A LINEAR OPERATOR FOR THE COMPUTATION OF SOUNDFIELD MAPS

L. Bianchi, V. Baldini Anastasio, D. Marković, F. Antonacci, A. Sarti, S. Tubaro

Dipartimento di Elettronica, Informazione e Bioingegneria - Politecnico di Milano via Ponzio 34/5 - 20133 Milano, Italia

ABSTRACT

In the process of soundfield imaging, as defined in the literature, a microphone array is subdivided into overlapping subarrays and soundfield images are obtained by juxtaposition of spatial spectra computed from individual subarray data. In this paper we show that the whole process can be conveniently seen as a linear transformation applied to array data. This linear transformation embeds a nonlinear mapping to cast the directional information in a more convenient domain: the ray space. We show by simulations that the proposed formulation is suitable for fast implementation of the soundfield imaging operation, and, more specifically, for the localization of acoustic sources.

Index Terms— Microphone arrays, soundfield imaging, plenacoustic function, array signal processing, ray space

1. INTRODUCTION

Soundfield imaging is the process of analyzing the spatial features of an acoustic field and depicting them in a form that is suitable for providing an intuitive source of information. Microphone arrays provide the technological mean to achieve the imaging goal, and they have been extensively studied in the past few years. In [1, 2, 3, 4, 5], approaches based on compact spherical microphone arrays are presented. Although these approaches provide accurate estimates of the direction of arrival of a sound source, they do not provide a direct way to infer the absolute location of the sound source.

In this work we adopt the *plenacoustic* representation of the sound field first presented in [6, 7] and then adopted for signal processing tasks in [8, 9, 10]. More specifically, in this paper we consider the soundfield imaging process as presented in [11, 12], regarded to as the process of acquiring directional sound field components impinging on a linear microphone array. This process can be conceptually divided into four steps: 1) a linear array of microphones is subdivided into overlapping subarrays composed by adjacent sensors; 2) data collected by each subarray are used to form an estimate of the magnitude of spatial spectrum through a beamscan operation [13]; 3) individual spatial spectra are mapped to the *ray space*, a geometric space where coordinates are the parameters of a line [14]; 4) the soundfield image is built by juxtaposition of the spatial spectra.

One of the main advantages of the soundfield imaging approach in [12] is that acoustic primitives appear in the soundfield image as linear patterns. This is a known property of the ray space, as demonstrated in [14], and a number of applications have been presented exploiting that property. In [12] the authors consider a two-dimensional propagation scenario composed by a single linear array of microphones; in this scenario they exploit the structure of the ray space to perform source localization through pattern analysis in the soundfield image. The use of high-resolution techniques to obtain estimates of the spatial spectra has been investigated in [15]. In [16] the authors study the use of soundfield images to estimate the radiation pattern of a violin. In [17] soundfield images are exploited for acoustic signal extraction purposes. In [18] an extension to multiple linear arrays is provided. In [19] the application to soundfield rendering is presented.

In this paper we revise the soundfield imaging process to show that it can be cast in the form of a linear transformation of array data. The simple scenario of a single linear array of microphones is considered here, and, following [12], we use a two-dimensional euclidean space as the domain of the soundfield image. We remark that the presented formulation can easily be extended to other scenarios (e.g. multiple linear arrays, as in [18]), as the structure of the transformation proposed here is not altered by the specific domain (the ray space) where directional information is mapped. In particular, choosing a different ray space as a domain for the soundfield image only affects the mapping operation, which is independent of the transformation of array data. Another difference with respect to [12] is that in this work we consider spatial spectra with amplitude and phase; in order to highlight this difference, in the following we adopt the term soundfield map to denote a complex-valued soundfield image.

The rest of the paper is structured as follows. Sec. 2 introduces the adopted model for array signals. Sec. 3 gives a review of the soundfield imaging process. Sec. 4 derives the linear transformation to obtain the soundfield map. Sec. 5 analyzes the computational complexity of the soundfield mapping operation. Sec. 6 presents a simulative localization application exploiting the presented formulation. Finally, Sec. 7 draws some conclusions and Sec. 8 shows prior work.



Fig. 1: Microphone array and notation.

2. SIGNAL MODEL

Consider the setup depicted in Fig. 1. A uniform linear array of M microphones is placed on the y axis between $y = -q_0$ and $y = q_0$, so that the *i*th microphone is located at $m_i = [0, q_0 - 2q_0(i-1)/(M-1)]^T$, i = 1, ..., M. The array is subdivided into I = M - W + 1 maximally overlapped subarrays, each composed by W (odd) adjacent microphones. The set $\{m_k\}, k = i, ..., i + W - 1$ denotes the locations of the microphones belonging to the subarray with reference point at \mathbf{m}_i . We remark that this is a slightly different choice with respect to previous works (e.g. [12]), where the reference point is set to the center of the subarray; nonetheless, the approach proposed here can be extended effortless to handle different choices for the subarray reference point. The distance between microphones is $d = 2q_0/(M-1)$.

The microphone array observes a sound scene composed by L, possibly moving, acoustic sources, whose location is denoted by $\mathbf{p}_{S,l}(t) = [x_{S,l}(t), y_{S,l}(t)]^T$, l = 1, ..., L, where t denotes time. A short-time analysis on microphone signals is performed, so that the signal acquired by the *i*th microphone is transformed in

$$x_i(\tau,\omega) = \int_{-\infty}^{\infty} x_i(t)w(t-\tau)e^{-j\omega t} dt, \qquad (1)$$

where w(t) is a suitable window function. In the following we omit the dependency on time and frequency.

Consider now the *i*th subarray. Short time-frequency signals acquired by microphones are collected in the vector

$$\mathbf{x}_i = [x_i, \dots, x_{i+W-1}]^T, \tag{2}$$

which can be modeled as

$$\mathbf{x}_i = \mathbf{G}_i \mathbf{s}_i + \mathbf{v}_i,\tag{3}$$

where $\mathbf{s}_i = [s_{1,i}, \ldots, s_{L,i}]^T \in \mathbb{C}^{L \times 1}$ is the vector containing the *L* source signals and $\mathbf{v} \in \mathbb{C}^{W \times 1}$ is additive noise. The matrix $\mathbf{G}_i \in \mathbb{C}^{W \times L}$ collects the acoustic transfer function between sources and microphones. Assuming that the size of the subarray is small enough so that the source is in the far field with respect to it, the acoustic transfer function can be modeled as [20, Eq. 6.2.21]

$$\mathbf{G}_{i} = [\mathbf{g}(\theta_{i,1}), \dots, \mathbf{g}(\theta_{i,L})],$$
$$\mathbf{g}(\theta_{i,l}) = [1, e^{-j\frac{\omega}{c}d\sin(\theta_{i,l})}, \dots, e^{-j(W-1)\frac{\omega}{c}d\sin(\theta_{i,l}))}]^{T}.$$
(4)

The angle $\theta_{i,l}$, according to Fig. 1, denotes the angle under which the *i*th subarray sees the *l*th source

$$\theta_{i,l} = \arctan\left(\frac{y_{\mathrm{S},l} - q_0 + 2q_0(i-1)/(M-1)}{x_{\mathrm{S},l}}\right).$$
 (5)

3. SOUNDFIELD MAPS

In [12], for the purpose of estimating the spatial spectrum from data acquired by the *i*th subarray, the authors employ multiple beamforming operations, steered towards a preselected grid of directions $\{\theta_{\gamma}\}, \gamma = 1, ... \Gamma$, which is the same for each subarray. In order to focus array data towards the grid of directions $\{\theta_{\gamma}\}$, the steering matrix $\mathbf{A} \in \mathbb{C}^{W \times \Gamma}$ is employed for all subarrays, and it can be expressed as [20, Eq. 6.2.21]

$$\mathbf{A} = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_{\Gamma})], \\ \mathbf{a}(\theta_{\gamma}) = [1, e^{-j\frac{\omega}{c}d\sin(\theta_{\gamma})} \dots, e^{-j(W-1)\frac{\omega}{c}d\sin(\theta_{\gamma}))}]^T.$$
(6)

Each beamformer output composes the spatial spectrum $p_i(\theta)$, which is then mapped onto the ray space \mathcal{P} . As it has been described in [14], sound sources appear in \mathcal{P} as lines: this feature makes it convenient to represent spatial spectra onto the ray space. In particular, each sample $p_i(\theta_{\gamma})$ of a spatial spectrum encodes the (complex) amplitude of the acoustic ray crossing m_i with angle θ_{γ} . We remark that this is a different choice with respect to [12], where the temporally-average power output of the multiple beamformers were considered. With this in mind, in this work we use the term *soundfield map* instead of the term *soundfield image* adopted in [12].

In [12] the authors identify an acoustic ray crossing the microphone array with the parameters (m, q) defined as

$$m_{\gamma} = \tan(\theta_{\gamma}), \quad q_i = q_0 - 2q_0(i-1)/(M-1), \quad (7)$$

which are the slope and the intercept on the y axis of the acoustic ray, respectively. We define the matrix $\mathbf{P} \in \mathbb{R}^{(M-W+1)\times\Gamma}$, which stores the values of the soundfield map as

$$[\mathbf{P}]_{i,\gamma} = p_i(\theta_\gamma). \tag{8}$$

In applications where a wideband estimate of the soundfield map is needed, the authors propose to compute the soundfield map for a discrete set of frequencies $\{\omega_k\}$, $k = 1, \ldots, K$, and then use the product of their geometric and harmonic means, as proposed in [21]. In the following, the narrowband soundfield map at temporal frequency ω_k is denoted by $\mathbf{P}(\omega_k)$, while the wideband soundfield map is denoted by $\tilde{\mathbf{P}}$.



Fig. 2: Magnitude $|\mathbf{P}|$ of an illustrative soundfield map.

For illustrative purposes, Fig. 2 shows the magnitude of the wideband soundfield map corresponding to two sound sources. The black lines $\mathcal{R}\boldsymbol{p}_{\mathrm{S},l}$, l = 1, 2, represent the linear pattern corresponding to the *l*th sound source, i.e. $\mathcal{R}\boldsymbol{p}_{\mathrm{S},l} : y_{\mathrm{S},l} = mx_{\mathrm{S},l} + q$.

4. DERIVATION OF A LINEAR OPERATOR FOR SOUNDFIELD MAPPING

With reference to the signal model introduced in Sec. 2, in this section we derive a linear operator that allows us to compute a soundfield map starting from array data. A sample of the spatial spectrum for the *i*th subarray is defined as the inner product between the subarray data and the corresponding steering vector [20, Eq. 6.3.4], i.e.

$$p_i(\theta_{\gamma}) = \frac{1}{W} \langle \mathbf{x}_i, \mathbf{a}(\theta_{\gamma}) \rangle = \frac{1}{W} \mathbf{a}^*(\theta_{\gamma}) \mathbf{x}_i.$$
(9)

From (9), the spatial spectrum for the ith subarray can be written as

$$\mathbf{p}_{i} = \begin{bmatrix} p_{i}(\theta_{1}) \\ \vdots \\ p_{i}(\theta_{\Gamma}) \end{bmatrix} = \frac{1}{W} \begin{bmatrix} \mathbf{a}^{*}(\theta_{1})\mathbf{x}_{i} \\ \vdots \\ \mathbf{a}^{*}(\theta_{\Gamma})\mathbf{x}_{i} \end{bmatrix} = \frac{1}{W} \mathbf{A}^{*}\mathbf{x}_{i}, \quad (10)$$

where \mathbf{A} is defined in (6).

As reviewed in Sec. 3, the soundfield map is defined as the matrix whose rows are the spatial spectra estimated from each subarray, i.e.

$$\mathbf{P} = \frac{1}{W} \begin{bmatrix} \mathbf{p}_1^* \\ \vdots \\ \mathbf{p}_{M-W+1}^* \end{bmatrix} = \frac{1}{W} \begin{bmatrix} (\mathbf{A}^* \mathbf{x}_1)^* \\ \vdots \\ (\mathbf{A}^* \mathbf{x}_{M-W+1})^* \end{bmatrix} = \frac{1}{W} \begin{bmatrix} \mathbf{x}_1^* \mathbf{A} \\ \vdots \\ \mathbf{x}_{M-W+1}^* \mathbf{A} \end{bmatrix},$$
(11)

Upon defining the matrix \mathbf{X} , which has subarray data on each column, i.e. $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_{M-W+1}]$, we can rewrite (11) in the more compact form

$$\mathbf{P} = \frac{1}{W} \mathbf{X}^* \mathbf{A}.$$
 (12)

As it provides a way to compute the soundfield map as a linear combination of array data, (12) can also be inverted in order to synthesize array data given a soundfield map. In



Fig. 3: Computational complexity as a function of W. M = 32, $\Gamma = 90$.

particular, we define the *right inverse* of the matrix **A**, such that $\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{I}$, as

$$\mathbf{A}^{\dagger} = \mathbf{A}^* \left(\mathbf{A} \mathbf{A}^* + \lambda \mathbf{I} \right)^{-1}, \qquad (13)$$

where Tikhonov regularization is adopted. Using the definition in (13), we can obtain array data as

$$\mathbf{X}^* = W \mathbf{P} \mathbf{A}^{\dagger}. \tag{14}$$

5. COMPUTATIONAL COMPLEXITY

Equation (12) is important since it provides a way to compute the soundfield map as a linear combination of array data. This fact has a relevant impact on all applications involving soundfield imaging, since it provides a simple and fast computational schema. In particular, the computational complexity of (12) is

$$\mathcal{O}((M - W + 1)W\Gamma) \tag{15}$$

We notice that the computational complexity is linear with respect to the total number of microphones M and with respect to the number Γ of angular directions. On the other hand, from (5) follows that the computational complexity depends on the number W of microphones in each subarray in a quadratic fashion. We observe that the computational complexity reaches its maximum when W = (M + 1)/2 (Modd) or for W = M/2 (M even). Figure 3 shows the computational complexity, expressed in FLOPS as a function of Wfor an illustrative array composed by M = 32 microphones, while Γ is set to 90: notice the maximum at W = 16.

6. SIMULATIONS

In order to validate the proposed formulation, we exploit the computational ease of (12) to simulate a localization system based on soundfield maps. In particular, we consider a moving sound source that follows a trochoidal trajectory

$$x_{\rm S}(t) = a \frac{t}{R} - b \sin\left(\frac{t}{R}\right) + 0.5, \quad y_{\rm S}(t) = a - b \cos\left(\frac{t}{R}\right), \tag{16}$$

with a = 0.1, b = 0.4, R = 1 m. The source signal is observed over a time interval of $T = 20 \,\mathrm{s}$ with a sample rate $F_s = 16 \,\mathrm{kHz}$. Wideband soundfield maps are computed in the frequency band between 300 Hz and 5 kHz. The source signal s(t) is a white random variable, so that its Fourier transform is $s(e^{j\omega}) = 1$. The microphone array is composed by M = 16 microphones between $-q_0 = -0.5$ m and $q_0 = 0.5 \,\mathrm{m}$, thus resulting in an interelement spacing equal to $d = 6.67 \,\mathrm{cm}$. The number of microphones in each subarray is fixed to W = 5. Localization is performed on short-time data segments obtained by (1) with $\tau = n0.25$ s, $n = 0, \ldots, N - 1, N = 80$; the time domain window function employed is a rectangular window. In this work we do not perform a tracking of the acoustic source, as, instead, it is done in e.g. [22] using particle filtering or in [23] using an extended Kalman filter. The simulative setup is shown in Fig. 5, where the black dots illustrate the microphone positions in the array and the red curve depicts the source trajectory.

The localization operation is performed as outlined in [12]. In particular, given the wideband soundfield map at time τ , denoted as $\tilde{\mathbf{P}}(\tau)$, peaks are identified in each row at locations $\hat{m}_i, i = 1, \ldots, M - W + 1$. For this purpose, we define the matrix $\hat{\mathbf{M}} = [-\hat{\mathbf{m}}, \mathbf{1}]$, with $\hat{\mathbf{m}} = [\hat{m}_1, \ldots, \hat{m}_{M-W+1}]^T$, being $\mathbf{1} = [1, \ldots, 1]^T \in \mathbb{R}^{(M-W+1)\times 1}$; we define also the vector $\mathbf{q} = [q_1, \ldots, q_{M-W+1}]^T$ collecting the ordinates of the reference microphones for all subarrays. With the above definitions, the source position is estimated with a least-squares approach as

$$\hat{\boldsymbol{p}}_{\mathrm{S}} = (\hat{\mathbf{M}}^T \hat{\mathbf{M}})^{-1} \hat{\mathbf{M}}^T \mathbf{q}.$$
 (17)

A Monte-Carlo simulation with $N_{\rm trials} = 100$ trials for each time step is performed to assess the localization accuracy. For this purpose, a Gaussian noise is added to microphone signals as in (3), such that the signal to noise ratio is set to 30 dB. Localization accuracy is evaluated through the root-mean-square localization error

$$\text{RMSE}(\tau) = \sqrt{\frac{1}{N_{\text{trials}}} \sum_{\eta=1}^{N_{\text{trials}}} |\boldsymbol{p}_{\text{S}}(\tau) - \hat{\boldsymbol{p}}_{\text{S}}(\tau)|^2}.$$
 (18)

Figure 4 shows the RMSE for each time step. We observe that, as expected, the localization is accurate when the sound source is close to the microphone array, while it degrades as the source moves away: this is an expected behavior, determined by the limited aperture of the microphone array.

Figure 5 shows the result of the localization procedure. In particular, the blu crosses depict the estimated source positions as a function of time, with reference to the actual source trajectory depicted by the red curve.

We remark that this simulation is not meant to be a detailed anlysis of the localization performance for the soundfield imaging approach. The motivation of this section is to depict the role of the linear operator proposed in this paper in



Fig. 4: Root-mean-square localization error. The blue line connects adjacent samples.



Fig. 5: Localized source positions for a single realization.

a more complex system, involving other tasks. In particular, we have shown that soundfield maps can be computed in an on-line fashion even on modest computing platforms.

7. CONCLUSIONS

In this paper we have shown that the whole process of soundfield mapping can be expressed as a linear transformation of the array data. We have reported an analysis of the computational complexity, which serves as a rule-of-thumb in guiding some design choices (e.g. total number of microphones, number of microphones in each subarray, number of angular directions). In order to show the applicability of the proposed formulation in a practical context, we have simulated a localization system based on soundfield maps.

8. RELATION WITH PRIOR WORK

The plenacoustic representation for sound fields has been introduced in [6, 7] and then extended in [8, 9, 10]. This paper presented a computationally efficient formulation for the soundfield imaging process introduced in [11, 12], based on a linear array of microphones, which allows us to estimate the absolute location of a sound source. This approach proved its validity for applicative purposes, as reported in [15, 16, 17].

9. REFERENCES

- B. N. Gover, J. G. Ryan, and M. R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound field," *J. Acoust. Soc. Am.*, vol. 112, no. 5, pp. 1980–1991, Nov. 2002.
- [2] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2002.
- [3] B. N. Gover, "Directional measurement of airborne sound transmission paths using a spherical microphone array," J. Audio Eng. Soc., vol. 53, no. 9, pp. 787–795, Sept. 2005.
- [4] D. Khaykin and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," J. Acoust. Soc. Am., vol. 132, no. 1, pp. 261–270, July 2012.
- [5] L. Bianchi, M. Verdi, F. Antonacci, A. Sarti, and S. Tubaro, "High resolution imaging of acoustic reflections with spherical microphone array," in *Proc. IEEE Workshop on Applications* of Signal Process. to Audio and Acoustics (WASPAA), 2015.
- [6] T. Ajdler and M. Vetterli, "The plenacoustic function, sampling and reconstruction," in *Proc. IEEE Int. Conf. on Acoustics*, *Speech, and Signal Process. (ICASSP)*, 2003.
- [7] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.
- [8] F. Pinto and M. Vetterli, "Space-time-frequency processing of acoustic wave fields: Theory, algorithms, and applications," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4608–4622, Sept. 2010.
- [9] F. Pinto and M. Vetterli, "Applications of short space-time fourier analysis in digital acoustics," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2011.
- [10] F. Pinto, M. Kolundzija, and M. Vetterli, "Digital acoustics: processing wave fields in space and time using DSP tools," *AP-SIPA Transactions on Signal and Information Processing*, vol. 3, pp. 1–21, 2014.
- [11] D. Marković, G. Sandrini, F. Antonacci, A. Sarti, and S. Tubaro, "Plenacoustic imaging in the ray space," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012.
- [12] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Sound-field imaging in the ray space," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2493–2505, Dec. 2013.
- [13] H. L. Van Trees, *Optimum Array Processing*, John Wiley & Sons, New York, NY, USA, 2002, Part IV of Detection, Estimation and Modulation Theory.
- [14] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 812–824, May 2008.
- [15] L. Bianchi, D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Deconvolution of plenacoustic images," in *Proc. IEEE Work-shop on Applications of Signal Process. to Audio and Acoustics* (WASPAA), 2013.

- [16] A. Canclini, L. Mucci, F. Antonacci, A. Sarti, and S. Tubaro, "Estimation of the radiation pattern of a violin during the performance using plenacoustic methods," in *Proc. AES 138th Conv.*, 2015.
- [17] L. Bianchi, F. D'Amelio, F. Antonacci, A. Sarti, and S. Tubaro, "A plenacoustic approach to acoustic signal extraction," in *Proc. IEEE Workshop on Applications of Signal Process. to Audio and Acoustics (WASPAA)*, 2015.
- [18] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "Multiview soundfield imaging in the projective ray space," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 6, pp. 1054–1067, June 2015.
- [19] L. Bianchi, F. Antonacci, A. Sarti, and S. Tubaro, "Modelbased acoustic rendering based on plane wave decomposition," *Applied Acoustics*, vol. 104, pp. 127–134, Mar. 2016.
- [20] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, NJ, USA, 2004.
- [21] 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," in *Proc. SPIE 5417, Unattended/Unmanned Ground, Ocean, and Air Sensor Technologies and Applications VI*, 2004.
- [22] 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. on Digital Signal Process.*, 2007.
- [23] T. G. Dvorkind and S. Gannot, "Speaker localization exploiting spatial-temporal information," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2003.