Reconstruction of the Virtual Microphone Signal Based on the Distributed Ray Space Transform

1st Mirco Pezzoli, *Politecnico di Milano*, Milan, Italy, mirco.pezzoli@polimi.it
2nd Federico Borra, *Politecnico di Milano*, Milan, Italy, federico.borra@polimi.it
3rd Fabio Antonacci, *Politecnico di Milano*, Milan, Italy, fabio.antonacci@polimi.it
4th Augusto Sarti, *Politecnico di Milano*, Milan, Italy, augusto.sarti@polimi.it
5th Stefano Tubaro, *Politecnico di Milano*, Milan, Italy, stefano.tubaro@polimi.it

***Abstract*—In this paper we propose a technique for the reconstruction of the sound field at arbitrary positions based on a parametric sound field description. The methodology consists in the estimation of the sources model parameters (source position, radiation pattern and source signal), starting from the signals acquired by arbitrarily distributed microphone arrays. Given the model parameters it is possible to synthesize the signal of a virtual microphone at an arbitrary position and with an arbitrary pick-up pattern.**

***Index Terms*—virtual microphone, distributed microphone networks, source localization**

I. Introduction

In the literature, the sound field reconstruction problem has been approached by different techniques [1]–[4]. The aim is to reproduce the sound image at the listener as it was at the recording location. The *Virtual Miking* (VM) is an emerging application in space-time processing which concerns the sound field reconstruction. The goal of VM is the synthesis of a *virtual microphone* signal arbitrarily placed in the space with an arbitrary pick-up pattern. This is possible only if the acoustic field is known or can be reconstructed at the desired location. One of the ways to analyze the acoustic scene is to use a network of circular compact microphone arrays arbitrarily placed around the sources of interest. This acquisition scheme allows us to adopt a geometry-based acoustic representation of the sound field [5] defined as a parametrization of the plenacoustic function [6] in terms of the acoustic rays. The acoustic information is later mapped to the ray space by means of a novel tool called Distributed Ray Space Transform (DRST). The DRST is introduced as a generalization of the Ray Space Transform, presented in [7], to multiple and arbitrarily shaped arrays.

Our approach to VM is similar to the ones presented in [4] and [8], it enables to keep into account the radiation patterns of the sources and to specify an arbitrary pick-up pattern of the virtual microphone. We adopt a parametric description of the acoustic scene in which the parameters are related to the source: the source position, the source radiance pattern and the source signal. These parameters can be estimated individually thanks to the adoption of the DRST. The VM signal at the desired location and with the desired pick-up pattern can be synthesized by applying the proper acoustic propagation model with the estimated parameters. In this paper a free field scenario is adopted, but the DRST can be employed also in reverberant environments.

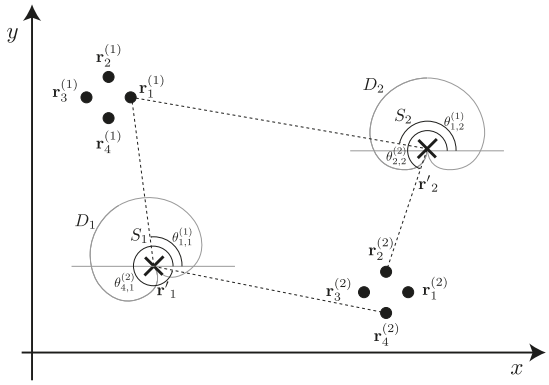

Fig. 1: Graphical representation of the analysis model

The paper is structured as follows: in Sec. II the parametric sound field model is introduced and the problem is formulated. Sec. III presents the estimation of the model parameters. In Sec. IV the synthesis of the VM signal is described. Sec. V reports the results obtained during simulations and experiments. Finally, Sec. VI draws conclusions.

II. Signal Model

Let us define the signal of a generic microphone m assuming, for the sake of simplicity, that sources and microphones lie all on the same plane:

$$P(\mathbf{r}_m,\omega)=\sum_{n=0}^{N-1} V_m(\theta_{n,m},\omega)D_n(\theta_{m,n},\omega)g(\mathbf{r}_m,\mathbf{r'}_n,\omega)S_n(\omega)+e_m(\omega), \tag{1}$$

where ω is the temporal frequency, $\mathbf{r}_m$ is the vector representing the mth microphone location, N is the number of sources, $\mathbf{r'}_n$ is the position of the nth source, $S_n(\omega)$ represents the signal emitted by the nth source and $e_m(\omega)$ is the microphone self-noise. The function $D_n(\theta_{m,n},\omega):\mathbb{R}^2\rightarrow(0,1)$ represents the directivity pattern, also known as radiation pattern of the nth source for the angle $\theta_{m,n}=\angle(\mathbf{r}_m-\mathbf{r'}_n)$, while the function $V_m(\theta_{n,m},\omega):\mathbb{R}^2\rightarrow(0,1)$ represents the pick-up pattern of the mth microphone for the angle $\theta_{n,m}=$

ISBN 978-90-827970-1-5

$\angle(\mathbf{r}'_n - \mathbf{r}_m)$. Finally, $g(\mathbf{r}_m, \mathbf{r'}_n, \omega)$ represents the well known Green's function [9] defined as:

$$g(\mathbf{r}_m, \mathbf{r'}_n, \omega) = \frac{e^{-jk\|\mathbf{r}_m - \mathbf{r'}_n\|}}{4\pi\|\mathbf{r}_m - \mathbf{r'}_n\|}, \tag{2}$$

where $k = \omega/c$ with c the speed of sound and $\|\cdot\|$ indicates the ℓ_2 norm. If we consider an analysis setup composed by a distributed microphone network of A compact arrays with M microphones each (see Fig. 1), we can write (1) as

$$\mathbf{p}^{(a)} = (\mathbf{V}^{(a)} \otimes \mathbf{D}^{(a)} \otimes \mathbf{G}^{(a)})\mathbf{s} + \mathbf{e}^{(a)}, \quad \forall a = 0, \dots, A-1, \tag{3}$$

where

$$\begin{aligned}
&\mathbf{p}^{(a)} = [P(\mathbf{r}_0^{(a)}), \dots, P(\mathbf{r}_{M-1}^{(a)})]^T \\
&\mathbf{s} = [S_0, \dots, S_{N-1}]^T \\
&\mathbf{e}^{(a)} = \left[e_0^{(a)}, \dots, e_{M-1}^{(a)}\right]^T \\
&\left[\mathbf{V}^{(a)}\right]_{m,n} = V_m(\theta_{n,m}^{(a)}) \\
&\left[\mathbf{D}^{(a)}\right]_{m,n} = D_n(\theta_{m,n}^{(a)}) \\
&\left[\mathbf{G}^{(a)}\right]_{m,n} = g(\mathbf{r}_m^{(a)}, \mathbf{r'}_n),
\end{aligned} \tag{4}$$

a is the array index and $\otimes$ is the Hadamard product. The variable ω has been omitted for convenience of notation. In the rest of the paper we consider omnidirectional microphones, i.e. $V_m(\theta_{n,m}^{(a)}, \omega) = 1\,\forall m, a, \omega$ while, in the synthesis stage, the pick-up patterns of the VMs are arbitrary. Hence, in order to compute the VM signals, we need to estimate three parameters for each source: its location, its radiance pattern and the emitted signal.

III. Sound field Analysis

A. Source localization

Assuming that the microphones locations are known, the transfer function model (2) requires the localization of multiple sources in the acoustic scene. In this paper we introduce the Distributed Ray Space Transform (DRST) as a tool for source localization. Recently, the Ray Space Transform (RST) [7] has proved to be a convenient tool for mapping the acoustic information in the ray space. In this domain, the main acoustic primitives (sources, array and reflectors) are mapped onto lines due to the fact that each point in the ray space corresponds to an acoustic ray in the geometric space, and the line parameters are uniquely related to the source location. As a consequence, we can easily perform a robust multiple source localization through pattern analysis.

The DRST inherits the core idea of the RST, but differently from the RST which is based on multiple beamforming operations performed on sub-arrays of a single extended linear array, the DRST carries out the beamforming independently for each array within a network of distributed compact arrays. Morever, DRST maps acoustic information in the projective ray space [10]. The acoustic rays are represented by their projective coordinates $\mathbf{l} = [l_1, l_2, l_3]^T$, which uniquely define the line on which the ray lies, i.e. $l_1 x + l_2 y + l_3 = 0$. Hence, in the projective ray space the acoustic primitives are mapped onto planes, whose parameters depend on the primitives position. The DRST consists in two steps: the computation of the pseudospectrum for each array a, as a function of the angle $\alpha \in [0, 2\pi)$ through a Delay and Sum beamforming operation [11] and the subsequent mapping in the projective ray space as

$$\begin{aligned}
&l_1^{(a)} = \gamma \sin(\alpha^{(a)}); \\
&l_2^{(a)} = \gamma \cos(\alpha^{(a)}); \\
&l_3^{(a)} = \gamma[y^{(a)} \cos(\alpha^{(a)}) - x^{(a)} \sin(\alpha^{(a)})], \gamma > 0,
\end{aligned} \tag{5}$$

where $x^{(a)}, y^{(a)}$ stand for the array reference point coordinates. Then, a robust linear regression (RANSAC [12]) is performed on the $N \times A$ highest peaks, in order to identify the planes relative to the sources. Finally, we can obtain an estimate of the sources positions $\hat{\mathbf{r}}'_n\,\forall n = 0, \dots N-1$ from the identified planes parameters. For further details on projective ray space and source localization please refer to [10].

B. Radiance pattern estimation

Let us assume the distance between sources and arrays being much greater than the arrays size. Therefore, in (3) we can assume $\theta_{m,n}^{(a)} \approx \theta_n^{(a)} = \angle(\mathbf{r}^{(a)} - \mathbf{r'}_n)$ for each microphone m in the ath array, where $\mathbf{r}^{(a)}$ stands for the reference point of the ath array (e.g center of gravity). With this assumption, the matrix $\mathbf{D}^{(a)}$ becomes $\left[\mathbf{D}^{(a)}\right]_{m,n} = D_n(\theta_n^{(a)})$. Hence, defining $\mathbf{P}^{(a)} = \mathrm{diag}(D_0(\theta_0^{(a)}), \dots, D_{N-1}(\theta_{N-1}^{(a)}))$, the model in (3) reduces to

$$\mathbf{p}^{(a)} = \mathbf{G}^{(a)}(\mathbf{P}^{(a)}\mathbf{s}) + \mathbf{e}^{(a)}, \quad \forall a = 0, \dots, A-1. \tag{6}$$

After source localization, we have an estimate $\hat{\mathbf{G}}^{(a)}$ of $\mathbf{G}^{(a)}$, thus we can estimate the vector $\mathbf{b}^{(a)} = \mathbf{P}^{(a)}\mathbf{s}$ through an LCMV beamformer [11] defined by the optimization problem

$$\mathbf{h}_n^{(a)} = \underset{\mathbf{h}}{\mathrm{argmin}}\ \mathbf{h}^H\mathbf{h} \qquad \text{s.t. } \mathbf{h}^H\hat{\mathbf{G}}^{(a)} = \mathbf{c}, \tag{7}$$

where $\mathbf{c} \in \mathbb{C}^{1\times N}$ with $[\mathbf{c}]_i = 1$ for $i = n$ and zero otherwise. The solution to the optimization problem given by [13] is

$$\mathbf{h}_n^{(a)} = \hat{\mathbf{G}}^{(a)}\left(\hat{\mathbf{G}}^{(a)^H}\hat{\mathbf{G}}^{(a)}\right)^{-1}\mathbf{c}^H. \tag{8}$$

Finally, the estimate of $\mathbf{b}^{(a)}$ is computed as

$$\hat{\mathbf{b}}^{(a)} = \mathbf{H}^{(a)^H}\mathbf{p}^{(a)} \quad \forall a = 0, \dots, A-1, \tag{9}$$

where $\mathbf{H}^{(a)} = \left[\mathbf{h}_0^{(a)}, \dots, \mathbf{h}_{N-1}^{(a)}\right]$. Let us define the vector $\mathbf{q}_n$

$$\begin{aligned}
\mathbf{q}_n &= \left[|\left[\hat{\mathbf{b}}^{(0)}\right]_n|, \dots, |\left[\hat{\mathbf{b}}^{(A-1)}\right]_n|\right]^T \\
&= |\hat{S}_n|\left[\hat{D}_n(\theta_n^{(0)}), \dots, \hat{D}_n(\theta_n^{(A-1)})\right]^T,
\end{aligned} \tag{10}$$

this represents an estimate of the nth source directivity pattern for the directions $\theta_n^{(a)}\,\forall a = 0, \dots, A-1$ scaled by the factor

ISBN 978-90-827970-1-5 © EURASIP 2018

$|\hat{S}_n|$. In order to reconstruct the sound field, we need the whole radiance pattern $D_n(\cdot)$. Hence, a model for $D_n(\cdot)$ is adopted

$$D_n(\theta_n^{(a)}) = \sum_{l=0}^{L-1} w_{n,l}\cos(l\theta_n^{(a)}) + r_{n,l}\sin(l\theta_n^{(a)}). \qquad (11)$$

Given the matrix $\mathbf{A}_n = [\mathbf{A}_{n,1}, \mathbf{A}_{n,2}]$, with $[\mathbf{A}_{n,1}]_{a,l} = \cos(l\theta_n^{(a)})$, $[\mathbf{A}_{n,2}]_{a,l} = \sin(l\theta_n^{(a)})$ and $\mathbf{y}_n = [w_{n,0}, \ldots, w_{n,L-1}, r_{n,0}, \ldots, r_{n,L-1}]^T$, an estimate of the coefficients $\mathbf{y}_n$ can be found by solving the optimization problem [14]

$$\hat{\mathbf{y}}_n = \underset{\mathbf{y}_n}{\operatorname{argmin}} \ \|\mathbf{q}_n - \mathbf{A}_n\mathbf{y}_n\|^2 \qquad \text{s.t. } \mathbf{F}\mathbf{y}_n \geq \mathbf{0}, \qquad (12)$$

where $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2] \in \mathbb{D}^{I\times 2L}$, $[\ \mathbf{F}_1]_{i,l} = \cos(l\phi_i)$, $[\ \mathbf{F}_2]_{i,l} = \sin(l\phi_i)$ with $\phi_i \in [0, 2\pi)$ $i = 0, \ldots, I-1$. It is worth noting that using the result of (12) in (11) we will obtain a scaled version of the radiance pattern $\hat{\mathbf{D}}_n$ due to (10).

C. Signal estimation

The estimation of the source signal $S_n(\omega)$ takes advantage of the previously estimated parameters. Specifically, an informed spatial filter is designed for each source n as

$$\mathbf{x}_n^\star = \underset{\mathbf{x}}{\operatorname{argmin}} \ \mathbf{x}^H\mathbf{x} \qquad \text{s.t. } \mathbf{x}^H\mathbf{Y} = \mathbf{d}, \qquad (13)$$

where $\mathbf{Y} = \left[\left(\hat{\mathbf{G}}^{(0)} \otimes \hat{\mathbf{D}}^{(0)}\right)^T, \ldots, \left(\hat{\mathbf{G}}^{(A-1)} \otimes \hat{\mathbf{D}}^{(A-1)}\right)^T\right]^T \in \mathbb{C}^{AM\times N}$ and $\mathbf{d} \in \mathbb{C}^{1\times N}$ with $[\mathbf{d}]_i = 1$ if $i = n$ and zero otherwise. Finally, we filter the microphones signals with $\mathbf{x}_n^\star$ in order to obtain the estimate $\hat{S}_n(\omega)$ of $S_n(\omega)$

$$\hat{S}_n(\omega) = (\mathbf{x}_n^\star)^H\,\mathbf{p}, \qquad (14)$$

where $\mathbf{p} = \left[\left(\mathbf{p}^{(0)}\right)^T, \ldots, \left(\mathbf{p}^{(A-1)}\right)^T\right]^T$. Note that the result of (14) is a scaled version of the signal since $\hat{\mathbf{D}}^{(a)}$ represents a scaled version of the radiance pattern (ref. (10)). However, this does not represent a problem in the framework, as discussed in the following section.

IV. Sound synthesis

Once the sound field analysis is completed as described in Sec. III, the estimated parameters are used to compute the signal of the jth virtual microphone as

$$\hat{P}(\mathbf{r}_{\mathrm{VM}j}, \omega) = \sum_{n=0}^{N-1} V_{\mathrm{VM}j}(\theta_{n,\mathrm{VM}j}, \omega)\hat{D}_n(\theta_{\mathrm{VM}j,n}, \omega)g(\mathbf{r}_{\mathrm{VM}j}, \hat{\mathbf{r}'}_n, \omega)\hat{S}_n(\omega) \qquad (15)$$

where $\mathbf{r}_{\mathrm{VM}j}$ is the jth VM location, $\theta_{n,\mathrm{VM}j} = \angle(\hat{\mathbf{r}'}_n - \mathbf{r}_{\mathrm{VM}j})$, $\theta_{\mathrm{VM}j,n} = \angle(\mathbf{r}_{\mathrm{VM}j} - \hat{\mathbf{r}'}_n)$ and $\hat{\mathbf{r}'}_n$ is the estimated source position. Note that $g(\mathbf{r}_{\mathrm{VM}j}, \hat{\mathbf{r}'}_n, \omega)$ is proportional to $1/\|\mathbf{r}_{\mathrm{VM}j} - \hat{\mathbf{r}'}_n\|$, hence in practice it is limited to the maximum value $g_{\max}$ to avoid amplifying the signal too much when the distance between the jth VM and the nth source becomes close to zero. An advantage of our approach is that it allows to assign an arbitrary pick-up pattern to the jth VM microphone defining the values of $V_{\mathrm{VM}j}(\cdot)$. This enables the simulation of different microphone models. The pick-up pattern of the common directional microphones e.g. cardioid, supercardioid and hypercardioid, can be easily modelled by a circular harmonics expansion (11). As an alternative, the nominal pick-up pattern of a real microphone can be adopted. It is worth noting that in general, any function can be assigned to $V_{\mathrm{VM}j}(\cdot)$, defining even non-physical characteristics to the VM. The block diagram of Fig. 2 summarizes the whole VM procedure.

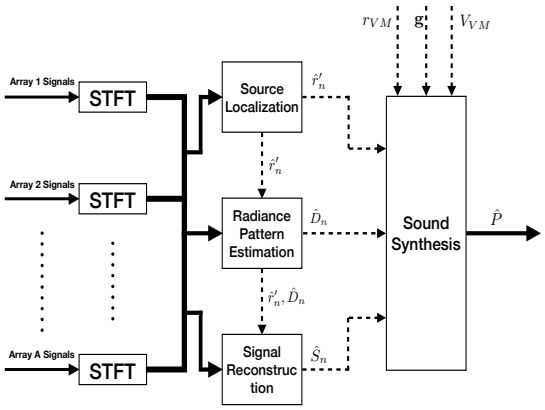

Fig. 2: The VM technique block diagram. Each block corresponds to the relative section of this paper.

V. Simulations and experiments

In order to validate the VM technique, our procedure is tested through an extensive software simulation campaign. We run the simulation using the setup shown in Fig. 3. Both the sources and the virtual microphones present a cardioid directivity pattern for all the frequency range. We employ six circular microphone arrays having a radius of $10\,\mathrm{cm}$, accommodating 4 omnidirectional microphones. The additive noise at the sensor is simulated using a random white Gaussian noise, whose variance is set in such a way to obtain a desired SNR w.r.t each microphone in the acoustic scene. The microphone signals are processed after performing a Short Time Fourier Transform (STFT) with a $20\mathrm{ms}$ Hann window and 75% overlap. Through the Inverse Short Time Fourier Transform of (15), the estimated time domain signal $\hat{p}_{\mathrm{VM}j}(t)$ of the jth VM is obtained using the estimated values (Sec. III), while the reference signal $p_{\mathrm{VM}j}(t)$ is computed adopting the actual theoretical values.

For the tests we employ speech signals taken from [15] in order to simulate a real scenario.

We devise five different metrics as a means to evaluate the proposed technique: the first three are devoted to the evaluation of the analysis stage while the last two to the synthesis stage. Here below we report a list of the aforementioned metrics.

Localization Error: In order to analyse the performance of the localization step, the Mean Squared Error (MSE) between estimated and actual positions of the sources has been adopted:

$$\mathrm{LE}(\mathbf{r}') = \frac{1}{N}\sum\nolimits_{n=0}^{N-1}\|\hat{\mathbf{r}}'_n - \mathbf{r}'_n\| \qquad (16)$$

Directivity Error: The source radiance pattern has been evaluated with an ad hoc metrics called Directivity Error (DE).

ISBN 978-90-827970-1-5 © EURASIP 2018

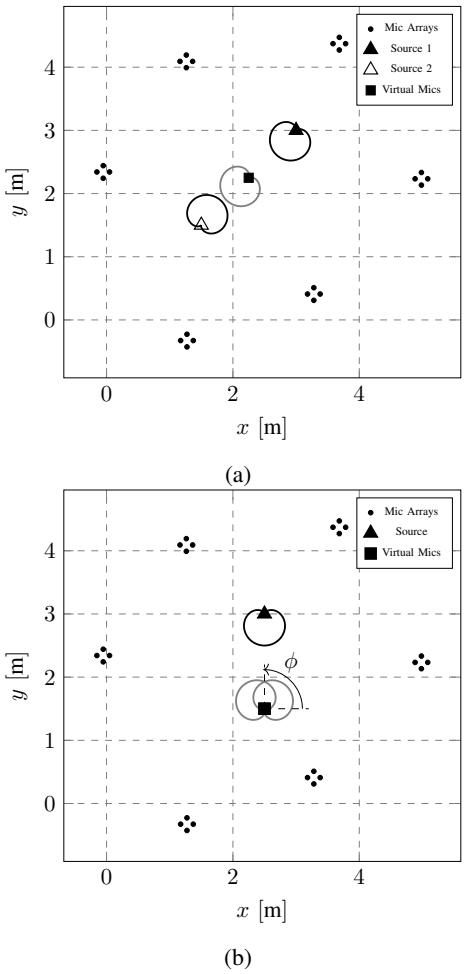

Fig. 3: Simulation setup

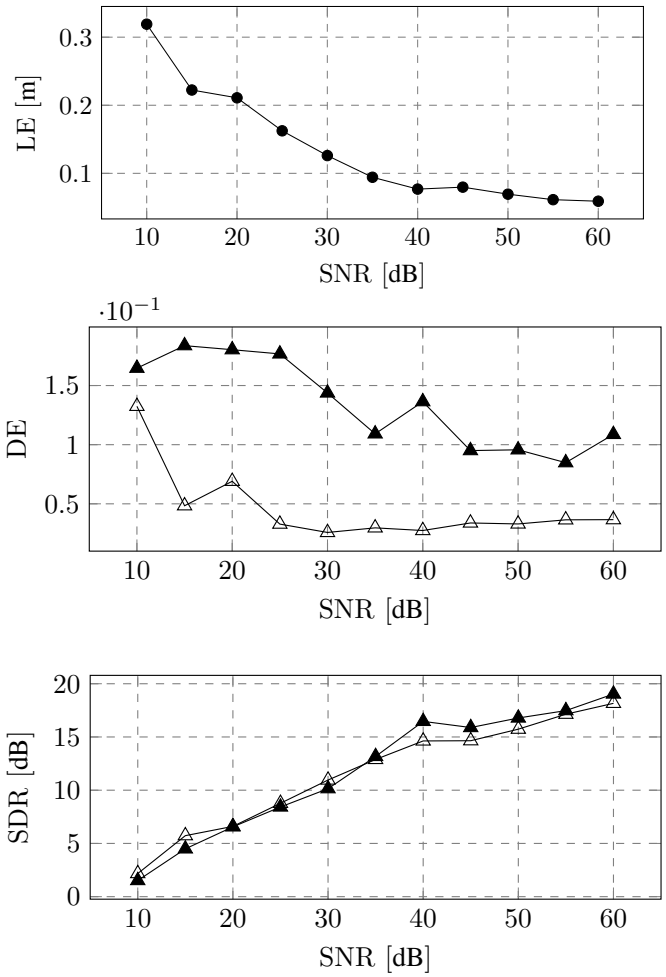

Fig. 4: Analysis metrics. The markers in DE and SDR curves correspond to the ones in setup Fig. 3a

DE is defined for each source n as follows:

$$\mathrm{DE}_n = \frac{1}{F}\sum_{f=0}^{\mathrm{F}-1}\frac{1}{I}\sum_{i=0}^{I-1}(\hat{D}_n(\theta_i,\omega_f) - D_n(\theta_i,\omega_f))^2 \quad (17)$$

where f is the frequency index and i is the angle index.

Source to Distortion Ratio: The Source to Distortion Ratio (SDR) is defined as the energy ratio between the desired signal and sources of distortion (i.e. interferers, noise and artifacts) and it has been used to evaluate the signal extraction performance using the methdology described in [16].

Synthesized Signal Error: The Synthesized Signal Error is defined as the Normalized Mean Squared Error (NMSE) between the jth VM estimated signal and the reference one:

$$\mathrm{SSE}_{\mathrm{VM}j} = 10\log_{10}\frac{\|\hat{p}_{\mathrm{VM}j}(t) - p_{\mathrm{VM}j}(t)\|^2}{\|p_{\mathrm{VM}j}(t)\|^2} \quad (18)$$

Intensity Level Difference: The ILD is defined as:

$$\mathrm{ILD}(\phi) = 10\log_{10}\frac{\|\hat{p}_{\mathrm{VM}j}(t,\phi)\|^2}{\|\hat{p}_{\mathrm{VM}i}(t,\phi)\|^2} \quad \forall i,j, \quad (19)$$

where $\hat{p}_{\mathrm{VM}j}(t,\phi)$ and $\hat{p}_{\mathrm{VM}i}(t,\phi)$ are the signals of two cardioid virtual microphones angled 90° in a XY stereo configuration directed toward ϕ (see Fig. 3b). We evaluate all the metrics but the ILD using the setup in Fig. 3a, while we use the setup in Fig. 3b to calculate the ILD.

Concerning the metrics related to the analysis stage, we report the results obtained by varying the SNR in Fig. 4. As we can see both the LE and the DE decrease as the SNR increases, while the SDR increases monotonically. It is worth noting that, although at very low SNR we have a localization error in the order of 30 cm, we are still able to maintain a low DE and a positive SDR.

As far as the metrics related to the synthesis stage are concerned, we can see from Fig. 5 that the SSE monotonically decreases as the SNR increases, but the error is in the order of -3 dB even when the SNR is very low. Moreover, the behaviour of this metric directly reflects that of the three analysis metrics. As an example, we report in Fig. 6 the comparison of the spectrograms of the estimated and reference virtual microphone signals for a SNR $= 40$ dB. We notice that the two spectrograms are very similar especially in the frequency range below 2 kHz. Above that frequency the spatial aliasing dominates affecting in particular the frequency dependent operations: the radiance pattern estimation Sec. III-B and the signal estimation Sec. III-C.

ISBN 978-90-827970-1-5 © EURASIP 2018

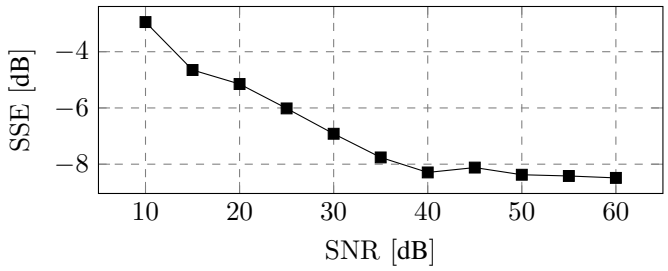

Fig. 5: Synthesized Signal Error of the VM in Fig. 3a for different level of the Signal to Noise Ratio at the microphones

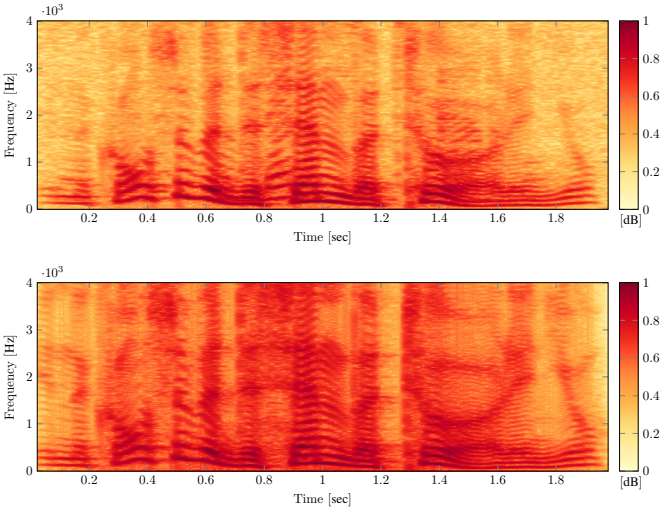

Fig. 6: Comparison of the spectrograms of the estimated and reference virtual microphone signals

Concerning the ILD, we consider a stereo recording scenario using two directional VMs with cardioid pattern and an angular difference of 90°, as shown in Fig. 3b. The difference in sound pressure level between the two microphones is an important spatial cue to be reproduced. In Fig. 7 the ILD of the two cardioid VMs is illustrated as a function of the stereo VMs look direction ϕ, where the microphone pair is rotated from 0° to 360°. As expected, the behaviour of the ILD curve reflects the orientation of the microphones, with the maximum and minimum in correspondence of a zero of the cardioid pattern pointing towards the source (225° and 315°). The zero crossings of the curve occur at 90° and 270° when the signal of the source is sensed with the same intensity from the two VMs. The curve of Fig. 7 corresponds to the expected behavior of a of a physical stereo pair microphones with the same characteristics.

VI. Conclusions

In this paper we presented a methodology for sound field reconstruction based on the estimation of the signal acquired at a virtual microphone location. The virtual microphone can be arbitrarily positioned in space and is characterized by an arbitrary pick-up pattern. We analyze the sound field through a parametric model and we estimate the parameters starting from the signals acquired by a network of microphone arrays. The effectiveness of the technique has been evaluated through a simulation campaign. It is foreseen an estension of the

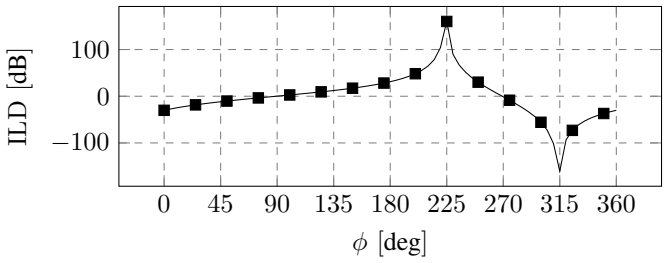

Fig. 7: ILD as a function of the rotation of the two cardioid virtual microphones in a stereo recording scenario

presented procedure to the case of reverberant environments, exploiting a priori information related to the geometry of the environment [17] [18].

References

[1] Juha Vilkamo, Tapio Lokki, and Ville Pulkki, "Directional audio coding: Virtual microphone-based synthesis and subjective evaluation," *J. Audio Eng. Soc*, vol. 57, no. 9, pp. 709–724, 2009.

[2] Richard Schultz-Amling, Fabian Kuech, Oliver Thiergart, and Markus Kallinger, "Acoustical zooming based on a parametric sound field representation," in *Audio Engineering Society Convention 128*. Audio Engineering Society, 2010.

[3] Miriam A Doron and Eyal Doron, "Wavefield modeling and array processing. i. spatial sampling," *IEEE Transactions on signal Processing*, vol. 42, no. 10, pp. 2549–2559, 1994.

[4] Oliver Thiergart, Giovanni Del Galdo, Maja Taseska, and Emanuël AP Habets, "Geometry-based spatial sound acquisition using distributed microphone arrays," *IEEE transactions on audio, speech, and language processing*, vol. 21, no. 12, pp. 2583–2594, 2013.

[5] Dejan Markovic, Fabio Antonacci, Augusto Sarti, and Stefano Tubaro, "Soundfield imaging in the ray space," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 12, pp. 2493–2505, 2013.

[6] Thibaut Ajdler, Luciano Sbaiz, and Martin Vetterli, "The plenacoustic function and its sampling," *IEEE transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, 2006.

[7] 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.

[8] Mirco Pezzoli, Federico Borra, Fabio Antonacci, Augusto Sarti, and Stefano Tubaro, "Estimation of the sound field at arbitrary positions in distributed microphone networks based on distributed ray space transform," 2018, Accepted for publication.

[9] E. G. Williams, *Fourier Acoustics*, Academic Press, London, UK, 1999.

[10] Dejan Marković, Fabio Antonacci, Augusto Sarti, and Stefano 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, 2015.

[11] Harry L Van Trees, *Optimum array processing: Part IV of detection, estimation and modulation theory*, vol. 1, Wiley Online Library, 2002.

[12] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[13] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, NJ, USA, 2004.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, MA, USA, seventh edition, 2009.

[15] "Sound quality assessment material recording for subjective tests," 2008.

[16] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1462–1669, July 2006.

[17] D Aprea, Fabio Antonacci, Augusto Sarti, and Stefano Tubaro, "Acoustic reconstruction of the geometry of an environment through acquisition of a controlled emission," *Signal Processing Conference, 2009 17th European*, pp. 710–714, 2009.

[18] Jason Filos, Antonio Canclini, Fabio Antonacci, Augusto Sarti, and Patrick A Naylor, "Localization of planar acoustic reflectors from the combination of linear estimates," IEEE, 2012, pp. 1019–1023.

ISBN 978-90-827970-1-5 © EURASIP 2018