A GEOMETRICAL APPROACH TO ROOM COMPENSATION FOR SOUND FIELD RENDERING APPLICATIONS

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

Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano

ABSTRACT

In this paper we propose a method for reducing the impact of room reflections in sound field rendering applications. Our method is based on the modeling of the acoustic paths (direct and reflected) from each of the loudspeakers of the rendering system, and a set of control points in the listening area. From such models we derive a propagation matrix and compute its least-squares inversion. Due to its relevant impact on the spatial impression, we focus on the early reflections part of the Room Impulse Response, which is conveniently estimated using the fast beam tracing modeling engine. A least squares problem is formulated in order to derive the compensation filter. We also demonstrate the robustness of the proposed solution against errors in geometric measurement of the hosting environment.

Index Terms— Soundfield rendering, geometrical acoustics, room reflections, room compensation

1. INTRODUCTION

In the past years several techniques for soundfield rendering have been presented in the literature. Relevant examples are WaveField Synthesis (WFS) [1, 2] and Higher Order Ambisonics (HOA) [3,4]. They rely on different assumptions and are used for different geometric configurations, but guarantee a pleasant and accurate spatialization in the listening area (sweet spot).

Soundfield rendering techniques are becoming more and more popular also for consumer applications. Several problems, however, arise when spatialization is performed in non-specialized environments. Among these, one of the most relevant is the negative impact of the reverberations of the hosting rooms, affecting the spatial impression. In order to attenuate this effect, two different choices can be adopted. The first is based on the use of diffusive and absorptive panels, which attenuate the reflections over the walls. However, this solution is costly and unaffordable in most of the cases. Room compensation [5–7], on the other hand, is based on a preconditioning of the speaker signals, i.e. the signals are filtered so that the reflections coming from the walls of the environment are attenuated. Due to its reduced invasivity, this kind of technique has become quite popular. Most of the methods in this category are based on the use of a multiplicity of microphones which capture the impulse response of the room and, possibly, adapt the compensation to modifications of the environment (door opening, people moving, change in the temperature, etc.). The presence of microphones in proximity of the listening area [8] could be considered cumbersome, however, in most of the cases.

In a typical scenario, what corrupts the spatial impression, is the early reflections part of the RIR [9], where echoes are distinguishable. On the other hand, the reverberant tail, where echoes are close each other in time, alters the timbre of the signal. In this paper we focus on the suppression of the early reflections. Given the geometry of the hosting environment, the most relevant reflections of the room can be predicted without microphones in the picture. More specifically, we use a fast beam tracing technique to estimate the propagation condition between the loudspeakers and a set of control points within the listening area. This information is analyzed in a frequency-subband fashion and organized into the propagation matrices. A least squares problem is setup, which is aimed at finding the filters to be applied to the speaker signals to attenuate the early reflections.

No specific assumption is made about the wavefield rendering technique in use. This makes the proposed methodology viable in a wide range of scenarios. In particular, in this paper we compensate for reflections on HOA and WFS, which can be considered poles apart in terms of number of speakers needed and extension of the rendering area. In both cases the compensation technique guarantees a good suppression of the main reflections.

2. PROBLEM FORMULATION AND BACKGROUND

Consider the setup in Fig. 1. A set of speakers is located in $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_M$ and it is demanded to render a set of virtual sources, positioned in $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_V$, in the listening area, represented by the shaded region. The size of the listening area changes according to the rendering technique adopted and to the frequency range of the signal. As an example, for WFS theoretically the listening area is extended to the whole half-space opposite to that of the virtual sources. On the other

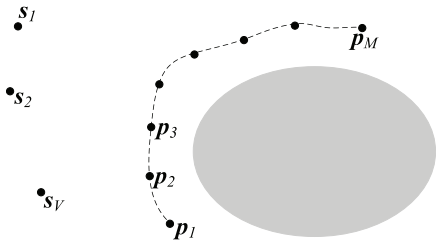

Fig. 1. The setup of a generic rendering system.

hand, HOA requires the speakers to surround the listener and the size of the sweet spot jointly depends on the frequency range and the order used in the spherical harmonics decomposition. Rendering techniques based on a numerical approximation [10, 11] require to define a set of control points within the listening area.

Regardless of the specific rendering technique adopted, they operate a filtering on the signal fed to the speakers in order to provide the desired spatial impression in the listening area. These filters are computed under the assumption of ideal free-field propagation, i.e. no reflective obstacle is present in the environment. We denote the rendering filter with the symbol $\mathbf{h}_{\mathrm{NC}}(\omega) = [h_{\mathrm{NC}_1}(\omega), h_{\mathrm{NC}_2}(\omega), \ldots, h_{\mathrm{NC}_M}(\omega)]^T$, where the subscript "NC" stands for non-compensated. The goal is to compute a set of modified "room-compensated" (RC) filters $\mathbf{h}_{\mathrm{RC}}(\omega) = [h_{\mathrm{RC}_1}(\omega), h_{\mathrm{RC}_2}(\omega), \ldots, h_{\mathrm{RC}_M}(\omega)]^T$ that dampen the early reflections coming from the walls of the hosting environment. Due to their corruptive effect on the spatial impression, we limit our attention to the compensation of the early reflections. The proposed room compensation technique is numerical, and therefore requires to define of a set of control points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ within the compensation area A_C. The compensation technique acts at its best in A_C, even if it could be effective also in a slightly wider area. We use the tools of geometrical acoustics for modeling the reflections from the hosting environment. Geometrical acoustics require the reflective surfaces to be large enough with respect to the wavelenght of the incident waves. In this context it is possible to model the reflected wavefronts as being generated by the image loudspeakers, whose location and orientation are obtained by mirroring the real speakers against the walls generating the reflections. Notice that image loudspeakers iteratively generate higher order image sources. We denote the image sources of the mth speaker with the symbol $\boldsymbol{p}'_{m,i}, i = 1, \ldots, Q_m$. It is easy to conclude that the number of image speakers grows more than quadratically with the reflection order. For each image source we need to evaluate its visibility from each point in A_C. This is a costly operation, as it consists in evaluating the occlusion of the image source on the part of every reflector in the environment. In order to speed up this process, we use fast beam tracing [12], which inherently computes the visibility information for each image source. The propagation condition between the speaker $\boldsymbol{p}_m$ and the control point $\boldsymbol{a}_n$ is given by the reverberant Green's function

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

where

$$g_{nm}(\omega) = \frac{e^{-j\frac{\omega}{c}\|\boldsymbol{a}_n - \boldsymbol{p}_m\|}}{4\pi\|\boldsymbol{a}_n - \boldsymbol{p}_m\|} \tag{2}$$

is the free-field Green's function and $V(\boldsymbol{a}_n, \boldsymbol{p}'_{m,i})$ is a binary function: $V(\boldsymbol{a}_n, \boldsymbol{p}'_{m,i}) = 1$ if $\boldsymbol{p}'_{m,i}$ is visible from $\boldsymbol{a}_n$ and $V(\boldsymbol{a}_n, \boldsymbol{p}'_{m,i}) = 0$ if it is occluded. $\beta_{m,i}$ is the attenuation coefficient for the image source $\boldsymbol{p}'_{m,i}$ and depends on the reflection order of the loudspeaker and on the reflective properties of the walls.

The reverberant Green's functions at frequency ω are organized in the $N \times M$ matrix

$$\mathbf{P}(\omega) = \begin{bmatrix} \gamma_{1,1}(\omega) & \ldots & \gamma_{1,M}(\omega) \\ \vdots & \ddots & \vdots \\ \gamma_{N,1}(\omega) & \ldots & \gamma_{N,M}(\omega) \end{bmatrix}. \tag{3}$$

Similarly, the free-field Green's functions at frequency ω are organized in the matrix

$$\mathbf{G}(\omega) = \begin{bmatrix} g_{1,1}(\omega) & \ldots & g_{1,M}(\omega) \\ \vdots & \ddots & \vdots \\ g_{N,1}(\omega) & \ldots & g_{N,M}(\omega) \end{bmatrix}. \tag{4}$$

3. COMPENSATION OF EARLY REFLECTIONS

We formulate the problem of room compensation as the computation of the matrix $\mathbf{C}(\omega)$ so that

$$\mathbf{P}(\omega)\mathbf{C}(\omega) = \mathbf{G}(\omega)\,. \tag{5}$$

Notice that the matrix $\mathbf{P}(\omega)$ is typically ill-conditioned with clear consequences on the robustness of the solution. A reconditioning of the solution of the least-squares problem in (5) is therefore in order. We resort to a regularization based on the Truncated Singular Value Decomposition (TSVD) of the least-squares inverse of the matrix $\mathbf{P}(\omega)$. The SVD of $\mathbf{P}(\omega)$ is given by

$$\mathbf{P}(\omega) = \mathbf{V}(\omega)\mathbf{\Sigma}(\omega)\mathbf{U}^H(\omega)\,, \tag{6}$$

where $\mathbf{\Sigma}(\omega)$ is the matrix containing the singular values $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_M^2$ of $\mathbf{P}(\omega)$. We retain only the K greatest singular values, so that

$$\sigma_1/\sigma_K \leq T\,, \tag{7}$$

where T is a prescribed threshold value. The regularized pseudo-inverse of $\mathbf{P}(\omega)$ is

$$\mathbf{P}_K^+(\omega) = \mathbf{U}(\omega)\mathbf{\Sigma}_K^+\mathbf{V}^H(\omega)\,, \tag{8}$$

where $\mathbf{\Sigma}_K^+ = \mathrm{diag}(1/\sigma_1^2, \ldots, 1/\sigma_K^+)$.

The compensation matrix is then obtained as

$$\mathbf{C}(\omega) = \mathbf{P}_K^+(\omega)\mathbf{G}(\omega) \; . \tag{9}$$

Finally, the room compensated filters are

$$\mathbf{h}_{\mathrm{RC}}(\omega) = \mathbf{C}(\omega)\mathbf{h}_{\mathrm{NC}}(\omega) \, . \tag{10}$$

Notice that no assumption about the geometry of the rendering system nor the rendering area is done. The only requirement to be met is that the output of the rendering system is written in terms of space-time filters $\mathbf{h}_{\mathrm{NC}}(\omega)$. The proposed methodology can be used, therefore, for a broad range of rendering techniques. In order to extend the proposed methodology to wideband signals, we compute the filters in a sub-band fashion.

As far as the threshold is concerned, we empirically assessed that choosing a higher value of T corresponds to increasing the robustness against possible errors in the data model. We found that a reasonable value is $T = 5$, used throughout all the experiments in the next section.

4. RESULTS

In this Section we show the results of some simulations, with the aim of verifying the effectiveness of the room compensation methodology and of testing its robustness against errors in the geometrical model of the hosting environment.

Setup With reference to Figure 2, we consider a uniform circular array with radius $1.5\,\mathrm{m}$ composed of $M = 48$ loudspeakers. A virtual point source, located at (r_s, θ_s) in the polar coordinate system in Fig. 2, is rendered with both WFS and HOA. Notice that we adopt the same setup for both the rendering techniques, in order to make comparable the results. Room compensation is performed in a circular region inside the array (gray-shaded area), with radius $1.35\,\mathrm{m}$ and sampled with $N = 600$ control points on a uniform rectangular grid. The hosting environment presents an irregular shape, whose perimeter is depicted by the external continuous black line in Fig. 2. We set the reflection coefficient at 0.8 for all the walls. Different offsets Δ of the nominal position of the bottom wall are considered, in order to verify the robustness of the compensation technique against geometrical errors of the room model. In all the tests, room compensation is performed up to the 3rd order of reflections, while for the evaluation of the system we modeled early reflections up to the 10th order.

Evaluation metric We evaluate the room compensation system analyzing the angular components of the wavefield, for all the possible directions of arrival $\theta \in [0°, 360°]$. To do so, we compute the power of the plane wave components [5]

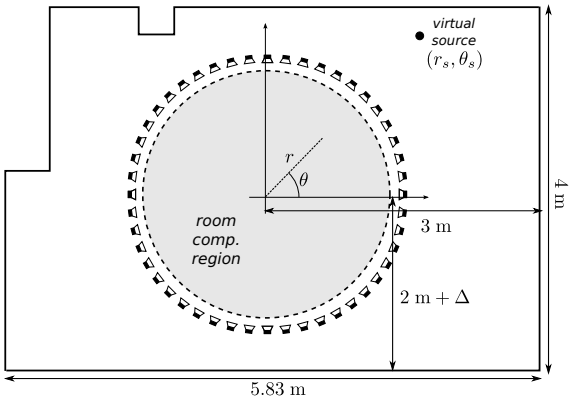

Fig. 2. Simulation setup.

$$W(\theta) = \int_0^{f_a} |L(\omega,\theta)|^2 \, d\omega = \int_0^{f_a} \left| \sum_{\mu=-\infty}^{+\infty} j^{-\mu} C_\mu(\omega) e^{j\mu\theta} \right|^2 d\omega \; , \tag{11}$$

where $L(\omega, \theta)$ is the plane wave decomposition [5], ω being the angular frequency; j is the imaginary unit; and $C_\mu(\omega)$ is the μth circular harmonics [5]. The spatial aliasing frequency f_a is approximately $900\,\mathrm{Hz}$ for the array under consideration.

Simulations As a first test, we consider the rendering of a virtual source located at a distance $r_s = 3\,\mathrm{m}$ from the array, with angular position $\theta_s = 150°$. For the moment, we consider no errors in the geometrical model, i.e. we pose $\Delta = 0$. Figure 3 shows the impulse response at the center of the listening area, for WFS and HOA. In particular, it compares the non-compensated response (i.e., obtained when no room compensation is performed) with the room-compensated one (i.e., obtained when the proposed room compensation technique is active). The effect of early reflections is clearly visible in the non-compensated response, which is characterized by several peaks occurring after the first one (related to the direct path between the source and the point of observation). Conversely, when room compensation is enabled, early reflections turn to be significantly attenuated for both WFS and HOA.

The results in Figure 3 are purely qualitative and restricted to a single observation point. We now consider the metric introduced by (11). Indeed, being defined over a source-free region of space [5], the plane wave decomposition turns to be informative of the entire rendering area. Figure 4 shows the function $W(\theta)$ for WFS and HOA, for different testing conditions. In particular, the blue line shows the non-compensated response; the red line is the error-free ($\Delta = 0$) room compensated response; the colored regions represent groups of room compensated responses, computed for different errors Δ. More specifically, we considered the following sets: $I_1 := \{\Delta : 1\,\mathrm{cm} \leq |\Delta| < 3\,\mathrm{cm}\}$, $I_2 := \{\Delta : 3\,\mathrm{cm} \leq |\Delta| < 6\,\mathrm{cm}\}$

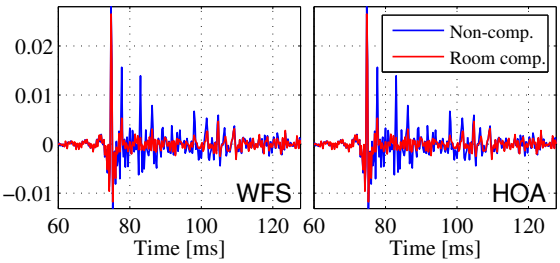

Fig. 3. Impulse response at the center of the listening area relative to the rendering of a virtual source at $(r_s = 3\,\mathrm{m}, \theta_s = 150°)$.

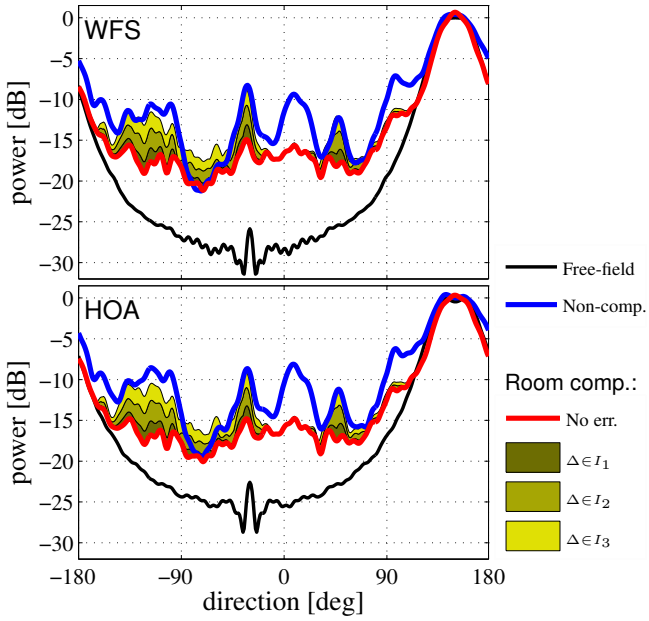

Fig. 4. Angular components of the wavefield relative to the rendering of a virtual source at $(r_s = 3\,\mathrm{m}, \theta_s = 150°)$.

and $I_3 := \{\Delta : 3\,\mathrm{cm} \leq |\Delta| \leq 9\,\mathrm{cm}\}$. Finally, the black curve denotes the ideal free-field response, i.e. as obtained if the hosting environment were completely anechoic. Notice in Figure 4 that the the rendering technique does not introduce relevant differences in both the non-compensated and room compensated responses. Therefore, for the sake of compactness, in the following discussion we limit to provide numerical details for the WFS case only. All the curves are normalized with respect to the free-field response, such that the main peak has a power of 0 dB. We first observe that the highest peak in all the responses corresponds to the angular position of the rendered source ($\theta_s = 150°$). As expected, it represents the only relevant plane-wave component in the free-field response[1]. The effect of the room is evident in the non-compensated response, which exhibits several peaks corresponding to the main reflections: $\theta_A = -117°$ (-10.5 dB), $\theta_B = -32°$ (-8.2 dB), $\theta_C = 9°$ (-9.3 dB), $\theta_D = 48°$ (-12.2 dB), $\theta_E = 97°$ (-7.1 dB). Room compensation attenuates their power. In particular, in error-free conditions ($\Delta = 0$) the resulting attenuation is of 7 dB, 6.7 dB, 6.3 dB, 5.7 dB, 4.8 dB, respectively. When we introduce an error on the position of the bottom wall, room compensation turns to be less effective. However, the degradation smoothly increases with the error, and is limited to the directions $\Theta = \{-135° < \theta < -25° \cup 40° < \theta < 55°\}$, which correspond to reflections involving the bottom wall. For $\Delta \in I_1$ (maximum error $\Delta_{MAX} = \pm 3\,\mathrm{cm}$), the degradation is almost negligible; for $\Delta \in I_2$ ($\Delta_{MAX} = \pm 6\,\mathrm{cm}$), we observe that the attenuation of early reflections is approximately halved in the region Θ; finally, for $\Delta \in I_3$ ($\Delta_{MAX} = \pm 9\,\mathrm{cm}$), room compensation does not introduce any substantial attenuation of reflections lying in Θ. Figure 5 details the attenuation of early reflections from the directions $\theta_A \ldots \theta_E$, for WFS and HOA, as a function of the error Δ. We observe that the attenuation of reflections at θ_A, θ_B and θ_D (that belong to the region Θ) is maximum for $\Delta = 0$ and monotonically decreases approaching 0 dB for $\Delta \approx \pm 10\,\mathrm{cm}$, following a bell-shaped symmetric behavior. As expected, for the reflections corresponding to θ_C and θ_E, the attenuation is almost independent from Δ.

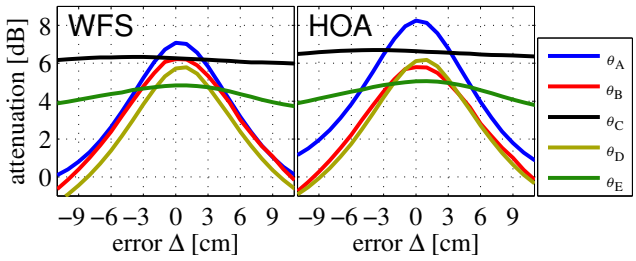

Fig. 5. Attenuation of early reflections, for a source at $(r_s = 3\,\mathrm{m}, \theta_s = 150°)$, as a function of the error Δ.

We now consider the rendering of a point source located far from the array, at a distance $r_s = 10\,\mathrm{m}$ and with angular position $\theta_s = 90°$. If we introduce some error Δ in the position of the bottom wall, we expect room compensation to be more challenging in this scenario. Indeed, due to its distance from the array, the source generates planar wavefronts that directly propagates towards and bounce off the bottom wall. Moreover, second order reflections are generated by the opposite wall (top wall in Figure 2), which are then further reflected from the bottom wall, and so on. The results relative to this situation are drawn in Figure 6, which plots the function $W(\theta)$ for WFS and HOA. As before, the curves are scaled so that the reference free-field response is 0 dB at $\theta_s = 90°$. Also in this case the effect of room compensation is very sim-

[1]The peak at $-30°$ is due to numerical errors introduced by the computation of the plane wave decomposition, and it is not related to any plane wave from that direction.

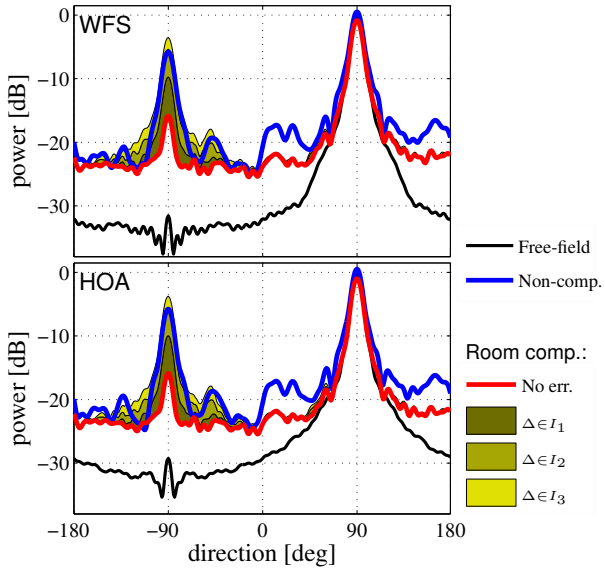

Fig. 6. Angular components of the wavefield relative to the rendering of a virtual source at $(r_s = 10\,\mathrm{m}, \theta_s = 90^\circ)$

ilar for both WFS and HOA, thus confirming its independence from the adopted rendering technique. We first observe that, since the rendered wavefronts are almost planar, the direct path from the virtual source produces a narrow peak centered at 90° in all the curves. A replica of the main peak is present at -90°, which accounts for the multiple (first and higher order) reflections generated by the bottom wall. When no compensation is performed, the power of the reflected peak is $-5.7\,\mathrm{dB}$, which is attenuated by approximately $11\,\mathrm{dB}$ when room compensation is active (for $\Delta = 0$). Moreover, it is interesting to notice that the peak at 90° in the non-compensated curve is slightly above $0\,\mathrm{dB}$, since it includes the second (and higher order) reflections from the top wall, while room compensation properly reduces its power to $0\,\mathrm{dB}$. Finally, we analyze the responses of room compensation considering different errors Δ, grouped, as before, in the error sets I_1, I_2 and I_3. The results in Figure 6 show that, even in a more critic scenario, room compensation is reasonably robust against the error in positioning the bottom reflector. Indeed, on one hand we notice that the attenuation of the main reflection at -90° decreases more rapidly in this case, for increasing values of $|\Delta|$. On the other hand, a big error ($\Delta_{\mathrm{MAX}} = \pm 9\,\mathrm{cm}$) only slightly worsens the non-compensated behavior. Furthermore, in this case the degradation results to be even more localized in space, being confined in the region $-135^\circ < \theta < -45^\circ$.

5. CONCLUSIONS

In this paper we proposed a technique for compensating early reflections in sound field rendering applications, which relies on a geometrical model of the room hosting the loudspeaker system. Tools of geometrical acoustics are used for predicting the reverberant Green's functions from the loudspeakers to the listening area, which are then encoded into a propagation matrix, whose least square inversion provides room compensation filters. The proposed method is independent from the adopted rendering algorithm. Simulations on WFS- and HOA-based systems prove the effectiveness of the method, which turned to be reasonably robust against errors in the geometrical model of the hosting environment. We are currently envisioning the possibility of increasing the robustness of the method by introducing additional constraints in the regularization of the propagation matrix.

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 and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, pp. 2038 – 2050, 2010.

[3] Jerome Daniel, Sebastien Moreau, and Rozenn Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Audio Engineering Society Convention 114*, 3 2003.

[4] Rozenn Nicol, "Sound spatialization by higher order ambisonics: Encoding and decoding a sound scene in practice from a theoretical point of view.," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, 2010.

[5] S. Spors, A. Kuntz, and R. Rabenstein, "An Approach to Listening Room Compensation with Wave Field Synthesis," in *AES 24th International Conference on Multichannel Audio*, Banff, Canada, June 2003, pp. 49–52.

[6] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *Journal of the American Society of Acoustics*, vol. 117, no. 4, pp. 2100–2111, 2005.

[7] M. Schneider and W. Kellermann, "Adaptive Listening Room Equalization Using a Scalable Filtering Structure in the wave domain," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012.

[8] 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.

[9] M. Kahrs and K. Brandenburg, *Applications of Digital Signal Processing to Audio and Acoustics*, Kluwer International Series in Engineering and Computer Science. Kluwer, 1998.

[10] F. Antonacci, A. Calatroni, A. Canclini, A. Galbiati, A. Sarti, and S. Tubaro, "Soundfield rendering with loudspeaker arrays through multiple beam shaping," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA '09*, New Paltz, New York, USA, October 2009, pp. 313 –316.

[11] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using the lasso," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 1902–1912, 2010.

[12] D. Markovic, A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Visibility-based beam tracing for soundfield rendering," in *2010 IEEE International Workshop on Multimedia Signal Processing (MMSP)*, oct. 2010, pp. 40 –45.