ROOM IMPULSE RESPONSE ESTIMATION BY ITERATIVE WEIGHTED L_1-NORM

Marco Crocco and Alessio Del Bue

Pattern Analysis and Computer Vision (PAVIS)
Istituto Italiano di Tecnologia (IIT)
Via Morego 30, 16163 Genova, Italy

ABSTRACT

This paper presents a novel method to solve for the challenging problem of acoustic Room Impulse Response estimation (RIR). The approach formulates the RIR estimation as a Blind Channel Identification (BCI) problem and it exploits sparsity and non-negativity priors to reduce ill-posedness and to increase robustness of the solution to noise. This provides an iterative procedure based on a reweighted l_1-norm penalty and a standard l_1-norm constraint. The proposed method guarantees the convexity of the problem at each iteration, it avoids drawbacks related to anchor constraints and it enforces sparsity in a more effective way with respect to standard l_1-norm penalty approaches. Experiments show that our approach outperform current state of the art methods on speech and non-speech real signals.

Index Terms— Room Impulse Response, Blind System Identification, Sparsity, Non-negative Priors, TDOA Estimation

I. INTRODUCTION

The estimation of the Room Impulse Response (RIR) is a problem at the basis of several applications in signal processing and providing remarkable theoretical challenges for the community. The computation of the RIR [1] or the closely related reverberation time [2], is the fundamental step for different applications such as room geometry reconstruction [3], [4], [5], [6], [7], room aware sound reproduction [8], speech enhancement [9], and dereverberation [10]. Since the emitting audio source is often unknown, RIR estimation is formulated as a Blind Channel Identification (BCI) problem with two main procedures. The first relies on the specific source statistics [11] while the second uses the diversity of multiple channels given different spatial locations where the audio signal is acquired [12], [13].

The first approach has a series of requirements, the main one involving long acquisition times to build up the statistic and the presence of stationary sources that often limits its use to custom real environments. The second approach avoids these drawbacks since it relies on the implicit properties of the RIR using, in its simplest form, the cross-relation identity in a Single Input Multi Output (SIMO) system that can be easily formulated as an eigenvalue problem [12]. This formalization of the problem has been extended in several ways in order to provide more robustness and to reduce the a priori information required for the solution [14], [15], [16].

In this paper we propose a novel formulation that aims at further improving the accuracy and robustness of RIR estimation by devising a novel iterative optimization strategy. The key features of our method are the elimination of strong constraints related to previous approaches (e.g. the anchor constraint [14], [15]) and the ability to promote sparsity and non-negativity to increase robustness and accuracy. Moreover, each iteration of the proposed method results in a convex problem easily solvable with quadratic programming techniques. The experimental tests based on synthetic and real acoustic signals (both speech and non-speech) compare favorably to our approach, achieving superior performance in respect to the state of the art.

The next Sec. II presents the formalisation of RIR estimation as an optimization problem together with the details of previous approaches. Sec. III proposes our method formalisation and optimization strategy. Experimental results in Sec. IV shows the performance of several methods compared to ours while Sec. V concludes with further considerations.

II. PROBLEM STATEMENT

Let us consider M microphones in a room and let us define $h_i(k)$ as the discrete time RIR from a single source to the i-th microphone. The signal $x_i(k)$ received at microphone i can be written as the discrete convolution between the transmitted signal $s(k)$ and the i-th RIR such that:

$$x_i(k) = \sum_{l=1}^{L} h_i(l)s(k-l), \qquad i = 1, \cdots, M \tag{1}$$

where L is the channel length. The BCI problem aims at recovering $h_i(l)$ for every i and l without knowing the transmitted signal $s(k)$. A family of methods is based on the cross-relation identity for which, in absence of noise, $x_i(k) * h_j(k) = x_j(k) * h_i(k)$ for every couple of i, j where $*$ denotes convolution. This principle is used in [12] by introducing a least squares minimization of the squared cross relation error as:

$$\min_{\mathbf{h}} J(\mathbf{h}) \quad s.t. \quad \|\mathbf{h}\|_2^2 = 1, \tag{2}$$

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

with

$$J(\mathbf{h}) = \sum_{i \neq j} \|\mathtt{X}_i \mathbf{h}_j - \mathtt{X}_j \mathbf{h}_i\|_2^2 \tag{3}$$

where $\mathbf{h}_i = [h_i(1), \ldots, h_i(L)]^\top$, $\mathtt{X}_i$ is the Toeplitz matrix whose first row and column are given by $[x_i(k-N+1), x_i(k-N), \ldots, x_i(k-N-L+2)]$ and $[x_i(k-N+1), x_i(k-N+2), \ldots, x_i(k), 0, \ldots, 0]^\top$ respectively, with N being the signal length, and $\mathbf{h} = [\mathbf{h}_1^\top, \ldots \mathbf{h}_M^\top]^\top$. The unitary constraint on the l_2-norm of $\mathbf{h}$ is needed to avoid the trivial zero solution. Furthermore, by rearranging the matrices $\mathtt{X}_i$, it can be shown that $J(\mathbf{h})$ is a quadratic form of $\mathbf{h}$, i.e. $J(\mathbf{h}) = \mathbf{h}^\top \mathtt{Q} \mathbf{h}$, where $\mathtt{Q}$ is an $ML \times ML$ positive semi-definite matrix. Thus, the solution to the above minimization problem is given by the singular vector corresponding to the smallest singular value of $\mathtt{Q}$ [12]. Unfortunately there are underlying assumptions to this formalisation that are not easily verified in a real situations. In particular, the channels have to be co-prime; hence their length L has to be known in advance. This information is not available in real situations and if the channel length is wrongly estimated, the problem is ill-conditioned and therefore highly sensitive to environmental noise.

This method has been subsequently improved exploiting the a priori knowledge of the RIR shape. To this end, the RIR can be modelled as a sequence of pulses, each one corresponding to the direct path and/or a reflection from a wall with positive amplitude coefficients [17]. Hence, sparsity [14], [16] and non-negativity [15] have been imposed to improve robustness. Even if the sparsity assumption may not hold for the "tail" of the RIR, applications concerning environment geometry [3], [4], [5], [6], [7], [8] require just the recovery of lower order reflections, i.e. the sparse portion of the RIR. Likewise, applications concerning speech enhancement [9] and dereverebration [10] have proven to work under the sparse RIR assumption.

In [16] a sparsity inducing l_1-norm penalty was added to the quadratic cost function $J(\mathbf{h})$, yielding the following minimization problem under the l_2-norm equality constraint:

$$\min_{\mathbf{h}} \; J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_2^2 = 1. \tag{4}$$

In general, it is well known that l_1 penalty is used in a wide range of problems as a surrogate of the more suited but computationally intractable l_0 penalty. Even if, under opportune hypotheses, the two norms yields equivalent solutions in terms of sparsity, this is not true in general: the main reason is that larger coefficients are penalized more heavily than smaller coefficients, whereas l_0 norm penalizes all non-zero coefficients in the same manner [18].

Moreover, the problem in Eq. (4) is affected by two fundamental drawbacks. First, the domain of the problem is non-convex due to the quadratic equality constraint, so making the minimization of $J(\mathbf{h})$ prone to local solutions. Second, if the signal spectrum contains "holes", i.e. in the likely case where we have no white noise signals, we might have a weighting of the power spectrum by zero or very small values. This is a subtle but disruptive effect, since the additional energy constraint given by the l_2-norm may overstress frequency components related to holes in the spectrum since these will contribute little in the cost function.

To address these issues, a single anchor constraint can be used to substitute the l_2-norm one [14] giving:

$$\min_{\mathbf{h}} \; J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad |h_1(a)| = 1, \tag{5}$$

where a is the anchor index that has to be chosen greater than the maximum of the differences $\tilde{k}_i - \tilde{k}_1$ over $i = 1, \cdots, M$ where $\tilde{k}_i$ is the index of the first non zero entry of the channel i. The single anchor constraint makes the problem convex [19] and less sensitive to noise in respect to the l_2-norm constraint as in [14]. However, the anchor constraint together with the l_1-norm penalty leads generally to an amplitude distortion in the first channel reconstruction, by placing an overly enhanced peak at sample a.

A further constraint has been exploited in [15] where the non-negative properties of the RIR [17] are included in the cost function leading to:

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad h_1(a) = 1, \quad \mathbf{h} \geq 0 \tag{6}$$

where $\mathbf{h} > 0$ means $h(l) > 0$ for $l = 1, \ldots, L$. Non-negative constraints provide increased robustness to noise by further regularizing the problem, as explained in [20] and [21].

III. PROPOSED METHOD

To solve the drawbacks related to the anchor constraint in Eqs. (5) and (6), we introduce a l_1-norm equality constraint. In this way, all the channel elements are equally taken into account without privileging the one corresponding to the anchor:

$$\min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_1 = 1, \quad \mathbf{h} \geq 0. \tag{7}$$

At the same time, differently from the l_2-norm constraint, the problem remains convex. Unfortunately, setting such l_1-norm equality has the undesired side effect of hampering the sparsity inducing property of the l_1 penalty: if the same l_1 term appears both as a penalty and as a constraint, the penalty becomes a constant that does not influence the argument of the minimum of the cost function. As a consequence, no term remains to promote sparsity since the l_1 constraint is not able *per se* to impose it. This effect can be intuitively seen, looking at the two-dimensional toy problem depicted in Fig. 1 where $\mathbf{h} \in \Re^2$. Here, in absence of measurement noise, let the exact sparse solution be at $h(1) = 0, \; h(2) = 1$ for which $J(\mathbf{h}) = \mathbf{h}^\top \mathtt{Q} \mathbf{h} = 0$, under the l_1 constraint $\|\mathbf{h}\|_1 = 1$. However, due to noisy measurements, the matrix $\mathtt{Q}$ becomes full rank and the corresponding family of ellipses are non-degenerate and with axes not aligned to the canonical basis of $\mathbf{h}$. As a consequence, the

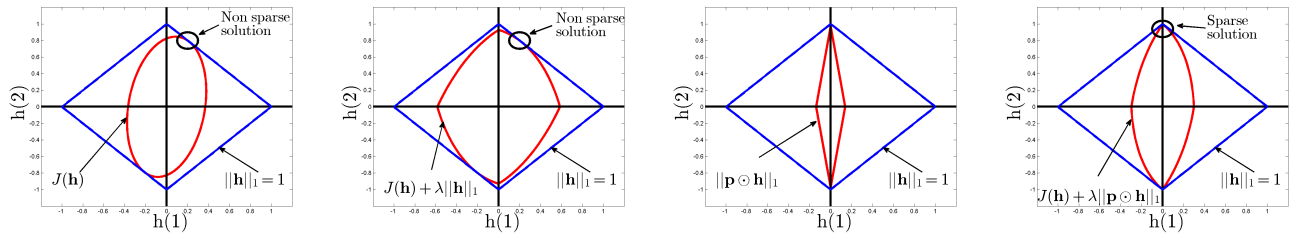

Fig. 1. Toy example: the blue line represents the l_1 equality constraint, while the red line represents respectively: (a) quadratic cost function, (b) quadratic cost function plus l_1 penalty, (c) weighted l_1 penalty, (d) quadratic cost function plus weighted l_1 penalty.

solution to the problem $\min_{\mathbf{h}} J(\mathbf{h}) \;\; s.t. \;\; \|\mathbf{h}\|_1 = 1, \;\; \mathbf{h} \geq 0$ without penalties, is non sparse, as can be seen in panel Fig. 1(a). The inclusion of a l_1 penalty modifies the shape of the cost function, but does not improve the sparsity of the solution (Fig. 1(b)).

To counter this problem, let us suppose to know in advance that the exact solution has $h(2) > h(1)$. We can compensate for the influence of the magnitude of $\mathbf{h}$ elements on the l_1 penalty, introducing a weighted l_1 penalty, $\|\mathbf{p} \odot \mathbf{h}\|_1$, where the two elements $p(i)$ of the weight vector $\mathbf{p}$ are such that $p(2) < p(1)$ and $\odot$ denotes the Hadamard product. The resulting weighted penalty is displayed in Fig. 1(c). By adding such weighted l_1 penalty to the quadratic term, we obtain a cost function pinched toward the $h(2)$ component and so a sparse solution equal to the exact one (Fig. 1(d)). From this intuition, we propose to solve a sequence of $z = 1, \cdots, Z$ minimization problems of the form:

$$\hat{\mathbf{h}}^{(z)} = \min_{\mathbf{h}} J(\mathbf{h}) + \lambda \|\mathbf{p}^{(z)} \odot \mathbf{h}\|_1 \quad s.t. \quad \|\mathbf{h}\|_1 = 1, \quad \mathbf{h} \geq 0, \tag{8}$$

where $\mathbf{p}^{(z)}$ is a $L \times 1$ weight vector. Notice that each problem is convex and solvable with quadratic programming techniques [19]. At each iteration, the weight vector $\mathbf{p}^{(z)}$ is updated as the inverse of the solution of the problem at the previous step $(z-1)$ giving:

$$p(l)^{(z)} = \frac{1}{\hat{h}(l)^{(z-1)} + \epsilon} \quad for \;\; l = 1, \cdots, L \tag{9}$$

where $p(l)^z$ and $\hat{h}(l)^{(z-1)}$ denote the l-th element of $\mathbf{p}^{(z)}$ and $\hat{\mathbf{h}}^{(z-1)}$ respectively and ϵ is a regularization parameter useful to avoid numerical instability, typically set at some orders of magnitude lower than the expected values of non-zero elements of $\hat{\mathbf{h}}^{(z-1)}$. Concerning the algorithm starting guess $\hat{\mathbf{h}}^{(0)}$, we adopt the solution provided by Eq. (6), i.e. the standard l_1 penalty with anchor and non-negativity constraints. Even if such initialization does not assure a very sparse solution, still the elements of $\hat{\mathbf{h}}^{(0)}$ corresponding to zero entries in the exact solution will be lower than the other elements. Hence, in the subsequent steps, such elements will be more penalized by the weighted l_1, yielding increasingly sparser solutions, as it can be seen by looking at the three iterations in the example displayed in Fig. 2. Also note that the real channel is not perfectly non-negative since the Dirac pulses, corresponding to the walls reflections, typically fall off the grid of samples resulting in sinc functions. Despite this slight mismatch between theoretical assumptions and real data, the position of the estimated peaks reproduces the positions of the ground truth peaks with remarkable precision[1]. Moreover the energy of each reflection is fairly preserved by the proposed method by enhancing the peaks corresponding to off-the-grid reflections, as can be observed comparing the first estimated peak (on the grid reflection) with the second, third and fourth (off the grid reflections).

When the solution converges, i.e. $\hat{\mathbf{h}}^{(z)} \approx \hat{\mathbf{h}}^{(z-1)}$, the weighted l_1 norm becomes equivalent to the l_0 norm, as $p(l)^{(z)}\hat{h}(l)^{(z)} \approx 1$ for non-zero entries of $\hat{\mathbf{h}}^{(z)}$ and

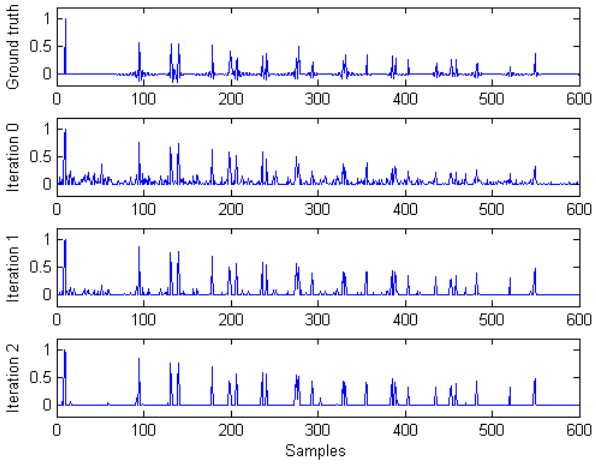

Fig. 2. Example of the iterative process for our proposed approach with SNR = 9 dB.

[1]A further mismatch is given by the finite wall dimension that makes the image method of [17] not accurate at low frequencies. However, this effect can be reduced by using a high-pass filtering of the signals at the microphones.

$p(l)^{(z)}\hat{h}(l)^{(z)} \approx 0$ for zero entries. This fact contributes to explain the superiority of weighted l_1 in inducing sparsity, with respect to its unweighted counterpart.

IV. EXPERIMENTS

The proposed algorithm, named Iterative L1 Penalty (IL1P), was evaluated against the eigenvalue decomposition (EIG) [12] and the non-negative l_1-norm method [15] (L1NN) through a set of 50 Monte Carlo simulations with realistic signals. The experimental setup is given by a rectangular room of $6 \times 5 \times 4$ m with microphones and sources position generated at each trial according to a uniform distribution. To avoid configurations in which the source is too close to the microphones, the x coordinate ranged from 0 m to 2 m for the microphones and from 4 m to 6 m for the source. The number of microphones was fixed to 2 and 4. The RIRs were simulated according to the image method [17] assuming a reflection coefficient from the walls of 0.8. Simulations were repeated setting the Signal to Noise Ratio (SNR) at the microphones to 40 dB, 20 dB, 14 dB 6 dB and 0 dB. All the parameters (λ, a and ϵ) were optimized through cross-validation and the following values were set: $\lambda = 0.001$, $a = 150$, $\epsilon = 10^{-9}$. For IL1P, the number of iterations was set to 3, including the initialization step. In Fig. 4 an example of the real and estimated channel for the four methods is displayed, for an SNR of 9 dB. With such noise, EIG completely fails in reconstructing the signal due to its extreme noise sensitivity. L1NN provides a reasonable solution but with very noisy reconstructions showing several spurious peaks and amplitude distortions due to the anchor constraint. The best reconstruction is achieved with IL1P: the spurious peaks have disappeared and almost all the true peaks are correctly matched both in position and energy. The quantitative summary of the results over the 50 Monte Carlo trials is reported in Table 1 for synthetic, non-speech and speech sources. In each box the first term is the Average Peak Position Mismatch (A_{PPM}) in samples and the second term, in square brackets, is the Average Percentage of Unmatched Peaks (A_{PUP}). These values are both calculated on the peaks corresponding to direct path and the six 1-st order reflections from the walls. A ground truth peak has been considered unmatched if the closest estimated peak is more than 20 samples away from it. In detail, the A_{PPM} and the A_{PUP} are calculated as it follows:

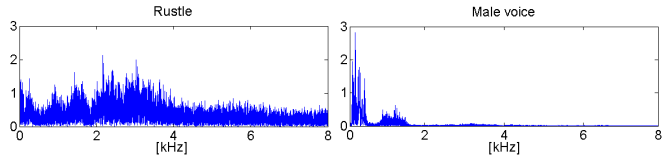

Fig. 3. Spectra of non-voice (left) and voice (right) recorded signal used in the experiments.

$$A_{PPM} = \sum_{i=1}^{50} \sum_{p=1}^{P_i} \frac{|\tau_{p,i}^{gt} - \tau_{p,i}^{e}|}{50P_i} \quad A_{PUP} = \sum_{i=1}^{50} \frac{K - P_i}{50K} \tag{10}$$

where $\tau_{p,i}^{gt}$ and $\tau_{p,i}^{gt}$ are respectively the p-th ground truth peak location and its corresponding estimation in samples for the $i-th$ Monte Carlo trial, P_i is the number of ground truth peaks for which a matching has been found among the estimated ones and K is the number of ground truth peaks. Such metrics allow to decouple the effect of the outliers, quantified by A_{PUP}, from the overall peak position accuracy expressed by A_{PPM}. The matching between ground truth peaks and estimated ones is performed, for each channel, using a Nearest Neighbour procedure repeated for each ground truth peak. EIG definitely fails in recovering the true peak locations, as can be seen looking at the high peak mismatching and the high ratio of outliers (Table 1, first column), substantially independent of the SNR. Differently, L1NN manages to obtain reasonable results for all the tested conditions. Evident improvements both in term of A_{PUP} and A_{PPM} are achieved by the proposed algorithm for a great part of SNRs tested. The improvement is particularly evident for SNRs equal or lower than 20 dB. In just one case, (0 dB, speech signal) L1NN moderately overcomes IL1P in terms of A_{PPM}; however this is compensated by the lower number of outliers (A_{PUP}) given by IL1P.

Similar tests have been run using 4 microphones, again confirming the overall superiority of IL1P: as an example, at SNR = 6 dB and with a synthetic signal, the performance is 0.47 [0.03] for our IL1P and 0.90 [0.09] for L1NN. Results in general get worse, moving from synthetic to non-speech and finally to speech, likely due to the progressively decreasing flatness of the source spectrum (check Fig. 3) that limits the amount of frequencies available. Nevertheless the very good performance of the proposed method witnesses its applicability in noisy environments.

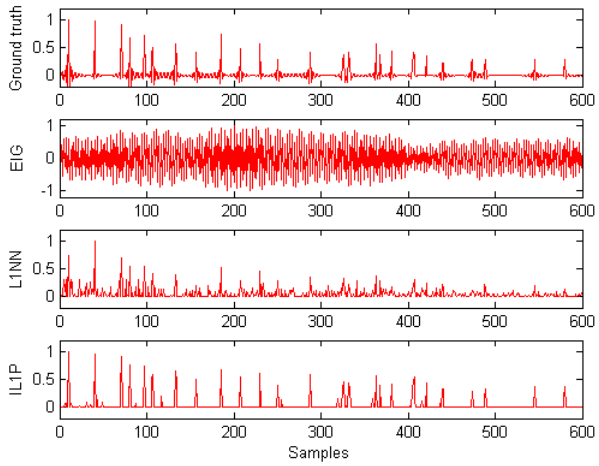

Fig. 4. Results for the four tested methods and ground truth. SNR = 9 dB. Synthetic source.

	SNR	EIG [12]	L1NN [15]	IL1P
SYNTH	0 dB	4.51 [0.34]	1.96 [0.16]	**1.17 [0.11]**
	6 dB	4.24 [0.34]	0.77 [0.10]	**0.40 [0.04]**
	14 dB	3.86 [0.31]	0.34 [0.03]	**0.28 [0.01]**
	20 dB	3.42 [0.30]	0.28 [0.03]	**0.27 [0.01]**
	40 dB	3.20 [0.28]	**0.26 [0.00]**	0.27 [**0.00**]
REAL	0 dB	4.53 [0.39]	2.46 [0.18]	**2.23 [0.14]**
	6 dB	4.50 [0.37]	1.29 [0.12]	**0.54 [0.02]**
	14 dB	3.82 [0.38]	0.39 [0.02]	**0.28 [0.00]**
	20 dB	3.65 [0.35]	0.29 [0.01]	**0.28 [0.00]**
	40 dB	3.31 [0.29]	**0.28 [0.00]**	**0.28 [0.00]**
SPEECH	0 dB	4.66 [0.38]	**2.87** [0.23]	3.31 [**0.16**]
	6 dB	4.91 [0.37]	1.96 [0.19]	**1.36 [0.08]**
	14 dB	4.44 [0.40]	0.98 [0.10]	**0.58 [0.01]**
	20 dB	4.30 [0.35]	0.50 [0.04]	**0.39 [0.01]**
	40 dB	3.64 [0.35]	0.32 [0.01]	**0.29 [0.00]**

Table I. Monte Carlo simulations for different sources.

V. CONCLUSIONS

The proposed method has proven to increase the accuracy of RIR estimation in respect to state of the art methods, in challenging realistic conditions. Future work will test the benefits of the method when applied to applications such as room reconstruction [7] and microphone localisation [22].

VI. REFERENCES

[1] A. Benichoux, L. Simon, E. Vincent, and R. Gribonval, "Convex regularizations for the simultaneous recording of room impulse responses," *Sig. Proc., IEEE Trans. on*, vol. 62, no. 8, pp. 1976–1986, April 2014.

[2] B. Dumortier and E. Vincent, "Blind rt60 estimation robust across room sizes and source distances," in *Acoust., Speech and Sig. Proc. (ICASSP), 2014 IEEE Int. Conf. on*, May 2014, pp. 5187–5191.

[3] I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proceedings of the National Academy of Sciences*, vol. 110, no. 30, pp. 12 186–12 191, 2013.

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

[5] F. Ribeiro, D. Florêncio, D. Ba, and C. Zhang, "Geometrically constrained room modeling with compact microphone arrays," *Audio, Speech, and Lang. Proc., IEEE Trans. on*, vol. 20, no. 5, pp. 1449–1460, 2012.

[6] A. H. Moore, M. Brookes, and P. A. Naylor, "Room geometry estimation from a single channel acoustic impulse response," in *21st European Sig. Proc. Conf. (EUSIPCO), Marrakech, Morocco*, 2013, pp. 1–5.

[7] M. Crocco, , A. Trucco, V. Murino, and A. Del Bue, "Towards fully uncalibrated room reconstruction with sound," in *22nd European Sig. Proc. Conf. (EUSIPCO), Lisbon, Portugal*, 2014.

[8] T. Betlehem and T. D. Abhayapala, "A modal approach to soundfield reproduction in reverberant rooms," in *2005 IEEE Int. Conf. on Acoust., Speech, and Sig. Proc., ICASSP, Philadelphia, USA*, 2005, pp. 289–292.

[9] M. Yu, W. Ma, J. Xin, and S. Osher, "Multi-channel l_1 regularized convex speech enhancement model and fast computation by the split bregman method," *Audio, Speech, and Lang. Proc., IEEE Trans. on*, vol. 20, no. 2, pp. 661–675, Feb 2012.

[10] Y. Lin, J. Chen, Y. Kim, and D. D. Lee, "Blind channel identification for speech dereverberation using l1-norm sparse learning," in *Advances in Neural Information Processing Systems*, 2007, pp. 921–928.

[11] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. Wiley & Sons, 2004, vol. 46.

[12] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *Information Theory, IEEE Transactions on*, vol. 40, no. 2, pp. 340–349, Mar 1994.

[13] Y. A. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *Signal Processing, IEEE Transactions on*, vol. 51, no. 1, pp. 11–24, 2003.

[14] Y. Lin, J. Chen, Y. Kim, and D. D. Lee, "Blind channel identification for speech dereverberation using l1-norm sparse learning," in *Advances in Neural Information Processing Systems*, 2007, pp. 921–928.

[15] Y. Lin, J. Chen, Y. Kim, and D. Lee, "Blind sparse-nonnegative (bsn) channel identification for acoustic time-difference-of-arrival estimation," in *Applications of Sig. Proc. to Audio and Acoust., 2007 IEEE Workshop on*, Oct 2007, pp. 106–109.

[16] K. Kowalczyk, E. Habets, W. Kellermann, and P. Naylor, "Blind system identification using sparse learning for tdoa estimation of room reflections," *Sig. Proc. Letters, IEEE*, vol. 20, no. 7, pp. 653–656, July 2013.

[17] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *The Journal of the Acoust. Soc. of America*, vol. 65, no. 4, pp. 943–950, 1979.

[18] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, USA: Cambridge Univ. Press, 2004.

[20] L. Benvenuti and L. Farina, "A tutorial on the positive realization problem," *Automatic Control, IEEE Transactions on*, vol. 49, no. 5, pp. 651–664, 2004.

[21] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[22] M. Crocco, A. Del Bue, and V. Murino, "A bilinear approach to the position self-calibration of multiple sensors," *Sig. Proc., IEEE Trans. on*, vol. 60, no. 2, pp. 660–673, Feb 2012.