A ROOM-COMPENSATED VIRTUAL SURROUND SYSTEM EXPLOITING EARLY REFLECTIONS IN A REVERBERANT ROOM

A. Canclini, D. Marković, F. Antonacci, A. Sarti, S. Tubaro

Dipartimento di Elettronica ed Informazione
Politecnico di Milano, p.zza Leonardo da Vinci 32
20133 Milano, Italy

ABSTRACT

In this paper we propose a novel methodology for rendering a surround sound in a prescribed area using a loudspeaker array, which exploits the early reflections on the walls of the room. With respect to the state of the art, the proposed technique is able to take advantage of the knowledge of the acoustic scene (i.e. room geometry and reflective properties of the walls) in order to reduce undesired wall reflections. Simulative results show that the use of room compensation enables an improved accuracy of the virtual surround.

Index Terms— Loudspeaker array, sound field rendering, virtual surround, room compensation

1. INTRODUCTION

Multichannel audio rendering is a research field that has been active for over four decades and has produced several commercial applications. Its aim is to increase the spatial impression of sound, at the cost of having to use a large number of speakers in the room. This, unfortunately, is often perceived as a nuisance and sometimes architectural constraints prevent us from adopting such solutions. Furthermore, the perceptual effectiveness of multi-channel systems tends to be strongly dependent on the level of reverberation in the room.

A method that addresses both such issues was first presented in [1, 2] and is today a consumer product. This approach uses a single speaker array for generating a plurality of independent acoustic beams. If the geometry of the surrounding room matches one of the prescribed ones, the produced beams will be reflected by the walls in such a way to reach the listener from the correct directions (left, right, surround left and surround right). An initial calibration procedure through a microphone positioned in prescribed positions enables to infer some clues of the acoustics of the room, which are then useful for system tuning. However, this system exploits only a limited knowledge of the environment in which it is operating.

Authors acknowledge the funding of the European Project Scenic under GA: 226007.

In this work we intend to show that knowing more about room geometry and acoustics could trigger new and more advanced functionalities. In particular, we assume that the geometry of the environment is known or it is inferred using techniques such as [3, 4]. We also know in advance the reflection coefficient (assumed to be equal for all the walls). With this knowledge at hand, the design of the acoustic beams emitted by the speaker array can be improved in order to reach more precisely the listening area. More interesting, however, is the possibility to perform a room compensation exploiting the room geometry. Commercial systems, in fact, exploit first and second-order reflections useful to render the presence of the image arrays. At the same time, however, they are sensitive to undesired reflections, which negatively affect the spatial impression.

In this work the synthesis of the virtual sources is conceived in two interconnected steps. First, the synthesis of the acoustic beams is accomplished by means of a least squares approach. The propagation of the sound from each loudspeaker to a set of control points is modeled through the Green's function. By imposing the desired soundfield and inverting the resulting equation in the least squares sense, we obtain a set of spatial filters for the loudspeaker array, one for each acoustic beam of interest.

The second step consists in the compensation of the undesired reflections. The idea of room compensation is not novel in the literature of sound rendering. In [5] the authors incorporate the acoustic transfer function in the data model. In [6] the authors introduce a feedback loop. Microphones measure the room response and the wavefield is analyzed and controlled by means of an orthogonalization of the room response. For the problem under analysis, we follow a different approach, in which we attenuate some reflections, while retaining others, as they are useful for reproducing left, right and rear channels of the virtual surround system. Tools of traditional acoustics are not suitable for this purpose, as reflections are not explicitly modeled, but are a side-effect of boundary conditions of the D'Alembert equations. We need, instead, a representation of the wavefield in which walls explicitly become part of the system. We resort, therefore, to geometrical acoustics,

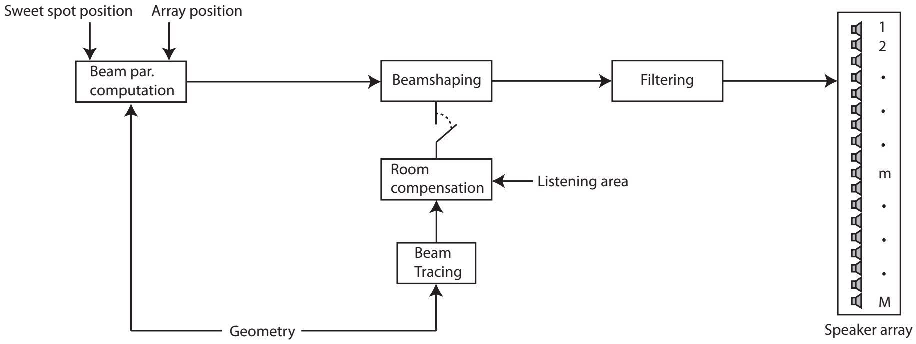


Fig. 1. Block diagram of the proposed system.

which models the propagation of wavefronts through acoustic rays, which travel and bounce off the walls of the room. In the past few years several methodologies for source localization, room inference and soundfield rendering have been developed in the realm of geometrical acoustics. For the problem under consideration, geometrical acoustics comes at our help because it models the propagation of the sound by handling each reflection separately, and therefore allowing to easily model the soundfield resulting from the reflections that must survive. Reflections from all the image sources (up to a prescribed order) are modeled through the beam tracing technique [7] and the resulting propagation model is stored in a matrix $\mathbf{P}(\omega)$. The matrix $\mathbf{D}(\omega)$ is computed, which contains the contribution only from the desired reflective paths. The transformation $\mathbf{C}(\omega)$ is found, which drives from the actual propagation matrix $\mathbf{P}(\omega)$ to the desired propagation matrix $\mathbf{D}(\omega)$. This transformation is applied to the spatial filters to compensate for the undesired reflections. We demonstrate, through equations and results, that this approach is effective for suppressing the undesired reflections. The rest of the paper is organized as follows: section 2 provides an overview of the system. Section 3 describes the methodology used for the shaping of the acoustic beam, introduced in [7]. Given the geometry of the room and the listening area, in section 4 we discuss geometrical issues related to the emission of beams from the loudspeaker array. Section 5 discusses the room compensation technique. Section 6 presents some results to show the effectiveness of the proposed methodology. Finally, Section 7 draws some conclusions.

2. OVERVIEW

Consider the geometry in Figure 2, in which a floormap of the acoustic scene is illustrated: a loudspeaker array in a reverberant room emits acoustic beams and the listener is assumed to be located in the dashed area. A central beam (C) is directed towards the listening area. Two lateral beams (L and R) exploit the bouncing off the side walls in order to render left and right speakers. Finally, two beams (RL and RR) exploit the bouncing on the side and bottom walls in order to render the presence of the rear-left and rear-right speakers. The idea of exploiting reflections to reproduce a surround sound is not novel in the literature, and also commercial systems exist which implement this idea. However, they do not compensate further bouncing of the beams over the walls, which corrupt the rendering effect. In this paper we propose a methodology for the suppression of such undesired reflections, while preserving the desired ones.

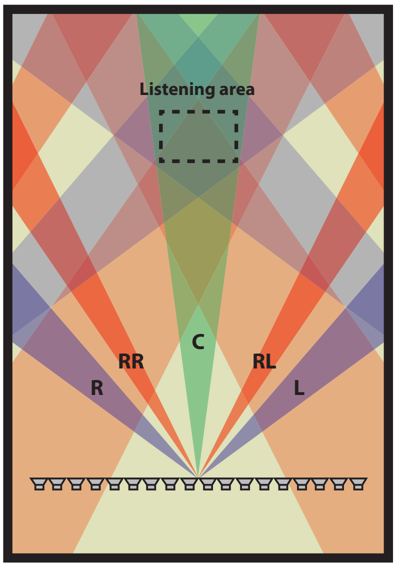


Fig. 2. Geometry of the virtual home theater system.

The block diagram that describes our approach to the virtual surround is shown in Fig. 1. The beamshaping block allows to render an angularly limited acoustic source (beam) through a loudspeaker array. The parameters of the beam (i.e. its origin, orientation and angular aperture) are tunable. The output of the beamshaping system are five spatial filters, one for each beam to be rendered. Through the knowledge of the geometry of the environment, the position of the array and of the listening area, the system estimates through simple geometric considerations the orientation and the aperture of the beams. The room compensation block modifies the design of the spatial filters in order to attenuate the effect of the undesired reflections. In order to attain this result, it leverages on the knowledge of the geometry and of the position of the listening area. Given this information, the beam tracing block [7] is able to compute the position of the image speakers to be compensated. Finally, the filtering block receives as input the five spatial filters and the dry signal and feeds the filtered signals to the loudspeaker array.

3. BEAMSHAPING

In this section we summarize the beamshaping methodology formerly discussed in [7]. Consider the geometry in Figure 3. The speakers are placed in positions $\mathbf{p}_m,\ m = 1, \dots M$ and the reference frame is centered in the speaker array. Control points $\mathbf{a}_n,\ n = 1, \dots N$ describe in which area we intend to control the directivity of the array, i.e. the portion of the room in front of it. The beam is defined through its orientation θ and its aperture ϕ, and it is cast from the center of the array. The Green's function from $\mathbf{p}_m$ to $\mathbf{a}_n$ at the frequency ω is

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

All the $N \times M$ Green's functions are collected in a matrix $\mathbf{G}(\omega)$. The N elements vector $\mathbf{d}(\omega)$ contains the desired soundfield at all the control points. The spatial filter $\mathbf{h}(\omega)$ should satisfy

$$\mathbf{G}(\omega)\mathbf{h}(\omega) = \mathbf{d}(\omega) . \quad (1)$$

The least squares approximate solution $\hat{\mathbf{h}}(\omega)$ of (1) contains the coefficients of the spatial filter for the frequency ω. The design of the spatial filter is repeated for all the five channels to be rendered, i.e. C, L, R, RL, RR. In the following we will refer to $\hat{\mathbf{h}}^{(\cdot)}(\omega)$ as the spatial filter for a generic channel.

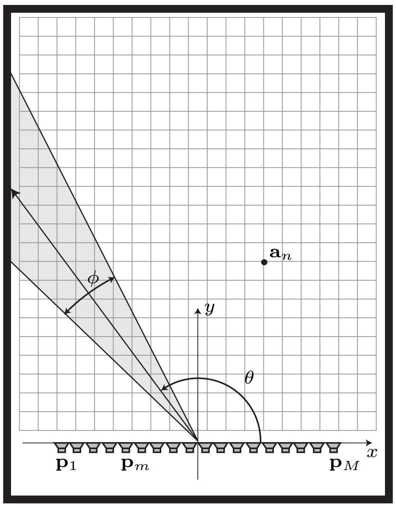


Fig. 3. Geometric notation of the beamshaping system.

4. BEAM PARAMETER ESTIMATION

Consider the geometry in Figure 4. The listener is assumed to be placed in the listening area with extension $w_z \times l_z$. For the sake of simplicity in the derivation, we assume the speaker array and the listening area to be on the axis of symmetry of a rectangular room. According to the geometric notation described in Figure 4, in order to reach the listener the five beams must be characterized by the following angles of departure:

$$\begin{cases} \theta_{\mathrm{C}} &= \pi/2 \\ \theta_{\mathrm{L}} &= \arctan\frac{h-d}{w} \\ \theta_{\mathrm{R}} &= \pi - \theta_{\mathrm{L}} \\ \theta_{\mathrm{RL}} &= \arctan\frac{h+d}{w} \\ \theta_{\mathrm{RR}} &= \pi - \theta_{\mathrm{RL}} \end{cases} . \quad (2)$$

We control the aperture of the beams R, L, RL and RR in such a way that their reflections over the walls overcome in the listening area the contribution from the direct path by a predetermined magnitude. We notice that, as in a traditional surround system, the five channels are correctly perceived in a single point (sweet-spot). However, the compensation of undesired reflections can be performed in the entire listening area, as described in the next Section.

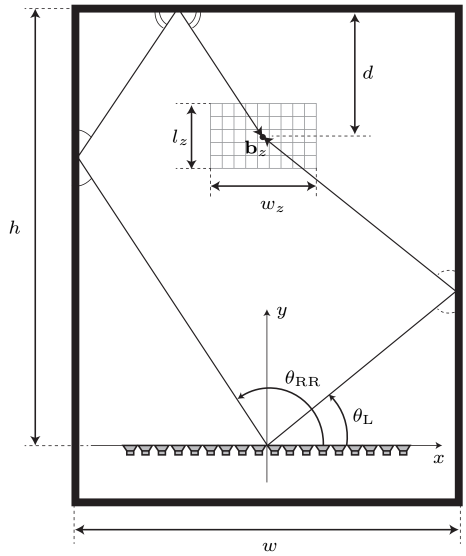


Fig. 4. Geometric notation for the estimation of the parameters of the beams and for the room compensation.

5. ROOM COMPENSATION

With reference to Fig. 4, we are interested in compensating undesired reflections in the listening area, in which we define a set of potential listening points $\mathbf{b}_z,\ z = 1, \dots Z$, independent from the control points defined for the beamshaping. By incorporating the reflections from the walls in the data model, the Green's function between the loudspeaker located at $\mathbf{p}_m$ and the listening point $\mathbf{b}_z$ becomes

$$\widetilde{g}_{zm}(\omega) = g_{zm}(\omega) + \gamma_{zm}(\omega) \ , \tag{3}$$

where $g_{zm}(\omega)$ is the Green's function of the direct path from $\mathbf{p}_m$ to $\mathbf{b}_z$, and

$$\gamma_{zm}(\omega) = \sum_{i=1}^{Q_m} \beta_{m,i} V(\mathbf{b}_z, \mathbf{p}'_{m,i}) \frac{e^{-j\frac{\omega}{c}\|\mathbf{b}_z - \mathbf{p}'_{m,i}\|}}{4\pi\|\mathbf{b}_z - \mathbf{p}'_{m,i}\|}, \tag{4}$$

are the Green's functions of the paths from $\mathbf{p}_m$ to $\mathbf{b}_z$ that include at least a reflection. The index i spans over the wall reflected images $\mathbf{p}'_{m,i}$ of the speaker $\mathbf{p}_m$; Q_m is the number of considered speaker images; $\beta_{m,i}$ is the attenuation due to the reflections over the walls; $V(\mathbf{b}_n, \mathbf{p}'_{m,i})$ is an index that evaluates if the image speaker $\mathbf{p}'_{m,i}$ is visible from the listening point $\mathbf{b}_z$ and it is evaluated through the beam tracing technique. The Green's functions $\widetilde{g}_{zm}(\omega)$ are collected in a $Z \times M$ matrix $\mathbf{P}(\omega)$.

We now define a second Green's function $\overline{g}^{(\cdot)}_{zm}(\omega)$, which describes the desired behavior of the beam in the listening area, i.e. containing only the reflections to be exploited. As an example, let us focus our attention on the beam L. In this case we have to retain the reflection from the left-side wall, while suppressing all the others. The Green's function of the desired soundfield is therefore

$$\overline{g}^{(\mathrm{L})}_{zm}(\omega) = g_{zm}(\omega) + \overline{\gamma}^{(\mathrm{L})}_{zm}(\omega) \ ,$$

where $\overline{\gamma}^{(\mathrm{L})}_{zm}(\omega)$ contains a subset of the image speakers included in $\gamma_{zm}(\omega)$ relative to the desired reflective path, which in this case is the reflection from the left-side wall. The Green's functions of the desired soundfield are organized in a $Z \times M$ matrix $\mathbf{D}^{(L)}(\omega)$. Similarly, this matrix is defined for the remaining beams RL, RR, R and C. As far as the central beam (C) is concerned, no reflection shall be audible in the listening area, therefore in this case $\overline{\gamma}^{(\mathrm{C})}_{zm}(\omega) = 0$ for all the speakers and listening points. Always with reference to the left channel, we are interested in finding a compensation matrix $\mathbf{C}^{(\mathrm{L})}(\omega)$ so that

$$\mathbf{P}(\omega)\mathbf{C}^{(\mathrm{L})}(\omega) = \mathbf{D}^{(\mathrm{L})}(\omega) \ . \tag{5}$$

The inversion of (5) in the least squares sense provides the approximation $\hat{\mathbf{C}}^{(\mathrm{L})}(\omega)$ of the compensation matrix, which leads to the room compensated spatial filter $\hat{\mathbf{h}}^{(\mathrm{L})}_{\mathrm{RC}}(\omega) = \hat{\mathbf{C}}^{(\mathrm{L})}(\omega)\hat{\mathbf{h}}^{(\mathrm{L})}(\omega)$. The response of the non-compensated beamshaper in the listening area is

$$\mathbf{r}^{(\mathrm{L})}_{\mathrm{NC}}(\omega) = [r^{(\mathrm{L})}_{\mathrm{NC}_1}(\omega), \dots r^{(\mathrm{L})}_{\mathrm{NC}_Z}(\omega)]^T = \mathbf{P}(\omega)\hat{\mathbf{h}}^{(\mathrm{L})}(\omega) \ , \tag{6}$$

while the response of the room compensated spatial filter is

$$\begin{aligned} \mathbf{r}^{(\mathrm{L})}_{\mathrm{RC}}(\omega) &= [r^{(\mathrm{L})}_{\mathrm{RC}_1}(\omega), \dots r^{(\mathrm{L})}_{\mathrm{RC}_Z}(\omega)]^T = \mathbf{P}(\omega)\hat{\mathbf{h}}^{(\mathrm{L})}_{\mathrm{RC}}(\omega) \\ &= \mathbf{P}(\omega)\hat{\mathbf{C}}^{(\mathrm{L})}(\omega)\hat{\mathbf{h}}^{(\mathrm{L})}(\omega) \approx \mathbf{D}^{(\mathrm{L})}(\omega)\hat{\mathbf{h}}^{(\mathrm{L})}(\omega) \ . \end{aligned} \tag{7}$$

We notice that (7) approximates the desired behavior in the listening area, i.e. the suppression of all the reflections but the desired ones.

6. RESULTS

In this section we show some simulative results aimed at showing the capabilities of the room compensation. In particular, we will focus on two kind of comparisons:

- the time-space response of the compensated and non compensated spatial filters compared with the desired response in the listening area. The generic time-space response is defined as $\boldsymbol{r}(t) = [\mathrm{r}_1(t), \dots \mathrm{r}_Z(t)]^T$. The response $\mathrm{r}_z(t)$ at each control point is approximated through an Inverse Discrete Fourier Transform applied to a sampled version of the response $r_z(\omega)$. In particular, in the simulations we considered 1024 frequency samples in the range from 0 to 4 kHz;
- the frequency behavior of the Normalized Mean Square Error (NMSE) between the desired and the non compensated / compensated soundfields averaged on the control points in the listening area. Given the generic frequency-space response (i.e. compensated or non-compensated) $\mathbf{r}(\omega)$ and the corresponding desired frequency-space response $\mathbf{r}_d(\omega) = [r_{d_1}(t), \dots r_{d_Z}(t)]^T$, the Normalized Mean Square Error (NMSE) is defined as

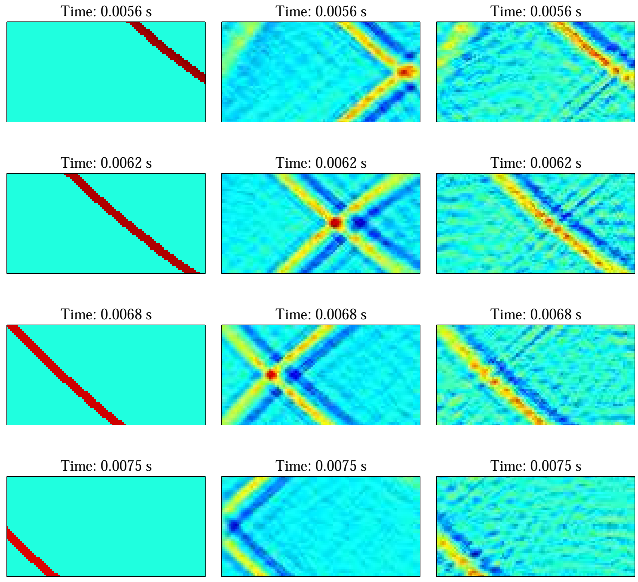
$$\mathrm{NMSE}(\omega) = \frac{\sum_{z=1}^{Z} |r_z(\omega) - r_{d_z}(\omega)|^2}{\sum_{z=1}^{Z} |r_{d_z}(\omega)|^2} \ .$$

We simulated the presence of a $2\,\mathrm{m}$ long linear array of $M = 32$ ideal loudspeakers in a reverberant room, with reflection coefficient equal to 0.75 for all the walls. The resulting spacing between the loudspeakers is about $6.5\,\mathrm{cm}$, which guarantees alias-free reproduction up to approximately $2.6\,\mathrm{kHz}$. With reference to the notation introduced in Fig. 4, the principal dimensions of the room and of the listening area are

$$\begin{cases} w = 3\ \mathrm{m} \\ h = 3\ \mathrm{m} \end{cases} \quad \text{and} \quad \begin{cases} d = 0.5\ \mathrm{m} \\ w_z = 1\ \mathrm{m} \\ l_z = 0.5\ \mathrm{m} \end{cases} ,$$

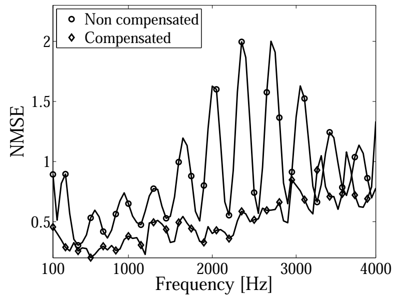
which give $\theta_{\mathrm{L}} = 39.8°$ and $\theta_{\mathrm{RL}} = 49.4°$. The angle of departure of beams θ_{R} and θ_{RR} are supplementary to θ_{L} and θ_{RL}, respectively. Room compensation is performed up to the 3rd order of reflection, while, for evaluation purposes, acoustic propagation is modeled up to the 10th order of reflection.

Fig.5 shows the time-space responses relative to the channel RL at different time instants within the listening area. In particular, the first column is the desired response; the second column is the non-compensated response, which well approximates the result achieved through the commercial system described in [2] and [1]; finally the third column refers to the compensated response. Notice that the desired wave front is present in both the non-compensated and compensated responses. However, the non-compensated one exhibits also a wave front resulting from a second order reflection coming from the bottom and the left-side walls. In the plots of the soundfield at time $t = 0.0056$s we also notice the presence of the direct wavefront, which only partially crosses the listening area. The influence of the direct wavefront appears more relevant in the compensated soundfield, but this is due to a different normalization.

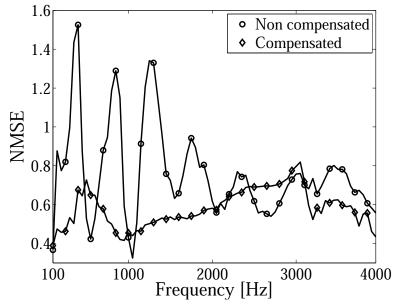


(a) Desired (b) Non-compensated (c) Compensated

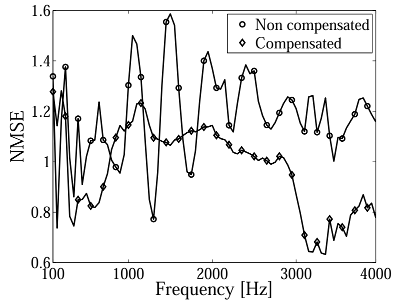
Fig. 5. Time-space response, in the listening area, of the channel RL at different time instants.



(a) Channel C



(b) Channel L



(c) Channel RL

Fig. 6. Normalized Mean Square Error of central, left and rear-left channels.

Fig.6 shows the frequency response of the Normalized Mean Square Error for the channels C, L and RL. Results for channels R and RR are not shown for reasons of symmetry. Notice that room compensation allows to reduce the Normalized Mean Square Error and consequently improve the accuracy of rendering.

7. CONCLUSIONS

In this paper we presented a technique for rendering a five channel surround system using a loudspeaker array positioned in a reverberant room. The system exploits the reflections coming from the walls of the environment in order to render the presence of loudspeakers placed around the listener. Moreover, the proposed technique is also able to compensate undesired reflections. The theoretical development is supported by experimental results which validate the proposed technique.

8. REFERENCES

[1] K. Suzuki, K. Kumagai, and S. Takumai, “Surround sound outputting device and surround sound outputting method,” Aug. 2009.

[2] Susumu Takumai, “Loudspeaker array device and method for setting sound beam of loudspeaker array device,” Mar. 2007.

[3] A. Canclini, P. Annibale, F. Antonacci, A. Sarti, R. Rabenstein, and S. Tubaro, “From direction of arrival estimates to localization of planar reflectors in a two dimensional geometry,” in *Proc. IEEE Int Acoustics, Speech and Signal Processing (ICASSP) Conf*, 2011, pp. 2620–2623.

[4] A. Canclini, F. Antonacci, M. R. P. Thomas, J. Filos, A. Sarti, P. A. Naylor, and S. Tubaro, “Exact localization of acoustic reflectors from quadratic constraints,” in *Proc. IEEE Workshop Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 17–20.

[5] T. Betlehem and T. D. Abhayapala, “A modal approach to soundfield reproduction in reverberant rooms,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP ’05)*, 2005, vol. 3.

[6] S. Spors, H. Buchner, and R. Rabenstein, “A novel approach to active listening room compensation for wave field synthesis using wave-domain adaptive filtering,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP ’04)*, 2004, vol. 4.

[7] D. Markovic, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, “Visibility-based beam tracing for soundfield rendering,” in *Proc. IEEE Int Multimedia Signal Processing (MMSP) Workshop*, 2010, pp. 40–45.