ESTIMATION OF ROOM DIMENSIONS FROM A SINGLE IMPULSE RESPONSE

Dejan Marković, Fabio Antonacci, Augusto Sarti, Stefano Tubaro

Politecnico di Milano
Dipartimento di Elettronica, Informazione e Bioingegneria

ABSTRACT

In this paper we propose a methodology for the estimation of the geometry of an environment based on a single Acoustic Impulse Response (AIR). The estimation algorithm makes use of tools for the modeling of propagation based on geometrical acoustics. A suitable cost function evaluates the distance between the simulated and measured AIRs. The room minimizing the cost function is chosen as the correct one. The cost function is strongly non linear. As a consequence, in order to reduce the complexity of the minimization problem, the algorithm needs a hypothesis about the class of geometry of the environment under analysis, such as rectangular or L-shaped rooms. We prove the effectiveness of the proposed algorithm with a number of simulations with increasing complexity.

Index Terms— Room geometry estimation, acoustic impulse response

1. INTRODUCTION

In many applications of space-time processing such as source localization [1], wave-field rendering [2, 3] or estimation of the radiative properties of acoustic sources [4], the knowledge of the geometry of the environment is needed or it could be beneficial.

In the last few years, several techniques have been presented in the literature for the estimation of the geometry of the environment from acoustic signals. Relevant examples are [5, 6, 7, 8]. Even if different assumptions are made, the above methodologies require that multiple acoustic impulse responses are acquired at different locations of sources and/or microphones, thus preventing their use in scenarios where a limited number of sensors and transducers is available. In [9], authors have presented a methodology that uses a single impulse response to estimate the room geometry. Knowledge of the delays of first and second-order reflections is converted into information about orientation and location of walls. Authors also demonstrate the uniqueness of the impulse response for a given geometry. An inaccurate knowledge of the delay of even a single echo in the AIR, however, could lead the algorithm to wrong estimations.

In this paper we want to keep the advantage of using a single impulse response, while gaining robustness against errors in the knowldge of the delays of the AIR. In the proposed methodology, a cost function compares the location of the echoes in the simulated and measured impulse responses and estimates the geometry as the one that minimizes the distance function. For this purpose we use a suitable tool for the modeling of the propagation. Since the cost function is based only on the comparison of the location of echoes in measured and simulated AIRs, we are interested in propagation simulation tools that accurately estimate the time of arrival of echoes. Simulation tools based on geometrical acoustics are suitable for this scenario. Among all the available solutions, in particular, we have chosen beam tracing [10]. The cost function non-linearly depends on the dimensions of the room. As a consequence, minimization algorithms based on the gradient of the cost function could get trapped in local minima. In order to prevent this situation we adopt genetic algorithms to perform the minimization. Moreover, we assume that information about the shape of the environment is available, while dimensions of the walls are unknown, thus simplifying the complexity of the minimization. This fact, along with the formulation of the cost function makes the algorithm robust against spurious echoes. Therefore, even if it provides an estimate of the floormap of the environment only, it can work also when reflections from floor and ceiling are relevant.

We prove the accuracy of the geometry estimation algorithm for two classes of room geometries: rectangular and L-shaped. The latter environment, in particular, is challenging. In fact, depending on the location of source and receiver, some walls could be visible only through second-order reflections.

The rest of the paper is organized as follows: section 2 formulates the problem of room geometry estimation for the considered scenario. Section 3 discusses the methodology. Section 4 presents some experiments and simulations aimed at proving the effectiveness of the proposed algorithm. Finally, section 5 draws some conclusive remarks.

2. PROBLEM STATEMENT

Let us consider environments whose size is unknown but whose shape can be determined through a visual inspection. Relevant examples are rectangular and L-shaped rooms. It is known in the literature that the relationship between the geometry of the environment and the acoustic impulse response between source and receiver in it, is complex and just for a few cases it can be predicted in a closed form. One of these cases is the rectangular room, but even in this simple case it is hard to determine the geometry with a single impulse response.

Let us consider the example in Fig. 1, where the source S and the receiver R are located in $\mathbf{x}_S = [x_S, y_S]$ and $\mathbf{x}_R = [x_R, y_R]$, respectively. Once the shape of the room is fixed, the remaining parameters of the room are the lengths of the walls, denoted with $d_1, \ldots, d_n$. One arbitrary corner of the room is chosen as the center of the reference frame. With this knowledge and the sound speed c, it is possible to predict the delays of the echoes associated to early reflections of the AIR. Moreover, if the acoustic properties of the walls (modeled by the frequency-dependent reflection coefficients $\alpha_1(f), \ldots, \alpha_n(f)$) are known, also the amplitude of the echoes can be predicted. We denote the predicted early reflections of the AIR with

$$h(d_1, \alpha_1(f), \ldots, d_n, \alpha_n(f), \mathbf{x}_S, \mathbf{x}_R, t) \ , \qquad (1)$$

where t is the time variable. In the following, for the sake of com-

978-1-4799-0972-8/13/$31.00 ©2013IEEE

pactness, we use the symbol

$$\boldsymbol{\lambda} = [d_1, \alpha_1(f), \dots, d_n, \alpha_n(f), \mathbf{x}_S, \mathbf{x}_R]$$

to denote the parameters of the room. It is worth noticing that the

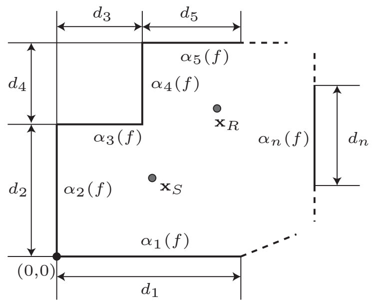


Figure 1: Floormap of an environment: $\mathbf{x}_S$ and $\mathbf{x}_R$ denote source and receiver locations; walls have lengths $d_1, \dots, d_n$ and reflection coefficients $\alpha_1(f), \dots, \alpha_n(f)$.

room geometry estimation algorithm proposed in this paper does not take into account the amplitude of the echoes but only their location on the time axis. In what follows, therefore, we drop the dependency of h on $\alpha_1(f), \dots, \alpha_n(f)$. The measured early-refletions part of the AIR between $\mathbf{x}_S$ and $\mathbf{x}_R$ is denoted by $\bar{h}(t)$.

We introduce a function $f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}$, which measures the similarity between modeled and measured early reflections. The room geometry is estimated as

$$\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda}} f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}\ . \tag{2}$$

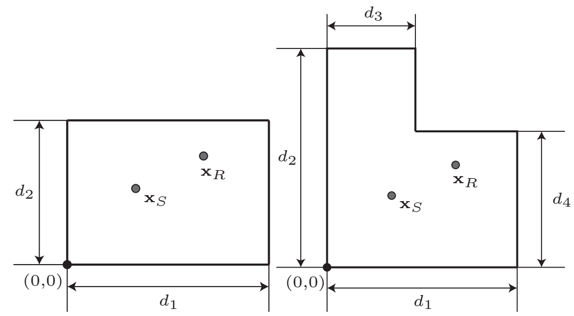
It is easy to realize that the dependency of $f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}$ on $\boldsymbol{\lambda}$ is non linear and it may present several local minima. As a consequence, minimization algorithms based on gradient descent are not suitable for this specific problem. Moreover, also the definition of the similarity function $f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}$ is crucial for a successful estimation. Therefore, in order to estimate the geometry of the environment given a single impulse response we need:

- a modeling engine that, given the geometry of the environment, efficiently computes the early reflections $h(\boldsymbol{\lambda}, t)$ between the source and receiver;
- a suitable cost function $f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}$ that measures the similarity between simulated $h(\boldsymbol{\lambda}, t)$ and measured early reflections $\bar{h}(t)$;
- a global optimization algorithm that is able to minimize $f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}$ for an accurate estimate of the room geometry $\boldsymbol{\lambda}$.

These requirements are discussed in the next section.

3. GEOMETRY ESTIMATION

In this manuscript we focus on the two types of environments depicted in Fig. 2. Rectangular rooms are characterized by the dimensions $\mathbf{d} = [d_1, d_2]$, while L-shaped ones by $\mathbf{d} = [d_1, d_2, d_3, d_4]$. These classes of geometries are important because many everyday environments fall in one of these two classes or combinations thereof. Also other shapes of environments can, however, be accommodated in the methodology. The mutual position of $\mathbf{x}_R$ and $\mathbf{x}_S$ are assumed to be known, thus the offset $\Delta\mathbf{x} = \mathbf{x}_R - \mathbf{x}_S$ is known. We consider in the following the location $\mathbf{x}_S$ as being either known or unknown. In the first case the optimization variables are $\boldsymbol{\lambda} = \mathbf{d}$, while in the second case the optimization variables are $\boldsymbol{\lambda} = [\mathbf{d}, \mathbf{x}_S]$. Results will provide examples for both cases.



(a) Rectangular room (b) L-shaped room

Figure 2: Examined classes of environments.

3.1. Cost function

Given the signal acquired with a single microphone and the signal emitted by the loudspeaker, there are techniques in the literature that extract the impulse response between them, e.g. [11]. Due to the complex nature of acoustic propagation in enclosures, echoes in $h(\boldsymbol{\lambda}, t)$ and $\bar{h}(t)$ may exhibit different relative magnitudes. In defining the cost function we decide, therefore, to focus on the temporal location of the echoes only. Techniques of propagation modeling based on geometrical acoustics can accurately predict the delays of echoes. We decided, in particular, to adopt fast beam tracing [10] for the computation of $h(\boldsymbol{\lambda}, t)$. Given the early reflections $h(\boldsymbol{\lambda}, t)$, we extract the corresponding Times Of Arrival (TOAs) as

$$T = \{\tau \mid h(\boldsymbol{\lambda}, \tau) > \Delta\}, \tag{3}$$

where Δ is a prescribed threshold. The TOAs $\tau_j \in T$ depend on room dimensions $\mathbf{d}$ as well as on the source and receiver positions, $\mathbf{x}_S$ and $\mathbf{x}_R$. Therefore, in order to estimate the unknown parameters $\boldsymbol{\lambda}$, our objective is to match the TOAs extracted from simulation $\tau_j \in T$ with TOAs $\bar{\tau}_i \in \bar{T}$ obtained from the early reflections part in $\bar{h}(t)$, obtained as in (3).

When using a single microphone (no information on directions of arrival), it is not possible to assign echoes to walls that generated them and, therefore, the function $h(\boldsymbol{\lambda}, t)$ is not invertible in general. Furthermore, due to occlusions, some reflective paths may be present in $\bar{h}(t)$ but not in $h(\boldsymbol{\lambda}, t)$ and vice versa. As an example, spurious peaks could appear in $\bar{h}(t)$ due to the noise or reflections from floor and ceiling. For these reasons, it is difficult to pair $\bar{\tau}_i \in \bar{T}$ with corresponding $\tau_j \in T$ in an element-wise fashion.

In order to make the estimation algorithm robust against wrong matches of TOAs, we propose a cost function $f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}$ defined as

$$f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\} = \sum_{\bar{\tau}_i \in \bar{T}} \min_{\tau_j \in T} |\bar{\tau}_i - \tau_j| + \sum_{\tau_j \in T} \min_{\bar{\tau}_i \in \bar{T}} |\bar{\tau}_i - \tau_j|, \tag{4}$$

which is the sum of distances (in time) of each measured TOA, $\bar{\tau}_i \in \bar{T}$, from the closest simulated TOA, $\tau_j \in T$, plus the sum of distances of each simulated TOA, $\tau_j \in T$, from the closest measured TOA, $\bar{\tau}_i \in \bar{T}$.

3.2. Minimization algorithm

As stated previously, methods based on gradient descent are not suitable for minimization of (4) as they could get stuck in a local minimum. Therefore, global optimization algorithms such as simulated annealing or genetic algorithms [12] should be used. In this paper we propose the adoption of the genetic algorithm for the minimization purpose. In the following we list the steps of the algorithm.

Step 1 - The initial population of rooms $\Lambda^{(0)} = \{\boldsymbol{\lambda}_m^{(0)}, m = 1, ..., M\}$, where M is the number of rooms in a single generation, is generated randomly with a uniform distribution over the space of feasible values $[\boldsymbol{\lambda}_{\min}, \boldsymbol{\lambda}_{\max}]$

$$\boldsymbol{\lambda}_m^{(0)} \sim \mathcal{U}(\boldsymbol{\lambda}_{\min}, \boldsymbol{\lambda}_{\max}). \tag{5}$$

The terms $\boldsymbol{\lambda}_{\max}$ and $\boldsymbol{\lambda}_{\min}$ are determined through a preliminary analysis of TOAs of relevant echoes in $\bar{h}(t)$.

Step 2 - The value of the cost function $f\{h(\boldsymbol{\lambda}_m^{(k)}, t), \bar{h}(t)\}$ is evaluated for each room of the kth generation, $\Lambda^{(k)} = \{\boldsymbol{\lambda}_m^{(k)}, m = 1, ..., M\}$, and the best N rooms are migrated to the next generation $\Lambda^{(k+1)}$. The remaining $M - N$ rooms of $\Lambda^{(k+1)}$ are generated from $\Lambda^{(k)}$ either through the crossover (fraction p) of two rooms or through the mutation (fraction $1 - p$) of a single room. The rooms that undergo crossover/mutation are chosen in a series of tournament selections: a random set of w rooms is extracted from the population, $\Omega \subset \Lambda^{(k)}$, and the room $\boldsymbol{\lambda}_m^k \in \Omega$ with the best value of the cost function is selected.

Crossover - Given two rooms selected from a previous generation $\boldsymbol{\lambda}_n^{(k)}, \boldsymbol{\lambda}_l^{(k)} \in \Lambda^{(k)}$, a good but not optimum value of their cost function may be due to the fact that both rooms approximate well some but not all parameters $\boldsymbol{\lambda}$ of the measured environment. However, as the function $h(\boldsymbol{\lambda}, t)$ is not invertible, we are unable to determine which parameters have a good/bad match. As a consequence, the crossover function Φ makes a random weighed mean of the two rooms

$$\begin{aligned}\boldsymbol{\lambda}_m^{(k+1)} &= \Phi(\boldsymbol{\lambda}_n^{(k)}, \boldsymbol{\lambda}_l^{(k)}) \\ &= \boldsymbol{\beta} \odot \boldsymbol{\lambda}_n^{(k)} + (\mathbf{1} - \boldsymbol{\beta}) \odot \boldsymbol{\lambda}_l^{(k)}, \ \ \boldsymbol{\beta} \sim \mathcal{U}(\mathbf{0}, \mathbf{1}),\end{aligned} \tag{6}$$

where $\odot$ denotes element-wise multiplication.

Mutation - The mutation $\boldsymbol{\lambda}_m^{(k+1)} = \Psi(\boldsymbol{\lambda}_n^{(k)})$ of the room $\boldsymbol{\lambda}_n^{(k)} \in \Lambda^{(k)}$ has the objective to prevent falling of all elements in the population to a local minimum. We design different types of mutation with different probabilities to occur: the perturbation

$$\Psi_a(\boldsymbol{\lambda}_n^{(k)}) = \boldsymbol{\lambda}_n^{(k)} + \boldsymbol{\delta}, \ \ \boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}), \tag{7}$$

where σ^2 is the perturbation variance; the generation of a new room

$$\Psi_b(\boldsymbol{\lambda}_n^{(k)}) = \boldsymbol{\lambda}_{\min} + (\boldsymbol{\lambda}_{\max} - \boldsymbol{\lambda}_{\min}) \odot \boldsymbol{\beta}, \ \ \boldsymbol{\beta} \sim \mathcal{U}(\mathbf{0}, \mathbf{1})\ ; \tag{8}$$

and the exchanging of vertical and horizontal axis

$$\Psi_c(\boldsymbol{\lambda}_n^{(k)}) = \overset{\sim}{\boldsymbol{\lambda}}{}_n^{(k)}, \tag{9}$$

where the $\tilde{\cdot}$ symbol denotes the fact that the values of dimensions and coordinates corresponding to x and y axes have been swapped. The small changes, such as Ψ_a for small variances σ^2, are designed to refine the estimation of room parameters and have higher probability to occur. Bigger changes are designed to prevent trapping in local minima and have smaller probability to occur. In particular, the mutation Ψ_c was designed to prevent trapping in a frequent type of local minimum.

Step 3 - The step 2 is repeated until the cost function reaches a prescribed threshold γ; the best cost function does not change for Δk generations; or a prescribed number of generations is reached. Let us denote with K the index of the last generation. The parameters $\boldsymbol{\lambda}_m^{(K)} \in \Lambda^{(K)}$ with the smallest value of the cost function provides the estimate of the room geometry, i.e.

$$\hat{\boldsymbol{\lambda}} = \arg \min_{\boldsymbol{\lambda}_m^{(K)} \in \Lambda^{(K)}} f\{h(\boldsymbol{\lambda}_m^{(K)}, t), \bar{h}(t)\}. \tag{10}$$

4. RESULTS

We performed a number of simulations to assess the performance and robustness of the proposed approach. In particular we consider the following scenarios:

- Rectangular room with $\mathbf{d} = [6.7, 9]$ m, $\mathbf{x}_S = [1.3, 1.5]^T$ m, $\mathbf{x}_R = [1.8, 0.634]^T$ m, with known or unknown source position, i.e. $\boldsymbol{\lambda}_1 = \mathbf{d}$ and $\boldsymbol{\lambda}_2 = [\mathbf{d}, \mathbf{x}_S]$;
- L-shaped room with $\mathbf{d} = [8, 6.5, 3.33, 10]$ m, $\mathbf{x}_S = [1.3, 1.5]^T$ m, $\mathbf{x}_R = [1.8, 0.634]^T$ m, with known and unknown source position, i.e. $\boldsymbol{\lambda}_1 = \mathbf{d}$ and $\boldsymbol{\lambda}_2 = [\mathbf{d}, \mathbf{x}_S]$.

In all the simulations the estimation algorithm considers the environment as two-dimensional, i.e. the simulated impulse response $h(\boldsymbol{\lambda}, t)$ is obtained modelling the propagation in a 2D environment. However, in order to match real-world conditions, the measured impulse response $\bar{h}(t)$ is obtained simulating the propagation in a 3D environment with height 3m, microphone/loudspeaker placed at height 1m, and an additive white Gaussian noise with $SNR = 20$dB. The measured impulse response is therefore obtained as

$$\bar{h}(t) = \{(h_{3D}(t) * s(t)) + w(t)\} \star s(t), \tag{11}$$

where $h_{3D}(t)$ indicates the impulse response of the 3D environment; $*$ denotes convolution; $s(t)$ is the source signal (white noise in the audio bandwidth); $w(t)$ is the additive noise; and $\star$ is the cross-correlation. It is important to notice that the estimation algorithm can be easily extended to account for the 3D environments adding a new dimension (height) to the unknown parameters $\boldsymbol{\lambda}$. However, we intentionally limit the estimation to a 2D floormap in order to test the robustness of the algorithm to spurious peaks in the measured impulse response. Reflections in $h(\boldsymbol{\lambda}, t)$ and $\bar{h}(t)$ are simulated up to the 4th order and 10th order, respectively.

We start with the simplest case: rectangular room and unknowns $\boldsymbol{\lambda}_1 = [d_1, d_2]$. The population size is $M = 20$; maximum number of generations $K = 25$; crossover fraction $p = 0.5$; number of rooms migrated unchanged to next generation $N = 4$; tournament size $w = 4$. The search space, determined from a preliminary analysis of $\bar{h}(t)$ is between $\boldsymbol{\lambda}_{\min} = [1, 1]$ m and $\boldsymbol{\lambda}_{\max} = [12, 12]$ m. Average value and standard deviations of $\boldsymbol{\lambda}$ for 50 independent realizations are shown in the first part of Table 1.

In this case we have just two unknowns, i.e. the room dimensions d_1 and d_2. Given the limited size of the search space we could make an exhaustive search, i.e. uniformly sample the space and choose $\boldsymbol{\lambda}$ with best value of $f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}$. Fig. 3 shows the contours of $f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}$ evaluated on a 300×300 grid. We notice that, although the measured impulse response is noisy and

Table 1: Rectangular room: parameters of the room geometry (dimensions and coordinates) and estimated values (50 realizations).

	d_1[m]	d_2[m]	x_S[m]	y_S[m]
$\boldsymbol{\lambda}$	6.7	9	1.3	1.5
$E[\hat{\boldsymbol{\lambda}}_1]$	6.7187	9.0096	–	–
$std(\hat{\boldsymbol{\lambda}}_1)$	0.0535	0.0223	–	–
$E[\hat{\boldsymbol{\lambda}}_2]$	6.7716	9.0345	1.3642	1.5491
$std(\hat{\boldsymbol{\lambda}}_2)$	0.0921	0.0362	0.1381	0.1024

Table 2: L-shaped room: parameters of the room geometry (dimensions and coordinates) and estimated values (50 realizations).

	d_1[m]	d_2[m]	d_3[m]	d_4[m]	x_S[m]	y_S[m]
$\boldsymbol{\lambda}$	8	6.5	3.33	10	1.3	1.5
$E[\hat{\boldsymbol{\lambda}}_1]$	8.0066	6.5687	4.0137	9.6144	–	–
$std(\hat{\boldsymbol{\lambda}}_1)$	0.0131	0.1377	0.7952	0.7202	–	–
$E[\hat{\boldsymbol{\lambda}}_2]$	7.6148	7.1589	4.0894	9.2198	1.2452	1.7511
$std(\hat{\boldsymbol{\lambda}}_2)$	0.6407	0.5719	1.0029	0.7914	0.3541	0.2987

exhibits spurious peaks, the cost function defined in (4) exhibits a minimum in correspondence of true dimensions $\mathbf{d} = [6.7, 9]$ m and it is reasonably smooth. However, as previously stated, the cost function shows also a secondary minimum. In particular the local minimum is at $[9, 6.7]$ m, and it is due to the symmetry of the configuration. This fact justifies the introduction of the mutation (9). As the number of unknowns increases, the function

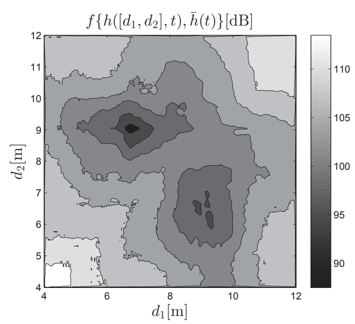


Figure 3: Cost function evaluated for different dimensions of the rectangular room.

$f\{h(\boldsymbol{\lambda}, t), \bar{h}(t)\}$ becomes more complex, discontinuous and with a higher number of local minima. The gradient descent methods could become trapped in a local minimum and an exhaustive search of the parameter space becomes prohibitively expensive. In fact, in case of just two unknowns and feasible values within $10\text{m} \times 10\text{m}$, we should evaluate the cost function 10000 times for an error below 5 cm. Using genetic algorithm just $M \cdot K = 500$ cost function evaluations are needed. The second part of Table 1 refers to the case of $\mathbf{x}_S$ unknown. For this simulation M has been increased to 50. The search space is determined by $\boldsymbol{\lambda}_{\min} = [1, 1, 0.5, 0.5]$ m and $\boldsymbol{\lambda}_{\max} = [12, 12, 2.5, 2.5]$ m. As expected, the accuracy decreases but it is suitable for many applications.

In the next simulations we considered a L-shaped room. We tested both cases of $\mathbf{x}_S$ known and unknown. Parameters of the genetic algorithm are the same as in the previous simulation. Results are shown in Table 2. We notice that the estimate is quite accurate for the case of $\mathbf{x}_S$ known, while it becomes less accurate in the second case. Due to the smoothness of the cost function around the global minimum, however, the geometry provided by the genetic algorithm can be used as an initial estimate of a gradient method to achieve an improved accuracy.

5. CONCLUSIONS

In this work we have proposed a methodology that, provided a priori knowledge on the shape of the environment, estimates its dimensions and the source location from a single impulse response. A cost function is defined for this purpose, which is based on the difference of the TOAs between simulated and measured impulse responses. The dimensions of the room that provide the minimum difference determine the estimated room geometry. Simulations prove the effectiveness of the approach in the case of two common types of environments. Estimation accuracy, achieved with the proposed minimization algorithm, is suitable for most potential applications and it could be further improved if used in combination with a gradient descent method.

6. REFERENCES

[1] F. Ribeiro, C. Zhang, D. Florencio, and D. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1781–1792, Sept. 2010.

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

[3] A. Canclini, D. Markovic, F. Antonacci, A. Sarti, and S. Tubaro, "A room-compensated virtual surround system exploiting early reflections in a reverberant room," in *proc. of European Signal Processing Conference (EUSIPCO)*, 2012, pp. 1–5.

[4] A. Brutti, M. Omologo, and P. Svaizer, "An environment aware ml estimation of acoustic radiation pattern with distributed microphone pairs," *Signal Processing*, vol. 93, no. 4, pp. 784–796, 2013.

[5] F. Antonacci, A. Sarti, and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoutic emissions," in *in proceedings of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, USA, Mar. 2010, pp. 2822–2825.

[6] S. Tervo and T. Korhonen, "Estimation of reflective surfaces from continuous signals," in *2010 IEEE International Conference onAcoustics Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, March 2010, pp. 153–156.

[7] D. Ba, F. Ribeiro, C. Zhang, and D. Florencio, "L1 regularized room modeling with compact microphone arrays," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, Mar. 2010, pp. 157–160.

[8] F. Antonacci, J. Filos, M. Thomas, E. Habets, A. Sarti, P. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 10, pp. 2683 – 2695, 2012.

[9] I. Dokmanic, Y. Lu, and M. Vetterli, "Can one hear the shape of a room: The 2-D polygonal case," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 321–324.

[10] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Transactions on Audio, Speech, and Language Processing,*, vol. 16, no. 4, pp. 812–824, May 2008.

[11] J. Borish and J. B. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," *J. Audio Eng. Soc.*, vol. 31, pp. 478–488, 1983.

[12] R. Duda, P. Hart, and D. Stork, *Pattern classification*, ser. Pattern Classification and Scene Analysis: Pattern Classification. Wiley, 2001.