EUSIPCO 2013 1569740629

ROOM GEOMETRY ESTIMATION FROM A SINGLE CHANNEL ACOUSTIC IMPULSE RESPONSE

Alastair H. Moore, Mike Brookes and Patrick A. Naylor

Centre for Law Enforcement Audio Research
Imperial College London, Department of Electrical and Electronic Engineering

ABSTRACT

For a 2D rectangular room of unknown dimensions and with unknown source and microphone positions, the times of arrival of reflections can be described in terms of image source positions. Adopting a microphone-centred coordinates system, it is shown that to satisfy certain combinations of arrival times imposes constraints on the possible room geometry: a second-order reflection from adjacent walls determines the source-microphone distance; a second-order reflection from opposite walls in a given dimension determines the source displacement in that dimension as a function of the source-receiver distance. Given a subset of time differences of arrival, the extent to which the geometry can be determined is related to these constraints. The geometry estimation is further posed as a least squares optimisation problem whose results verify the analytical results.

Index Terms— geometry estimation, acoustic impulse response, time of arrival, TOA, room identification

1. INTRODUCTION

Knowledge of an acoustic environment's geometry can benefit signal processing problems such as dereverberation [1][2] and room identification [3][4]. Information about the position of a source and microphone relative to surrounding walls is contained in the acoustic impulse response in the form of the time of arrival (TOA) of reflected wavefronts. The room geometry can be determined exactly if multiple impulse responses are obtained, either from an array of transducers [5][6] or from sequential measurements with different source/microphone positions [7]. With only a single channel impulse response, the geometry is ambiguous due to reciprocity and symmetries. Nevertheless, the single channel case is still of interest in applications such as audio forensics where the available data may be inherently limited [8].

The geometry of a convex polygon bounded plane can be estimated using TOAs from a single channel impulse response in the special case where the source and microphone are co-located and a complete set of first- and second-order reflections are available [9]. Here we consider the more general case of arbitrarily located source and microphone but under a more restrictive room model. It does not assume that an absolute time reference is available and it investigates the extent to which the geometry can be inferred if some reflections are missing.

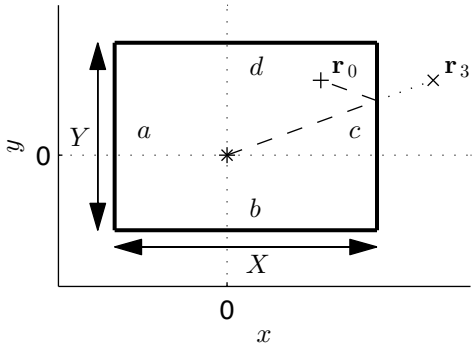

Fig. 1. Problem geometry showing source, $\mathbf{r}_0$ $(+)$, and receiver $(*)$ in a plane bounded by four walls. $\mathbf{r}_3$ $(\times)$ is the image of $\mathbf{r}_0$ in wall $w = c$.

In Section 2 we define the geometry model and the problem we address. In Section 3 we derive expressions for the constraints imposed on the solution by specific combinations of reflections and formulate an error function which in the general case can be minimised numerically to find an estimate of the geometry. In Section 4 we present illustrative examples before drawing conclusions in Section 5.

2. PROBLEM DESCRIPTION

2.1. Geometry

The geometry we wish to estimate is shown in Fig. 1 and consists of an acoustic source and a microphone lying in a plane which is bounded by four ideal reflectors (a.k.a. walls) arranged in a rectangle. The origin is placed at the microphone and the source is at $\mathbf{r}_0 = [x_0\, y_0]^T$. The reflectors are labelled sequentially in an anti-clockwise direction, $w = \{a, b, c, d\}$, and the shortest distance from the origin to reflector w is l_w. Thus, the dimensions of the rectangle are $X \times Y = (l_a + l_c) \times (l_b + l_d)$. The room geometry is there-

i	s_i	x_i	y_i
0	[]	x_0	y_0
1	$[a]$	$-(2l_a + x_0)$	y_0
2	$[b]$	x_0	$-(2l_b + y_0)$
3	$[c]$	$2l_c - x_0$	y_0
4	$[d]$	x_0	$2l_d - y_0$
5	$[a\ b]$	$-(2l_a + x_0)$	$-(2l_b + y_0)$
6	$[a\ c]$	$2l_a + 2l_c + x_0$	y_0
7	$[a\ d]$	$-(2l_a + x_0)$	$2l_d - y_0$
8	$[b\ c]$	$2l_c - x_0$	$-(2l_b + y_0)$
9	$[b\ d]$	x_0	$2l_b + 2l_d + y_0$
10	$[c\ a]$	$-(2l_a + 2l_c - x_0)$	y_0
11	$[c\ d]$	$2l_c - x_0$	$2l_d - y_0$
12	$[d\ b]$	x_0	$-(2l_b + 2l_d - y_0)$

Table 1. For each image source, i, of up to second order, the table lists the sequence of reflecting walls, s_i, and the image source position $[x_i\ y_i]^T$.

fore defined by

$$\mathbf{\Omega} = [x_0,\ y_0,\ l_a,\ l_b,\ l_c,\ l_d]. \tag{1}$$

The acoustic impulse response $h(t)$ between the source and microphone is given by

$$h(t) = \sum_{i=0}^{I-1} \alpha_i \delta(t - \tau_i) \tag{2}$$

and is the sum of I reflected copies of the input signal, $\delta(t)$, where each reflection is characterised by its amplitude, α_i, and time of arrival (TOA), τ_i. Using this notation, $i = 0$ corresponds to the direct path from the source to the microphone.

Each reflection can be considered as arising from a so-called image source [10] lying outside the rectangle. Thus, in Fig. 1, the image source at $\mathbf{r}_3$ has the same path length to the microphone as the path from $\mathbf{r}_0$ reflected in wall c. In general the image in wall w of a source at $\mathbf{r}$ is

$$\mathbf{r}' = \mathbf{r} + 2\left\langle \mathbf{p}_w - \mathbf{r}.\mathbf{n}_w \right\rangle \mathbf{n}_w \tag{3}$$

where $\mathbf{p}_w$ is any point on wall w and $\mathbf{n}_w$ is the outward facing unit normal of wall w. Image sources may be calculated recursively; the images of $\mathbf{r}_0$ are first-order image sources and the images of first-order image sources are second-order image sources. Table 1 defines a vector s_i associated with each image source $\mathbf{r}_i$ which specifies the sequence of reflecting walls. The reflections associated with $\mathbf{r}_6$ and $\mathbf{r}_{10}$ both include walls a and c but the sequence is reversed. Since a and c are opposite in the x direction they are referred to as opposite-x reflections. Similarly the reflections for $\mathbf{r}_9$ and $\mathbf{r}_{12}$ are called opposite-y reflections. Reflections off two adjacent walls are called adjacent reflections; reversing the sequence of an adjacent reflection does not affect the TOA and for a given geometry, $\mathbf{\Omega}$, only one of the two sequences is possible. Accordingly each adjacent wall pair is listed only once in the table.

For an image at $\mathbf{r}_i = [x_i\, y_i]^T$ the path length R_i to the microphone is

$$R_i = \|\mathbf{r}_i\| = \sqrt{x_i^2 + y_i^2} \tag{4}$$

and the TOA is

$$\tau_i = \frac{R_i}{c} \tag{5}$$

where c is the speed of sound. Table 1 lists expressions for x_i and y_i.

We assume that our measurement system lacks an absolute time reference and so a measurement of the acoustic impulse response yields $h(t - \kappa)$ where κ is an unknown constant. Estimating the reflection TOAs from $h(t - \kappa)$ also includes this offset

$$\gamma_i = \tau_i + \kappa. \tag{6}$$

Taking the difference between two TOAs gives the time difference of arrival (TDOA) which is independent of κ

$$\gamma_{i,j} = \gamma_i - \gamma_j \tag{7}$$

and using the substitution $-\kappa = -\gamma_0 + R_0/c$ allows an expression for each of the image sources of the form

$$\left(\gamma_{i,0} + \frac{R_0}{c}\right)^2 = \frac{|x_i|^2 + |y_i|^2}{c^2}. \tag{8}$$

2.2. Problem Statement

We assume that we are able to measure $h(t - \kappa)$, the acoustic impulse response from the source to the microphone, subject to an unknown time offset, κ. We also assume that from the measured impulse response we are able to identify the times of arrival of a subset of the reflections whose image sources are listed in Table 1.

In practical measurements it is unlikely that all 13 reflections up to second-order will be discriminable. We assume that the direct sound is always observable and if a second order reflection is observable both the corresponding first order reflections are also observable. We also assume that TOAs are labelled (i.e. we know which entry in Table 1 each corresponds to). In practice a search of feasible label assignments will be required.

The remainder of this paper is concerned with the extent to which $\mathbf{\Omega}$ can be inferred given particular sets of estimated times of arrival γ_i. Even when $\mathbf{\Omega}$ cannot be found it is sometimes possible to determine the source-microphone distance, R_0, and one or both of the room dimensions, X and Y. In Section 3.1 we show that given an adjacent constraint R_0 can be calculated directly. In Section 3.2 we derive an implicit

equation for R_0 and x_0 from an opposite constraint. In Section 3.3 it is demonstrated graphically that by satisfying multiple constraints x_0 and y_0 can be specified uniquely. Finally, in Section 3.4, an error metric is formulated which allows $\mathbf{\Omega}$ to be estimated using numerical methods.

3. CONSTRAINTS ON SOLUTIONS IMPOSED BY SPECIFIC REFLECTIONS

In this section we examine the constraints imposed on $\mathbf{\Omega}$ by various combinations of reflections.

3.1. Adjacent reflections

Consider a path involving adjacent walls, e.g. $s_5 = [a\ b]$. Expressing (8) using the values from Table 1 for $i = \{1, 2, 5\}$ and $j = 0$ gives three equations

$$(\gamma_{1,0} + R_0/c)^2 = \left((2l_a + x_0)^2 + y_0^2\right)/c^2 \quad (9)$$

$$(\gamma_{2,0} + R_0/c)^2 = \left(x_0^2 + (2l_b + y_0)^2\right)/c^2 \quad (10)$$

$$(\gamma_{5,0} + R_0/c)^2 = \left((2l_a + x_0)^2 + (2l_b + y_0)^2\right)/c^2. \quad (11)$$

Multiplying out the right hand side of (9), (10) and (11), making the substitution $R_0^2 = x_0^2 + y_0^2$ and simplifying gives

$$\gamma_{1,0}^2 c^2 + 2\gamma_{1,0} c R_0 = 4l_a^2 + 4l_a x_0 \quad (12)$$

$$\gamma_{2,0}^2 c^2 + 2\gamma_{2,0} c R_0 = 4l_b^2 + 4l_b y_0 \quad (13)$$

$$\gamma_{5,0}^2 c^2 + 2\gamma_{5,0} c R_0 = 4l_a^2 + 4l_a x_0 + 4l_b^2 + 4l_b y_0. \quad (14)$$

Substituting (12) and (13) into (14) and rearranging leads to

$$R_0 = \frac{1}{2c} \frac{\gamma_{1,0}^2 + \gamma_{2,0}^2 - \gamma_{5,0}^2}{(\gamma_{5,0} - \gamma_{1,0} - \gamma_{2,0})}. \quad (15)$$

Thus, knowledge of any adjacent reflection's TOA together with the corresponding first-order reflections' TOAs determines R_0, the distance between source and microphone via an expression analogous to (15). We refer to this as an adjacent constraint.

3.2. Opposite reflections

Given an opposite reflection in the x-dimension (e.g. $s_6 = [a\ c]$), the TDOA constraints which must be satisfied with $i = \{1, 3, 6\}$ and $j = 0$ are

$$(\gamma_{1,0} + R_0/c)^2 = ((2l_a + x_0)^2 + y_0^2)/c^2 \quad (16)$$

$$(\gamma_{3,0} + R_0/c)^2 = ((2l_c - x_0)^2 + y_0^2)/c^2 \quad (17)$$

$$(\gamma_{6,0} + R_0/c)^2 = ((2l_a + 2l_c + x_0)^2 + y_0^2)/c^2. \quad (18)$$

Multiplying out the squares on the left of (16) and (17) and rearranging with the substitution $R_0^2 - y_0^2 = x_0^2$ gives

$$2l_a = \sqrt{\gamma_{1,0}^2 c^2 + 2\gamma_{1,0} c R_0 + x_0^2} - x_0 \quad (19)$$

$$2l_c = \sqrt{\gamma_{3,0}^2 c^2 + 2\gamma_{3,0} c R_0 + x_0^2} + x_0 \quad (20)$$

which after substitution back into (18) leads to

$$0 = \left(\sqrt{\gamma_{1,0}^2 c^2 + 2\gamma_{1,0} c R_0 + x_0^2} + \sqrt{\gamma_{3,0}^2 c^2 + 2\gamma_{3,0} c R_0 + x_0^2} + x_0\right)^2 - \left(\gamma_{6,0}^2 c^2 + 2\gamma_{6,0} c R_0 + x_0^2\right). \quad (21)$$

3.3. Illustrative example

The formulae presented in the preceding sections define constraints which the TDOAs of particular sets of image sources impose on the geometry. To view these constraints graphically an example set of TOAs were calculated for the geometry

$$\mathbf{\Omega} = [0.5,\ 0.5,\ 1.1,\ 2.1,\ 1.9,\ 2.9] \quad (22)$$

with $\kappa = 4$ and units chosen such that $c = 1$.

The plot area of Fig. 2 represents the potential solution space for the source position, $[x_0\ y_0]^T$ in which the constraints from (15) and (21) are labelled with the index, i, of the corresponding second order image source in Table 1. From (15) an adjacent reflection restricts the source position to a circle around the origin. A curve showing possible positions this allows is labelled 5. The opposite-x image source $i = 6$ led to (21) which also defines a locus of allowable positions, shown as curve 6. The other opposite image sources ($i = 9, 10, 12$)

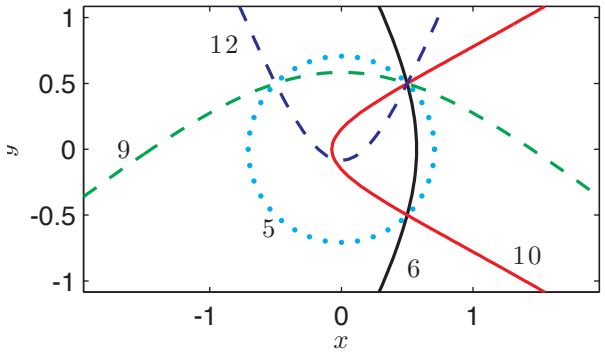

Fig. 2. Loci of source positions which satisfy individual constraints. Dotted line: adjacent constraint, solid lines: opposite-x constraints, dashed lines: opposite-y constraints. Labels 5, 6, 10, 9 and 12 indicate the row of Table 1 which lists the image source associated with that constraint.

each impose similar constraints that are labelled as 9, 10 and 12, respectively.

The intersection of two or more curves in Fig. 2 represents a source position which satisfies the intersecting constraints. The only point at which all the curves intersect is (0.5, 0.5) which is the true source position. Once the source position is known it is trivial to calculate the distance to each wall using null- and first-order image source TOAs, if they are available (see, for example, (19)).

When reflections are missing there are fewer constraints so alternative source positions are possible. For example, without the curves associated with the two opposite-x constraints (6 and 10), the remaining three loci intersect at (-0.5,0.5) indicating an ambiguity in the sign of x_0. As such it is not possible to calculate l_a or l_c from the first-order image source TOAs. However by summing (19) and (20) we get

$$X = (l_a + l_c) = \Big(\sqrt{\gamma_{1,0}^2 c^2 + 2\gamma_{1,0} c R_0 + x_0^2} + \sqrt{\gamma_{3,0}^2 c^2 + 2\gamma_{3,0} c R_0 + x_0^2}\Big)/2. \quad (23)$$

Recalling (4), all terms in x_0 and y_0 are squared so to find X it is sufficient to know $|x_0|$ and $|y_0|$.

Table 2 shows a taxonomy of which parts of the geometry can be determined depending on which reflections are present. In the interest of space the distances from the microphone to the walls, $l_{s_i},\ i = 1, \ldots, 4$ are excluded because there is a straightforward relationship that for the cases listed if x_0 is known so are l_a and l_c and if y_0 is known so are l_b and l_d.

3.4. Least squares formulation

Given a set of N measured TOAs, γ_i, $i = \{q_1, q_2, \ldots, q_N\}$, where $\mathbf{q} = [q_1, q_2, \ldots, q_N]$ is a vector of observable image sources, a vector of all $1/2N(N-1)$ possible TDOAs is defined

$$\Gamma_{\mathbf{\Omega}} = \left[\gamma_{q_1,q_2},\ \gamma_{q_1,q_3}, \ldots, \gamma_{q_{N-1},q_N}\right]^T. \quad (24)$$

For an estimate of the geometry, $\widetilde{\mathbf{\Omega}}$, reflection arrival times, $\tilde{\tau}_i$, $i = \{q_1, q_2, \ldots, q_N\}$, are calculated . The corresponding TDOAs, $\tilde{\tau}_{i,j} = \tilde{\tau}_i - \tilde{\tau}_j$ are assembled into a vector

$$\Gamma_{\widetilde{\mathbf{\Omega}}} = \left[\tilde{\tau}_{q_1,q_2},\ \tilde{\tau}_{q_1,q_3}, \ldots, \tilde{\tau}_{q_{N-1},q_N}\right]^T. \quad (25)$$

We evaluate a room geometry estimate using the sum of the squared differences between the elements of (24) and (25). Therefore, the optimal solution in the least squares sense is

$$\mathbf{\Omega_{ls}} = \arg\min_{\widetilde{\mathbf{\Omega}}} \left\| \Gamma_{\mathbf{\Omega}} - \Gamma_{\widetilde{\mathbf{\Omega}}} \right\|_2^2. \quad (26)$$

Constraints				Resolved parameters				
1	2-a	2-x	2-y					
2	1			R_0				
3	1	1		R_0	X		x_0	$\lvert y_0\rvert$
2		2		R_0	X		x_0	$\lvert y_0\rvert$
3	1		1	R_0		Y	$\lvert x_0\rvert$	y_0
2			2	R_0		Y	$\lvert x_0\rvert$	y_0
4	1	1		R_0	X	Y	x_0	$\lvert y_0\rvert$
4		2		R_0	X	Y	x_0	$\lvert y_0\rvert$
4	1		1	R_0	X	Y	$\lvert x_0\rvert$	y_0
4			2	R_0	X	Y	$\lvert x_0\rvert$	y_0
4		2	1	R_0	X	Y	x_0	y_0
4		1	2	R_0	X	Y	x_0	y_0

Table 2. Taxonomy of parameters which are resolved according to how many of each type of constraint are imposed. The first column represents constraints due to first-order image sources, $i = \{1, 2, 3, 4\}$. The second, third and fourth columns represent adjacent ($i = \{5, 7, 8, 11\}$), opposite-x ($i = \{6, 10\}$) and opposite-y ($i = \{9, 12\}$) constraints respectively. Each second-order constraint assumes two associated first-order constraints.

4. EXPERIMENTAL VALIDATION

The aim of the experimental validation was to verify the analytical results from Section 3.3 by finding all possible source positions which can produce a given set of TOAs $\mathbf{q}$. Using the geometry defined in (22) three test cases were considered. The first had an adjacent constraint $i = 5$ so $\mathbf{q} = [0, 1, 2, 5]$ which is representative of the first four TOAs which would be observed for a source-microphone positioned near a corner. The second had two symmetric opposite constraints $i = \{6, 10\}$ so $\mathbf{q} = [0, 1, 3, 6, 10]$ which is representative of a source-microphone positioned in a long narrow space (e.g. a corridor). The final case had an adjacent constraint $i = 5$, an opposite constraint 6 and all first-order constraints $i = \{1, 2, 3, 4\}$ so $\mathbf{q} = [0, 1, 2, 3, 4, 5, 6]$. This is representative of what might be observable for a source-microphone located towards the middle of a room.

In each case a grid of co-ordinates were specified as candidate source positions, $\mathbf{r}_0$. For each candidate source position the remaining geometry parameters (l_a, l_b, l_c, l_d) were adjusted to minimise the least squares error as defined in (26). As the TOAs are a non-linear function of the geometry this was achieved by numerical optimisation using MATLAB's `lsqnonlin` function, which implements the trust region reflective algorithm [11].

The log error achieved is represented by the shade of the grid location in the plots in Figure 3. The error tends to zero when the candidate source position allows a solution to be found which reproduces exactly the TDOAs. Superimposed on each plot are the analytical solutions to the imposed con-

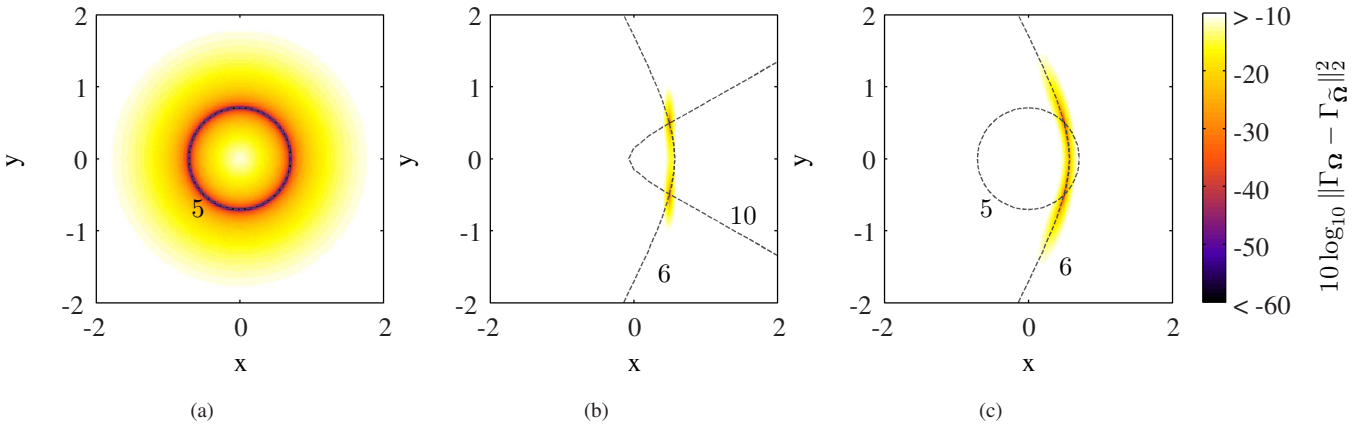

Fig. 3. Lowest error obtained when optimising geometry $\widetilde{\mathbf{\Omega}}$ with x_0 and y_0 fixed using the reflection subsets (a) $\{0, 1, 2, 5\}$, (b) $\{0, 1, 3, 6, 10\}$, (c) $\{0, 1, 2, 3, 4, 5, 6\}$. The theoretical constraints are superimposed as dashed lines.

straints. In Fig. 3(a) this is a circle around the origin while for Fig. 3(b) and (c) it is two points at (0.5, 0.5) and (0.5,-0.5). In each case there is a perfect match between the analytical loci and the points where minimum error is achieved.

5. CONCLUSIONS

Using a rectangular room model the image source method has been used to predict reflection time of arrivals. By considering the constraints imposed by certain combinations of reflections the extent to which the geometry can be known has been determined. The results have been confirmed using numerical optimisation to find the geometry which produces time difference of arrivals which best match those of an unknown geometry.

6. REFERENCES

[1] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*, Springer, 2010.

[2] Xiang (Shawn) Lin, Andy W. H. Khong, and Patrick A Naylor, "A forced spectral diversity algorithm for speech dereverberation in the presence of near-common zeros," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 3, pp. 888–899, Mar. 2012.

[3] Nils Peters, Howard Lei, and Gerald Friedland, "Name that room: Room identification using acoustic features in a recording," in *Proceedings of the 20th ACM international conference on Multimedia*, 2012, pp. 841–844.

[4] Noam R. Shabtai, Yaniv Zigel, and Boaz Rafaely, "Feature selection for room volume identification from room impulse response," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009, pp. 249–252.

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

[6] S. Tervo and T. Tossavainen, "3D room geometry estimation from measured impulse responses," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.

[7] F. Antonacci, A. Sart, and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2010, pp. 2822–2825.

[8] M. Slaney and P. A. Naylor, "Audio and acoustic signal processing," *IEEE Signal Process. Mag.*, vol. 28, 2011.

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

[10] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[11] T. F. Coleman and Y. Li, "An interior, trust region approach for nonlinear minimization subject to bounds," *SIAM J. on Optimization*, vol. 6, pp. 418–445, 1996.