Multiview Soundfield Imaging in the Projective Ray Space

Dejan Marković, *Member, IEEE*, Fabio Antonacci, *Member, IEEE*, Augusto Sarti, *Senior Member, IEEE*, and Stefano Tubaro, *Senior Member, IEEE*

I. Introduction

RECONSTRUCTING an acoustic scene through sound analysis is a goal that has captivated the research community for decades. The literature is rich with solutions that use arrays of microphones for capturing, analyzing and characterizing the objects that the acoustic scene is made of. Numerous algorithms have been proposed for acoustic source localization and tracking [1], [2], [3], [4], [5]; as well as for localizing reflectors from measurements of TOA [6], [7], TDOA [8], [9] and DOA [10], [11], [12]. More recently, a novel soundfield imaging method has been proposed [13], which is inspired by the concept of plenoptic analysis [14], [15]. Its goal is to capture the acoustic counterpart of the plenoptic function, the directional plenacoustic function [16], [17], defined as the contribution of the sound field at a given position coming from a given direction. The spatial region in which the plenacoustic function is measured is called "Observation Window" (OW). The result is a new image-like representation of the sound field. The generation of the soundfield image does not require other than standard array processing techniques, widely used for many different tasks.

The key advantage of soundfield images is in the fact that they gather and organize at once and in a single representation all the information that we need in order to develop a wide range of acoustic scene analysis applications. The generation of the soundfield image is, therefore, a highly parallelizable application-independent processing stage that could be easily implemented in hardware form. The application-dependent processing stage that follows takes advantage of the fact that the objects that constitute the acoustic scene are always mapped onto our soundfield representation as spatially extended linear features which are easier to extract and identify using algorithms taken from the wide literature of pattern and image analysis. The image of such features carries an inherent representational redundancy that is exploited to improve accuracy. With this approach, the soundfield image can be used to facilitate various tasks that entail the localization and the analysis of acoustic sources in the scene: multiple acoustic source localization, tracking and separation; estimation of the room geometry; estimation of the radiance pattern of a source; estimation of reflection properties of walls. Furthermore, the range of possible applications of the soundfield imaging is not limited to acoustic scene analysis. These techniques could also be applied for wave field rendering [18], [19].

The soundfield imaging method proposed in [13] works with planar geometry, therefore the OW of the acoustic scene is a simple line segment. The device that captures the soundfield image is the soundfield camera and it is implemented by an array of microphones subdivided into overlapped sub-arrays. Using a beamforming technique each sub-array estimates the plenacoustic function in its center, i.e. the acoustic image from that position. Each acoustic image becomes one row of the soundfield image as shown in Fig. 1. The domain of the soundfield image is called "ray space" because each point in the domain corresponds to a ray crossing the OW. More specifically, rays are identified by the parameters of the line on which they lie, referred to a frame that is attached to the OW itself. In particular, the OW lies on the y axis with the origin in its middle. The image coordinates are therefore the slope m and the intercept q with the y axis of this frame. As the equation of the line does not specify a direction, the ray will be conventionally assumed as crossing the OW in only one of the two possible directions. This means that half of all possible rays as well as the rays that

Manuscript received December 12, 2014; revised March 02, 2015; accepted March 25, 2015. Date of publication April 01, 2015; date of current version April 15, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Rongshan Yu.

The authors are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milano, Italy (e-mail: dejan.markovic@polimi.it; fabio.antonacci@polimi.it; augusto.sarti@polimi.it; stefano.tubaro@polimi.it).

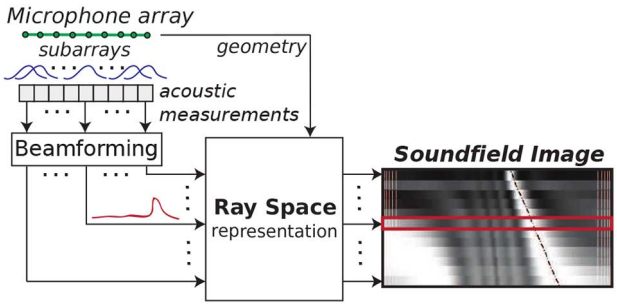

Fig. 1. Creating the soundfield image.

are parallel to the OW cannot be parameterized. This representational limitation is not so crucial in the case of a single soundfield camera, since rays parallel to the y axis cannot be sensed. However, this limitation becomes overly restricting when the goal is to jointly use multiple soundfield cameras and we have to represent all possible rays.

In this paper we redefine the representation adopted in [13] by introducing a new ray parameterization that enables the simultaneous use of multiple soundfield cameras in different locations that observe the same acoustic scene. This more general setup poses new challenges but, as we will show in this paper, brings relevant benefits to applications of soundfield imaging. The parameterization of rays that we define in this paper is projective, as it uses three homogeneous (scalable) parameters to describe the line on which the ray lies. In particular, the representation will be based on "oriented projective geometry", in order to keep track not just of orientations but also of directions of rays. We will see that a wide range of geometric transformations of interest, such as change of reference frame, projection, reflection, etc., can be represented as linear transformations (homographies) [20] using the projective parameterization.

Depending on the geometric configuration of the array that implements the soundfield camera, the soundfield image exhibits a loss of resolution (blurring) that changes from ray to ray, according to its orientation with respect to the array. We will show that one immediate advantage of defining a ray space that does not have to be "attached" to a specific OW is that we can better control the loss of resolution as all the rays are equally visible in the newly defined representation. This fact has important benefits on the applications of soundfield imaging. In particular, we will show that the accuracy of source localization from soundfield images captured from multiple smaller OW's greatly improves with respect to the single-OW case. The fact that observing the scene from different viewpoints improves the ability to assess the distance of a source, is well-known in the literature. Examples of methodologies that exploit this very fact are Global Coherence Field [21], Steered Response Power and variations thereof [22] [23]. Despite the change of representation, we will see that the advantages of the previous approach described in [13] for source localization are fully preserved. Acoustic primitives are still mapped onto linear (planar) patterns, therefore the localization of multiple sources consists in finding such linear patterns in the soundfield images and inferring their parameters. The projective parameterization also allows us to address the problem of determining the mutual locations of the soundfield cameras from acoustic measurements, known in the literature as self-calibration, and approached in various ways [24][25][26]. In this paper we approach self-calibration as the estimation of the homographies that map the reference frames of the individual cameras onto the global reference frame.

The rest of the paper is organized as follows: Section II describes the projective ray space; its relationship with the Euclidean ray space of [13]; and the homographies of the most typical camera configurations. Section III describes the process of acquisition of soundfield images, from both theoretical and implementation standpoints. Section IV discusses two possible applications of multi-view soundfield imaging, namely that of self-calibration of arrays and that of source localization. In order to validate the proposed methodologies, we show the results of an extensive simulation campaign and experiments on real data. Finally, Section V draws some conclusions and offers a perspective on the next research steps.

II. The Projective Ray Space

In order to capture the plenacoustic function, in this paper we consider two dimensional geometries, i.e. microphone arrays lie on the same plane of the acoustic primitives in the scene. The planar arrays are practical to implement and, at the same time, retain validity for a variety of applications, e.g. tracking and separation of acoustic sources, acquisition of signals for data-based rendering, etc. The parameterization of rays in 3D requires the adoption of a different ray space, for example based on Plucker coordinates [27]. Such parameterizations, however, work in projective spaces. The results achieved in this paper, therefore, can be considered as a prerequisite for soundfield imaging in 3D space. Let us consider a region of space $\mathcal{V} \subset \mathbb{R}^2$ that is free of scatterers, and let us denote with $\boldsymbol{x} = [x, y]^T$ a point in $\mathcal{V}$. In planar geometry, the sound-field can be written as $p(\boldsymbol{x}, \omega, t)$, but since we are particularly interested in the dependency on position (x, y), in what follows we will omit t. The sound field $p(\boldsymbol{x}, \omega)$ is a solution of the homogeneous Helmholtz equation

$$\nabla^2 p(\boldsymbol{x}, \omega) + \|\mathbf{k}\|^2 p(\boldsymbol{x}, \omega) = 0, \quad \forall \boldsymbol{x} \in \mathcal{V}, \tag{1}$$

where $\mathbf{k}$ is the wavenumber oriented as the propagation vector and $\|\mathbf{k}\| = \omega/c$. A widely accepted decomposition for an arbitrary solution of (1) requires the explicit modeling of every directional contribution to the sound field at point $\boldsymbol{x}$, i.e. the sound field is modeled as a superposition of plane waves with wavenumbers $\hat{\mathbf{k}}(\theta)$. This representation is known as Plane-Wave Decomposition [28]

$$p(\mathbf{x}, \omega) = \frac{1}{2\pi} \int_0^{2\pi} exp(j\langle \boldsymbol{x}, \hat{\mathbf{k}}(\theta) \rangle) \phi(\theta, \omega) d\theta, \tag{2}$$

where the integration is taken over the contour of a notional unit circle [29]. The term $\phi(\theta, \omega)$ is a complex-valued function that modulates each plane wave component in amplitude and phase. This function is usually referred to as *spatial spectrum* [30] or *Herglotz density function* [31], [32].

We are interested in estimating the contribution from direction θ to the sound field in $\boldsymbol{x}$ at frequency ω. At this purpose, the *Plenacoustic Function* is defined as the integrand in (2), i.e.

$$f(x, y, \theta, \omega) \triangleq e^{j\frac{\omega}{c}\langle \boldsymbol{x}, \hat{\mathbf{k}}(\theta) \rangle} \phi(\theta, \omega). \tag{3}$$

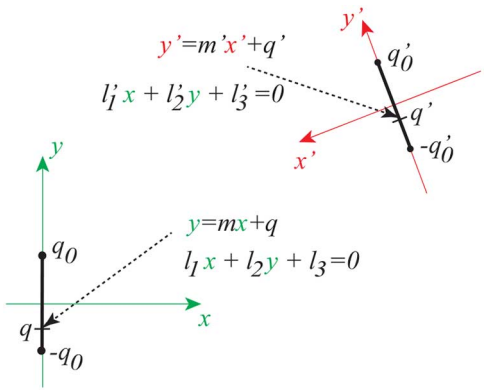

Fig. 2. The local (Euclidean) and the global (projective) parameterization of rays. The two OWs have reference frames attached to them, (x, y) and (x', y'). The Euclidean parameterization of rays (m, q) is defined with respect to the local reference frame. The line incident on the second OW, $y' = m'x' + q'$, cannot be represented in the reference frame of the first camera (x, y) as it does not allow the representation of rays with directions coming from the negative half-space $x < 0$. Using the projective parameterization of rays (l_1, l_2, l_3) we can use any reference frame as a global reference frame and therefore we are able to write the line incident on the second OW in the (x, y) frame as $l'_1 x + l'_2 y + l_3 = 0$.

In what follows, where not differently specified, we also omit the dependency of the plenacoustic function on ω. We remark that the estimation of the directional contributions of the sound field from array processing is not novel in the literature. In fact, it has been shown in [33], [34], [29] that the plane-wave components of the sound field in a point $\mathbf{x}$ can be estimated through beamforming. On the other hand, in [13], the output of multiple beamformers is used for estimating and representing the modulus of the plenacoustic function over an extended observation window in a suitable fashion.

Under the hypotheses of validity of geometrical acoustics [35], the plenacoustic function can be thought of and expressed as a function of the acoustic rays. An acoustic ray is an oriented line that identifies a planar wavefront with wavenumber $\mathbf{k}(\theta)$, and is locally orthogonal to it. A beam of acoustic rays fanning out of an acoustic source, therefore, identifies an infinite combination of infinitesimal planar wavefront contributions. One key fact of geometrical acoustics is that we can rely on the Radiance Invariance Law (RIL) [36] to reduce the dimensionality of our representation. The RIL states that the acoustic radiance (the absolute value of $f(x, y, \theta)$) remains constant along the acoustic path, which implies that the planar plenacoustic function only has two degrees of freedom instead of three. In the next paragraph we shortly summarize the parameterization adopted in [13], identifying its drawbacks when multiple observation windows are in use. With reference to Fig. 2, the y axis is aligned with the OW with the origin in the middle (the OW is between $-q_0$ and q_0). A ray crossing the OW is represented by the line of equation $y = mx + q$, with parameters m and q, and a crossing direction. We conventionally assign to all lines the direction coming from the positive half-space $x > 0$. This Euclidean definition of the ray space (m, q) has the advantage of being simple as well as compatible with the parameterization defined in [37], [38]. However, it has the disadvantage of being OW-dependent (local) and exhibiting "blind spots". This is why we need to generalize it in order to render it suitable for multi-OW operation.

A. Parameterization

As mentioned above, we want to define a parameterization that accommodates all possible rays irrespective of the reference frame of choice. Through this parameterization we must be able to tell which way a ray is pointing, so that we can easily work with one-directional OWs. Oriented projective geometry [39] allows us to do that.

The equation of a line referred to any reference frame is $l_1 x + l_2 y + l_3 = 0$, which can be written in vector form as

$$\mathbf{p}^T \mathbf{l} = 0, \quad \mathbf{l} = [l_1, l_2, l_3]^T, \quad \mathbf{p} = [x, y, 1]^T. \tag{4}$$

This representation is homogeneous as all vectors of the form $\mathbf{l} = k[l_1, l_2, l_3]^T, k \neq 0$ represent the same line, and therefore they form a class of equivalence (projective space). The homogeneous coordinates $\mathbf{l}$ are suitable for describing rays with arbitrary orientation. In order to distinguish between two oppositely directed rays, all we need to do is limit the range of the scaling factor k to either positive or negative values only. This is how we define the oriented projective space $\mathbb{T}^2$ [39]. As a generic point $[l_1, l_2, l_3]^T$ corresponds to a ray in the geometric space, the oriented projective space will be here referred to as the *Projective Ray Space* $\mathcal{P}$.

In order to simplify the visualization of this ray space we will often reduce its dimensionality by slicing $\mathcal{P}$ with a prescribed plane. The resulting section is called in the paper *Reduced Ray Space*. The choice of the plane that we use for slicing it, however, must be made in such a way that all the rays crossing the OW's in the right direction are visible in the resulting reduced ray space.

B. Acoustic Primitives in the Projective Ray Space

1) Rays: As discussed above, an acoustic ray in the geometric space is a projective point in $\mathcal{P}$, visualized as a half-line of coordinates $k[l_1, l_2, l_3]^T, k > 0$ passing through the origin in the ray space.

2) Sources: An acoustic source is a point of "outward" orientation in the geometric space, as it can be thought of as the set of all possible rays that originate from it. Let us consider a source of coordinates $\mathbf{p}_A = k[x_A, y_A, 1], k > 0$. A ray passes through this point if and only if

$$\mathbf{p}_A^T \mathbf{l} = 0. \tag{5}$$

The parameters of all such rays are therefore given by the set $\mathcal{I}_{\boldsymbol{p}_A} = \{\mathbf{l} \in \mathcal{P} | \mathbf{p}_A^T \mathbf{l} = 0\}$, which in the following will be referred to as the image of $\mathbf{p}_A$. In the projective ray space, the image of this point corresponds to a projective line, which is visualized as a plane passing through the origin (the parameters of this plane are the projective coordinates $\mathbf{p}_A$). This plane divides $\mathcal{P}$ into the two half-spaces $\mathcal{P}^+_{\boldsymbol{p}_A} = \{\mathbf{l} \in \mathcal{P} | \mathbf{p}_A^T \mathbf{l} > 0\}$ and $\mathcal{P}^-_{\boldsymbol{p}_A} = \{\mathbf{l} \in \mathcal{P} | \mathbf{p}_A^T \mathbf{l} < 0\}$. The former corresponds to all the rays that pass by $\boldsymbol{p}_A$ leaving it on their left, whereas the latter identifies those that leave $\boldsymbol{p}_A$ on their right. Fig. 3 shows $\mathcal{I}_{\boldsymbol{p}_A}$, $\mathcal{P}^+_{\boldsymbol{p}_A}$ and $\mathcal{P}^-_{\boldsymbol{p}_A}$ in the reduced ray space.

3) Segments: Oriented segments are important elements of the acoustic scene because they model both observation windows and acoustic reflectors. Consider a segment whose endpoints are $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$, as shown in Fig. 4(a). The ray space rep-

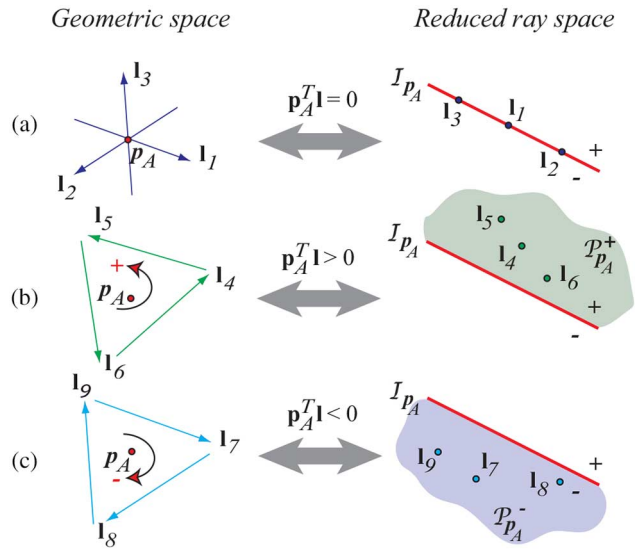

Fig. 3. Representation $\mathcal{I}_{\boldsymbol{p}_A}$ of a point $\boldsymbol{p}_A$ in the reduced ray space and half-spaces $\mathcal{P}^+_{\boldsymbol{p}_A}$ and $\mathcal{P}^-_{\boldsymbol{p}_A}$ in the reduced ray space.

resentation of this segment is shown in Fig. 4(b). As we can see, the images of $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$ are the lines $\mathcal{I}_{\mathbf{p}_A}$ and $\mathcal{I}_{\mathbf{p}_B}$, respectively. We want to be able to tell which direction a ray is crossing the window $\boldsymbol{p}_A\boldsymbol{p}_B$. We can do so by keeping track of those that leave $\mathbf{p}_A$ on their left and $\mathbf{p}_B$ on the right (rays of type 1, e.g. $\mathbf{l}_1$ in Fig. 4) and those that leave $\mathbf{p}_A$ on the right and $\mathbf{p}_B$ on the left (rays of type 2, e.g. $\mathbf{l}_4$ in Fig. 4). Recalling the definition of the half-spaces $\mathcal{P}^+_{\boldsymbol{p}}$ and $\mathcal{P}^-_{\boldsymbol{p}}$ for a generic point $\boldsymbol{p}$ given above, it is quite easy to verify that the image of rays of type 1 is

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

whereas the image of rays of type 2 is

$$\mathcal{I}^-_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{P}^-_{\boldsymbol{p}_A} \cap \mathcal{P}^+_{\boldsymbol{p}_B}.$$

The images $\mathcal{I}^+_{\boldsymbol{p}_A\mathbf{p}_B}$ and $\mathcal{I}^-_{\boldsymbol{p}_A\mathbf{p}_B}$ are wedge-shaped regions, bounded by the planes $\mathcal{I}_{\mathbf{p}_A}$ and $\mathcal{I}_{\mathbf{p}_B}$, which meet in $\mathbf{l}_{\boldsymbol{p}_A\boldsymbol{p}_B}$. In the geometric space, the ray $\mathbf{l}_{\boldsymbol{p}_A\boldsymbol{p}_B}$ corresponds to the oriented line that passes through the segment $\boldsymbol{p}_A\boldsymbol{p}_B$. In the example of Fig. 4, the rays $\mathbf{l}_1$ and $\mathbf{l}_4$ are in $\mathcal{I}^+_{\boldsymbol{p}_A\boldsymbol{p}_B}$ and $\mathcal{I}^-_{\boldsymbol{p}_A\boldsymbol{p}_B}$, respectively. The other two rays, $\mathbf{l}_2$ and $\mathbf{l}_3$, do not cross the OW, therefore they lie outside of both $\mathcal{I}^+_{\boldsymbol{p}_A\boldsymbol{p}_B}$ and $\mathcal{I}^-_{\boldsymbol{p}_A\boldsymbol{p}_B}$. The image of a non-oriented segment is the union of the two images of the oppositely oriented segments

$$\mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{I}^+_{\boldsymbol{p}_A\boldsymbol{p}_B} \cup \mathcal{I}^-_{\boldsymbol{p}_A\boldsymbol{p}_B}.$$

Observation windows–We are interested in OW's that are one-sided, i.e. able to "sense" the rays that cross them in one of the two directions only. Furthermore, we want to be able to manage multiple OW's, each corresponding to a different soundfield camera. In what follows, we will use the superscript (i) to identify the ith OW. The ray space region $\mathcal{V}^{(i)}$ that identifies the rays crossing the ith OW in the correct direction is called "visibility" of that OW. From the previous discussion, $\mathcal{V}^{(i)}$ is one of the two wedge-shaped regions that combine into the image of the segment that the OW lies upon.

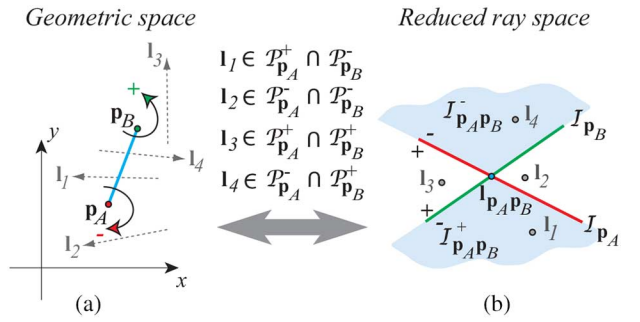

Fig. 4. A reflector in the geometric space and its image in the reduced ray space.

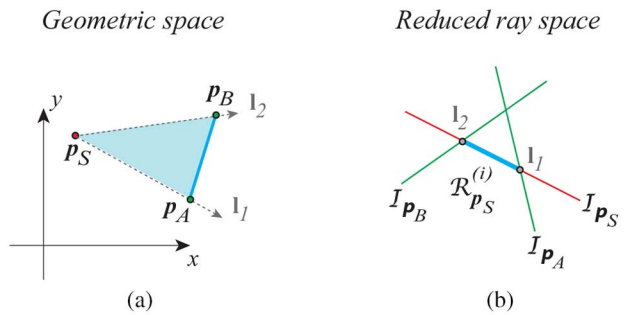

Fig. 5. A source in the geometric space sensed by the OW $\boldsymbol{p}_A\boldsymbol{p}_B$ and its ROI in the reduced ray space.

Acoustic reflectors–The wavefront reflected by a planar wall can be thought of as originating from an image source, whose location is determined by mirroring the source about the reflector line. As all rays coming from the image source are bound to pass *through* the reflector, we can think of the it as an "illuminating window", i.e. an aperture that casts reflected acoustic radiance onto the scene. With reference to Fig. 4, if the segment $\boldsymbol{p}_A\boldsymbol{p}_B$ is a reflector, the set of rays that can originate from a reflection against its left-face is completely contained in $\mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B}{}^-$. Conversely, $\mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B}{}^+$ includes the rays that originate from a reflection against the right face of $\boldsymbol{p}_A\boldsymbol{p}_B$. Notice that a reflector casts (reflected) acoustic radiance in front of it, but it also casts an acoustic shadow behind it. This will be discussed later in Section II-C2 when handling occlusions.

4) Regions of Interest: We define the Region Of Interest (ROI) of a primitive (a source or of a reflector) on the ith Observation Window as the portion of the image of that primitive visible from the OW. The ROI can be readily obtained by intersecting $\mathcal{V}^{(i)}$ with the image of the primitive.

Sources–The ROI of a source located in $\boldsymbol{p}_S$ in the visibility $\mathcal{V}^{(i)}$ is given by

$$\mathcal{R}^{(i)}_{\boldsymbol{p}_S} = \mathcal{I}_{\boldsymbol{p}_S} \cap \mathcal{V}^{(i)}. \tag{6}$$

Fig. 5(b) shows the ROI of the source $\mathbf{p}_S$, corresponding to the intersection between the image of the source $\mathcal{I}_{\boldsymbol{p}_S}$ and the visibility of the OW (i.e. the wedge delimited by the images of the points $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$). The resulting segment parameterizes all the ray originated by the source and passing through the OW (shaded region in Fig. 5(a)).

Reflectors–The ROI of the reflector $\boldsymbol{p}_A\boldsymbol{p}_B$ in the visibility $\mathcal{V}^{(i)}$ is

$$\mathcal{R}^{(i)}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B} \cap \mathcal{V}^{(i)}. \tag{7}$$

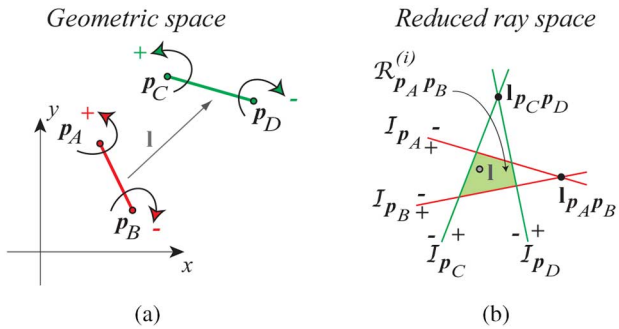

Fig. 6. The reflector $\boldsymbol{p}_A\boldsymbol{p}_B$ and the OW $\boldsymbol{p}_C\boldsymbol{p}_D$ in the geometric space and the ROI of the reflector in the reduced ray space as the intersection of the visibility region of the OW and the image of the reflector.

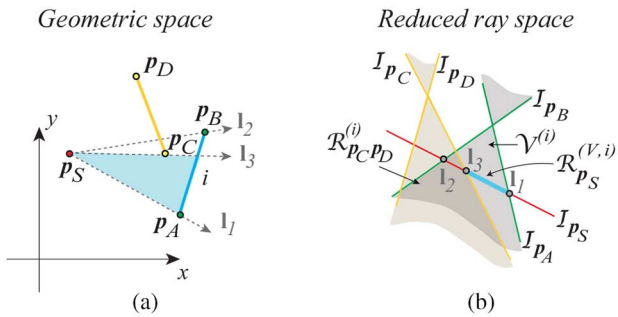

Fig. 7. Region of Visibility of the source.

Fig. 6 shows the ROI of the reflector $\boldsymbol{p}_A\boldsymbol{p}_B$ in the visibility of the OW $\boldsymbol{p}_C\boldsymbol{p}_D$, given by the intersection of the image $\mathcal{I}_{\boldsymbol{p}_A\boldsymbol{p}_B}$ of the reflector with the visibility of the OW $\mathcal{V}^{(i)}$.

C. Managing Multiple Primitives

In general, an acoustic scene is made of multiple reflectors and/or sources. In this case ROIs frequently overlap in the projective ray space. When this happens, it becomes important to understand which ROI covers which. This problem was already discussed in [13] and led to the determination of the Region Of Visibility (ROV), based on the culling of the respective ROIs. In this paragraph we revisit the definition of ROVs in the projective space. When working with multiple OW's, however, the situation becomes more complex, and will be discussed later.

1) Sources: Consider the setup in Fig. 7, corresponding to the acoustic scene of Fig. 5 with the added reflector $\boldsymbol{p}_C\boldsymbol{p}_D$. The segment $\boldsymbol{p}_C\boldsymbol{p}_D$ acts like an obstacle for $\boldsymbol{p}_S$, therefore the acoustic rays departing from $\boldsymbol{p}_S$ and visible from $\boldsymbol{p}_A\boldsymbol{p}_B$ range from $\mathbf{l}_1$ to $\mathbf{l}_3$.

Using the above notation, the rays produced by the source $\boldsymbol{p}_S$ and visible from the OW $\boldsymbol{p}_A\boldsymbol{p}_B$ can be identified by the region of visibility $\mathcal{R}^{(V,i)}_{\boldsymbol{p}_S}$ of the source $\boldsymbol{p}_S$ from the ith OW

$$\mathcal{R}^{(V,i)}_{\boldsymbol{p}_S} = \mathcal{R}^{(i)}_{\boldsymbol{p}_S} \cap \overline{\mathcal{R}}^{(i)}_{\boldsymbol{p}_C\boldsymbol{p}_D} \tag{8}$$

where $\overline{\mathcal{R}}^{(i)}_{\boldsymbol{p}_C\boldsymbol{p}_D}$ is the region of $\mathcal{V}^{(i)}$ that is not occupied by $\mathcal{R}^{(i)}_{\boldsymbol{p}_C\boldsymbol{p}_D}$, i.e.

$$\overline{\mathcal{R}}^{(i)}_{\boldsymbol{p}_C\boldsymbol{p}_D} = \mathcal{V}^{(i)} - \mathcal{R}^{(i)}_{\boldsymbol{p}_C\boldsymbol{p}_D}. \tag{9}$$

This definition of Region of Visibility holds valid for image sources as well. In fact, an image source is only visible through the reflector that generated it. We need to remember, however, that this very reflector will act like an occluder for all the other sources.

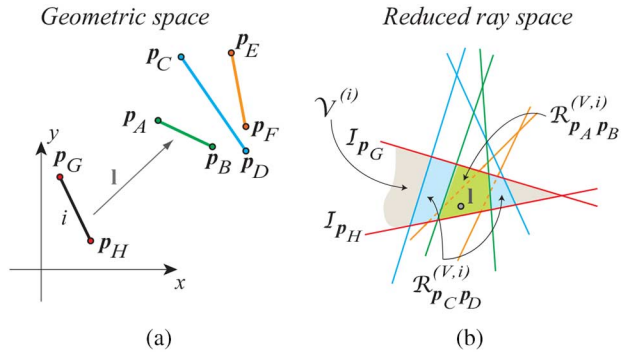

Fig. 8. The region of visibility of the reflectors as visibility culling of the ROIs in the visibility region of the OW.

2) Reflectors: The situation becomes more complicated when the acoustic scene under study has multiple reflectors. Let us consider the example of Fig. 8. Here the ith OW, with endpoints $\boldsymbol{p}_G$ and $\boldsymbol{p}_H$, senses an acoustic scene in which the three reflectors $\boldsymbol{p}_A\boldsymbol{p}_B, \boldsymbol{p}_C\boldsymbol{p}_D, \boldsymbol{p}_E\boldsymbol{p}_F$ are present. Notice that there are directions for which $\boldsymbol{p}_A\boldsymbol{p}_B$ occludes $\boldsymbol{p}_C\boldsymbol{p}_D$ and both $\boldsymbol{p}_A\boldsymbol{p}_B$ and $\boldsymbol{p}_C\boldsymbol{p}_D$ occlude $\boldsymbol{p}_E\boldsymbol{p}_F$. The region of visibility $\mathcal{R}^{(V,i)}_{\boldsymbol{p}_C\boldsymbol{p}_D}$ of the reflector $\boldsymbol{p}_C\boldsymbol{p}_D$ in the visibility region $\mathcal{V}^{(i)}$ is given by

$$\mathcal{R}^{(V,i)}_{\boldsymbol{p}_C\boldsymbol{p}_D} = \mathcal{R}^{(i)}_{\boldsymbol{p}_C\boldsymbol{p}_D} - \left(\mathcal{R}^{(i)}_{\boldsymbol{p}_C\boldsymbol{p}_D} \cap \mathcal{R}^{(i)}_{\boldsymbol{p}_A\boldsymbol{p}_B}\right). \tag{10}$$

Similarly, the region of visibility $\mathcal{R}^{(V,i)}_{\boldsymbol{p}_E\boldsymbol{p}_F}$ of the reflector $\boldsymbol{p}_E\boldsymbol{p}_F$ in the visibility region $\mathcal{V}^{(i)}$ is

$$\mathcal{R}^{(V,i)}_{\boldsymbol{p}_E\boldsymbol{p}_F} = \mathcal{R}^{(i)}_{\boldsymbol{p}_E\boldsymbol{p}_F} - \left(\mathcal{R}^{(i)}_{\boldsymbol{p}_E\boldsymbol{p}_F} \cap \left(\mathcal{R}^{(i)}_{\boldsymbol{p}_C\boldsymbol{p}_D} \cup R^{(i)}_{\boldsymbol{p}_A\boldsymbol{p}_B}\right)\right). \tag{11}$$

In this specific case $\mathcal{R}^{(V,i)}_{\boldsymbol{p}_E\boldsymbol{p}_F} = \emptyset$, while $\mathcal{R}^{(V,i)}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \mathcal{R}^{(i)}_{\boldsymbol{p}_A\boldsymbol{p}_B}$. When a source is present in the acoustic scene, only a linear portion of the ROV is visible. Indeed, only the rays departing the image source, obtained by mirroring the real source against the considered reflector, and not occluded by other reflectors are visible in the soundfield image. These rays are organized in the Ray Space on a linear pattern.

D. Managing Multiple OW's

When multiple OW's are present, we can define the global region of visibility of an acoustic primitive as the union of all the ROVs relative to the individual soundfield cameras that are present in the acoustic scene. In order to do so, however, we need to make sure that all ROV's are referred to the same (global) reference frame. As for the sources, the global region of visibility $\mathcal{R}^{(V)}_{\boldsymbol{p}_S}$ of a source $\boldsymbol{p}_S$ when viewed by all OW's $i = 1, \ldots, N$ is defined as

$$\mathcal{R}^{(V)}_{\boldsymbol{p}_S} = \bigcup_{i=1}^{N} \mathcal{R}^{(V,i)}_{\boldsymbol{p}_S}. \tag{12}$$

As for reflectors, the global region of visibility $\mathcal{R}^{(V)}_{\boldsymbol{p}_A\boldsymbol{p}_B}$ of the reflector $\boldsymbol{p}_A\boldsymbol{p}_B$ when viewed by all OW's $i = 1, \ldots, N$ is

$$\mathcal{R}^{(V)}_{\boldsymbol{p}_A\boldsymbol{p}_B} = \bigcup_{i=1}^{N} \mathcal{R}^{(V,i)}_{\boldsymbol{p}_A\boldsymbol{p}_B} \tag{13}$$

E. Relationship with the (m, q) Ray Space

The parameterization defined in [13] identifies a ray with the pair (m, q) of the equation $y = mx + q$ of the line that the ray lies upon, with a conventionally assigned direction (see Fig. 2). This Euclidean parameterization, therefore, is not of global validity, as it only accommodates rays that cross the y axis in one direction and are not parallel to it. This, indeed, was not an issue when working with a single OW, but becomes a strong limitation when working with multiple OWs. Nonetheless, it has the advantage of displaying the soundfield image in a "normalized" fashion. The Euclidean ray space (m, q) can be readily derived from the projective ray space, as it represents a special case of reduced ray space, obtained by setting

$$m = -\frac{l_1}{l_2} \tag{14}$$

$$q = -\frac{l_3}{l_2}. \tag{15}$$

F. From Local to Global Projective Ray Spaces

When working with multiple OW's, one simple choice could require each device to acquire a soundfield image that is referred its own (local) reference frame. These images can then be mapped onto each other's reference frame or onto a global reference frame. In this section we derive the equations that map the local projective ray spaces onto the global one. This analysis will become particularly useful later on, when using data coming from multiple camera or when performing self-calibration.

Each OW corresponds to a soundfield camera that works on a local reference frame. As mentioned before, the superscript (i) identifies the OW. In what follows, unless differently specified, we will assume that the reference frame is the local one, i.e. that attached to the OW.

Let $\mathbf{p}_A = [x_A, y_A, 1]^T$ be the homogeneous coordinates of the point A in the global reference frame and $\mathbf{p}_A^{(i)}$ the coordinates of the same point in the reference frame relative to the ith OW. It is well known [20] that $\mathbf{p}_A = [x_A, y_A, 1]^T$ and $\mathbf{p}_A^{(i)}$ are related by

$$\mathbf{p}_A = \mathbf{H}^{(i)} \mathbf{p}_A^{(i)}, \tag{16}$$

where

$$\mathbf{H}^{(i)} = \begin{bmatrix} \mathbf{R}^{(i)} & \mathbf{t}^{(i)} \\ \mathbf{0} & 1 \end{bmatrix}, \tag{17}$$

and $\mathbf{R}^{(i)}$ and $\mathbf{t}^{(i)}$ are the rotation matrix and translation vector, respectively, that characterize the rigid motion from local to global frame. Eq. (4) can be rewritten as

$$\mathbf{l}^T \mathbf{p}_A = \mathbf{l}^{(i)^T} \mathbf{p}_A^{(i)}, \tag{18}$$

therefore, using Eq. (16) in Eq. (18) we obtain

$$\mathbf{l} = (\mathbf{H}^{(i)})^{-T} \mathbf{l}^{(i)}. \tag{19}$$

III. Soundfield Images in the Projective Ray Space

In this Section we describe the formation of soundfield images from both theoretical and implementation standpoints.

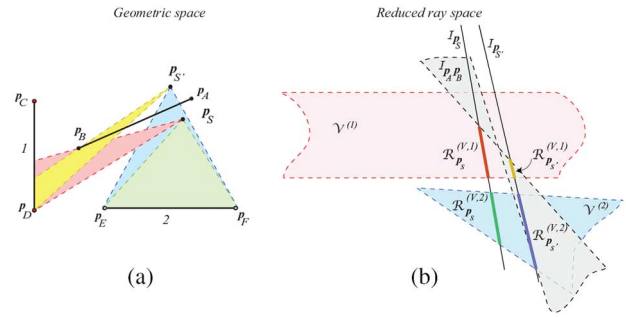

Fig. 9. Acoustic scene with the source $\boldsymbol{p}_S$, the reflector $\boldsymbol{p}_A\boldsymbol{p}_B$, the image source $\boldsymbol{p}_{S'}$ and two OW's.

A. The Ideal Scenario

In this section we aim at mapping the plenacoustic function $f(x, y, \theta)$ onto the global visibility $\mathcal{V}$. In order to progressively introduce the discussion, we first assume that our soundfield cameras are ideal, meaning that they are able to capture the acoustic radiance (plenacoustic function) of all rays crossing the OW from the correct side, with no resolution losses or aliasing phenomena.

A ray passing through a generic point $\boldsymbol{p} = [x, y]^T$ with direction θ has parameters

$$\begin{aligned} l_1 &= k \sin(\theta) \\ l_2 &= -k \cos(\theta) \\ l_3 &= k[y \cos(\theta) - x \sin(\theta)], k > 0. \end{aligned} \tag{20}$$

If we parameterize the plenacoustic function using these parameters, we obtain the soundfield map $p(\mathbf{l})$. The domain of this function is now the projective ray space. Using Eq. (20), we can write that

$$p(\mathbf{l}) = \begin{cases} f\left(x, -\frac{l_1 x + l_3}{l_2}, -\arctan\frac{l_1}{l_2}\right), & l_2 \neq 0 \\ f\left(-\frac{l_1}{l_3}, y, \pi/2\right), & l_2 = 0. \end{cases} \tag{21}$$

Notice that, if we scale x (or y in the second case), the plenacoustic function is picked at a different point that lies on the same ray. Thanks to the Radiance Invariance Law (RIL), however, its value will not change. Notice also that the plenacoustic function is complex-valued as it carries the phase information at the considered frequency.

For example, consider the acoustic scene of Fig. 9: the source $\boldsymbol{p}_S$ faces the reflector located on $\boldsymbol{p}_A\boldsymbol{p}_B$. The OW's 1 and 2 (lying on the segments $\boldsymbol{p}_C\boldsymbol{p}_D$ and $\boldsymbol{p}_E\boldsymbol{p}_F$, respectively) observe the scene. For the sake of simplicity, and with no loss of generality, we assume the reference frame to be centered on the first soundfield camera, with the y axis aligned with $\boldsymbol{p}_C\boldsymbol{p}_D$. In this perpendicular configuration, the visibility region of the first camera takes on the shape of a strip in the reduced ray space. This was the choice of frame adopted in [13]. The obstacle causes a reflection, which is modeled by the image source $\boldsymbol{p}_{S'}$. Shaded regions denote the acoustic beams that depart from $\boldsymbol{p}_S$ and $\boldsymbol{p}_{S'}$ and cross the OW's. Fig. 9(b) shows the regions of visibility of $\boldsymbol{p}_S$ and $\boldsymbol{p}_{S'}$. Each acoustic source is characterized by its radiance pattern (denoted with $b_\cdot(\mathbf{l})$, where $\cdot$ is replaced by the source name), which describes the way in which the source radiates the sound in space. For the scene in Fig. 9, $b_{\boldsymbol{p}_S}(\mathbf{l})$ and $b_{\boldsymbol{p}_{S'}}(\mathbf{l})$

parameterize the radiance of the soundfield emitted by the direct and image sources, respectively, along the ray $\mathbf{l}$. Under the hypotheses of validity of the Radiance Invariance Law, we can express the contribution of the source in $\boldsymbol{p}_S$ to the soundfield image $p_{\boldsymbol{p}_S}(\mathbf{l})$ as

$$p_{\boldsymbol{p}_S}(\mathbf{l}) = \begin{cases} b_{\boldsymbol{p}_S}(\mathbf{l}) & \mathbf{l} \in \mathcal{R}^{(V)}_{\boldsymbol{p}_S} \\ 0 & \text{elsewhere} \end{cases}. \tag{22}$$

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

$$p_{\boldsymbol{p}_{S'}}(\mathbf{l}) = \begin{cases} b_{\boldsymbol{p}_{S'}}(\mathbf{l}) & \mathbf{l} \in \mathcal{R}^{(V)}_{\boldsymbol{p}_{S'}} \\ 0 & \text{elsewhere} \end{cases}. \tag{23}$$

In the presence of multiple reflectors we can expect the soundfield image to exhibit higher-order reflections as well. Although, due to the lower amplitudes and resolution limits of a real soundfield camera, the linear patterns associated to higher-order image sources are more difficult to detect and extract, there are methods that allow us to account for such contributions [37], [38]. Using direct and first-order contributions, for example, we can estimate the corresponding reflector [13], and higher order reflection paths could, in principle, help us estimate the environment geometry [11]. However, as mapping higher order reflection paths is out of the scope of this paper, for reasons of illustrational simplicity we will limit our analysis to first-order reflections.

B. Capturing Soundfield Images

Multiple soundfield cameras, in principle, could be designed to operate and process data in the same (global) projective ray space. In order to do so, however, such camera would need to exchange information about their mutual locations. Requiring the cameras to work in their own (local) reference frame simplifies things a lot. The change of frame can be done afterwards through some calibration process. We will see later that calibration can be performed acoustically, using (local) soundfield images. For the acquisition of soundfield images referred to a local frame we can proceed as described in [13]. This approach is briefly summarized here for the reader's convenience.

We assume, for illustrational simplicity, that all the cameras have identical geometries. The discussion can be straightforwardly extended to the case of cameras with different geometry. Each local reference frame is defined in such a way that the relative OW lies is on the y axis between $y^{(i)} = -q_0$ and $y^{(i)} = q_0$. Each OW is spatially sampled with a uniform arrangement of M microphones. The soundfield image $P^{(i)}(m^{(i)}, q^{(i)})$ acquired by the ith camera is related to the (complex valued) soundfield map $p(m^{(i)}, q^{(i)})$ through

$$P^{(i)}(m^{(i)}, q^{(i)}) = \|p(m^{(i)}, q^{(i)})\|^2.$$

We divide the ith array into $M - W + 1$ overlapping sub-arrays, each made of W microphones. We implement a bank of wideband MVDR beamformers, each producing a pseudospectrum $H^{(i,j)}(\theta)$, where (i, j) identifies the jth sub-array on the ith camera. The pseudospectrum can be thought of as the angular power distribution of the acoustic rays passing through the center of the sub-array. We can therefore write

$$P^{(i)}(m^{(i)}, q^{(i)}_j) = H^{(i,j)}(\arctan(m^{(i)})), \tag{24}$$

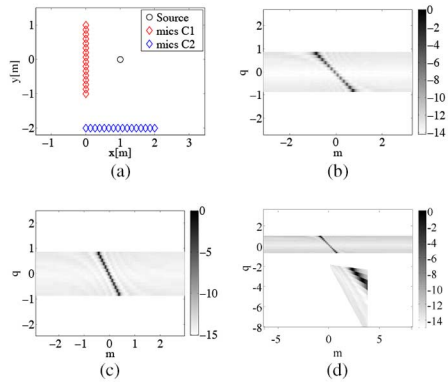

Fig. 10. Example of a soundfield image acquired by two cameras (C1 and C2) made of 16 microphones each. Values of the soundfield images are expressed in dB (a) Setup (b) $P^{(1)}(m, q)$ (c) $P^{(2)}(m^{(2)}, q^{(2)})$ (d) $P_G(m, q)$.

$q^{(i)}_j$ being the y coordinate of the center of the jth sub-array in the reference frame of the ith camera.

C. Using Multiple Soundfield Images

With no loss of generality, we assume the global reference frame to be centered on the first camera, i.e. $[x, y]^T = [x^{(1)}, y^{(1)}]^T$. We also choose the reduced ray space to match the normalized (m, q) ray space of the first camera. This will help us assess the advantages of our new representation with respect to the old one. In order to jointly use the information coming, for example, from two cameras, we need to be able to map $P^{(i)}(m^{(i)}, q^{(i)})$ onto $P^{(i)}(\mathbf{l}), i = 1, \ldots, N$. This can be done in two steps. We first need to turn the Euclidean soundfield image $P^{(i)}(m^{(i)}, q^{(i)})$ into a projective one $P^{(i)}(\mathbf{l}^{(i)})$, i.e. "lift" the reduced ray space to a (local) projective ray space. In practice, with reference to Eq. (14), this consists of adopting homogeneous coordinates

$$P^{(i)}(\mathbf{l}^{(i)}) = P^{(i)}\left(-l^{(i)}_1/l^{(i)}_2, -l^{(i)}_3/l^{(i)}_2\right). \tag{25}$$

Notice that rays that are parallel to the axis $y^{(i)}$ cannot be represented in the form $P^{(i)}(m^{(i)}, q^{(i)})$, but they can be readily represented in the form $P^{(i)}(\mathbf{l}^{(i)})$, though this has no practical impact, as OW cannot sense rays with that orientation. We can now use the transformation (19) to map the local ray space $\mathbf{l}^{(i)}$ onto the global one

$$P^{(i)}_G(\mathbf{l}) = P^{(i)}\left(\left(\mathbf{H}^{(i)}\right)^T \mathbf{l}\right), \tag{26}$$

where $P^{(i)}_G(\mathbf{l})$ represents the contribution of the ith image in the global ray space.

Merging the information coming from individual soundfield images is rather straightforward when the visibility regions of the related OWs are disjoint, i.e. $\mathcal{V}^{(i)} \cap \mathcal{V}^{(j)} = \emptyset, i, j = 1, \ldots, N, i \neq j$, in which case we have

$$P_G(\mathbf{l}) = \bigcup_{i=1}^{N} P^{(i)}_G(\mathbf{l}). \tag{27}$$

A more interesting case is given by overlapping visibility regions of the cameras. In this situation multiple local soundfield images convey information about the same acoustic rays. The reliability of the information brought by the individual cameras

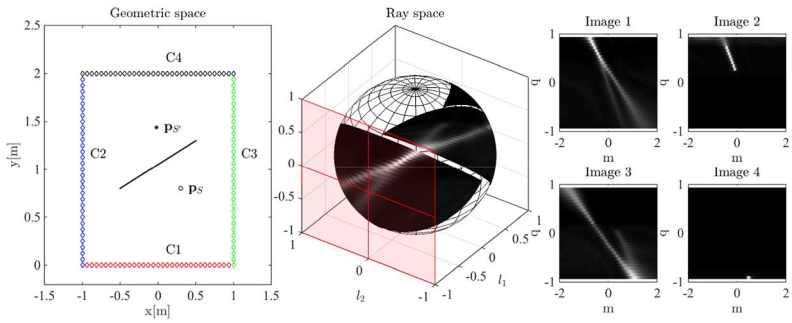

Fig. 11. Acoustic scene composed by a source ($\boldsymbol{p}_S$) and a reflector that generates an image source ($\boldsymbol{p}_{S'}$), observed by four soundfield cameras (C1, C2, C3 and C4); the global ray space; and the individual soundfield images captured by the four cameras.

in these regions of overlap could be different, due to differences in resolution depending on the relative positioning of the camera and the acoustic source. A typical situation where visibility regions heavily overlap is that of a single moving OW that captures a static acoustic scene at a steady pace. In this paper, however, we will be primarily concerned with the case of multiple static cameras with non-overlapping regions, i.e. acquiring different rays. We will not consider the problem of fusing images in the possible overlaps, because this would require us working with (complex-valued) soundfield maps instead of (real-valued) soundfield images.

Fig. 10 shows an example of a multiview soundfield image with non-overlapping visibility regions. The scene is made of a single acoustic source, with no obstacles/reflectors. The two cameras are made of 16 microphones each. The distance between adjacent microphones is 13.3 cm, corresponding to an aliasing frequency of about $f_a = 1.3$ kHz [13]. We simulated the source emitting a narrowband signal with a center frequency of $f_a/2$. Fig. 10 shows the individual soundfield images (i.e. in the local reference frames) and the global one. Notice that (as expected) $\mathcal{R}^{(1)}_{\boldsymbol{p}_S}$ and $\mathcal{R}^{(2)}_{\boldsymbol{p}_S}$ lie on small segments of $\mathcal{I}^{(1)}_{\boldsymbol{p}_S}$ and $\mathcal{I}^{(2)}_{\boldsymbol{p}_S}$. In a single-camera case, we would localize the source by determining the parameters of the line on which $\mathcal{R}^{(1)}_{\boldsymbol{p}_S}$ or $\mathcal{R}^{(2)}_{\boldsymbol{p}_S}$ lie. The resulting accuracy would become worse as the distance from the source increases. Using multiple cameras simplify this task a lot, as the parameters of the line $\mathcal{I}_{\boldsymbol{p}_S}$ can be estimated from multiple segments. Notice also that the linear pattern associated to the source looks sharper in the normalized image referred to the local ray space of the second camera (Fig. 10(c)) than in its remapped version of Fig. 10(d). This is due to the choice of the reduced ray space, which introduces a distortion. This distortion, however, is only in the visualization. If we develop a localization algorithm that inherently works in the global projective ray space, this distortion will not have an impact on the localization accuracy.

In order to better understand the usefulness of the global projective ray space and its relation with local reduced ray spaces we consider the ideal scenario of Fig. 11, in which four soundfield cameras capture an acoustic scene made of a source and a reflector. This is an interesting case because, although the global projective ray space is able to accommodate all rays of interest, there is no single reduced ray space that allows us to simultaneously display all of them. We recall that a reduced ray space is obtained by slicing the space of homogeneous coordinates with a plane and there is no plane that slices through all visibility regions at the same time. One alternate way to visualize these data is to slice the space of homogeneous coordinates with a spherical surface instead. This is easily done by constraining the norm of the rays ($k\mathbf{l}$ kept constant), as shown in Fig. 11 and visualizing the radiance of the individual rays on this new spherical "reduced" ray space. Different cameras capture different views of the scene: the first camera sees both the source and its reflected image; the second camera sees the source as partially occluded by the reflector; the third camera sees the source and a part of the image source within the reflector's region of visibility; and for the fourth camera the source is nearly completely occluded. All this data cannot be represented in a single reduced ray space. For example, the shaded plane $l_1 = -1$ of the Fig. 11 (in the middle) represents the reduced ray space of the first camera. The rays parallel to this plane map to infinity on the reduced ray space and cause the distortion we observed in Fig. 10 when we try to represent data of other cameras. Furthermore, the rays with oppositely oriented rays are not distinguishable, which means we can not represent the data captured by the fourth camera in the (m, q) ray space.

IV. Examples of Application

In this Section we propose two examples of application of multiview soundfield imaging, aimed at illustrating the effectiveness of the projective soundfield representation: localization of acoustic sources and autocalibration of arrays, i.e. the estimation of the relative position and orientation of all arrays using acoustic data only.

A. Source Localization

In [13] we showed that single-camera soundfield images enable the localization of multiple acoustic sources. Here we summarize the procedure and extend it to the multiple camera scenario.

The sources map to linear features on the soundfield image and, therefore, they can be localized estimating the parameters of the corresponding lines. In particular, in a first stage, the soundfield image is analyzed in order to find relevant features. This is done finding peaks above a prescribed threshold. These features of the soundfield image are then matched to sources present in the scene. For this purpose the Hough transform ([40], [41]) is used to cluster the selected features into linear patterns

and discard outliers. Finally, the location of the sources are estimated through linear regression on the estimated linear patterns. The features of the soundfield image correspond to the acoustic rays and, therefore, to directions-of-arrival (DOAs). As a consequence, this is essentially a DOA-based localization. However, the use of the ray space representation brings a number of important benefits. The usually difficult problem of matching acoustic measurements to multiple acoustic sources and discarding outliers can be performed in a robust and efficient way using methods found in the rich literature of computer vision. Furthermore, localization becomes a linear estimation problem. The use of multiple cameras, however, poses some challenges. The same source could be seen under very different angles from the two arrays, causing the resolution of the two soundfield images to differ a lot. This means that the accuracy of peak localization could be uneven for the two cameras and a suitable data fusion strategy becomes necessary.

Let us consider a source with position given in projective coordinates $\overline{\mathbf{p}} = k[\overline{x}, \overline{y}, 1]^T$. We know that all the lines that pass through $\overline{\mathbf{p}}$ must satisfy the equation $\mathbf{l}^T\overline{\mathbf{p}} = 0$. Through the Hough transform we find the features that are aligned in the soundfield image. The location of these features is denoted by $\mathbf{l}_1, \ldots, \mathbf{l}_P$. They represent peaks in the soundfield images, i.e. the rays produced by the acoustic source and captured by the soundfield cameras. In an ideal condition, these rays pass exactly through $\mathbf{p}$, so that we can write

$$\begin{cases} \mathbf{l}_1^T\overline{\mathbf{p}} & = 0 \\ \ldots & \\ \mathbf{l}_P^T\overline{\mathbf{p}} & = 0 \end{cases} \tag{28}$$

These equations can be rearranged in a matrix form as $\mathbf{L}\overline{\mathbf{p}} = 0$, where $\mathbf{L} = [\mathbf{l}_1, \ldots, \mathbf{l}_P]^T$. Estimating $\mathbf{p}$ is based on obtaining the null-space of $\mathbf{L}$. In order to do so, we first define $\mathbf{L}_w = \mathbf{W}\mathbf{L}$ where $\mathbf{W} = \text{diag}(1/e_1, \ldots, 1/e_P)$ is a weighting diagonal matrix, and $e_i, i = 1, \ldots, P$ is the error introduced in the system of Eq. (28) by the peak-picking algorithm. We compute the singular value decomposition of the 3×3 matrix $\mathbf{L}_w^T\mathbf{L}_w = \mathbf{U}_L\mathbf{D}_L\mathbf{V}_L^T$, where $\mathbf{U}_L$ and $\mathbf{V}_L^T$ are the singular vectors matrices of the decomposition and $\mathbf{D}_L = \text{diag}(\sigma_{1L}, \sigma_{2L}, \sigma_{3L}), \sigma_{1L} > \sigma_{2L} > \sigma_{3L}$, contains the singular values of $\mathbf{L}_w^T\mathbf{L}_w$. The estimate $\hat{\mathbf{p}}$ (in homogeneous coordinates) of the source location is given by the singular vector of $\mathbf{V}_L$ related to the smallest singular value of $\mathbf{D}_L$

$$\hat{\mathbf{p}} = \mathbf{V}_L|_3, \tag{29}$$

where $\mathbf{V}_L|_3$ is the third column of the matrix $\mathbf{V}_L$.

As for the matrix $\mathbf{W}$, we estimate the error $e_i, i = 1, \ldots, P$, of the peak location as the width at -3 dB of the lobe containing the local maximum $\mathbf{l}_i$. This is based on the observation that the peak location becomes more sensitive to errors as the width of the lobe increases.

In order to validate the proposed solution and show the benefits with respect to the single-camera case, we conducted a simulation campaign using the setup of Fig. 12. The two cameras are made of 9 microphones each, uniformly spaced at a step of 0.06 m. In the simulations we vary the distance d of the source from the center of the two cameras; and the SNR of the signal

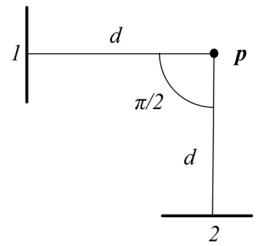

Fig. 12. Simulation setup used for the localization.

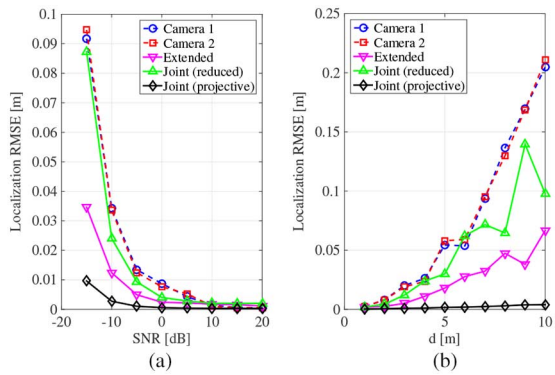

Fig. 13. RMSE of the localization error, expressed as distance between the estimated and actual position, shown for different values of SNR (a) and d (b). Localization is performed using: only camera 1; only camera 2; single extended camera; both cameras in reduced ray space; and both cameras in the projective ray space.

acquired by microphones. The angle formed by the lines joining the center of the two cameras and $\boldsymbol{p}$ is set to $\alpha = \pi/2$. For all the experiments, the source signal is a white noise in the bandwidth [300 Hz, 2.5 kHz]. We tested the localization accuracy as the distance between the estimated and actual position over 200 realizations. In these tests, the performance of the localization from two cameras in the projective ray space is compared with:

- localization with the camera 1 or 2 only;
- localization with a single extended camera of 18 microphones spaced by 6 cm, whose center coincides with the center of camera 1;
- localization from the joint soundfield image of camera 1 and 2 in the reduced ray space.

The first simulation is aimed at testing the robustness of the localization algorithm against noise. In this experiment the Signal to Noise Ratio varies within the range $[-15 \text{ dB}, 20 \text{ dB}]$ (see Fig. 13(a)). The other parameters of the simulation are kept constant: $d = 2$ m, $\alpha = \pi/2$. Notice that the localization in the projective ray space outperforms the other methodologies, including the localization in the reduced ray space using data from camera 1 and 2. This is due to the fact that the transformation from the projective ray space to the reduced one introduces a relevant distortion in the resulting image, as already shown in Fig. 10. After peak detection in the reduced ray space, performing localization in the global projective ray space significantly reduces the error and improves the localization accuracy. Notice also that the proposed algorithm outperforms the localization using the extended camera.

In the second experiment we vary the distance d of the source from the center of the cameras, while keeping $\alpha = \pi/2$

and SNR $= 0$ dB. Results are shown in Fig. 13(b). Also in this case the localization accuracy turns out to outperform the other techniques, when done in the projective ray space. The advantage over the single-camera case, especially at large distances, can be explained from the fact that using two soundfield images we are effectively performing a triangulation. This becomes particularly helpful when the source is far. In this case the estimation of the source distance is typically affected by relevant errors. Using multiple arrays that are spatially distributed increases the baseline and helps us reduce this error. Finally, notice that the error relative to the localization using joint cameras in the reduced ray space (green line) suffers from outliers, which cause the irregular trend from $d = 8$ m to $d = 10$ m.

B. Self-calibration

The problem of self-calibration date back to 90's for what concerns video signals (e.g. [42]). Self-calibration of multiple microphone arrays has been approached more recently [24]–[43]. We can divide self-calibration algorithms in two classes, according to the specific goal. Some self-calibration algorithms are aimed at localizing microphones of an array (intrinsic calibration) and others are aimed at estimating the mutual position and orientation of different arrays (extrinsic calibration). In this paragraph we refer to the second class, and we aim at estimating the homography $\mathbf{H}$ that maps the second reference frame onto the first one. The advantage brought by the plenacoustic analysis on self-calibration lies in the ability of the source localization algorithm to work with multiple sources; and perform clustering in the ray space to match acoustic events with sources. This allows us to perform self-calibration when multiple sources are simultaneously active. For illustrational simplicity, we focus on the case of two cameras, although a generalization to more cameras is possible. In our case, the homography $\mathbf{H}$ is a simple isometry of the form

$$\mathbf{H} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & \Delta x \\ \sin(\theta) & \cos(\theta) & \Delta y \\ 0 & 0 & 1 \end{bmatrix}, \tag{30}$$

where θ and $[\Delta x, \Delta y]^T$ are the rotation angle and the translation vector that make the second camera move onto the first one. We begin with localizing S sources with the cameras 1 and 2. The source locations in homogeneous coordinates referred to the local reference frame of the ith camera are $\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_S^{(i)}$ and $\mathbf{p}_s^{(i)} = [x_s^{(i)}, y_s^{(i)}, w_s^{(i)}]^T$, while their estimates are $\hat{\mathbf{p}}_1^{(i)}, \ldots, \hat{\mathbf{p}}_S^{(i)}$ and $\hat{\mathbf{p}}_s^{i} = [\hat{x}_s^{(i)}, \hat{y}_s^{(i)}, \hat{w}_s^{(i)}]^T$. We adopt the notation $\hat{\mathbf{p}}_i^{(1)} \leftrightarrow \hat{\mathbf{p}}_i^{(2)}$ to indicate that $\hat{\mathbf{p}}_i^{(1)}$ and $\hat{\mathbf{p}}_i^{(2)}$ are estimates of the same source in the reference frames of the two cameras. In accordance with (16) we can write that $\mathbf{p}_i^{(1)} = \mathbf{H}\mathbf{p}_i^{(2)}$. Notice, however, that we cannot treat this equation in a conventional fashion, because homogeneous coordinates are inherently scalable therefore that equality is to be intended up to a scaling factor. The method that we propose is based on the Direct Linear Transformation (DLT), typically used in applications of 3D vision [20]. An alternate way of rewriting $\mathbf{p}_i^{(1)} = \mathbf{H}\mathbf{p}_i^{(2)}$ that does not suffer from the problem of being valid up to a scaling factor is $\mathbf{p}_i^{(1)} \times \mathbf{H}\mathbf{p}_i^{(2)} = 0$,

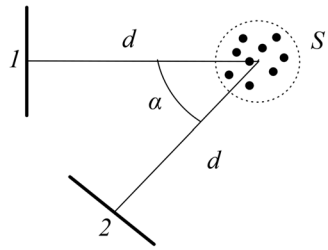

Fig. 14. Setup of the simulations for the autocalibration.

where "$\times$" denotes the vector product. This constraint can be reformulated into

$$\mathbf{p}_i^{(1)} \times \mathbf{H}\mathbf{p}_i^{(2)} = \begin{bmatrix} y_i^{(1)}\mathbf{h}^{3T}\mathbf{p}_i^{(2)} - w_i^{(1)}\mathbf{h}^{2T}\mathbf{p}_i^{(2)} \\ w_i^{(1)}\mathbf{h}^{1T}\mathbf{p}_i^{(2)} - x_i^{(1)}\mathbf{h}^{3T}\mathbf{p}_i^{(2)} \\ x_i^{(1)}\mathbf{h}^{2T}\mathbf{p}_i^{(2)} - y_i^{(1)}\mathbf{h}^{1T}\mathbf{p}_i^{(2)} \end{bmatrix}, \tag{31}$$

which can be rewritten into

$$\begin{bmatrix} \mathbf{0}^T & -w_i^{(1)}\mathbf{p}_i^{(2)} & y_i^{(1)}\mathbf{p}_i^{(2)} \\ w_i^{(1)}\mathbf{p}_i^{(2)} & \mathbf{0}^T & -x_i^{(1)}\mathbf{p}_i^{(2)} \\ -y_i^{(1)}\mathbf{p}_i^{(2)} & x_i^{(1)}\mathbf{p}_i^{(2)} & \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{bmatrix} = \mathbf{0}, \tag{32}$$

where $\mathbf{h}^{jT}$ is the jth row of $\mathbf{H}$. Notice that only two out of the three rows of the coefficient matrix in (32) are linearly independent. The third row, in particular, can be obtained by summing the first row multiplied by $x_i^{(1)}$ and the second row by $y_i^{(i)}$. If we drop the third equation, we find that each source generates two equations in the form

$$\begin{bmatrix} \mathbf{0}^T & -w_i^{(1)}\mathbf{p}_i^{(2)^T} & -y_i^{(1)}\mathbf{p}_i^{(2)^T} \\ w_i^{(1)}\mathbf{p}_i^{(2)^T} & \mathbf{0}^T & -x_i^{(1)}\mathbf{p}_i^{(2)^T} \end{bmatrix} \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{bmatrix} = \mathbf{A}_i\mathbf{h} = \mathbf{0}. \tag{33}$$

If we consider all the sources, we obtain the system $\mathbf{A}\mathbf{h} = \mathbf{0}$, where $\mathbf{A} = [\mathbf{A}_1^T\mathbf{A}_2^T \ldots \mathbf{A}_S^T]^T$. Notice that the above system is homogeneous. This fact reduces the degrees of freedom of the vector $\mathbf{h}$. We observe that $S = 4$ sources generate eight constraints on $\mathbf{h}$. We also impose that $\|\mathbf{h}\| = 1$ in order to prevent the trivial solution $\mathbf{h} = \mathbf{0}$ and to remove the scalability of the solutions. Notice that $\mathbf{H}$ has the structure in (30), since it is an isometry. If we also use this constraint, the number of unknowns drops, and consequently the number of sources needed for the autocalibration is reduced to $S = 3$, (as long as they are not collinear). We remark, however, that the dependency of $\mathbf{H}$ on θ is non-linear, thus making the estimation more complex. When more than $S = 4$ sources are available, we resort to the least squares solution of $\mathbf{A}\mathbf{h} = \mathbf{0}$. In this case the estimate $\hat{\mathbf{h}}$ of the vector $\mathbf{h}$ is given by the singular vector associated to the least singular value of the SVD of $\mathbf{A}^T\mathbf{A}$, similarly to the localization. In particular, we define $\mathbf{A}^T\mathbf{A} = \mathbf{U}_A\mathbf{D}_A\mathbf{V}_A^T$, where $\mathbf{D}_A$ contains the singular values, and $\mathbf{U}_A$ and $\mathbf{V}_A$ are the singular vector matrices of dimensions 9×9. The estimate $\hat{\mathbf{h}}$ is obtained as

$$\hat{\mathbf{h}} = \mathbf{V}_A|_9. \tag{34}$$

In order to validate the autocalibration algorithm, we conducted an extensive simulation campaign. The setup of the simulation is shown in Fig. 14. The two cameras are identical and accommodate each 9 microphones spaced by 0.06 m, for an overall

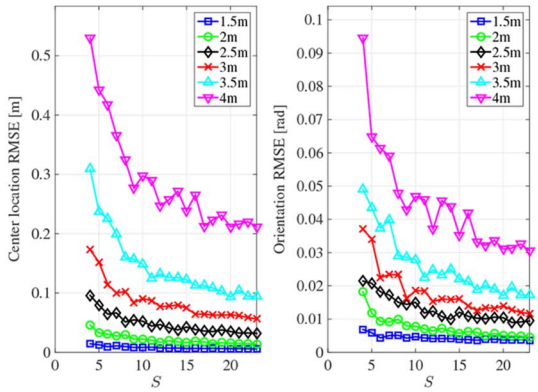

Fig. 15. RMSE of the autocalibration errors, expressed as distance between the estimated camera center and its actual location and error on the orientation of the second camera. With reference to Fig. 14, 1.5 m $< d <$ 4 m and $\alpha = \pi/4$.

length of the array of 0.48 m. In the simulations we varied the number of sources S, the distance d between the cameras and the centroid of the cluster of sources, and the angle α formed by the lines joining the center of the cameras and the centroid of the cluster of sources. The sources have been simulated to produce a white noise in the bandwidth [300 Hz, 2.5 kHz]. For all the simulations we set SNR $=$ 10 dB.

In the first experiment we varied the distance d between the cameras and the cluster of sources in the range [1.5 m,4 m] and we tested the accuracy of the autocalibration algorithm for different number S of sources. The angle α was fixed at $\pi/4$. The error of the autocalibration is expressed as the distance between the estimated center of the second camera and its actual location and the absolute value of the difference between the estimated angle α and the actual value. Results are averaged over 50 repetitions of the experiment and shown in Fig. 15. Notice that the error increases with d. This is due to the limited baseline of the cameras. In fact, when sources are distant from the camera center, the localization error increases. This, in turn, has an impact on the accuracy of the estimation of $\mathbf{h}$. Notice also that $S = 10$ is sufficient, for most of the distances, to guarantee an accurate autocalibration. In the second set of simulations we vary the angle α between the cameras in the range $[\pi/4, \pi]$, while $d = 3$ m. The other conditions are identical to the previous simulations. The estimation accuracy for $\alpha = \pi/4$ can be seen in Fig. 15 ($d = 3$ m). The results for other values of α are not shown for reasons of space and clarity of visualization as they are pretty constant for all tested angles. This behavior can be explained by the fact that the cameras are always pointed towards the cluster of sources and are at the same distance from it, achieving, as a consequence, the same localization accuracy independently of the angle between the two cameras. As in the previous case, $S = 10$ sources are sufficient to get an accurate estimation.

C. Experimental Results

In order to validate the results of the proposed algorithm on real data, we conducted an experiment in which we evaluate, in an integrated fashion, the accuracy of autocalibration and localization algorithms.

The experimental setup consists of two cameras of eight microphones each, spaced by 0.055 m. In order to prevent any influence of reverberation and other external factors on the assessment, the two cameras have been placed in an environment with absorptive walls, where the reverberation time has been estimated being $T_{60} \approxeq 50$ ms. The rotation angle θ and displacement vectors $\mathbf{t}$ between the arrays (see Eq. (30)) have been hand-measured using the Multi Dimensional Scaling [44] to obtain $\theta = 1.51$ rad, $\mathbf{t} = [-1.793 \text{ m}, -1.987 \text{ m}]^T$. The geometry of the setup is shown in Fig. 16. In this figure the locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ of the sources used for the assessment of the localization accuracy are shown as well.

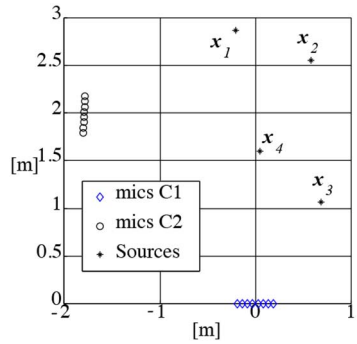

Fig. 16. Geometry of the arrays used for localization and autocalibration experiments and location of the sources used for the localization experiment.

TABLE I
ACTUAL LOCATION OF SOURCES USED FOR THE REAL EXPERIMENTS

Source #	x [m]	y [m]
1	-0.21	2.875
2	0.58	2.555
3	0.685	1.065
4	0.05	1.605

As for the autocalibration, a total number of fifteen source locations arbitrarily placed has been used. As done for the simulations, the sources reproduced a white noise sequence in the bandwidth [300 Hz,2.5 kHz]. The errors $\Delta\theta$ and $\Delta\mathbf{t}$ on the rotation angle and translation vectors of the estimate are given, respectively, by $\Delta\theta = -0.051$ rad, $\Delta\mathbf{t} = [0.002 \text{ m}, 0.004 \text{ m}]^T$. Such results confirm the validity of the self-calibration algorithm also in real scenarios. Notice, moreover, that due to measurement errors, also the ground-truth data could be affected by some error.

The localization experiment adopts the same setup used for the autocalibration. For the convenience of the reader, the locations of the sources to be localized, which constitutes a different set from the sources used for the autocalibration, are given in Table I. Notice from Fig. 16 that, as seen from camera 1, sources at $\boldsymbol{x}_2$ and $\boldsymbol{x}_4$ are visible under the same angle. We can expect, therefore, that localization based only on camera 1 for this pair of sources fails. The same holds for $\boldsymbol{x}_3$ and $\boldsymbol{x}_4$ for camera 2. Using the self-calibration obtained from acoustic measurements, we evaluated the accuracy of the localization for different configurations, namely: 1) one source active at any time; 2) one of the 6 combinations of two sources active at any time; 3) one of the 4 combinations of three sources active at any time; 4) all sources active at any time.

Notice that the maximum frequency of 2.5 kHz is below the spatial Nyquist frequency. Beamforming is not affected, therefore, by spatial aliasing. Table II shows the localization error

TABLE II
LOCALIZATION RESULTS FOR ALL THE POSSIBLE COMBINATIONS OF SOURCES

Sources	Cameras	ε_1 [m]	ε_2 [m]	ε_3 [m]	ε_4 [m]
1	C1	0.38	-	-	-
	C2	0.117	-	-	-
	Joint	0.116	-	-	-
2	C1	-	0.275	-	-
	C2	-	0.306	-	-
	Joint	-	0.186	-	-
3	C1	-	-	0.073	-
	C2	-	-	0.189	-
	Joint	-	-	0.117	-
4	C1	-	-	-	0.07
	C2	-	-	-	0.09
	Joint	-	-	-	0.067
1&2	C1	0.22	0.411	-	-
	C2	0.357	1.048	-	-
	Joint	0.111	0.162	-	-
1&3	C1	0.764	-	0.062	-
	C2	0.101	-	0.091	-
	Joint	0.1	-	0.061	-
1&4	C1	0.148	-	-	1.437
	C2	0.096	-	-	0.109
	Joint	0.132	-	-	0.108
2&3	C1	-	0.247	0.117	-
	C2	-	0.3	0.206	-
	Joint	-	0.149	0.105	-
2&4	C1	-	0.36	-	0.09
	C2	-	0.421	-	0.11
	Joint	-	0.217	-	0.075
3&4	C1	-	-	0.07	0.09
	C2	-	-	0.883	0.11
	Joint	-	-	0.07	0.075
&2&3	C1	0.477	0.168	0.083	-
	C2	0.109	0.777	0.691	-
	Joint	0.106	0.167	0.085	-
1&2&4	C1	0.288	0.261	-	0.827
	C2	0.152	0.856	-	0.1
	Joint	0.115	0.26	-	0.063
1&3&4	C1	0.51	-	0.064	0.791
	C2	0.109	-	0.862	0.093
	Joint	0.096	-	0.064	0.086
2&3&4	C1	-	0.187	0.151	0.818
	C2	-	0.48	0.881	0.15
	Joint	-	0.122	0.151	0.104
1&2&3&4	C1	0.734	0.14	0.13	0.5
	C2	0.142	0.84	0.78	0.07
	Joint	0.13	0.14	0.13	0.085

$\varepsilon_i = ||\hat{\boldsymbol{x}}_i - \boldsymbol{x}_i||, i = 1, \ldots, 4$, where $\hat{\boldsymbol{x}}_i$ is the estimated location of the ith source, for all the possible combinations. The localization has been performed using the camera 1, the camera 2, or with the joint localization algorithm presented before.

Results confirm that the joint localization outperforms the localization based on a single camera. As an example, let us consider the case of the four sources active at the same time (last three rows of Table II). Notice that sources at $\boldsymbol{x}_1$ and $\boldsymbol{x}_4$ are poorly localized by camera 1; conversely camera 2 is not able to localize $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$. Joint localization, finally, allows us to accurately localize all four sources. We also notice that increasing the number of active sources does not imply a reduction in the localization accuracy.

V. CONCLUSIONS

In this paper we have proposed a new projective parameterization for the ray space, which generalizes the domain of soundfield images defined in [13]. The new representation is based on oriented projective geometry and uses homogeneous coordinates, with multiple advantages:

- a single reference frame accommodates all rays of interest with no blind spots;
- no loss of accuracy can be attributed to a bad choice of reference frame;
- it is now possible to optimize the spatial configuration of sensors with control over the achieved resolution.

Given the recent progress in integrated microphone arrays and the wide range of possible applications, the soundfield imaging becomes an interesting approach for organizing, managing, displaying and processing the data that such devices will be able to collect. In this paper we choose to address two possible applications that take advantage of the new parameterization. On one hand, the projective ray space representation allowed us to simultaneously use multiple arrays with significant improvement in localization accuracy. At the same time, in order to take advantage of this new global representation at best, we developed a self-calibration methodology, operating entirely in the acoustic domain, which enables the georeferentiation of the various soundfield cameras. We are currently focusing on how to derive and work with (complex-valued) soundfield maps instead of soundfield images, with the goal of fusing soundfield images in the most general case of overlapping regions of interest and, most of all, with the goal of developing a rendering system based on plenacoustic principles.

REFERENCES

[1] D. Ward and R. Williamson, "Particle filter beamforming for acoustic source localization in a reverberant environment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2002, vol. 2, pp. 1777–1780.

[2] E. Lehmann and R. Williamson, "Particle filter design using importance sampling for acoustic source localisation and tracking in reverberant environments," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 168–168, 2006.

[3] F. Antonacci, D. Riva, D. Saiu, A. Sarti, M. Tagliasacchi, and S. Tubaro, "Tracking multiple acoustic sources using particle filtering," in *Proc. Eur. Signal Process. Conf. (EUSIPCO'06)*, Sep. 2006.

[4] F. Antonacci, M. Matteucci, D. Migliore, D. Riva, A. Sarti, M. Tagliasacchi, and S. Tubaro, "Tracking multiple acoustic sources in reverberant environments using regularized particle filter," in *Proc. 15th Int. Conf. Digital Signal Process.*, Jul. 2007, pp. 99–102.

[5] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 2, pp. 439–443, Feb. 2013.

[6] F. Antonacci, A. Sarti, and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Dallas, TX, USA, Mar. 2010, pp. 2822–2825.

[7] S. Tervo, J. Patynen, and T. Lokki, "Acoustic reflection localization from room impulse responses," *Acta Acust. united with Acust.*, vol. 98, no. 3, pp. 418–440, 2012.

[8] F. Antonacci, J. Filos, M. Thomas, E. Habets, A. Sarti, P. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 10, pp. 2683–2695, Dec. 2012.

[9] I. Dokmanic, R. Parhizkar, A. Walther, Y. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," in *Proc. Nat. Acad. Sci.*, 2013, vol. 110, no. 30, pp. 12 186–12 191.

[10] H. Sun, E. Mabande, K. Kowalczylc, and W. Kellermann, "Joint DOA and TDOA estimation for 3D localization of reflective surfaces using eigenbeam MVDR and spherical microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 113–116.

[11] 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 *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 2620–2623.

[12] E. Mabande, H. Sun, K. Kowalczyk, and W. Kellermann, "On 2d localization of reflectors using robust beamforming techniques," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 153–156.

[13] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Soundfield imaging in the ray space," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 12, pp. 2493–2505, Dec. 2013.

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

[15] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, Feb. 1992.

[16] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.

[17] T. Ajdler, L. Sbaiz, A. Ridolfi, and M. Vetterli, "On a stochastic version of the plenacoustic function," in *Proc. IEEE Conf. Acoust., Speech, Signal Process.*, 2006, vol. 4, pp. 1125–1128.

[18] F. Antonacci, A. Canclini, A. Galbiati, A. Calatroni, A. Sarti, and S. Tubaro, "Soundfield rendering with loudspeaker arrays through multiple beamshaping," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, 2009, pp. 313–316.

[19] L. Bianchi, F. Antonacci, A. Sarti, and S. Tubaro, "Rendering of directional sources through loudspeaker arrays based on plane wave decomposition," in *Proc. IEEE Int. Workshop Multimedia Signal Process. (MMSP' 13)*, Pula, Italy, Sep.–Oct. 30–2, 2013, pp. 13–18.

[20] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[21] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'94)*, 1994, vol. 2, pp. 273–276.

[22] H. Do, H. Silverman, and Y. Yu, "A real-time srp-phat source location implementation using stochastic region contraction (src) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'07)*, 2007, vol. 1, pp. 121–125.

[23] M. Cobos, A. Marti, and J. Lopez, "A modified srp-phat functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 71–74, Jan. 2011.

[24] M. Hennecke and G. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays (HSCMA)*, May 2011, pp. 127–132.

[25] P. Pertila, M. Mieskolainen, and M. Hamalainen, "Closed-form self-localization of asynchronous microphone arrays," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays (HSCMA)*, May 2011, pp. 139–144.

[26] N. Gaubitch, B. Kleijn, and R. Heusdens, "Auto-localization in ad-hoc microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2013, pp. 106–110.

[27] S. Charneau, L. Aveneau, and L. Fuchs, "Exact, robust and efficient full visibility computation in plucker space," *Vis. Comput.*, vol. 23, no. 9, pp. 773–782, 2007.

[28] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography*. New York, NY, USA: Academic, 1999.

[29] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 2–16, Jan. 2010.

[30] M. Guillaume and Y. Grenier, "Sound field analysis based on analytical beamforming," *EURASIP J. Adv. Signal Process.*, vol. 2007, pp. 1–15, 2007.

[31] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*. New York, NY, USA: Springer, 2013.

[32] F. M. Fazi, M. Noistering, and O. Warufsel, "Representation of sound fields for audio recording and reproduction," in *Proc. Acoustics'12: 11me Congr. Français d'Acoust. Annu. Meeting Inst. Acoust.*, Nantes, France, Apr. 23–27, 2012.

[33] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on orthonormal decomposition of the soundfield," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'02)*, Orlando, FL, USA, May 13–17, 2002, vol. 2, pp. 1781–1784.

[34] H. Teutsch and W. Kellerman, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," *J. Acoustic Soc. Amer.*, vol. 120, no. 5, pp. 2724–2736, 2006.

[35] H. Kuttruff, *Room Acoustics, Fifth Edition*. Abingdon, U.K.: Spon, 2009.

[36] F. Everest and K. Pohlmann, *Master Handbook of Acoustics*. New York, NY, USA: McGraw-Hill Education, 2009.

[37] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Real time modeling of acoustic propagation in complex environments," in *Proc. 7th Int. Conf. Digital Audio Effects*, Oct. 2004, pp. 274–279.

[38] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 812–824, May 2008.

[39] J. Stolfi, *Oriented Projective Geometry: A Framework for Geometric Computations*. New York, NY, USA: Academic, 1991.

[40] P. V. C. Hough, "Method and means for recognizing complex patterns," U.S. patent 3,069,654, Dec. 18, 1962.

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

[42] R. K. M. Pollefeys and L. V. Gool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 1998.

[43] C. Schindelhauer, Z. Lotker, and J. Wendeberg, "Network synchronization and localization based on stolen signals," in *Proc. 30th Annu. ACM SIGACT-SIGOPS Symp. Principles Distrib. Comput. (PODC)*, 2011, pp. 223–224.

[44] T. Cox and M. Cox, *Multidimensional Scaling*. Boca Raton, FL, USA: Chapman & Hall/CRC, 2001.