IEICE TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

VOL. E97-A NO. 9
SEPTEMBER 2014

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY
The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

IEICE TRANS. FUNDAMENTALS, VOL.E97–A, NO.9 SEPTEMBER 2014

PAPER Special Section on Spatial Acoustic Signal Processing and Applications

A Robust Geometric Approach to Room Compensation for Sound Field Rendering

Antonio CANCLINI[†], **Dejan MARKOVIĆ**[†], **Lucio BIANCHI**[†], **Fabio ANTONACCI**[†a)], ***Nonmembers***, **Augusto SARTI**[†], ***Member***, ***and*** **Stefano TUBARO**[†], ***Nonmember***

SUMMARY In this manuscript we present a methodology for reducing the impact of the hosting room reflections in sound field rendering applications based on loudspeaker arrays. The problem is formulated in a least-squares sense. Since matrices involved in the problem are ill-conditioned, it is important to devise a suitable technique for the regularisation of the pseudo-inverse. In this work we adopt a truncated SVD method. The truncation, in particular, aims at reducing the impact of numerical errors and also errors on the knowledge of the sound speed. We include a wide set of experimental results, which validate the proposed technique.

key words: *active room compensation, sound field rendering, early reflections*

1. Introduction

Acoustic Wave Field (WF) rendering techniques are aimed at producing a desired soundfield in a prescribed region of space (rendering region or "sweet spot") using a spatial distribution of loudspeakers. Popular rendering techniques such as WaveField Synthesis (WFS) [1], [2] or Higher Order Ambisonics (HOA) [3], [4]) are designed to operate in the presence of low or no reverberation. When this condition is not met, the resulting rendering accuracy turns out to be limited and so does the rendering impact on the listener. Different solutions are available for reducing the impact of the environment. The most obvious one is, indeed, passive acoustic conditioning, but for home entertainment and multimedia applications, this may easily become a too expensive and invasive option. Active acoustic conditioning is an alternate option that aims at reducing the impact of the environment by driving the loudspeakers with signals that reduce the reverberant component in a prescribed region of space. These techniques are known as *room compensation* [5] or *room equalization* [6]. From a general standpoint, most techniques are based on the use of a number of microphones for measuring the soundfield at a set of control points within the listening area. Measured and target soundfields are then compared, and suitable compensation filters are generated and applied to the loudspeaker signals. Some solutions are based on the off-line computation of such filters, while others accomplish this task in an adaptive fashion, in order to accommodate changes in the environment. In [7] multichannel filtering has been proposed for correcting room effects in selected points within the listening area. The accuracy of such systems depends on the number of sensing points, which raises issues of cost and effectiveness. Less expensive is the solution proposed in [8], where the wave field measurements are limited to a circle, which significantly reduces the number of measurement positions required for an accurate equalization. Room compensation is accomplished by matching the modal decomposition of the desired and the rendered soundfields through a least-squares procedure. In [9] and [10] wavefield analysis is performed adaptively, to accommodate possible changes in the environment (temperature changes, geometric modifications, etc.). Generalizations of this method have been recently proposed in [6]. Although the above methods require fewer microphones, they could be prohibitively expensive for low-cost multimedia applications. Furthermore, the presence of microphones could limit the actual size of the rendering area.

As widely described in the literature, the spatial impression that the listener experiences is to be mostly attributed to early reflections [11], [12]. In fact, echoes in late reverberations are so densely packed to prevent the hearing system from correctly extracting any useful spatial information from them. We can also reasonably assume that the echoes that are caused by diffusion on the part of small objects are of small magnitude, and that typical environments of interest have a reasonably large clarity index (i.e. early reflections dominate over late reverberations). We can therefore focus on compensating low-order reflections on dominant walls. These assumptions, in fact, tend to greatly lower the cost of a compensation strategy, as early reflections can be predicted with a good accuracy using computational acoustics modeling. The method that we discuss in this manuscript is based on the modeling of the acoustic paths (direct and reflected) from each one of the loudspeakers of the rendering system, to the set of control points in the listening area. From such models we derive a propagation matrix and compute its least-squares inverse. This requires a suitable modelling engine for computing the sound field component associated to early reflections. As a large number of Room Impulse Responses (RIRs) needs to be computed, the modelling engine should be fast, while retaining a significant degree of accuracy. Solutions based on Finite Element Methods (FEMs) or Boundary Element Methods (BEMs) are not suitable for this purpose, as they do not enable to selectively compute and focus on the first part of the

Manuscript received December 10, 2013.
Manuscript revised March 20, 2014.

[†]The authors are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, piazza Leonardo da Vinci, 32, 20133 Milano, Italy.
a) E-mail: fabio.antonacci@polimi.it

DOI: 10.1587/transfun.E97.A.1884

impulse response. Among all the many propagation modeling techniques, fast beam tracing [13] turns out to be effective and suitable for our purpose of simulating the propagation in the hosting environment.

The component of the RIR relative to early reflections is inverted in order to match the desired (free-field) propagation. The idea presented in this manuscript was initially developed in [14] in a different context. In fact, in [14] the goal was of suppressing only certain reflections while keeping others. With this methodology the authors showed that exploiting wall reflections, it is possible to emulate the presence of virtual loudspeakers surrounding the listener. With respect to [14], in this manuscript we include a detailed analysis on a new regularisation methodology that guarantees an improved robustness against errors of the sound speed. We also include extensive real data experiments aimed at assessing the accuracy of the rendering and its robustness against errors of sound speed and of the room geometry. For all the measurements, compensated (possibly with errors on the sound speed), non-compensated sound fields rendered through WFS are compared with the reference free-field sound field acquired in a nearly anechoic environment.

The rest of the work is organised as follows. Section 2 provides a common background for all the rendering techniques and formulates the problem of room compensation. Section 3 describes the developed compensation technique. Section 4 shows experimental results and simulations aimed at showing the effectiveness of the proposed technique. Finally, Sect. 5 offers conclusive remarks.

2. Background and Notation

In this section we revisit some of the techniques existing in the literature of sound field rendering. The goal is not to provide a detailed description of the theory underlying the individual algorithms, rather to develop a common notation on rendering. In the second part of the section we formulate the problem of room compensation.

A generic rendering system, shown in Fig. 1, is composed by an arbitrary distribution of M loudspeakers located at $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_M$. The goal of this system is to reproduce the wave field generated by a set of V virtual point sources located at $\mathbf{s}_1, \ldots, \mathbf{s}_V$ by means of the loudspeakers. The rendering system is designed for reproducing the target wave field inside the listening area, denoted by the grey-shaded region shown in Fig. 1. Wave Field Synthesis (WFS) [1] is

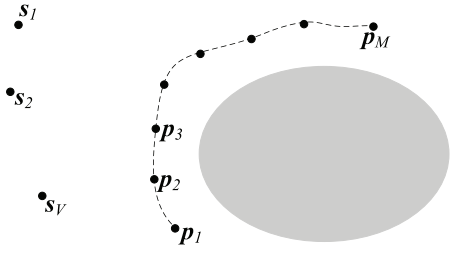

Fig. 1 A general model of a rendering system.

based on the synthesis of the wavefronts by means of the loudspeakers. In WFS the listening area is inherently extended to the whole region opposite to the virtual sources and this fact represents a great advantage over other techniques. Techniques based on Least Squares [15] are based on an approximation of the sound field based in a least squares sense for a set of points in the rendering region. Finally, techniques based on Ambisonics require the loudspeakers to be located regularly on a spherical section and the listening region is restricted to a sweet spot located at the centre of the sphere. The size of the sweet spot is related to the order of the spherical harmonic decomposition.

If we look at the above rendering techniques as "black-box" systems, their output can be represented as a set of filters $\mathbf{h}_{\mathrm{NC}}(\omega) = [h_{\mathrm{NC}_1}(\omega) \ldots, h_{\mathrm{NC}_M}(\omega)]^T$ to be applied to the signals fed to the M loudspeakers. These filters are typically computed under the assumption of ideal (free-field) propagation between the loudspeakers and the listening area.

When a rendering system is installed in a real room (hosting environment), however, the wavefield rendered through the filters $\mathbf{h}_{\mathrm{NC}}(\omega)$ is modified by the presence of reverberations. We formulate the problem of room compensation as the computation of a modified set of rendering filters $\mathbf{h}_{\mathrm{RC}}(\omega) = [h_{\mathrm{RC}_1}(\omega) \ldots h_{\mathrm{RC}_M}(\omega)]^T$, with the aim of dampening the effect of reverberations. More specifically, we focus our attention on the early reflections. Room compensation is performed in a region A_C. The region A_C is sampled by a set of N control points located at $\mathbf{a}_n$, $n = 1, \ldots, N$.

We use the tools of geometrical acoustics, which are valid for sufficiently large reflective surfaces. Reflections are modelled by the presence of image loudspeakers, whose positions are obtained by iteratively mirroring the loudspeakers against the walls of the hosting environment. In this work, for the modelling of the propagation, we employ the beam tracer method in [13], which efficiently computes the visibility between each image loudspeaker and each control point. We denote with $\mathbf{p}'_{m,i}$ the set of image loudspeakers corresponding to the mth loudspeaker, where $i = 1, \ldots, Q_m$; Q_m is the number of considered image loudspeakers associated to $\mathbf{p}_m$. The acoustic propagation between the mth loudspeaker and the nth control point is modeled by

$$\gamma_{nm}(\omega) = g_{nm}(\omega) + \sum_{i=1}^{Q_m} \beta_{m,i} V(\mathbf{a}_n, \mathbf{p}'_{m,i}) \frac{e^{-j\frac{\omega}{c}\|\mathbf{p}'_{m,i} - \mathbf{a}_n\|}}{4\pi\|\mathbf{p}'_{m,i} - \mathbf{a}_n\|}, \tag{1}$$

where

$$g_{nm}(\omega) = \frac{e^{-j\frac{\omega}{c}\|\mathbf{p}_m - \mathbf{a}_n\|}}{4\pi\|\mathbf{p}_m - \mathbf{a}_n\|}$$

is the free-field Green's function from $\mathbf{p}_m$ to $\mathbf{a}_n$; $V(\mathbf{a}_n, \mathbf{p}'_{m,i})$ is a binary function that maps the visibility of $\mathbf{a}_n$ from $\mathbf{p}'_{m,i}$. In particular, $V(\mathbf{a}_n, \mathbf{p}'_{m,i}) = 1$ if the image loudspeaker $\mathbf{p}'_{m,i}$ is visible from $\mathbf{a}_n$ and 0 otherwise. The term $\beta_{m,i}$ is the attenuation coefficient associated to the image loudspeaker at

$\mathbf{p}'_{m,i}$. Its value depends on the reflective properties of the walls and on the reflection order of the image loudspeaker. Theoretically one could consider in Eq. (1) as many image loudspeakers as desired. It is worth noticing, however, that high-order reflections are strongly dampened. Moreover, the presence effect is mainly due to the early reflections. In what follows, we consider reflections up to the third order. Experimental results will confirm the validity of this choice.

3. Room Compensation

Under ideal free-field propagation conditions (i.e., in completely absence of reverberation) the wavefield at the control points would be equal to $\mathbf{r}_{\mathrm{FF}}(\omega) = \mathbf{G}(\omega)\mathbf{h}_{\mathrm{NC}}(\omega)$, where

$$\mathbf{G}(\omega) = \begin{bmatrix} g_{11}(\omega) & \dots & g_{1M}(\omega) \\ \vdots & \ddots & \vdots \\ g_{N1}(\omega) & \dots & g_{NM}(\omega) \end{bmatrix}$$

is the $N \times M$ matrix collecting the free-field Green's functions.

Due to the reverberations introduced by the hosting environment, however, the reproduced wavefield is given by $\mathbf{r}_{\mathrm{NC}}(\omega) = \mathbf{P}(\omega)\mathbf{h}_{\mathrm{NC}}(\omega)$, where the ideal propagation matrix $\mathbf{G}(\omega)$ is replaced by

$$\mathbf{P}(\omega) = \begin{bmatrix} \gamma_{11}(\omega) & \dots & \gamma_{1M}(\omega) \\ \vdots & \ddots & \vdots \\ \gamma_{N1}(\omega) & \dots & \gamma_{NM}(\omega) \end{bmatrix}, \tag{2}$$

which contains the reverberant Green's functions defined in (1).

We aim at computing a matrix $\mathbf{C}(\omega)$, called in the following as *compensation matrix*, such that

$$\mathbf{P}(\omega)\mathbf{C}(\omega) = \mathbf{G}(\omega). \tag{3}$$

The trivial least-squares solution of (3) would lead to a solution of the problem that is not robust against numerical errors, due to the ill-conditioning of the matrix $\mathbf{P}(\omega)$. We resort to a regularisation based on the Truncated Singular Value Decomposition (TSVD), which presents the advantage of an easy-tuning of the regularization parameter. We denote with $\mathbf{P}(\omega) = \mathbf{V}(\omega)\mathbf{\Sigma}(\omega)\mathbf{U}^H(\omega)$ the SVD of $\mathbf{P}(\omega)$, and with $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_M^2$ the associated singular values. The regularized least-squares solution is then

$$\mathbf{C}_K(\omega) = \mathbf{P}(\omega)_K^+\mathbf{G}(\omega), \tag{4}$$

where

$$\mathbf{P}(\omega)_K^+ = \mathbf{U}(\omega)\mathbf{\Sigma}_K^+(\omega)\mathbf{V}^H(\omega), \tag{5}$$

and $\mathbf{\Sigma}_K^+(\omega) = \mathrm{diag}(1/\sigma_1^2, \ldots, 1/\sigma_K^2, 0, \ldots, 0)$ retains the first K singular values of $\mathbf{\Sigma}(\omega)$. The modified set of space-frequency filters is finally obtained as

$$\mathbf{h}_{\mathrm{RC}}(\omega) = \mathbf{C}_K(\omega)\mathbf{h}_{\mathrm{NC}}(\omega). \tag{6}$$

The wavefield reproduced at the control points through the application of the filters $\mathbf{h}_{\mathrm{RC}}(\omega)$ is

$$\mathbf{r}_{\mathrm{RC}}(\omega) = \mathbf{P}(\omega)\mathbf{h}_{\mathrm{RC}}(\omega) = \mathbf{P}(\omega)\mathbf{C}_K(\omega)\mathbf{h}_{\mathrm{NC}}(\omega). \tag{7}$$

Since the solution of (4) implies that $\mathbf{P}(\omega)\mathbf{C}_K(\omega) \approx \mathbf{G}(\omega)$, it follows that $\mathbf{r}_{\mathrm{RC}}(\omega) \approx \mathbf{r}_{\mathrm{FF}}(\omega)$. In other words, the wavefield $\mathbf{r}_{\mathrm{RC}}(\omega)$, rendered in the hosting environment through the modified set of filters $\mathbf{h}_{\mathrm{RC}}(\omega)$, approximates the desired wavefield $\mathbf{r}_{\mathrm{FF}}(\omega)$ ideally obtained in free-field conditions.

It is important to determine the number K of singular values so that the ill-conditioning of the matrix $\mathbf{P}(\omega)$ is mitigated and, at the same time, the accuracy of the room compensation is not sacrificed. In [16] and subsequent works we proposed to use a threshold value v such that

$$\sigma_1^2/\sigma_K^2 \leq v. \tag{8}$$

In this manuscript we complement the regularisation criterion adopted in [16], with another one, which comes from different considerations. Quite a few contributions in the last years have discussed the impact of errors on the sound speed on space-time processing applications, e.g. [17]. In this manuscript we aim at discussing a regularisation technique that reduces the influence of errors of the sound speed to room compensation. At this regard, in [18] the author found important theoretical results on the robustness of the regularised pseudo-inverse $\mathbf{P}_K^+(\omega)$ when the matrix $\mathbf{P}(\omega)$ is affected by an additive error. In this manuscript we apply the theory in [18] to errors on the sound speed. In order for the results of [18] to be applicable to our case, however, we need the additive error on sound speed to result in a linear perturbation on $\mathbf{P}(\omega)$. In the following paragraphs, therefore, we investigate the impact of sound speed errors on the matrix $\mathbf{P}(\omega)$. For completeness, we also derive similar equations for the free-field propagation matrix $\mathbf{G}(\omega)$.

With the goal of deriving manageable equations, we rewrite the free-field Green function as

$$g_{nm}(\omega) = \frac{e^{-j\omega s\|\mathbf{p}_m - \mathbf{a}_n\|}}{4\pi\|\mathbf{p}_m - \mathbf{a}_n\|},$$

where $s = 1/c$ is the *slowness*. We model the error on the slowness as

$$s = \overline{s} + \delta s,$$

where δs models the additive error on the slowness and $\overline{s}$ is its nominal value.

If we assume $\overline{s} \gg \delta s$, we can write $g_{nm}(\omega)$ in terms of its first-order Taylor series

$$g_{nm}(\omega) \approx \overline{g}_{nm}(\omega) - j\omega\delta s\|\mathbf{p}_m - \mathbf{a}_n\|\overline{g}_{nm}(\omega), \tag{9}$$

where $\overline{g}_{nm}(\omega)$ is the Green function corresponding to the nominal sound speed. Similar derivations also apply to the reverberant Green's function $\gamma_{nm}(\omega)$. If we write the matrix $\mathbf{G}(\omega)$ in terms of the perturbed Green functions, we get

$$\mathbf{G}(\omega) = \overline{\mathbf{G}}(\omega) + \delta s\mathbf{A}(\omega) \odot \overline{\mathbf{G}}(\omega), \tag{10}$$

where

$$\mathbf{A}(\omega) = -j\omega \begin{bmatrix} \|\mathbf{p}_1 - \mathbf{a}_1\| & \dots & \|\mathbf{p}_M - \mathbf{a}_1\| \\ \vdots & \ddots & \vdots \\ \|\mathbf{p}_1 - \mathbf{a}_N\| & \dots & \|\mathbf{p}_M - \mathbf{a}_N\| \end{bmatrix}$$

and $\odot$ denotes the Hadamard (element-wise) multiplication.

As for the reverberant propagation matrix $\mathbf{P}(\omega)$, we obtain

$$\mathbf{P}(\omega) = \overline{\mathbf{P}}(\omega) + \delta s \mathbf{B}(\omega) \odot \overline{\mathbf{P}}(\omega), \tag{11}$$

where $\overline{\mathbf{P}}(\omega)$ is the reverberant propagation matrix with the nominal sound speed and

$$\mathbf{B}(\omega) = -j\omega \begin{bmatrix} \sum_{i=1}^{Q_m} \|\mathbf{p}'_{1,i} - \mathbf{a}_1\| & \dots & \sum_{i=1}^{Q_m} \|\mathbf{p}'_{1,i} - \mathbf{a}_N\| \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{Q_m} \|\mathbf{p}'_{M,i} - \mathbf{a}_1\| & \dots & \sum_{i=1}^{Q_m} \|\mathbf{p}'_{M,i} - \mathbf{a}_N\| \end{bmatrix}.$$

We conclude that for sufficiently low frequencies an additive perturbation of the slowness implies a linear perturbation of the free-field propagation matrices $\mathbf{G}(\omega)$ and $\mathbf{P}(\omega)$.

We aim at finding a singular value σ_K that imposes an upper bound on the error of the compensation matrix under uncertainties of the sound speed. For compactness in the derivation, in the following expressions we omit the dependency on the frequency ω. We rewrite (11) as

$$\mathbf{P} = \overline{\mathbf{P}} + \mathbf{E}, \qquad \text{with} \quad \mathbf{E} = \delta s \mathbf{B} \odot \overline{\mathbf{P}}. \tag{12}$$

From Theorem 3.1 in [18] we can readily write that if $\|\mathbf{E}\| < \sigma_K$, then

$$\|\mathbf{P}_K^+\| \leq \frac{1}{\sigma_K - \|\mathbf{E}\|}. \tag{13}$$

Equation (13) states that when $\|\mathbf{E}\|$ increases, and approaches σ_K, $\|\mathbf{P}_K^+\|$ increases monotonically. Let us define with $\overline{\mathbf{C}}_K$ the compensation matrix obtained with the nominal sound speed, with $\mathbf{C}_K$ the compensation matrix with perturbation on the sound speed and with $\mathbf{F}$ the perturbation on $\mathbf{G}$. In [18] author demonstrates that

$$\begin{aligned} &\|\mathbf{C}_K - \overline{\mathbf{C}}_K\| \leq \\ &\|\overline{\mathbf{C}}_K\| \frac{\frac{\sigma_1}{\sigma_K}}{1 - \frac{\|\mathbf{E}\|}{\sigma_K}} \left[\frac{\|\mathbf{E}\|}{\|\mathbf{P}\|} + \frac{\|\mathbf{F}\|}{\|\mathbf{G}\|} + \frac{\frac{\|\mathbf{E}\|}{\sigma_K}}{1 - \frac{\|\mathbf{E}\|}{\sigma_K} - \frac{\sigma_{K+1}}{\sigma_K}} \frac{\|\mathbf{C}_K - \overline{\mathbf{C}}_K\|}{\|\mathbf{G}\|} \right] \\ &+ \frac{\frac{\|\mathbf{E}\|}{\sigma_K}}{1 - \frac{\|\mathbf{E}\|}{\sigma_K} - \frac{\sigma_{K+1}}{\sigma_K}}. \end{aligned} \tag{14}$$

If we assume an upper bound $\|\delta s\|$ on the slowness and that the location of speakers is exactly known, we can also estimate an upper bound on the norm of all the matrices appearing in (13) and (14). If we also impose an upper bound on $\|\mathbf{C}_K - \overline{\mathbf{C}}_K\|$ and $\|\mathbf{P}_K^+\|$, we can find the singular value σ_K that satisfies (13) and (14). These two constraints are used along with (8). In particular, only the singular values that satisfy at the same time (8), (13) and (14) are retained.

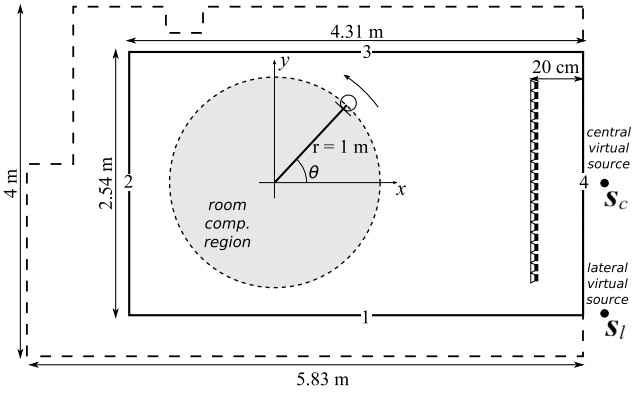

Fig. 2 Experimental setup.

4. Results

In this section we present a set of simulations and experiments that confirm the validity of the proposed room compensation methodology and its robustness against errors in modeling the acoustic propagation.

4.1 Setup

The experimental setup is shown in Fig. 2. The rendering system consists of a 2 m-long uniform linear array of 32 loudspeakers whose aliasing frequency is, therefore, $f_a \approx 2.7$ kHz. For our tests we used two room configurations:

1. a rectangular room whose walls are drawn with bold lines in Fig. 2. This testing environment was built using a set of highly reflecting vertical panels placed in various geometric configurations within a low-reverberation room. The reflectors were laminated MDF panels ($250 \times 60 \times 4$cm), forming four vertical walls (numbered from 1 to 4), with absorbing floor and ceiling. This environment allowed us to test the system in the presence of a variable number of walls (i.e. 2 or 3 reflective walls). The reflection coefficient of the panels, averaged in the frequency range [500 Hz, f_a], was estimated to be 0.8, using the technique proposed in [19];
2. a real office room with a complex geometry and reflective floor and ceiling, whose floormap is as described by the dashed lines of Fig. 2. The height of the ceiling is 2.9 m and the average reflection coefficient of the walls is 0.7.

We used WaveField Synthesis (WFS) [2] to render a virtual source placed in two possible location behind the array; in particular we consider a "central" virtual source at $\mathbf{s}_c = [3.5\,\text{m}, 0\,\text{m}]$; and a "lateral" source at $\mathbf{s}_l = [3.5\,\text{m}, -1.3\,\text{m}]$ (with respect to the reference frame shown in Fig. 2). The room compensation region is shown in Fig. 2 as a gray-shaded circle in front of the array whose radius is $r = 1$ m and whose center is 2.5 m from the array. It accommodates $N = 1000$ control points. In order to measure the wave

field within the room compensation region we used a measurement system based on a cardioid microphone mounted at the tip of a rotating arm, controlled by a step motor. This rig allows us to samples the wavefield over Z regularly spaced points over the border of the compensation region. For all the tests, we set the regularisation parameter in (8) to $v = 5$, as we discovered, from preliminary experiments, that represents a good trade-off between accuracy and robustness against ill-conditioning. Moreover, we imposed $\|\mathbf{C}_K - \overline{\mathbf{C}}_K\| \leq 1.1\,\|\overline{\mathbf{C}}_K\|$ and $\|\mathbf{P}_K^+\| \leq 1.1\,\|\overline{\mathbf{P}}_K^+\|$.

4.2 Simulation Results

In this paragraph we show the results of a simulation aimed at investigating the effect of room compensation on the frequency response of the rendering system. We simulated a setup equivalent to the one depicted in Fig. 2, considering the 4.31 m × 2.54 m rectangular room without floor and ceiling. The array renders the central virtual source located at $\mathbf{s}_c$, in the frequency range [500 Hz, 2 kHz]. The loudspeakers are modeled as omni-directional point sources. We computed the ideal free-field response $\mathbf{r}_{\mathrm{FF}}(\omega)$ at the control points; similarly, we predicted the non-compensated response $\mathbf{r}_{\mathrm{NC}}(\omega)$, modeling reflections up to the 10th order; finally, we computed the response $\mathbf{r}_{\mathrm{RC}}^{(i)}(\omega)$, where room compensation was performed modelling and compensating reflections up to the ith order. Figure 3 presents a comparison of: the average[†] free-field frequency response $r_{\mathrm{FF}}(\omega) = \mathrm{mean}[\mathbf{r}_{\mathrm{FF}}(\omega)]$; the average non-compensated response $r_{\mathrm{NC}}(\omega) = \mathrm{mean}[\mathbf{r}_{\mathrm{NC}}(\omega)]$; and the average compensated response $r_{\mathrm{RC}}^{(i)}(\omega) = \mathrm{mean}[\mathbf{r}_{\mathrm{RC}}^{(i)}(\omega)]$, for $i = 1$ and $i = 3$. The effect of the room on the rendered wavefield is clearly visible in the non-compensated response $r_{\mathrm{NC}}(\omega)$, whose peaks match the resonant modes (denoted with gray vertical lines in Fig. 3) introduced by the 2 walls parallel to the array. In this configuration, in fact, the position of the virtual source is such that the rendered wavefronts mainly propagate in the direction orthogonal to the array, thus causing strong reflections onto walls 2 and 4 (see Fig. 2). The response $r_{\mathrm{NC}}(\omega)$ presents oscillations in the range [−6 dB, +4 dB]. When reflections are compensated up to the 1st order (top part of Fig. 3), we observe a partial attenuation of the resonant peaks, and the oscillations of the $r_{\mathrm{RC}}^{(1)}(\omega)$ are in the range [−2 dB, +2 dB]. Increasing the compensation order ($i = 3$, bottom part of Fig. 3), the response is flattened and $\left|r_{\mathrm{RC}}^{(3)}(\omega)\right| < 0.4$ dB on the average. In other words, the room compensation acts as an equalization system inside the compensation region. It is worth noticing, however, that the system in Eq. (3) admits only the least-squares approximate solution $\mathbf{C}_K(\omega)$. Consequently, a perfect equalization that equals the flat response $r_{\mathrm{FF}}(\omega)$ is not feasible, even if the compensation order i is arbitrarily increased. From our preliminary simulations, we verified that, for this configuration, setting $i > 3$ does not introduce any noticeable improvement.

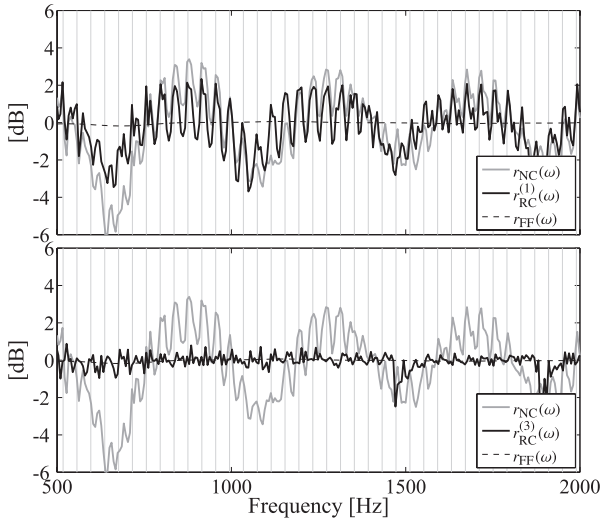

Fig. 3 Frequency response.

[†]The operator mean[**r**] refers to the arithmetic mean of the elements of the vector **r**.

4.3 Experimental Results

In this paragraph, after a description of the metrics adopted for evaluating the system, we show the results of a set of experiments conducted in the setup in Fig. 2.

4.3.1 Measurements

The evaluation metrics are based on the measurement of the wavefield produced when the loudspeakers emit a Maximum Length Sequence. The wavefield is measured by means of a rotating cardioid microphone, as shown in Fig. 2. Following the same approach as in [20], we estimate the frequency response $H_c(\omega, \theta_j)$ at Z angular positions θ_j, $j = 1 \ldots Z$ of the cardioid, which uniformly samples the circumference with radius $r = 1$ m enclosing the compensation region. From this measurement we obtain the μth circular harmonic

$$C_\mu(\omega) = 2H_c(\omega, \theta_j)\left[J_\mu(kr) - \mathrm{sgn}(\omega)J'_\mu(kr)\right]^{-1},$$

where $J_\mu(\cdot)$ is the Bessel function of first kind and order μ, and $k = \omega/c$, c being the speed of sound.

The first metric that we consider is the power of the plane wave components $L(\omega, \theta)$ in the frequency range of interest [$f_{\min} = 500$ Hz, f_a], computed as [9]

$$W(\theta) = \int_{f_{\min}}^{f_a} |L(\omega,\theta)|^2\, d\omega = \int_{f_{\min}}^{f_a} \left| \sum_{\mu=-\infty}^{+\infty} j^{-\mu} C_\mu(\omega) e^{j\mu\theta} \right|^2 d\omega.$$

From $C_\mu(\omega)$ we also recover the frequency response of the system, as it would be measured using an omnidirectional microphone:

$$\mathcal{FT}\{h(t,\theta)\} = H(\omega,\theta) = \sum_{\mu=-\infty}^{+\infty} C_\mu(\omega) J_\mu(kr) e^{j\mu\theta}. \quad (15)$$

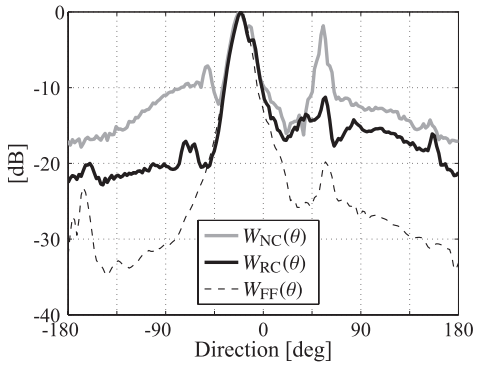

Fig. 4 Power of the plane wave components of the rendered wave fields for a lateral source when only walls 1 and 3 (see Fig. 2) are present.

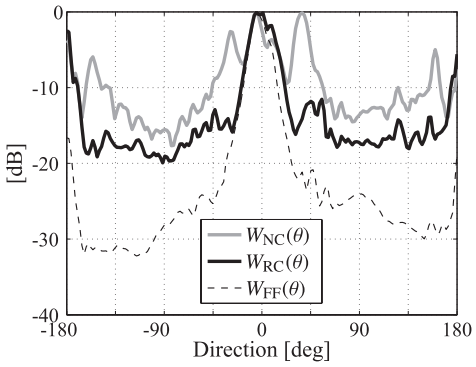

Fig. 5 Power of the plane wave components of the rendered wave fields for a central source when walls 1,2 and 3 (see Fig. 2) are present.

In order to offer additional insight on the evolution in time of the reflections we introduce a second metric, which is directly related to the impulse responses $h(t, \theta)$. This second metric is motivated by psychoacoustic considerations, as the temporal distribution of the reflections is known to have a relevant impact on perception. We define the echogram as

$$E(\theta, t) = \frac{h(\theta, t)}{|\max_t h(\theta, t)|}, -180^\circ < \theta < 180^\circ, 0 < t < 25\,\text{ms}.$$

We limit the analysis to the first 25 ms of the impulse response for clarity in the visualisation.

4.3.2 Experiments in the Controlled Room

We start considering the scenario of the room compensation system operating in the presence of a limited number of reflective walls in the reconfigurable room described in paragraph 4.1. In all the experiments, room compensation is applied up to the third order of reflection.

Figure 4 shows the power of the plane wave components of the rendered wave fields for a lateral source when walls 1 and 3 are present, according to the drawing in Fig. 2. In particular, the function $W(\theta)$ is shown in three cases: i) when no room compensation is applied ($W_{NC}(\theta)$); ii) when room compensation is active ($W_{RC}(\theta)$); iii) in free-field conditions ($W_{FF}(\theta)$), obtained measuring the rendered wavefield in a low-reverberation chamber.

We observe that, in all three cases the main direction of propagation is represented by the direct path at $\theta = -20^\circ$; in the non compensated response $W_{NC}(\theta)$ the two main reflections on walls 3 and 1 are noticeable at 55° and -51°, with a power of -1.8 dB and -7.1 dB respectively. In the room-compensated response such reflections are approximately 10 dB smaller. It is also interesting to observe the behavior of the free-field response $W_{FF}(\theta)$. While, in theory, it should only exhibit the direct path, in fact it exhibits some low-power (below -20 dB) reflective peaks, due to non ideal propagating conditions.

In a second test we introduce a third reflective panel, corresponding to the wall number 2 in Fig. 2. In this experiment the central source is rendered. Figure 5 shows the function $W(\theta)$ relative to the rendering of a central virtual source, for the setup in which walls 1, 2 and 3 are present (see Fig. 2). The direct path is characterized by $\theta = 0^\circ$. Observing $W_{NC}(\theta)$, we identify three groups of reflections: the first-order side reflections, coming from walls 3 and 1, at $\pm 35^\circ$; the first-order reflection at -180° coming from wall 2; and the second order reflections at $\pm 155^\circ$, caused by a first reflection on wall 3 or on wall 1, and then onto wall 2. All the reflections, except for the one at -180°, are attenuated of about 10 dB in the compensated response $W_{RC}(\theta)$. As for the reflection coming from $\theta = 180^\circ$, this is an acoustic path that propagates towards the array of speakers and cannot be compensated, as the array cannot synthesise anticausal wavefronts [9]. It is important to underline that this is not a limit of the proposed compensation technique, but it is due to the array geometry. The use of an array of different geometry (e.g. a circular array) would allow us to overcome this problem. Consistently, Fig. 5 shows that $W_{RC}(\theta)$ approaches $W_{NC}(\theta)$ for $\theta = \pm 180^\circ$. Notice a reflection at -17 dB coming from $\theta = \pm 180^\circ$, due to non-idealities of the low-reverberation chamber, also in the free-field response.

4.3.3 Real Room

We now examine the performance of the rendering system in the second testing environment, a real office space with no passive acoustic conditioning. In order to validate the robust regularisation technique proposed in Sect. 3, we also include the results when errors on the sound speed have been introduced in the data model on purpose.

Figure 6 shows the power of the plane wave components for a central source. For this configuration, the most relevant reflections are at $\theta_1 = -156^\circ, \theta_2 = -138^\circ, \theta_3 = -49^\circ, \theta_4 = 51^\circ, \theta_5 = 158^\circ$, as marked by the vertical lines in Fig. 6.

In nominal working conditions, the attenuation of the reflective paths (except that at 180°) ranges between 6.5 dB for the reflection at -156° and 12 dB for the reflection at 51°. It is interesting to notice that the maximally attenuated reflection is also the most relevant one. The plot of $W_{RCe}(\theta)$ refers to the power of the plane wave components when an error of +4 m/s has been introduced in the data

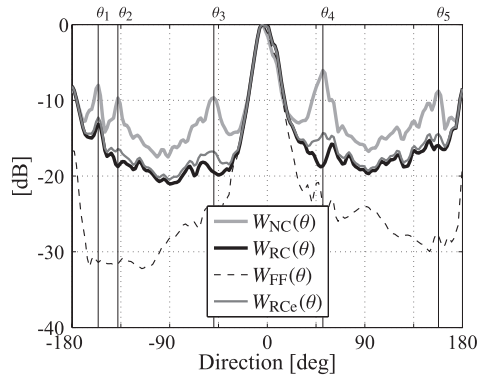

Fig. 6 Power of the plane wave components of the rendered wave fields for a central source in the real room (see dashed lines in Fig. 2).

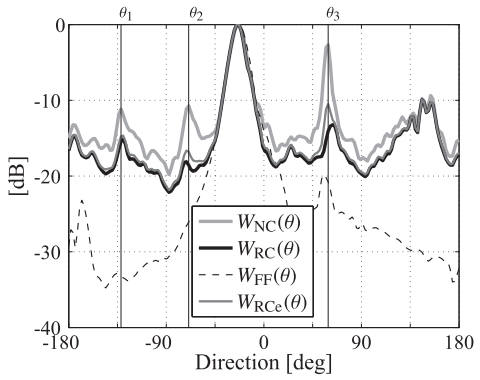

Fig. 8 Power of the plane wave components of the rendered wave fields for a lateral source in the real room (see dashed lines in Fig. 2)

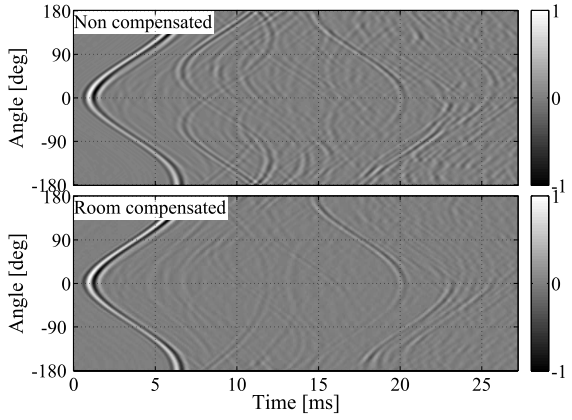

Fig. 7 Echograms $E(\theta, t)$ of compensated and non-compensated impulse responses for the central source in the real room.

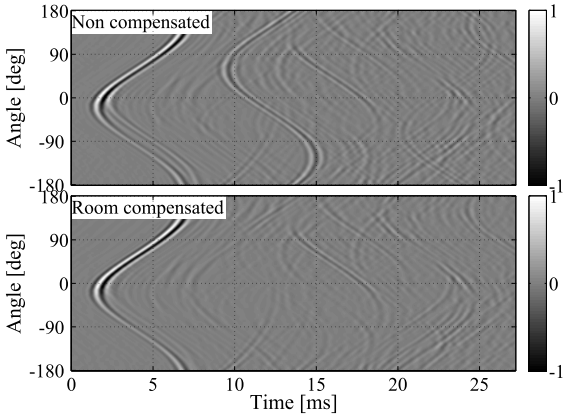

Fig. 9 Echograms $E(\theta, t)$ of compensated and non-compensated impulse responses for the lateral source in the real room.

model. Notice that the performance of the room compensation only slightly decreases in this case. Figure 7 shows the echograms of the compensated and non-compensated impulse responses for the same experiment. Notice that almost all reflections within 15 ms are effectively attenuated, except for that coming from the wall in front of the array. Some reflection with a delay greater than 15 ms survive in the compensated response. This has a twofold explanation: on one hand these are reflections of order higher than three, which are not compensated; on the other hand some of these reflections come from the wall in front of the array and cannot be compensated. Figure 8 shows the power of the plane wave components for the lateral source in the real room. The most relevant reflections are at $\theta_1 = -132°(-12\,\text{dB}), \theta_2 = -69°(-11\,\text{dB}), \theta_3 = 59°(-3\,\text{dB})$. Notice also an acoustic event at 150°, which is not compensated as it is due and diffraction onto a corner (not explicitly modelled in (2)) and a further reflection. Among all the tested configurations, this is the most critical one. Even in this challenging scenario, however, all the reflective paths, except that at 155°, are at least −13 dB below the direct one in the compensated response. An appropriate modelling of the diffractive paths within the beam tracing framework (see [21]) could be used to further improve the compensation effectiveness. As for the case of the central source, an error of +4 m/s on the sound speed does not impact significantly on the accuracy of the compensation. Figure 9 shows the corresponding echograms of the compensated and non-compensated impulse responses. Notice that in the first 15 ms an accurate compensation of the reflections is attained. On the other hand, some of the successive reflections are not compensated as they are not modelled (they are of fourth or higher-order) or cannot be compensated as they propagate in opposite direction with respect to the wavefronts originating from the array. Finally, notice the presence in the echogram of the non-compensated impulse response of a secondary wavefront that shortly follows the direct one. This is the wavefront coming from the wall behind the array. This echo severely impacts on the timbre of the rendered sound [20]. In fact the frequency response of the sum of the direct- and echo-related signals exhibits a behaviour similar to a comb filter. On the other hand, the echogram of the room-compensated impulse response shows that this reflection is greatly attenuated in the compensated response. A similar behaviour was not evident in Fig. 7 due to the different location of the virtual source.

In order to test the robustness of the compensation algorithm against different errors of the sound speed, we re-

Table 1 Attenuation (dB) of the reflections at angles $\theta_1, \ldots, \theta_5$ (see Fig. 6 for different errors of the sound speed.

Error	θ_1	θ_2	θ_3	θ_4	θ_5
0 m/s	6.5	8.3	9.4	12	7
+0.5 m/s	6.5	8.2	9	11.6	7
+1 m/s	6.5	8	9.2	11.4	6.5
+2 m/s	6.3	7.6	8.9	11.4	6.5
+3 m/s	6.2	7.2	8.7	10.3	5.9
+4 m/s	6.1	6.8	8.5	9.7	5.7

peated the experiment in Fig. 6 for different errors on the sound speed, ranging from 0.5 m/s to 4 m/s. For reasons of space, we provide in Table 1 only the attenuation of the reflections at the angles $\theta_1, \ldots, \theta_5$ marked in Fig. 6. Results confirm that the attenuation gracefully degrades when the error on the sound speed increases. We also tested the robustness of the room compensation system against errors in modeling the geometry of the hosting environment, considering differing scaling of its nominal size. We observed that, also in this case, the performances are only slightly degraded. This can be explained by the fact that a scaling of the room is equivalent to introduce a scaling in the distance between the image loudspeakers and the control area, which has a similar effect to considering a wrong propagation speed. The detailed results are not reported in the manuscript for reasons of space.

5. Conclusions

In this manuscript we have presented a method for the compensation of the most relevant reflections of the hosting environment in sound field rendering applications. The compensation technique is based on a least-squares equalization of the frequency response of the array-environment system. Two regularization criteria are adopted, one for the numerical stability of the inversion, one for the robustness of the inversion when uncertainty about the sound speed is present. The method is general enough to be applicable to a wide range of algorithms. Experimental results of compensation confirm the validity of the proposed methodology in normal working conditions.

References

[1] A.J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," J. Acoust. Soc. Am., vol.93, pp.2764–2778, 1993.
[2] J. Ahrens, R. Rabenstein, and S. Spors, "The theory of wave field synthesis revisited," Proc. 124th Conv. Audio Eng. Soc., 2008.
[3] J. Daniel, S. Moreau, and R. Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," Proc. 114th Conv. Audio Eng. Soc., 2003.
[4] R. Nicol, "Sound spatialization by higher order ambisonics: Encoding and decoding a sound scene in practice from a theoretical point of view," Proc. 2nd International Symposium on Ambisonics and Spherical Acoustics, 2010.
[5] S. Spors, H. Buchner, and R. Rabenstein, "A novel approach to active listening room compensation for wave field synthesis using wave-domain adaptive filtering," Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP), 2004.
[6] M. Schneider and W. Kellermann, "Adaptive listening room equalization using a scalable filtering structure in the wave domain," Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012.
[7] J. Lopez, A. Gonzalez, and L. Fuster, "Room compensation in wave field synthesis by means of multichannel inversion," Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp.146–149, 2005.
[8] T. Betlehem and T.D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," J. Acoust. Soc. Am., vol.117, no.4, pp.2100–2111, 2005.
[9] S. Spors, A. Kuntz, and R. Rabenstein, "An approach to listening room compensation with wave field synthesis," Proc. 24th Audio Eng. Soc. Int. Conf. on Multichannel Audio, pp.49–52, 2003.
[10] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," J. Acoust. Soc. Am., vol.122, pp.354–369, 2007.
[11] M. Kahrs and K. Brandenburg, Eds., Applications of Digital Signal Processing to Audio and Acoustics, Kluwer International Series in Engineering and Computer Science, Kluwer, 1998.
[12] J. Ahrens, Analytic Methods of Sound Field Synthesis, Tlabs Series in Telecommunication Services, Springer Verlag Berlin Heidelberg, 2012.
[13] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," IEEE Trans. Audio Speech Language Process., vol.16, no.4, pp.812–824, 2008.
[14] A. Canclini, D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "A room-compensated virtual surround system exploiting early reflections in a reverberant room," Proc. European Signal Processing Conference (EUSIPCO), pp.1029–1033, 2012.
[15] G.N. Lilis, D. Angelosante, and G.B. Giannakis, "Sound field reproduction using the lasso," IEEE Trans. Audio Speech Language Process., vol.18, no.8, pp.1902–1912, 2010.
[16] F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, and S. Tubaro, "Rendering of an acoustic beam through an array of loudspeakers," Proc. 12th International Conference on Digital Audio Effects, DAFx-09, Como, Italy, 2009.
[17] P. Annibale, J. Filos, P. Naylor, and R. Rabenstein, "Geometric inference of the room geometry under temperature variations," 5th International Symposium on Communications, Control, and Signal Processing (ISCCSP), pp.1–4, 2012.
[18] P. Hansen, "The truncated SVD as a method for regularization," BIT Numerical Mathematics, vol.27, no.4, pp.534–553, 1987.
[19] D. Marković, K. Kowalczyk, F. Antonacci, C. Hofmann, A. Sarti, and W. Kellermann, "Estimation of acoustic reflection coefficients through pseudospectrum matching," IEEE/ACM Trans. Audio Speech Language Process., vol.22, no.1, pp.125–137, 2014.
[20] L. Bianchi, F. Antonacci, A. Canclini, A. Sarti, and S. Tubaro, "A psychoacoustic-based analysis of the impact of pre-echoes and post-echoes in soundfield rendering applications," Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC), 2012.
[21] F. Antonacci, A. Sarti, and S. Tubaro, "Two-dimensional beam-tracing from visibility diagrams for real-time acoustic rendering," EURASIP Journal on Advances in Signal Processing, vol.1, pp.1–18, 2010.

Antonio Canclini was born in Sondalo, Italy, on December 19, 1983. He received the M.Sc. degree (cum laude) in computer science and engineering (2008) and the Ph.D. degree (cum laude) in information technology (2012) from the Politecnico di Milano, Milan, Italy. He is currently a Post-Doc Researcher at the Image and Sound Processing Group, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano. His research focuses on space-time audio signal processing (localization of acoustic sources and reflectors, spatial sound field rendering).

Dejan Marković was born in 1985. He received the M.Sc. degree cum laude in 2009 and the Ph.D. degree cum laude in 2013 from the Politecnico di Milano, Milan, Italy. Since January 2013, he is a Post Doctoral Research Assistant at Image and Sound Processing Group (ISPG), Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano. His research activity is mainly focused on analysis, geometric modeling and microphone array processing of acoustic wavefields for multimedia applications.

Lucio Bianchi received the Bachelor degree in Electronic Engineering and the Master of Science in Computer Engineering from Politecnico di Milano, Milano, IT, in 2010 and 2012 respectively. He is currently a Ph.D. student at Image and Sound Processing Group in Dipartimento di Elettronica, Informazione e Bioingegneria at Politecnico di Milano. His research focuses on sampling, processing and reconstruction of acoustic fields.

Fabio Antonacci was born in Bari (Italy) on July 26, 1979. He received Laurea degree in 2004 in Telecommunication Engineering and Ph.D. in Information Engineering in 2008, both at Politecnico di Milano, Italy. He is currently assistant professor at Politecnico di Milano and works at the Image and Sound Processing Group. His research focuses on space-time processing of audio signals, for both speaker and microphone arrays (source localization, acoustic scene analysis, rendering of spatial sound) and on modeling of acoustic propagation (visibility-based beam tracing). He published approximately 50 articles in proceedings of international conferences and on peer-reviewed journals.

Augusto Sarti received the M.S. and the Ph.D. degrees in electronic engineering, both from the University of Padua, Italy, in 1988 and 1993, respectively, with research on nonlinear communication systems. His graduate studies included a joint graduate program with the University of California at Berkeley, where he worked on nonlinear system theory. In 1993 he joined the Dipartimento di Elettronica e Informazione of the Politecnico di Milano, Milan, Italy, where he is currently an Associate Professor. His research interests are in the area of digital signal processing, with particular focus on sound analysis, processing and synthesis; space-time audio processing; geometrical acoustics; music information retrieval. He also worked on problems of multidimensional signal processing, vision-based 3D scene reconstruction; camera calibration; image analysis; motion planning and nonlinear system theory. He coauthored over 180 scientific publications on international journals and congresses as well as numerous patents in the multimedia signal processing area. He coordinates the Sound and Music Computing Lab of the Image and Sound Processing Group of the Politecnico di Milano. He promoted and coordinated or contributed to numerous (20+) EC-funded project. He is a member of the IEEE Technical Committee on Audio and Acoustics Signal Processing, and Associate Editor of IEEE Signal Processing Letters. He has served as guest editor for numerous special issues of international journals. He was co-chairman of the 2005 Edition of the IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS); Chairman of 2009 edition of the Digital Audio Effects conference, (DAFx); and in organizing committees of numerous other conferences in the area of signal processing.

Stefano Tubaro was born in Novara, Italy, in 1957. He received his Electronic Engineering degree at the Politecnico di Milano, Milano, Italy, in 1982. He then joined the Dipartimento di Elettronica e Informazione of the Politecnico di Milano, first as a researcher of the National Research Council; then as an Associate Professor (1991) and finally as a Full Professor (2004). He initially worked on problems related to speech analysis; motion estimation/compensation for video analysis/coding; and vector quantization applied to hybrid video coding. In the past few years, his research interests have focused on image and video analysis for the geometric and radiometric modeling of 3-D scenes; and advanced algorithms for video coding and sound processing. He has authored over 150 scientific publications on international journals and congresses. He coauthored two books on digital processing of video sequences. He also coauthored several patents on image processing techniques. He coordinates the research activities of the Image and Sound Processing Group (ISPG) at the Dipartimento di Elettronica e Informazione of the Politecnico di Milano; which is involved in several research programs funded by industrial partners, the Italian Government, and by the European Commission. He has been involved in the IEEE Technical Committee of Multimedia Signal Processing (2005–2009), and is currently involved in that of Image Video and Multidimensional Signal Processing.