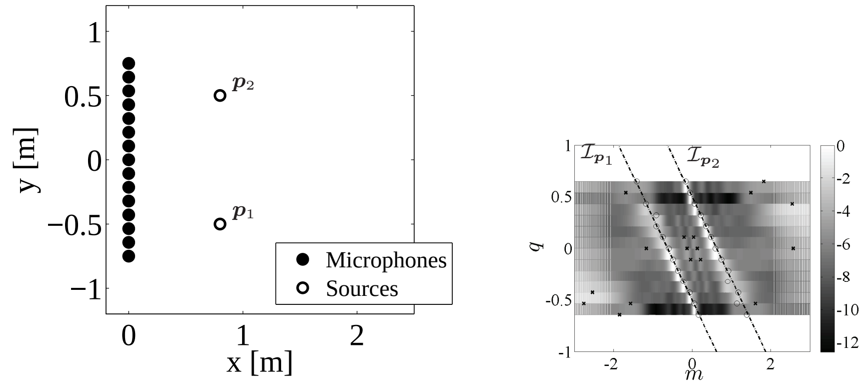
	MACs	AMs
STEP 1	$MT\frac{N}{2}\log_2 N$	MTN
STEP 2	SKW^2T	$SKWT$
STEP 3	$SK(\frac{1}{2}W^3 + \frac{3}{2}W^2 + W)$	SKW^2
STEP 4	$SKL(W^2 + W + 1)$	$SKL(W^2 + 2W)$
STEP 5	SLK	SLK

IV. EXAMPLES OF APPLICATION

In order to illustrate how to analyze and interpret the information gathered by the soundfield cameras defined in Section III, we now discuss two examples of applications: one focusing on the localization of multiple sources and one discussing the localization of reflectors.

A. Multiple Source Localization

Let us consider the problem of localizing multiple acoustic sources with the ULA and the quincunx cameras. The first step is the disambiguation of measurements (DOAs, TOAs, TDOAs) obtained from the arrays and their matching to the corresponding sources. Disambiguation of TDOAs, for example, can be performed as in [31]. A method for matching measurements and sources is proposed in [32], based on a Guassian likelihood function. When using ray-space imaging, the disambiguation and pairing of information is greatly simplified because the ray-space representation of the plenacoustic function enables the clustering of DOAs of the same source on linear patterns. Consider, for example, the setup shown in Fig. 11(a). The acoustic scene consists of two sources, placed in $\boldsymbol{p}_1 = [0.8 \text{ m}, -0.5 \text{ m}]^T$ and $\boldsymbol{p}_2 = [0.8 \text{ m}, 0.5 \text{ m}]^T$, and of a ULA of $N = 15$ sensors. The acquired ray-space image is shown in Fig. 11(b). The dashed lines represent the duals $\mathcal{I}_{\boldsymbol{p}_1}$ and $\mathcal{I}_{\boldsymbol{p}_2}$ of the sources. Circles mark the peaks of the pseudospectra (horizontal rows) corresponding to the two sources. Crosses are located in correspondence of secondary peaks. In order to localize multiple sources we need to distinguish between useful and spurious peaks in the ray-space image and, at the same time, assign useful peaks to one of the corresponding sources. This can be readily accomplished using a Hough transform ([33], [34]) on the ray-space image. The Hough transform, in fact, detects collinear local maxima and finds the parameters of the related lines, which become an estimate of the sources. In order to achieve sufficient accuracy in source localization, however, we need the grid density of the Hough map to be prohibitively large. This is why the Hough transform is here used only to find a first approximation of the source locations, which allows us to assign the peaks to the corresponding sources. Better estimates of the source locations can then be obtained through linear regression over measurements of the same source. Notice also that the Hough transform is not necessary if we focus on the single-source case in no-aliasing and moderately noisy conditions (no grating lobes and spurious peaks).



(a) Setup

(b) ray-space image

Fig. 11. Geometry of an acoustic scene with two sources and corresponding ray-space image.

Through the Hough transform we first obtain the approximate coordinates $(\overline{x}_j, \overline{y}_j)$ of the N_S sources $\boldsymbol{p}_j$, $j = 1, \ldots, N_S$ and determine the set I_j of indices that identify the rows of the ray-space image where the source j is present (visible). For each source $\boldsymbol{p}_j$, we then determine the set of maxima (one per row) on the image, which are matched to that source

$$\mathcal{L}_j = \left\{ (m_i, q_i) : \frac{\left| m_i \overline{x}_j - \overline{y}_j + q_i \right|}{\sqrt{1+m^2}} < \epsilon, \; i \in I_j \right\} , \tag{14}$$

where the index i identifies the subarray (row of the ray-space image) and ϵ is an appropriate threshold. Notice that the number of sources N_S could either be known in advance or be estimated by the Hough transform itself. Notice also that I_j can be used for estimating of the ROV of the source to be localized. Fig. 11(b) shows the detected lines in the case of two sources. Peaks belonging to $\mathcal{L}_j,\ j = 1, 2$ are marked with small circles, while outliers are marked with crosses.

Now that we have associated the maxima to the corresponding source, we can find a better estimate of the location of the sources using a least-squares technique. Let us consider an acoustic source in $\boldsymbol{p}_j = [x_j, y_j]^T$. From eq. (1) we know that all rays departing from $\boldsymbol{p}_j$ must satisfy the constraint $mx_j - y_j + q = 0$, which can be rewritten as $\mathbf{h}^T \boldsymbol{p}_j = -q$, where $\mathbf{h} = [m, -1]^T$. For each set of maxima $\mathcal{L}_j$ we can therefore define the system of equations

$$\begin{cases} \mathbf{h}_{i_1} \boldsymbol{p}_j & = -q_{i_1} \\ & \vdots \\ \mathbf{h}_{i_{N(j)}} \boldsymbol{p}_j & = -q_{i_{N(j)}} \end{cases} , \tag{15}$$

where the subscripts $i_1, \ldots, i_{N(j)}$ are the indices in I_j. Equation (15) can be written in the matrix form

$$\mathbf{H} \boldsymbol{p}_j = \mathbf{q} , \tag{16}$$

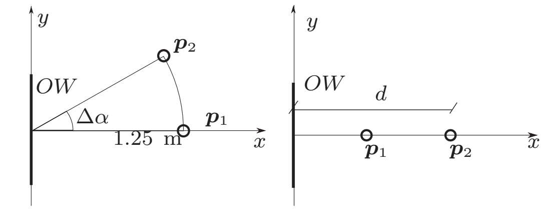
where $\mathbf{H} = [\mathbf{h}_{i_1} \ \ldots \ \mathbf{h}_{i_{N(j)}}]^T$ and $\mathbf{q} = [-q_{i_1} \ \ldots \ - q_{i_{N(j)}}]^T$. We find $\boldsymbol{p}_j$ using least squares, i.e.

$$\hat{\boldsymbol{p}}_j = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{q} . \tag{17}$$

The localization procedure is repeated for all the sets $\mathcal{L}_j$.

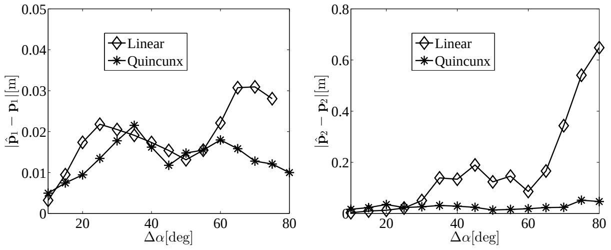
As we can see, source localization and, in particular, the problem of disambiguating measurements and matching them with sources is here turned into a pattern analysis problem performed on an image. The fact that the patterns are rectilinear, turns the localization algorithm into that of solving a system of linear equations, which is quite a desirable feature.

In order to numerically assess the accuracy of the above source localization algorithm we performed three simulations and three real acoustic experiments. The setup of the first simulation is shown in Fig. 12(a). The two sources, the first in $\boldsymbol{p}_1 = [1.25, 0]^T$ and the second in $\boldsymbol{p}_2 = [1.25\cos(\Delta\alpha), 1.25\sin(\Delta\alpha)]^T$, both expressed in meters, produce independent noises in the vocal bandwidth $(300 \div 4000\text{Hz})$. The signal acquired by the sensors is affected by an additive gaussian error with a SNR of 10 dB. The localization experiment is repeated for each location of $\boldsymbol{p}_2$ 100 times, each with a different noise realization. Simulations have been performed with both ULA and quincunx cameras, adopting the deployment of sensors shown in Fig. 8(a) and Fig. 8(c), respectively. Figs. 13(a) and (b) show the localization error of $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, respectively, as $\Delta\alpha$ varies from $10°$ to $80°$. The error on $\boldsymbol{p}_1$ is nearly constant and no noticeable difference between ULA and quincunx cameras can be noticed. As for $\boldsymbol{p}_2$, the quincunx camera, due to the higher resolution on the m axis, guarantees an almost constant localization error as $\Delta\alpha$ varies. On the other hand, the localization of $\boldsymbol{p}_2$ is possible using the ULA camera only for $\Delta\alpha \leq 60°$. Beyond that angle the resolution loss on the m axis becomes too relevant to guarantee a correct localization of peaks in $\widetilde{P}(m,q)$.



(a) (b)

Fig. 12. Setup for the simulations to assess the accuracy of multiple source localization.

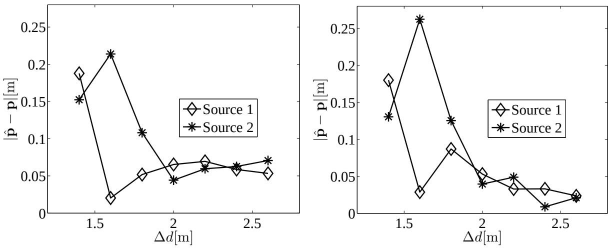


(a) $\boldsymbol{p}_1$ (b) $\boldsymbol{p}_2$

Fig. 13. Localization error of $\boldsymbol{p}_1$ (a) and $\boldsymbol{p}_2$ (b) for linear and quincunx arrays as a function of the angular difference $\Delta\alpha$ depicted in Fig.12(a). Notice that scales are different.

In the following simulations we focus on the same configuration that we employed for the real-world experiments, a ULA array of 16 microphones spaced of 0.06 m. Fig. 12(b) shows the setup of the second simulation. The sources are aligned on the line $y = 0$ m. In particular the first source is in $\boldsymbol{p}_1 = [1 \text{ m}, 0 \text{ m}]^T$, and the second one is in $\boldsymbol{p}_2 = [d, 0 \text{ m}]^T$, where d ranges between 1.4 m and 2.6 m. Notice that in this setup the ROIs $\mathcal{R}_{\boldsymbol{p}_1}$ and $\mathcal{R}_{\boldsymbol{p}_2}$ of the two sources meet in $q = 0$, which makes the localization more challenging due to mutual occlusion between sources. Localization results are shown in Fig. 14(a). If the sources are close to each other the corresponding lines on the ray-space image are not distinguishable due to resolution limits, which results in a higher localization error. However, the technique guarantees a good localization accuracy when $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are not too close to each other. The error on $\boldsymbol{p}_2$ increases slightly as it moves far away from the array, as the limited size of the observation window, compared to d, reduces the localization performance.

We also conducted experiments to verify the accuracy of the algorithm on real-world data. All the experiments are conducted in a low-reverberation room. The first experiment follows the setup of Fig. 12(b). Localization results are shown in Fig. 14 (b). As seen in the above simulations, also in this case the localization improves as the distance between $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ increases.

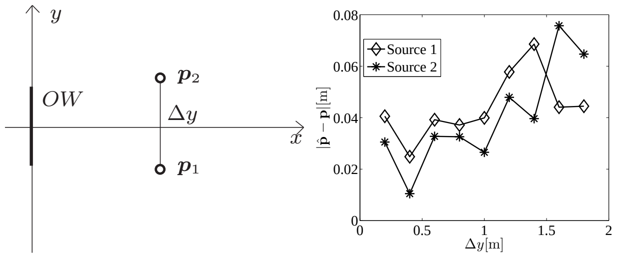


(a) Simulations (b) Experiments

Fig. 14. Simulation and experimental results for the setup shown in Fig. 12(b) for the linear array.

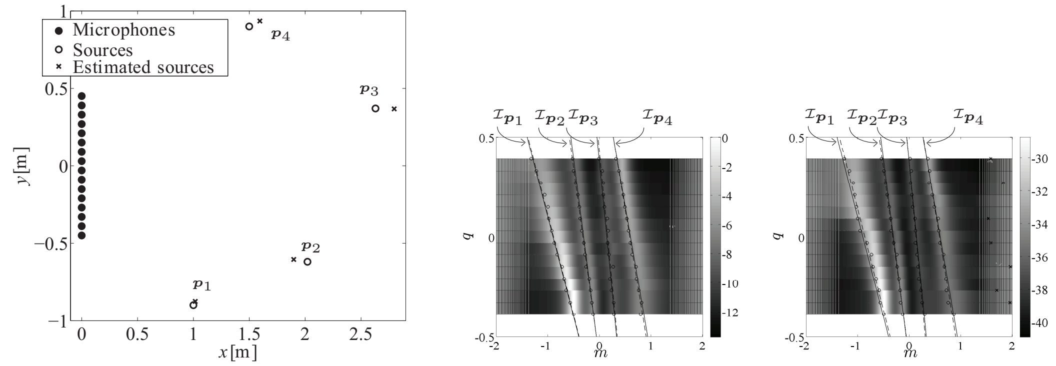
Fig. 15(a) shows the setup of the second experiment. Two acoustic sources are placed in $\boldsymbol{p}_1 = [1.5\ \mathrm{m}, \Delta y/2]^T$ (coordinates in meters) and $\boldsymbol{p}_2 = [1.5\ \mathrm{m}, -\Delta y/2]^T$, respectively. The distance Δy between sources ranges from $0.2\ \mathrm{m}$ to $1.8\ \mathrm{m}$. Results in Fig. 15(b) show that an accurate estimate is obtained even for $\Delta y = 0.2\ \mathrm{m}$, i.e. when the sources are very close to each other. As Δy increases the estimation error first diminishes and then increases again due to resolution loss.

In the third experiment we tested the system in a more challenging scenario of four acoustic sources. The setup and the estimated source positions are shown in Fig. 16(a). Fig. 16(c) shows the acquired ray-space image. In order to assess how well real data match simulative data, we performed a simulation for the same scenario. Fig. 16(b) shows the simulated ray-space image. The algorithm is able to correctly discriminate between contributions of different sources and estimate their positions due to the fact that the corresponding peaks naturally cluster on lines on the ray-space image, as shown in Figs. 16(b) and 16(c), thus enabling an accurate localization with both real-world and simulated data. The average localization error of the four sources is $0.1052\ \mathrm{m}$ and $0.1151\ \mathrm{m}$ for the real-world and simulated data, respectively.



(a) (b)

Fig. 15. Setup (a) and results of the second experiment (b). Two sources lie on a line that is parallel to the y axis at a varying distance Δy.



(a) (b) (c)

Fig. 16. Geometric domain (a), modeled ray-space image (b) and acquired ray-space image (c) with four sources present in the acoustic scene.

In the last simulation we test the robustness of the algorithm at different values of Signal-to-Noise Ratio. The Table II shows the average localization error and standard deviation for a single source placed in $\boldsymbol{p}_1 = [1\ \mathrm{m}, 0\ \mathrm{m}]^T$. What rules the ability of the algorithm to correctly localize the source is the fact that the peak of the pseudospectrum related to the source is distinguishable from the background noise. Under the acceptable assumption of spatially white noise, the computation of the pseudospectrum turns out to concentrate the energy of the signal at the Direction Of Arrival of the source, while spreading the energy of the noise at all the possible DOAs. As a consequence, the algorithm is robust against additive noise and the localization is still possible even at $-20\ \mathrm{dB}$. Furthermore, if the additive noise does not produce spurious peaks of magnitude comparable to the peaks related to the signal and does not alter their location, only small deviations in the localization can be observed as shown in Table II for higher values of SNR.

TABLE II

LOCALIZATION ERROR $|\hat{\mathbf{p}} - \mathbf{p}|$[M] AT DIFFERENT VALUES OF SNR: AVERAGE ERROR AND STANDARD DEVIATION.

SNR [dB]	**average error** [m]	**standard deviation** [m]
20	0.0016	0.0002
10	0.0017	0.0004
0	0.0017	0.0007
-10	0.0027	0.0018
-20	0.1406	0.5537
-30	1.8115	1.5839

B. *Reflector Localization*

Consider the acoustic scene of Fig. 17 which has a source in $\boldsymbol{p}_S$ and an acoustic reflector $\boldsymbol{p}_A\boldsymbol{p}_B$. In this case the array senses not only the contribution of the direct path from $\boldsymbol{p}_S$, but also the echo associated to the reflective path coming from the image source $\boldsymbol{p}_{S'}$. Notice that the line that the reflector lies on is the axis of the segment $\boldsymbol{p}_S\boldsymbol{p}_{S'}$. This means that by localizing $\boldsymbol{p}_S$ and $\boldsymbol{p}_{S'}$ we also localize the line that the reflector lies on.

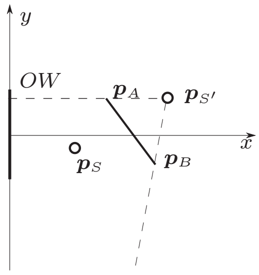


Fig. 17. A reflector causes a reflective path from the image source $\boldsymbol{p}_{S'}$ to be sensed by the microphone array.

The reflector localization procedure is, in principle, similar to the multiple source localization problem. However, as the reflective signal is a delayed replica of the direct signal, they are strongly correlated, which reduces the resolution of $\widetilde{P}(m, q_i)$ with respect to the case of multiple independent sources, thus affecting the localization accuracy. We also notice in Fig. 17 that $\boldsymbol{p}_{S'}$ can only be sensed by the portion of the array that falls within the reflective beam from $\boldsymbol{p}_{S'}$, delimited by dashed lines. We can therefore expect that the estimation of $\boldsymbol{p}_{S'}$ will suffer a loss of accuracy of some degree. The line joining the estimated sources $\hat{\boldsymbol{p}}_S$ and $\hat{\boldsymbol{p}}_{S'}$ has parameters

$$\overline{m} = \frac{y_S - y_{S'}}{x_S - x_{S'}} \ , \quad \overline{q} = y_S - x_S \frac{y_S - y_{S'}}{x_S - x_{S'}} \ .$$

The line that the reflector lies on is therefore given by

$$\widetilde{m} = -\frac{1}{\overline{m}} \ , \quad \widetilde{q} = \frac{1}{2}[y_S - y_{S'} + \frac{1}{\overline{m}}(y_S - y_{S'})] \tag{18}$$

We tested the accuracy of the reflector localization algorithm through simulations based on the setup of Fig. 18. The source is in $\boldsymbol{p}_S = [0.5 \text{ m}, 0 \text{ m}]^T$, and the reflector is at a distance D from the y axis, which ranges from 0.6 m to 1.5 m, and it is parallel to it. The source produces a white noise within the bandwidth $(300 \div 4000\text{Hz})$ with a SNR of 20 dB. We conducted the test using the ULA of Fig. 8(a), as a quincunx camera with a small number of microphones is not suitable for this setup, as shown in the previous paragraph. Figs. 19(a) and 19(b) plot the localization error of $\boldsymbol{p}_S$ and $\boldsymbol{p}_{S'}$, respectively. Figs. 19(c) and 19(d) plot the error on the distance and the angle of the estimated reflector with respect to the actual one, respectively. The error on $\boldsymbol{p}_S$ is nearly constant for all the distances. A different situation arises for $\boldsymbol{p}_{S'}$. In fact, when D is below 0.7 m, $\boldsymbol{p}_S$ and $\boldsymbol{p}_{S'}$ are close each other, and the algorithm exhibits a poor accuracy in localizing $\boldsymbol{p}_{S'}$. For intermediate distances the localization error decreases. If D is above 1.2 m the error on $\boldsymbol{p}_{S'}$ becomes larger, due to the limited extension of the array with respect to D.

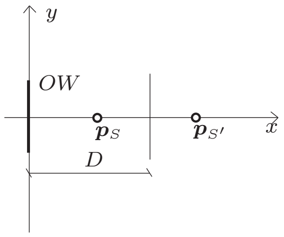
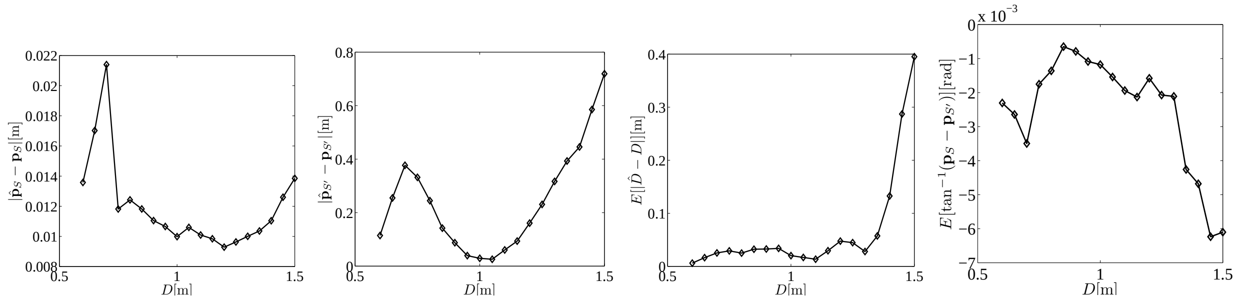


Fig. 18. Setup of the simulation for the reflector localization.

We also conducted an experiment to verify the accuracy of the algorithm on real-world data. Setup and results are shown in Fig. 20, based on a ULA camera of 13 microphones. The source, placed in $\boldsymbol{p}_S = [0.8, 0.62]^T$, (expressed in m) produces a white noise within the bandwidth $(300 \div 4000\text{Hz})$. The actual location of the reflector is marked by the solid segment, while the dashed line represents the estimated line that the reflector lies on. Stars and circles mark the estimated and actual locations of the direct and image sources, respectively. If we look at Fig. 20(a), we notice that the direct rays are sensed from all viewpoints within the OW. The image source, on the other hand, is only visible from those

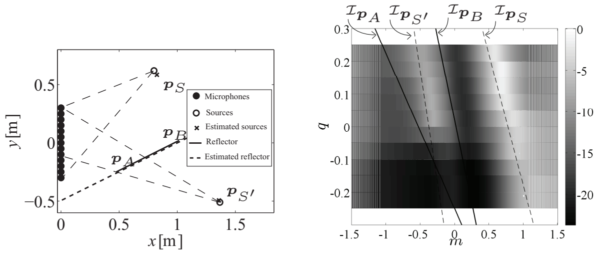


(a) Direct source $\boldsymbol{p}_S$ (b) Image source $\boldsymbol{p}_{S'}$ (c) Distance error (d) Angle error

Fig. 19. Localization error of source and image source and of the reflector for the setup in Figure 18

points of the OW that are "illuminated" by the beam that originates from $\boldsymbol{p}_{S'}$ and crosses the reflector $\boldsymbol{p}_A\boldsymbol{p}_B$. This situation is described in a dual fashion on the ray-space image of Fig. 20(b). Here we can see two dashed lines, corresponding to the dual of the source $\mathcal{I}_{\boldsymbol{p}_S}$ and of the image source $\mathcal{I}_{\boldsymbol{p}_{S'}}$. Of $\mathcal{I}_{\boldsymbol{p}_{S'}}$ we can only image its ROV, which is given by the intersection between $\mathcal{I}_{\boldsymbol{p}_{S'}}$ and the ROI of the reflector, delimited by the solid lines $\mathcal{I}_{\boldsymbol{p}_A}$ and $\mathcal{I}_{\boldsymbol{p}_B}$.

The resulting localization error of the direct and image sources are of 4.5 cm and approximately 1 cm, respectively, for the image source. This difference is due to the fact that the direct source is angled with respect to the array, while the image one is almost frontal.



(a) (b)

Fig. 20. Reflector localization. Continuous and dashed lines are the reflector and the estimated lying lines of the reflector, respectively.

V. Conclusions

In this manuscript we proposed a novel approach to acoustic scene analysis based on the concept of ray-space imaging. We first defined the soundfield camera as a device that captures the acoustic radiance along all the acoustic rays that cross an observation window. After discussing the process of ideal ray-space image formation we introduced two implementations based on arrays of different geometries, and discussed their behavior in terms of resolution and aliasing.

We found ray-space imaging to be a powerful analysis paradigm for multiple reasons

- it turns problems of acoustic analysis in space and time into problems of pattern analysis on images, which can be approached with methods found in the rich literature of pattern analysis and multidimensional signal processing;
- image generation becomes a pre-processing step that remains the same throughout a wide range of applications and is highly parallelizable, thus paving the way to the production of a shared hardware framework;
- objects in the acoustic scene correspond to the image patterns that are easily discerned and modeled, which simplifies pattern analysis/detection/extraction. Our definition of the ray space, in particular, makes such patterns linear, with clear advantages in terms of detection performance.

The experiments presented in this manuscript have the purpose of offering an initial proof of concept of these points and will be further explored and expanded in future works.

Indeed, the larger the number of microphones of the array, the greater the detail in the acquired images. Recent progress in MEMS and integrated electronics technology suggests that the number of microphones that can be managed in integrated arrays is on a growing trend. Ray-space imaging can therefore become particularly useful for managing and organizing the massive data that such devices will be able to collect. In the meantime, the two examples that we discussed in this manuscript show that even a limited number of microphones can be useful and very informative.

We believe that this approach to the analysis could enable the development of novel solutions for a wide class of applications such as wave field analysis/extrapolation; image fusion; image-based self-calibration; source separation; environment inference, etc. We are, in fact, currently working on these applications with encouraging results.

References

[1] D. Jarrett and E. Habets, “On the noise reduction performance of a spherical harmonic domain tradeoff beamformer,” *IEEE Signal Processing Letters*, vol. 19, no. 11, pp. 773–776, Nov. 2012.

[2] E. Habets, S. Gannot, I. Cohen, and P. Sommen, “Joint dereverberation and residual echo suppression of speech signals in noisy environments,” *IEEE Tr. on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1433–1451, Nov. 2008.

[3] W. Herbordt, H. Buchner, and W. Kellermann, “An acoustic human-machine front-end for multimedia applications,” *European Journal on Applied Signal Processing*, vol. 2003, no. 1, pp. 1–11, January 2003.

[4] M. Peltola, T. Lokki, and L. Savioja, “Augmented reality audio for location-based games,” in *Audio Engineering Society Conference: 35th International Conference: Audio for Games*, 2 2009.

[5] A. Brutti, M. Omologo, and P. Svaizer, “An environment aware ml estimation of acoustic radiation pattern with distributed microphone pairs,” *Signal Processing*, vol. 93, no. 4, pp. 784–796, 2013.

[6] T. Betlehem and T. D. Abhayapala, “Theory and design of sound field reproduction in reverberant rooms,” *J. of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, 2005.

[7] F. Ribeiro, C. Zhang, D. Florencio, and D. Ba, “Using reverberation to improve range and elevation discrimination for small array sound source localization,” *IEEE Tr. on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1781–1792, Sept. 2010.

[8] A. Canclini, D. Markovic, F. Antonacci, A. Sarti, and S. Tubaro, “A room-compensated virtual surround system exploiting early reflections in a reverberant room,” in *proc. of European Signal Processing Conference (EUSIPCO)*, 2012, pp. 1–5.

[9] E. H. Adelson and J. R. Bergen, *The Plenoptic Function and the Elements of Early Vision*. MIT Press, 1991, pp. 3–20.

[10] E. H. Adelson and J. Y. A. Wang, “Single lens stereo with a plenoptic camera,” *IEEE Tr. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, Feb. 1992.

[11] T. Ajdler, L. Sbaiz, and M. Vetterli, “The plenacoustic function and its sampling,” *IEEE Tr. on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.

[12] T. Ajdler and M. Vetterli, “The Plenacoustic Function, Sampling and Reconstruction,” in *IEEE Conf. on Acoustics, Speech and Signal Processing*, vol. 5, 2003, pp. 616–619.

[13] D. Marković, G. Sandrini, F. Antonacci, A. Sarti, and S.Tubaro, “Plenacoustic Imaging in the Ray Space,” in *proc. of IWAENC 2012 - Intl. Workshop on Acoustic Signal Enhancement*, Aachen, Germany, Sept. 2012, pp. 8–12.

[14] P. Stoica and R. Moses, *Spectral analysis of signals*. Prentice Hall, 2005.

[15] C. Zhang and T. Chen, “A survey on image-based rendering representation, sampling and compression,” *Signal Processing: Image Communication*, vol. 19, no. 1, pp. 1–28, 2004.

[16] A. Gelman, J. Berens, and P. Dragotti, “Layer-based sparse representation of multiview images,” *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 61, pp. 1–15, Mar. 2013.

[17] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, “Fast tracing of acoustic beams and paths through visibility lookup,” *IEEE Tr. on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 812–824, May 2008.

[18] E. Fisher and B. Rafaely, “Near-field spherical microphone array processing with radial filtering,” *IEEE Tr. on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 256–265, Feb. 2011.

[19] H. Sun, W. Kellermann, E. Mabande, and K. Kowalczyk, “Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing,” *J. Acoust. Soc. Am. (JASA)*, vol. 131, no. 4, pp. 2828–2840, April 2012.

[20] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, and S. Tubaro, “From direction of arrival estimates to localization of planar reflectors in a two dimensional geometry,” in *2011 IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 2620–2623.

[21] F. Everest and K. Pohlmann, *Master Handbook of Acoustics*. McGraw-Hill Education, 2009.

[22] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *Proc. of the 23rd annual conf. on Computer graphics and interactive techniques*, ser. SIGGRAPH ’96. New York, NY, USA: ACM, 1996, pp. 43–54.

[23] M. Levoy and P. Hanrahan, “Light field rendering,” in *Proc. of the 23rd annual conf. on Computer graphics and interactive techniques*, ser. SIGGRAPH ’96. New York, NY, USA: ACM, 1996, pp. 31–42.

[24] F. Antonacci, A. Sarti, and S. Tubaro, "Two-dimensional beam-tracing from visibility diagrams for real-time acoustic rendering," *EURASIP Journal on Advances in Signal Processing*, pp. 1–18, 2010.

[25] D. Markovic, C. Hofmann, F. Antonacci, K. Kowalczyk, A. Sarti, and W. Kellermann, "Reflection coefficient estimation by pseudospectrum matching," *proc. of Intl. Workshop on Acoustic Signal Enhancement, IWAENC 2012*, Sept. 2012.

[26] D. Marković, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Visibility-based beam tracing for soundfield rendering," in *2010 IEEE Intl. Workshop on Multimedia Signal Processing (MMSP)*, Saint Malo, France, Oct. 2010, pp. 40–45.

[27] M. Azimi-Sadjadi, A. Pezeshki, and N. Roseveare, "Wideband DOA estimation algorithms for multiple moving sources using unattended acoustic sensors," *IEEE Tr. on Aerospace and Electronic Systems*, vol. 44, no. 4, pp. 1585–1599, 2008.

[28] H. V. Trees and J. Wiley, *Optimum array processing*. Wiley–Interscience, 2002.

[29] F. P. Ribeiro and V. Nascimento, "Fast transforms for acoustic imaging - part I: Theory," *IEEE Tr. on Image Processing*, vol. 20, no. 8, pp. 2229–2240, 2011.

[30] ——, "Fast transforms for acoustic imaging - part II: Applications," *IEEE Tr. on Image Processing*, vol. 20, no. 8, pp. 2241–2247, 2011.

[31] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Tr. on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1479–1489, Nov. 2008.

[32] A. Lombard, Y. Zheng, H. Buchner, and W. Kellermann, "TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis," *IEEE Tr. on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1490 – 150, August 2011.

[33] P. V. C. Hough, "Method and means for recognizing complex patterns," USA Patent 3 069 654, December 18, 1962.

[34] R. Duda and P. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

Dejan Marković was born in 1985. He received the MSc degree cum laude in 2009 and the PhD degree cum laude in 2013 from the Politecnico di Milano, Italy.

Since January 2013 he is a Post Doctoral Research Assistant at Image and Sound Processing Group (ISPG), Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano. His research activity is mainly focused on analysis, geometric modeling and microphone array processing of acoustic wavefields for multimedia applications.

Fabio Antonacci was born in Bari (Italy) on July 26, 1979. He received Laurea degree in 2004 in Telecommunication Engineering and Ph.D. in Information Engineering in 2008, both at Politecnico di Milano, Italy.

He is currently a post-doc researcher at Image and Sound Processing Group in Dipartimento di Elettronica, Informazione e Bioingegneria at Politecnico di Milano, Italy. His research focuses on space-time processing of audio signals, for both speaker and microphone arrays (source localization, acoustic scene analysis, rendering of spatial sound) and on modeling of acoustic propagation (visibility-based beam tracing).

He published approximately 50 articles in proceedings of international conferences and on peer-reviewed journals.

Augusto Sarti (M ’04) received the M.S. and the Ph.D. degrees in electronic engineering, both from the University of Padua, Italy, in 1988 and 1993, respectively, with research on nonlinear communication systems. His graduate studies included a joint graduate program with the University of California at Berkeley, where he worked on nonlinear system theory.

In 1993, he joined the Dipartimento di Elettronica e Informazione of the Politecnico di Milano, Milan, Italy, where he is currently an Associate Professor. His research interests are in the area of digital signal processing, with particular focus on sound analysis, processing and synthesis; space-time audio processing; geometrical acoustics; music information retrieval. He also worked on problems of multidimensional signal processing, vision-based 3D scene reconstruction; camera calibration; image analysis; motion planning and nonlinear system theory.

He coauthored over 180 scientific publications on international journals and congresses as well as numerous patents in the multimedia signal processing area. He coordinates the Sound and Music Computing Lab of the Image and Sound Processing Group of the Politecnico di Milano. He promoted and coordinated or contributed to numerous (20+) EC-funded project. He is a member of the IEEE Technical Committee on Audio and Acoustics Signal Processing, and Associate Editor of IEEE Signal Processing Letters. He has served as guest editor for numerous special issues of international journals. He was co-chairman of the 2005 Edition of the IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS); Chairman of 2009 edition of the Digital Audio Effects conference, (DAFx); and in the organizing committees of numerous other conferences in the area of signal processing.

Stefano Tubaro (M'01) was born in Novara, Italy, in 1957. He received his Electronic Engineering degree at the Politecnico di Milano, Milano, Italy, in 1982. He then joined the Dipartimento di Elettronica e Informazione of the Politecnico di Milano, first as a researcher of the National Research Council; then as an Associate Professor (1991) and finally as a Full Professor (2004).

He initially worked on problems related to speech analysis; motion estimation/compensation for video analysis/coding; and vector quantization applied to hybrid video coding. In the past few years, his research interests have focused on image and video analysis for the geometric and radiometric modeling of 3-D scenes; and advanced algorithms for video coding and sound processing. He has authored over 150 scientific publications on international journals and congresses. He co-authored two books on digital processing of video sequences. He also co-authored several patents on image processing techniques.

He coordinates the research activities of the Image and Sound Processing Group (ISPG) at the Dipartimento di Elettronica e Informazione of the Politecnico di Milano; which is involved in several research programs funded by industrial partners, the Italian Government, and by the European Commission. He has been involved in the IEEE Technical Committee of Multimedia Signal Processing (2005-2009), and is currently involved in that of Image Video and Multidimensional Signal Processing.