

EXTRACTION OF ACOUSTIC SOURCES FOR MULTIPLE ARRAYS BASED ON THE RAY SPACE TRANSFORM

F. Borra, F. Antonacci, A. Sarti, S. Tubaro

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

ABSTRACT

In this paper we present a source extraction technique for multiple uniform linear arrays distributed in space. The technique adopts the Ray Space Transform representation of the sound field, which is inherently based on the Plane Wave Decomposition. The Ray Space Transform gives us an intuitive representation of the acoustic field, thus enabling the adoption of geometrically-motivated constraints in the spatial filter design. The proposed approach is semi-blind since it needs as input an estimate of the source positions. We prove the effectiveness of the proposed solution through simulations using both white noise and speech signals.

Index Terms— Microphone array, near-field beamforming, source extraction, Ray Space Transform

1. INTRODUCTION

The extraction of acoustic sources using microphone arrays is a challenging and widely investigated problem in the signal processing community, e.g. [1, 2]. The majority of the algorithms in the literature address this problem using spatial filtering techniques applied to a single microphone array that samples the acoustic scene. However, in the recent literature several techniques that use distributed microphone arrays have appeared [3, 4, 5, 6, 7, 8, 9]. When multiple microphone arrays are present, the planar waveforms assumption fails, hence suitable signal models for the near-field case have to be adopted [10, 11, 12]. In this paper we propose a filter design technique for this scenario. More precisely, we adopt the Ray Space Transform representation of the sound field recently introduced in [13]. The advantage of the Ray Space Transform is that it provides analysis and synthesis operators, which exhibit perfect reconstruction capabilities, thus making it possible to perform processing operations in the Ray Space in a lossless fashion, while keeping the geometrical intuitive representation of the Ray Space [14]. The spatial filter that performs the source extraction is designed in the Ray Space in a two-steps process: in the first one we select the plane wave components that must be kept through simple geometric rules that aim at discarding the components which either exhibit the influence of the interferer or are not related to the

desired source. The second step consists of a linearly constrained minimum variance beamformer designed in the ray space.

2. FROM RAY SPACE TRANSFORM TO MULTIVIEW RAY SPACE TRANSFORM

In this section we shortly revisit the Ray Space Transform (RST) and we extend its use to multiple linear arrays distributed in the space.

The RST is based on a short space-time Fourier transform of the signals captured by a microphone array and enables to evaluate the acoustics field in different points of the space along different directions. More precisely, the signals acquired by the microphones are first converted in the frequency domain. Then, for each subband centered around frequency ω , a spatial window is applied using shifted Gaussian windows modulated by complex exponential basis functions that span a set of prescribed directions. Assuming a continuous aperture displaced along the y axis from $-q_0$ to q_0 this can be formalized as

$$[\mathbf{Y}]_{i,w}(\omega) = \int_{-q_0}^{q_0} p(y, \omega) e^{-\frac{j\omega}{\sqrt{1+m_w^2}} y m_w} \psi_{i,w}^*(y) dy, \quad (1)$$

where ω is the temporal frequency, c is the speed of sound, $i = 0, \dots, I-1$ is the index of the spatial window, $\frac{m_w}{\sqrt{1+m_w^2}}$ with $w = 0, \dots, W-1$ encodes the modulation angle of the complex exponential, $p(y, \omega)$ is the signal acquired at position y and ψ^* is the window. Despite different windows can be employed [14], the Gaussian one offers perfect reconstruction capabilities.

The parametrization of $[\mathbf{Y}]_{i,w}(\omega)$ depends on the indexes i and w that control the position of the window along the array and the angle towards which the window is “tuned”, respectively. In [13], the ray space parameterization is adopted, in which a plane wave component is represented by an acoustic ray, i.e. the direction locally orthogonal to the plane wave. In particular, acoustic rays are parameterized by the parameters (m, q) , where m represents the slope and q the intercept of the line on which the acoustic ray lies [15]. However, as underlined in [16], this representation of the acoustic rays is

not suitable when multiple arrays are present in the acoustic scene due to the singularity of the mapping for rays parallel to y axis. A projective parameterization of the rays enables to remove this singularity. In [16] it is also shown how to convert the ray space attached to each array to the global projective ray space. However, the transformation of the array data operated in [16] is not invertible.

In this paper we apply the geometrical framework of the projective ray space to the Ray Space Transform, thus enabling processing of the array data in the ray space in a lossless fashion. As done in [13], we transform the integral equation in (1), which can be repurposed for the a th discrete array ($a = 1, \dots, A$) as

$$\mathbf{y}^{(a)}(\omega) = \left(\Psi^{(a)}(\omega) \right)^H \mathbf{p}^{(a)}(\omega), \quad (2)$$

where $\mathbf{p}^{(a)}(\omega)$ is a vector that contains the signals acquired by the microphones in the a th array, $\mathbf{y}^{(a)}(\omega)$ contains the transform coefficients associated to each pair $(m^{(a)}, q^{(a)})$ and the elements of the matrix $\Psi^{(a)}(\omega)$ are given by

$$[\Psi]_{l, i+wI^{(a)}+1}^{(a)}(\omega) = d^{(a)} e^{j \frac{\omega}{c} \frac{y m_w^{(a)}}{\sqrt{1+(m_w^{(a)})^2}}} e^{-\frac{\pi(y_l - \bar{q}_i^{(a)})^2}{\sigma^2}}, \quad (3)$$

where $l = 1, \dots, L^{(a)}$ is the index of the microphone and $d^{(a)}$ is the microphone spacing in the a th array, i.e. without loss of generality we consider uniform linear arrays in this work. We can now use the transformation described in [16] in order to map the value in $\mathbf{y}^{(a)}(\omega)$ to a global ray space where each point is identified by the parameters $(l_{1,G}, l_{2,G}, l_{3,G})$ of the homogeneous line equation $l_{1,G}x + l_{2,G}y + l_{3,G} = 0$. If we assume that each ray is visible only from at most a single array, we can construct a block diagonal matrix to obtain all transform coefficients for all the arrays in one go as

$$\Psi_G = \text{diag}(\Psi^{(1)}, \dots, \Psi^{(A)}) \quad (4)$$

Moreover, if we juxtapose all the vectors $\mathbf{p}^{(a)}$ in $\mathbf{p}_G = [\mathbf{p}^{(1)T} \dots \mathbf{p}^{(A)T}]^T$, we can write

$$\mathbf{y}_G = \Psi_G^H \mathbf{p}_G. \quad (5)$$

This equation allows us to transform the signals acquired by all the arrays to a global projective ray space domain. It is worth noticing that when we parametrize the rays in a projective form, we obtain infinitely many triplets that equivalently describe the same ray, i.e. $(l_{1,G}, l_{2,G}, l_{3,G}) \equiv \alpha(l_{1,G}, l_{2,G}, l_{3,G})$, $\alpha \in \mathbb{R}$. The parameter α can be arbitrarily chosen.

3. SIGNAL EXTRACTION USING MULTIPLE ARRAY

In this section, we go through a filtering technique developed in the projective Ray Space domain in order to extract the signal emitted by a desired source in the presence of interferers

using multiple arrays. The method is semi-blind since the position of the sources has to be known in advance or estimated for example with the method proposed in [16].

The model adopted for the signal received by the a th array is

$$\mathbf{p}^{(a)}(\omega) = \mathbf{F}^{(a)}(\omega) \mathbf{s}(\omega) + \mathbf{e}^{(a)}(\omega), \quad (6)$$

where $\mathbf{F}^{(a)} = [\mathbf{f}_1^{(a)}, \dots, \mathbf{f}_N^{(a)}]$ is the matrix that contains the transfer function vectors from the N sources to the sensors of the a th array, which counts $L^{(a)}$ microphones, while $\mathbf{s} = [S_1, \dots, S_N]^T$ contains the signals emitted by the N sources and $\mathbf{e}^{(a)} = [E_1^{(a)}, \dots, E_{L^{(a)}}^{(a)}]^T$ is the additive noise vector. The dependency on the temporal frequency ω will be omitted for brevity. If we want to model the overall output signal of all the microphone arrays present in the scene we can write

$$\begin{bmatrix} \mathbf{p}^{(1)} \\ \vdots \\ \mathbf{p}^{(A)} \end{bmatrix} = \begin{bmatrix} \mathbf{F}^{(1)} \\ \vdots \\ \mathbf{F}^{(A)} \end{bmatrix} \mathbf{s}(\omega) + \begin{bmatrix} \mathbf{e}^{(1)} \\ \vdots \\ \mathbf{e}^{(A)} \end{bmatrix} \quad (7)$$

$$\mathbf{p}_G = \mathbf{F}_G \mathbf{s} + \mathbf{e}_G.$$

As seen in (5), we can map the signal \mathbf{p}_G on the projective ray space by using the matrix Ψ_G and we can directly apply a filter \mathbf{h} in this domain to obtain

$$y = \mathbf{h}^H \Psi_G^H \mathbf{p}_G = \mathbf{h}^H \Psi_G^H \mathbf{F}_G \mathbf{s} + \mathbf{h}^H \Psi_G^H \mathbf{e}_G = y_p + y_e, \quad (8)$$

where $\mathbf{h} \in \mathbb{C}^{(\sum_{a=1}^A L^{(a)} W^{(a)}) \times 1}$ is the vector of filter coefficients. Notice that the number of filter coefficients depends on the number of arrays and how densely we sample the ray space, thus it can become very large.

In order to overcome this problem, the source extraction approach presented in this section is divided in two steps: the first one is used to select which elements in the projective ray space must be retained by setting some filter coefficients to zero, while in the second an optimization problem is solved in order to set the filter coefficients that have not been zeroed in the first step. The following two paragraphs describe the two steps separately.

3.1. Source Extraction Algorithm: first step

We take advantage of the intuitive projective ray space representation of a signal. In particular, we can think of setting to zero some components of the ray space that do not honor prescribed constraints. We assume the location of the desired source to be known (see [16] for efficient localization in the projective ray space), and we set to zero the elements of the ray space relative to rays that do not pass in its neighborhood. More precisely, we can think of drawing a circle of radius ϱ_n around each source. Formally, this is given by

$$\mathcal{S}_n = \left\{ \mathbf{l}_G \mid \frac{\langle \mathbf{l}_G, \mathbf{r}'_n \rangle}{\|\mathbf{l}_G\|} \leq \varrho_n \right\}, \quad (9)$$

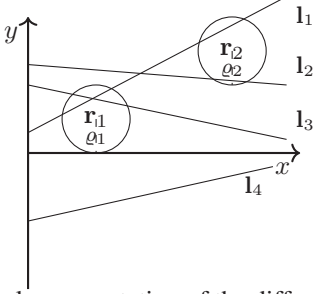


Fig. 1. Graphical representation of the different types of rays.

where $\mathbf{U} = \text{diag}(1, 1, 0)$, $\mathbf{r}'_n = [x_n, y_n, 1]$ is the position vector of the n th source in homogeneous coordinates. Each one of these sets contains all the rays whose distance from the position of the n th source is less than ϱ_n . If we assume, without loss of generality, that the first source is the desired one, we can define the set of rays to be retained as

$$\mathring{\mathcal{S}} = \mathcal{S}_1 - \left[\mathcal{S}_1 \cap \left(\bigcup_{n=2}^N \mathcal{S}_n \right) \right]. \quad (10)$$

If \mathcal{L}_G is the set of all rays, we can define

$$\check{\mathcal{S}} = \mathcal{L}_G - \mathring{\mathcal{S}}, \quad (11)$$

as the set of discarded rays. In Fig. 1 an illustration of the different types of rays is shown. In this particular case we have that $\{l_1, l_2, l_3, l_4\} \in \mathcal{L}_G$, $\{l_1, l_3\} \in \mathcal{S}_1$, $\{l_1, l_2\} \in \mathcal{S}_2$, $\{l_3\} \in \mathring{\mathcal{S}}$ and $\{l_1, l_2, l_4\} \in \check{\mathcal{S}}$. Once we have identified which rays belong to the set $\check{\mathcal{S}}$, we can set to zero all the coefficients of $\mathbf{h}(\omega)$ relative to the triplets in this set.

The radius ϱ_n is given by

$$\varrho_n = \min \left(\left(\min_{i=1, \dots, L_{\text{tot}}} \|\mathbf{r}_i - \mathbf{r}'_n\| \right)^{-1}, \ell_n \right), \quad (12)$$

where ℓ_n is an arbitrary threshold value, $L_{\text{tot}} = \sum_{a=1}^A L^{(a)}$ and $\mathbf{r}'_n = [x'_n, y'_n]^T$. The rationale behind this policy is that an array close to the source can “see” it easily and, at the same time, if this source is an interferer it “occludes” the view of the array. Hence, the radius increases with the reciprocal of the distance from the n th source to the nearest microphone. The threshold value ℓ_n is useful in order to avoid the situation $\mathring{\mathcal{S}} = \emptyset$ caused by a big value of ϱ_n .

3.2. Source Extraction Algorithm: second step

The design of the coefficients of the spatial filter in $\mathring{\mathcal{S}}$ is accomplished through a linearly constrained optimization problem. Let us denote with $\mathring{\mathbf{h}}(\omega)$ and $\mathring{\Psi}_G(\omega)$ the filter coefficients of $\mathbf{h}(\omega)$ and the matrix obtained by the columns of Ψ_G relative to the rays in $\mathring{\mathcal{S}}$, respectively. We impose that the desired signal remains unchanged while interferers are attenuated. Additionally, we aim at gaining robustness against

localization errors. A near-field model of the transfer function $\mathbf{F}_G(\omega)$ is adopted, i.e

$$\mathbf{f}_n(\omega) = \left[\frac{\rho_n e^{-j \frac{\omega}{c} (\|\mathbf{r}_1 - \mathbf{r}'_n\| - \rho_n)}}{\|\mathbf{r}_1 - \mathbf{r}'_n\|}, \dots, 1, \dots, \frac{\rho_n e^{-j \frac{\omega}{c} (\|\mathbf{r}_{L_{\text{tot}}} - \mathbf{r}'_n\| - \rho_n)}}{\|\mathbf{r}_{L_{\text{tot}}} - \mathbf{r}'_n\|} \right]^T, \quad (13)$$

where $\rho_n = \sqrt{x_n'^2 + y_n'^2}$ is the distance from the origin of the global reference frame. As in the classical LCMV beam-former design [17, Chap. 6], we minimize the variance of the noise component of the output of the filter $y_e(\omega)$, that in this case is defined as

$$E[y_e^2] = \mathbf{h}^H \Psi_G^H E[\mathbf{e}_G \mathbf{e}_G^H] \Psi_G \mathbf{h}. \quad (14)$$

If we assume that the noise components in the microphone signals are spatially white (i.e $E[\mathbf{e}_G \mathbf{e}_G^H] = \mathbf{I}$) and we consider only the filter coefficients that have not been zeroed in the first step, we can rewrite (14) as $E[y_e^2] = \mathbf{h}^H \mathring{\Psi}_G^H \mathring{\Psi}_G \mathbf{h}$. The distortionless constraint of the desired source imposes that

$$\mathring{\mathbf{h}}^H \mathring{\Psi}_G^H \mathbf{f}_1 = 1. \quad (15)$$

Conversely, interferers are attenuated by imposing that

$$\mathring{\mathbf{h}}^H \mathring{\Psi}_G^H \mathbf{f}_n = \epsilon_n \quad n = 2, \dots, N, \quad (16)$$

where $\epsilon_n \in \mathbb{R}^+$ is a small value less than one that sets the amplitude of the spatial response in the n th interferer location. This value can also be set to zero, but this is not a convenience choice because it decreases the performances in terms of robustness to localization errors [18]. In order to further gain robustness against localization errors we impose that

$$\begin{aligned} \frac{\partial \mathbf{f}_n}{\partial x} &= \left[\frac{\partial F_{1,n}}{\partial x} \dots, \frac{\partial F_{L_{\text{tot}},n}}{\partial x} \right]^T = 0 \quad n = 2, \dots, N \\ \frac{\partial \mathbf{f}_n}{\partial y} &= \left[\frac{\partial F_{1,n}}{\partial y} \dots, \frac{\partial F_{L_{\text{tot}},n}}{\partial y} \right]^T = 0 \quad n = 2, \dots, N, \end{aligned} \quad (17)$$

where

$$\begin{aligned} \frac{\partial F_{i,n}}{\partial x} &= \left[\frac{x'_n}{\rho_n} \left(\frac{1}{\rho_n} + j \frac{\omega}{c} \right) - \frac{x'_n - x_i}{\|\mathbf{r}_i - \mathbf{r}'_n\|} \left(\frac{1}{\|\mathbf{r}_i - \mathbf{r}'_n\|} + j \frac{\omega}{c} \right) \right] F_{i,n}, \\ \frac{\partial F_{i,n}}{\partial y} &= \left[\frac{y'_n}{\rho_n} \left(\frac{1}{\rho_n} + j \frac{\omega}{c} \right) - \frac{y'_n - y_i}{\|\mathbf{r}_i - \mathbf{r}'_n\|} \left(\frac{1}{\|\mathbf{r}_i - \mathbf{r}'_n\|} + j \frac{\omega}{c} \right) \right] F_{i,n}, \end{aligned} \quad (18)$$

with $\mathbf{r}_i = [x_i, y_i]^T$ being the coordinates of the i th microphone. The final optimization problem results to be

$$\begin{aligned} \mathring{\mathbf{h}}_0 &= \arg \min \quad \mathring{\mathbf{h}}^H \mathbf{A} \mathring{\mathbf{h}} \\ \text{subject to} \quad & \mathring{\mathbf{h}}^H \mathbf{B} = \mathbf{c}, \end{aligned} \quad (20)$$

where $\mathbf{A} = \mathring{\Psi}_G^H \mathring{\Psi}_G$ and all the constraints have been written in the matrix form $\mathring{\mathbf{h}}^H \mathbf{B} = \mathbf{c}$. In particular,

$$\mathbf{B} = \left[\mathring{\Psi}_G^H \mathbf{f}_1, \dots, \mathring{\Psi}_G^H \mathbf{f}_N, \frac{\partial \mathbf{f}_2}{\partial x} \Big|_{\mathbf{r}'=\mathbf{r}'_2}, \dots, \frac{\partial \mathbf{f}_N}{\partial x} \Big|_{\mathbf{r}'=\mathbf{r}'_N}, \frac{\partial \mathbf{f}_2}{\partial y} \Big|_{\mathbf{r}'=\mathbf{r}'_2}, \dots, \frac{\partial \mathbf{f}_N}{\partial y} \Big|_{\mathbf{r}'=\mathbf{r}'_N} \right], \quad (21)$$

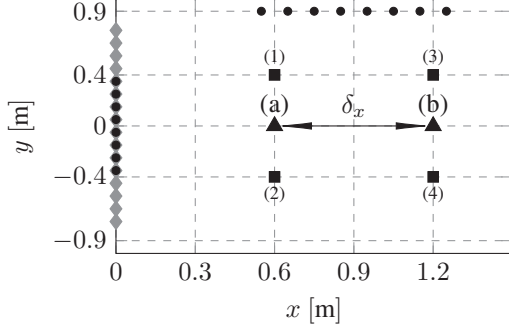


Fig. 2. Simulation configuration

Table 1. Filter parameters

Parameter	Value
ϵ_n	0.01 (40 dB)
ι_n	50 cm (target source)
	10 cm (interfering sources)
$(\sigma^{(1)})^2 = (\sigma^{(2)})^2 = \sigma^2$	$\sigma^2 = \max(\min(c/f, 0.8\ell), d)$
$\bar{q}^{(1)} = \bar{q}^{(2)}$	$d^{(1)} = d^{(2)}$
$\bar{m}^{(1)} = \bar{m}^{(2)}$	0.12
$I^{(1)} = I^{(2)}$	$L^{(1)} = L^{(2)}$
$W^{(1)} = W^{(2)}$	100

and

$$\mathbf{c} = [1, \epsilon_2, \dots, \epsilon_N, \mathbf{0}], \quad (22)$$

where $\mathbf{0} \in \mathbb{R}^{1 \times 2(N-1)}$ is a vector of all zeros. The solution to (20) is [19]

$$\mathbf{h}_0 = \mathbf{A}^{-1} \mathbf{B} (\mathbf{B}^H \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{c}^H. \quad (23)$$

Once obtained \mathbf{h}_0 from (23), we can construct the overall filter \mathbf{h} by placing the coefficients that have been set to zero in the first step and the ones obtained in the second step at the right place in the vector $\mathbf{h}(\omega)$.

4. SIMULATIONS

We validate the proposed method using two different simulations. The benchmark algorithm is a near-field beamformer robust against localization errors [20]. We use two metrics: the ΔSINR (Delta Signal to Interferences plus Noise Ratio) and the TPS (Target related perceptual score). Both these metrics have been computed using the freely available tool presented in [21].

The signal is transformed in the temporal frequency domain with an STFT (Short-Time-Fourier-Transform) of 1024 samples Hanning window and 256 samples of hop size. The setup of all the experiments is shown in Fig. 2 where the microphone spacing, both for the single and for the multiple arrays case is fixed at $d = 10$ cm. The filter parameters are summarized in Table. 1.

In the first simulation, we have studied the behavior of our approach in terms of the ΔSINR as a function of the distance δ_x between the two sources denoted by (a) and (b) in

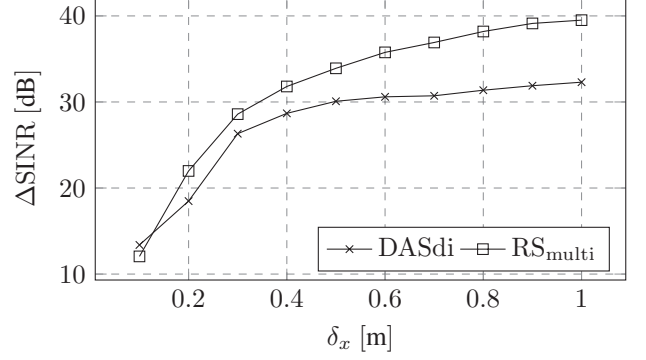


Fig. 3. ΔSINR as a function of the sources mutual distance

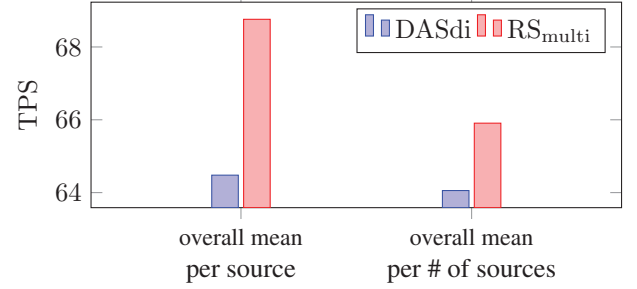


Fig. 4. Target related perceptual score

Fig. 2. In particular, we have considered the source denoted with (b) as the desired one while the other as interfering. The desired and interference signals are white noise signals with identical Power Spectral Density. As we can see from Fig. 3 the methodology proposed, namely RS_{multi} , outperforms the DASdi method arriving to obtain an improvement of about 10 dB when the distance between the two sources is 1 m.

In the second simulation, we have studied the behavior of our methodology in terms of TPS by varying both the number of sources and the desired source extracted. The signal used for the four sources are speech signals taken from [22]. In particular, with reference to the setup in Fig. 2, we have simulated all possible ordered selections with 2,3 and 4 elements. In Fig. 4 two bar plots report the TPS overall mean based on the selected desired source and on the number of sources active in the scene. As we can see for both the situations our approach improves, even if both algorithms perform quite well, ranging TPS from 0 to 100.

5. CONCLUSIONS

In this paper we have proposed a method for the extraction of sound sources using multiple microphone arrays. The design of the filter is done using an intuitive and unified representation of the sound field captured by the different arrays. An optimization problem is set up in a two stage fashion in order to minimize the undesired signal contributions and increase the robustness to localization errors in. We have provided numerical evaluations to prove the effectiveness of the proposed solution.

6. REFERENCES

- [1] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*, Springer-Verlag, Berlin, DE, 2005.
- [2] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer-Verlag, Berlin, DE, 2008.
- [3] Zicheng Liu, “Sound source separation with distributed microphone arrays in the presence of clock synchronization errors,” in *Proc. IWAENC*, 2008, pp. 1–4.
- [4] Axel Plinge and Gernot A Fink, “Multi-speaker tracking using multiple distributed microphone arrays,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 614–618.
- [5] Maja Taseska and Emanuël AP Habets, “Spotforming: Spatial filtering with distributed arrays for position-selective sound acquisition,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 7, pp. 1291–1304, 2016.
- [6] M. Taseska and E. A. P. Habets, “Informed spatial filtering for sound extraction using distributed microphone arrays,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 7, pp. 1195–1207, July 2014.
- [7] Toby Christian Lawin-Ore, Sebastian Stenzel, Jürgen Freudenberger, and Simon Doclo, “Alternative formulation and robustness analysis of the multichannel wiener filter for spatially distributed microphones,” in *Acoustic Signal Enhancement (IWAENC), 2014 14th International Workshop on*. IEEE, 2014, pp. 208–212.
- [8] Yuan Zeng and Richard C. Hendriks, “Distributed delay and sum beamformer for speech enhancement via randomized gossip,” *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 1, pp. 260–273, 2014.
- [9] S. Gannot and A. Bertrand, “Introduction to distributed speech enhancement algorithms for ad hoc microphone arrays & wireless acoustic sensor networks,” 2013, Tutorial presented at the 21st European Signal Processing Conference, *EUSIPCO*.
- [10] E. Fisher and B. Rafaely, “Near-field spherical microphone array processing with radial filtering,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 2, pp. 256–265, Feb. 2011.
- [11] S. Doclo and M. Moonen, “Design of far-field and near-field broadband beamformers using eigenfilters,” *Signal Processing*, vol. 83, pp. 2641–2673, 2003.
- [12] R. A. Kennedy, T. D. Abhayapala, and D. B. Ward, “Broadband nearfield beamforming using a radial beam-pattern transformation,” *IEEE Trans. Signal Process.*, vol. 46, no. 8, pp. 2147–2156, Aug. 1998.
- [13] Lucio Bianchi, Fabio Antonacci, Augusto Sarti, and Stefano Tubaro, “The ray space transform: A new framework for wave field processing,” *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5696–5706, 2016.
- [14] L. Bianchi, V. Baldini, D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, “A linear operator for the computation of soundfield maps,” in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, 2016.
- [15] D. Markovic, F. Antonacci, A. Sarti, and S. Tubaro, “Soundfield imaging in the ray space,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2493–2505, Dec 2013.
- [16] D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, “Multiview soundfield imaging in the projective ray space,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 6, pp. 1054–1067, June 2015.
- [17] H. L. Van Trees, *Optimum Array Processing*, John Wiley & Sons, New York, NY, USA, 2002, Part IV of Detection, Estimation and Modulation Theory.
- [18] Rui Li, Xin Zhao, and Xiao-Wei Shi, “Derivative constrained robust lcmv beamforming algorithm,” *Progress In Electromagnetics Research C*, vol. 4, pp. 43–52, 2008.
- [19] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, NJ, USA, 2004.
- [20] F. Borra, L. Bianchi, F. Antonacci, S. Tubaro, and A. Sarti, “A robust data-independent near-field beamformer for linear microphone arrays,” in *2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sept 2016, pp. 1–5.
- [21] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann, “Subjective and objective quality assessment of audio source separation,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 2046–2057, Sept. 2011.
- [22] “Sound quality assessment material recording for subjective tests,” <https://tech.ebu.ch/publications/sqamcd>, 2008.