PLENACOUSTIC IMAGING IN THE RAY SPACE

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

Dipartimento di Elettronica e Informazione, Politecnico di Milano
Piazza Leonardo da Vinci, 32, 20133 Milano, Italy

ABSTRACT

In this work we present a novel approach to acoustic scene analysis with microphone arrays. Acoustic measurements are here represented in the ray space, which is defined as the space of parameters that describe acoustic rays. The ray space can be thought of as the domain of the plenacoustic function, which defines the sound pressure field in all positions in space. We discuss sampling and distortion with respect to the ideal plenacoustic function introduced using a microphone array. Plenacoustic images can potentially be used for environment inference, source characterization and wavefield extrapolation.

Index Terms— Plenacoustic function, acoustic images.

1. INTRODUCTION

Space-time audio processing algorithms that exploit knowledge on the environment (geometry and reflectivity) to improve their performance are appearing more and more frequently in the literature (e.g. [2]). These algorithms rely on information provided by environment inference algorithms (e.g. [3]). The same information can also be explicitly used for modeling/predicting the soundfield in enclosures [4]. In alternative, acoustic measurements can be used for predicting the soundfield without explicitly estimating the environment [5].

The problems of environment inference and wavefield extrapolation could be formulated and addressed through the study of the "plenacoustic function" [5], which captures both geometric and radiometric properties of the acoustic scene.

The plenacoustic function was introduced in [5] as the acoustic counterpart of the plenoptic function [6], which describes the intensity of the light flow at every position for every direction at all frequencies and, for dynamic scenes, time. In this work we are interested in developing a "plenacoustic camera", which is able to take "snapshots" of the plenacoustic function and represent it in a parameter space where its regularity can be exploited through signal processing. We will use acoustic signals acquired with a microphone array, and process them to map the related measurements onto the parametric space of acoustic rays (Ray Space). We will call "plenacoustic image" the result of this mapping. Unlike plenoptic (lightfield) cameras, in the acoustic domain we cannot rely on devices for capturing rays coming from a given direction. We will therefore use space-time processing instead, and cope with the resulting degradation with respect to the ideal acoustic image.

The paper is organized as follows: in Section 2 we discuss the domain of the plenacoustic function and introduce the adopted parametrization of acoustic rays. In Section 3 we define an ideal plenacoustic image; show how to obtain an approximation of this image using a microphone array; and give some insight on the structure of the captured data. Section 4 shows an example of acquired plenacoustic image.

2. THE RAY SPACE PARAMETRIZATION

In defining a plenoptic function, several assumptions are often made (e.g. static scenes, grayscale images, reduction of degrees of freedom on camera locations) in order to reduce the dimensionality of the representation [7]. Popular parametrizations are the lumigraph [8] and the lightfield [9], whose domain is the *space of oriented lines*. The Plenoptic data is used in computer vision for localization, mapping, synthetic view generation (image based rendering), etc. An example of a commercial plenoptic camera is based on [10].

In [5] the plenacoustic function was defined as the instantaneous acoustic pressure at given location without the directional information as the longer wavelengths of acoustic waves make it difficult to measure. However, as observed also in [5], this omnidirectional function can be turned into a directional one using phase information. We think of the plenacoustic function as a function of position and direction. As a consequence, we represent geometric primitives of interest and the acoustic measurements in the space of oriented lines here referred to as the *Ray Space*. This is the same domain that was used in [4] for the modelling of acoustic propagation in enclosures. The purpose of this paper is to reverse the paradigm and use the same domain for analysis purposes.

A ray can be seen as an oriented line in the geometric space. A line in 2D space is represented by a linear equation $l_1x+l_2y+l_3 = 0$ and can be parametrized with the line coefficients $[l_1, l_2, l_3]^T$. These coordinates are homogeneous (scalable) as $\mathbf{l} = k[l_1, l_2, l_3]^T$, with $k \neq 0$, represent the same line. In order to distinguish two rays lying on the same line but with opposite orientations we limit the range of the scalar k to the positive or negative interval. Thus we define coordinates in an oriented projective space $\mathbb{P}^2$. As a generic point (l_1, l_2, l_3) corresponds to a ray in the geometric space, the Euclidean space ($\mathbb{R}^3$) spanned by such homogeneous coordinates of lines is called the Ray Space (see Figure 1).

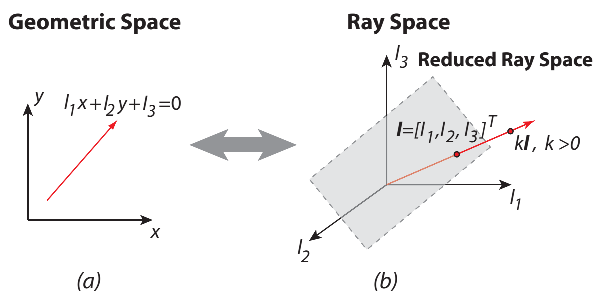


Fig. 1. A ray in geometric (a) and ray spaces (b).

The geometric primitives of interest are acoustic sources (i.e. loudspeakers), receivers (i.e. microphones) and reflectors (i.e. walls). Sources and receivers are assumed to be point-like while reflectors are assumed to be planar. We represent them in the Ray Space as sets of rays. For clarity of visualization we depict the primitives in a reduced 2D Ray Space obtained by intersecting the Ray Space with a prescribed plane, as shown in Figure 1 (b).

Acoustic rays: As seen above we parameterize acoustic rays with (homogeneous) coefficients of the lines they lie on. This parametrization defines the Ray Space. An acoustic ray in the geometric space corresponds to a half-line passing through the origin in the Ray Space as shown in Figure 1. Furthermore, using the line equation, a ray $\mathbf{l}$ is passing through a point A with homogeneous coordinates $\mathbf{x}_A = k[x_A, y_A, 1]^T$, $k > 0$, if $\mathbf{x}_A^T\mathbf{l} = 0$.

Sources and receivers: Acoustic sources and receivers can be seen as points in geometric space. A point is identified in the Ray Space by the set of all rays that pass through it. Using the condition $\mathbf{x}_A^T\mathbf{l} = 0$, all rays passing through the point A are $\overline{A} = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T\mathbf{l} = 0\}$. Thus a point in the geometric space corresponds to a plane passing through the origin in the Ray Space (see Figure 2 (a)). Furthermore, being a hyper-plane, it divides the Ray Space into two half-spaces. This allows us to test the orientation of a ray with respect to the point as follows: all rays that have the point on their left or right (with respect to travel direction) are given by: $\overline{A}_+ = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T\mathbf{l} > 0\}$ or $\overline{A}_- = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_A^T\mathbf{l} < 0\}$ (see Figure 2 (b) and (c)).

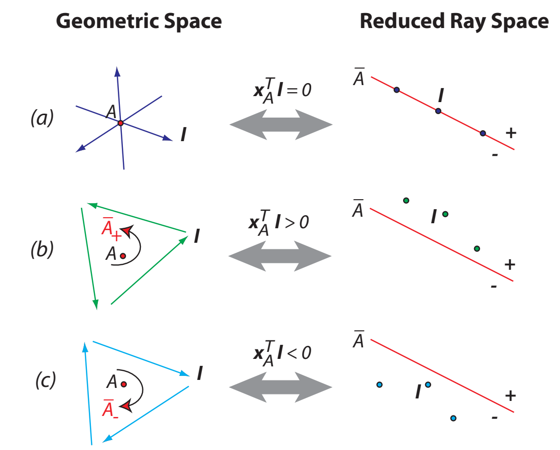


Fig. 2. The orientation of rays with respect to a point in the geometric space and their configuration in the ray space.

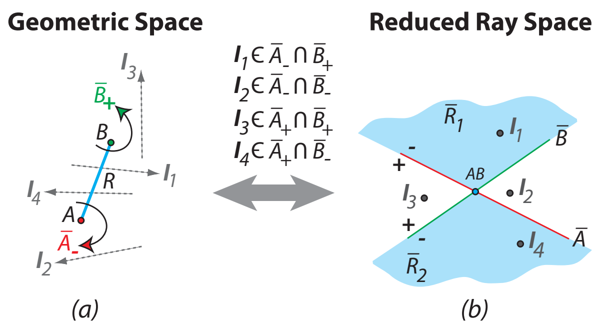


Fig. 3. Four subspaces defined by the two endpoints in the ray space (b) and corresponding rays in the geometric space (a).

Reflectors: In the geometric space the reflector is a line segment completely defined by the two endpoints. In the Ray Space we represent it with the set of all rays that intersect this line segment. Exploiting the orientation relations we can distinguish two oriented reflectors $\overline{R}_1$ and $\overline{R}_2$, i.e. the two reflectors defined by the same line segment but characterized by different directions of incident rays: $\overline{R}_1 = \overline{A}_- \cap \overline{B}_+$ and $\overline{R}_2 = \overline{A}_+ \cap \overline{B}_-$ (see Figure 3). Finally, the non oriented reflector is the union of two oriented reflectors that compose it, i.e. $\overline{R} = \overline{R}_1 \cup \overline{R}_2$.

3. PLENACOUSTIC IMAGES

The ideal plenacoustic camera – An ideal plenacoustic camera is a device that is able to capture the (complex) amplitudes of all acoustic rays that fall onto it. The term "plenacoustic" comes from the fact that there is no single point (camera center) that rays are bound to pass through. A camera of this sort can be thought of as a line segment in the geometric space (see Fig. 4 (a)) or, dually, the set of all rays that cross such segment. Its representation in the Ray Space is therefore given by the parameters of all such rays. This representation, in fact, corresponds to that of a reflector. In Fig.4 (b) the grey area represents the set of all rays that can be captured by the acoustic camera, and the red area represents the rays that intersect both reflector and acoustic camera.

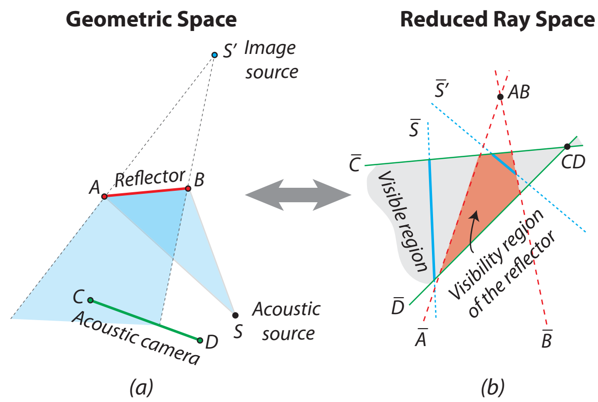


Fig. 4. The ideal acoustic camera (a) and the ray space representation of acoustic image (b).

So far we have discussed the representation of geometric primitives as a function of the acoustic rays that are generated (source), collected (receiver) or reflected (reflector) by them. Each ray that parametrizes the geometric primitive is associated to a magnitude. As soon as the acoustic source "lights up", a proliferation of acoustic rays populates the environment due to the presence of reflectors. A specular reflection originates from an image source (see Figure 4 (a)) whose visibility is spatially limited as the rays that it originates are bound to cross the reflector. Only a subset of these rays will, in fact, be captured by the ideal plenacoustic camera (see Fig. 4 (b)), as the camera itself is also characterized by a region of visibility (i.e. it has its own field of view), which corresponds to a "beam-shaped" region in Ray Space. It is interesting to notice how the presence of reflected rays contribute to enriching the plenacoustic view of the scene. For example, if the room includes scattering walls, the diffuse reflections will be captured by the acoustic camera and the whole reflector's visibility region will "brighten up". The case of Lambertian (equip-diffusive) surfaces, however, is far more common in the optical case than it is in the acoustic one. However, we can always acoustically illuminate the environment from different locations in space to extract more information from the acoustic scene.

The array approximation – Implementing a plenacoustic camera with a microphone array (see Fig. 5 (a)) means estimating the acoustic rays that intersect the array through space-time processing. In order to do so, for every microphone m_i, $i = 1, ..., M$ of the array we do the following:

1. spatially window the contributions of the microphones using a function w_i centered at m_i;
2. use a beamforming technique to obtain the spatial pseudospectrum $P_i(\theta)$ and associate it to the point m_i;
3. the plenacoustic image will be $P_i(\mathbf{l})$, which is obtained by mapping the values of $P_i(\theta)$ onto the Ray Space (see Fig. 5 (b)).

The use of a spatial windowing function allows us to trade between conflicting needs:

- in order to increase the field of view (a wider beam in Ray Space) the array must be spatially extended;
- in order to estimate rays that pass through a point m_i of the array, we need to work with just a few neighboring microphones.

Spatial windowing overcomes this problem but introduces an *aperture* phenomenon and other distortions. For example, a smaller spatial window reduces the angular resolution (it "blurs" the plenacoustic image) but increases the spatial resolution (information about rays passing through a single point m_i). On the other hand, a larger spatial window improves the angular resolution (assuming the far-field assumption made by most beamforming techniques is still valid) but decreases the spatial resolution. On the positive side, windowing suggests that an array of many microphones could be replaced with a smaller array that slides along its axis (under the hypothesis of stationary sound field).

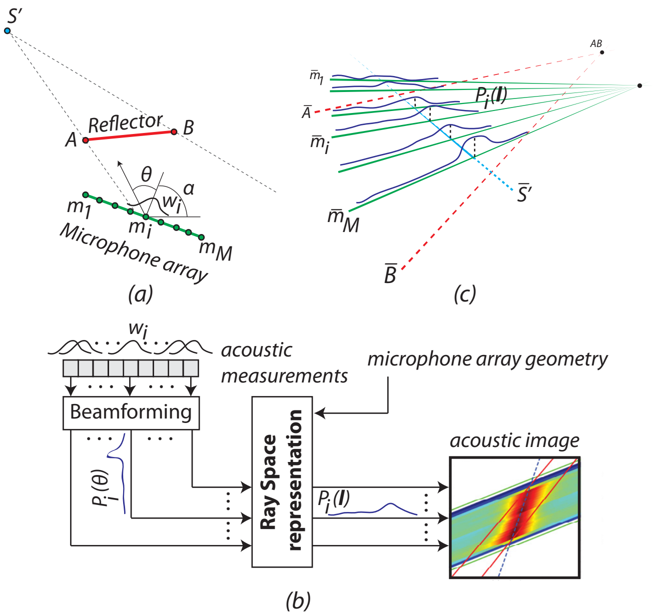


Fig. 5. The microphone array (a); schematic overview of the acoustic image acquisition process (b); the ray space representation of acoustic image (c).

As we can see in Fig. 5 (b), a microphone array is an approximation of a plenacoustic camera in that it samples the field of vision (each microphone corresponds to a "ray" of the beam that represents the camera field of vision in the Ray Space); and the acquired intensity profile along each sample is a "blurred" (due to the aperture) version of the ideal plenacoustic image. This is clearly visible in Fig. 5 (c), which is obtained for sample points m_i (microphone locations) and the impulses representing the sources are convolved with the aperture function of the adopted beamforming technique. The image configuration depends on the geometry (reflector and source positions) while the amplitudes vary in time and frequency according to the radiation pattern of the source, the polar pattern of the microphones, the reflection coefficients and traveling path distance. In case of wideband sources we can obtain a number of images for different frequency bands of interest or (if we are interested in extracting only geometric information) a single image combining images at different frequencies.

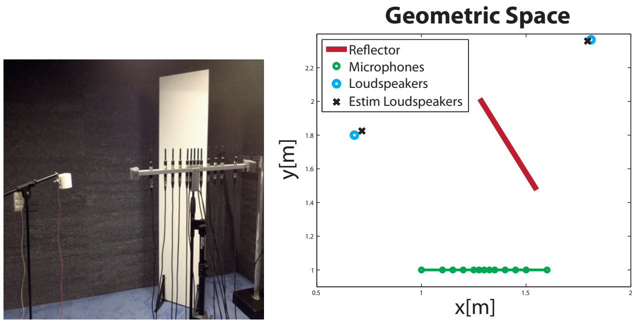
Although the acquisition process could be easily performed with an arbitrarily shaped array, the linear array has the advantage of maximal extension and, consequently, best field of view. It also has the advantage of being easily manageable in the ray space. We should bear in mind, however, that the field of view of this camera is limited to half of the geometric space.

Notice that an acoustic camera (in a more traditional sense) could be seen as a compact array of microphones (whose size is negligible compared with the size of the imaged scene). This camera would be able to estimate the magnitudes of all rays that pass through a single point in space (the location of the array), therefore its field of view in the Ray Space corresponds to a single ray (an extremely narrow beam). Along this ray we find the values of the pseudo spectrum corresponding to the directions of arrival of the rays. Examples of acoustic cameras are the cylindrical arrays (2D case) and spherical arrays (3D case). An alternative, more expensive, implementation of a plenacoustic camera could be a spatial distribution of small acoustic cameras of this sort, each taking care of estimating sound field magnitudes along all directions. This would have the effect of reducing the aperture problem; increasing the field of view (to a full angle); distributing the processing (over the small acoustic cameras) and compacting the information (each small acoustic camera would transmit a set of vectors and the related amplitudes). The expensive hardware requirements of this scenario could be overcome in the near future thanks to the availability of inexpensive integrated microphone arrays. This approach resembles the construction of some plenoptic cameras, which can be built as arrays of mini-cameras. With this perspective in mind, the use of the proposed method combined with robust superdirective beamforming techniques becomes particularly interesting.

4. EXPERIMENTS

In order to show an example of the captured acoustic image we performed an acquisition with a $0.6m$ long linear array of 13 microphones. The experiment was conducted within an low-reverberation room with a $60cm$ wide reflective surface placed in front of the array. The environment was "illuminated" by a white noise in the frequency band $1kHz - 10kHz$ emitted by a small loudspeaker. The experimental setup is shown in Fig. 6.

The signals acquired by the sensors were sampled at $44.1kHz$. We used a 3-sample spatial rectangular window w and the wideband Capon method [11] to perform beamforming. The absolute value of the resulting plenacoustic image is shown in the reduced Ray Space in Fig. 7. As expected, the aperture problem and the relatively small



(a) (b)

Fig. 6. Experimental setup and real/image source positions estimated from the acoustic image.

total number of microphones causes the plenacoustic image to be blurred (aperture) and sampled. Yet, the image clearly exhibits the lines that represent the loudspeaker and its (wall-reflected) image. In addition, the plenacoustic image clearly shows the visibility of the image loudspeaker (which is to be confined within the boundaries of the reflector).

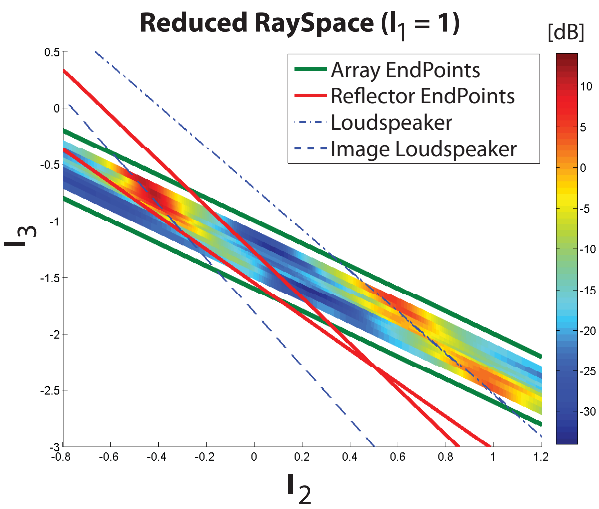


Fig. 7. Acoustic image showing the loudspeaker and the image loudspeaker limited by the extension of the reflector.

This image can be used for extracting a significant amount of information on the acoustic scene. We focus here on acoustic source localization. According to the adopted parametrization the source S with coordinates $\mathbf{x}_S$ corresponds in the plenacoustic image to the line $\overline{S} = \{\mathbf{l} \in \mathbb{P}^2 | \mathbf{x}_S^T \mathbf{l} = 0\}$. Therefore source localization can be approached as detection of lines. For this purpose, given the plenacoustic image, we look for peaks, we cluster them in sets belonging to the same line (i.e. using the Hough transform) and we apply the least square fitting in order to find the line parameters. An example of direct/image loudspeaker localization is shown in Fig. 6 (b). Even in the presence of noise; measurement errors on reference positions; finite aperture of the array; and discrete sampling of the ray space, we notice a good match between the reference locations $[x_S, y_S] = [0.68m, 1.8m]$, $[x_{S'}, y_{S'}] = [1.81m, 2.36m]$ and the estimated locations $[\hat{x}_S, \hat{y}_S] = [0.71m, 1.82m]$, $[\hat{x}_{S'}, \hat{y}_{S'}] = [1.79m, 2.35m]$.

5. CONCLUSIONS

In this work we showed how to acquire samples of the plenacoustic function using a microphone array. The acquired images capture the acoustic environments as “seen” from different points in space and the adopted parametrization maps the data in a space that shows a high degree or regularity. We showed that the data structure that we proposed for defining the plenacoustic image is, in fact, very effective for collecting and displaying information on the viewed scene. In particular, plenacoustic cameras have a relevant potential in the near field range, where the camera’s field of vision (viewing beam in the ray space) is wider. When this happens, the camera provides rich information on the acoustic propagation in the environment, including the extension of acoustic reflectors and the visibility of image sources.

We are currently working on using this data for applications of environment inference and soundfield extrapolation.

Acknowledgement

This work has been partially accomplished in the EU FET-Open SCENIC project under GA 226007.

6. REFERENCES

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

[2] F. Antonacci, A. Canclini, A. Galbiati, A. Calatroni, A. Sarti, and S. Tubaro, “Soundfield rendering with loudspeaker arrays through multiple beamshaping,” in *proc. of 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA*, 2009.

[3] F. Antonacci, A. Sarti, and S. Tubaro, “Geometric reconstruction of the environment from its response to multiple acoutic emissions,” in *proc. of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing , ICASSP-2010, Dallas, TX*, March 14-19 2010, pp. pp. 2822–2825.

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

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

[6] Edward H. Adelson and James R. Bergen, “The plenoptic function and the elements of early vision,” in *Computational Models of Visual Processing*. 1991, pp. 3–20, MIT Press.

[7] J. Berent and P.L. Dragotti, “Plenoptic manifolds,” *Signal Processing Magazine, IEEE*, vol. 24, no. 6, pp. 34 –44, nov. 2007.

[8] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen, “The lumigraph,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1996, SIGGRAPH ’96, pp. 43–54, ACM.

[9] Marc Levoy and Pat Hanrahan, “Light field rendering,” in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1996, SIGGRAPH ’96, pp. 31–42, ACM.

[10] Ren Ng, *Digital light field photography*, Ph.D. thesis, Stanford, CA, USA, 2006, AAI3219345.

[11] M. R. Azimi-Sadjadi, A. Pezeshki, L. L. Scharf, and M. Hohil, “Wideband doa estimation algorithms for multiple target detection and tracking using unattended acoustic sensors,” *Proc. of the SPIE04 Defense and Security Symposium*, vol. 5417, pp. 111, April 2004.