Inference of Room Geometry From Acoustic Impulse Responses

Fabio Antonacci, Jason Filos, *Member, IEEE*, Mark R. P. Thomas, *Member, IEEE*,
Emanuël A. P. Habets, *Senior Member, IEEE*, Augusto Sarti, *Member, IEEE*, Patrick A. Naylor, *Senior Member, IEEE*,
and Stefano Tubaro, *Member, IEEE*

Abstract—Acoustic scene reconstruction is a process that aims to infer characteristics of the environment from acoustic measurements. We investigate the problem of locating planar reflectors in rooms, such as walls and furniture, from signals obtained using distributed microphones. Specifically, localization of multiple two-dimensional (2-D) reflectors is achieved by estimation of the time of arrival (TOA) of reflected signals by analysis of acoustic impulse responses (AIRs). The estimated TOAs are converted into elliptical constraints about the location of the line reflector, which is then localized by combining multiple constraints. When multiple walls are present in the acoustic scene, an ambiguity problem arises, which we show can be addressed using the Hough transform. Additionally, the Hough transform significantly improves the robustness of the estimation for noisy measurements. The proposed approach is evaluated using simulated rooms under a variety of different controlled conditions where the floor and ceiling are perfectly absorbing. Results using AIRs measured in a real environment are also given. Additionally, results showing the robustness to additive noise in the TOA information are presented, with particular reference to the improvement achieved through the use of the Hough transform.

Index Terms—Inference, room geometry, acoustic impulse responses.

I. INTRODUCTION

ACOUSTIC scene reconstruction is a process that aims to establish an understanding of the acoustic environment in which space-time processing algorithms operate. Reflective boundaries and surfaces can be estimated from acoustic measurements, allowing the geometry of the room to be recon-

F. Antonacci, S. Tubaro, and A. Sarti are with the Dipartimento Di Elettronica E Informazione, Politecnico Di Milano, 20133 Milan Italy (e-mail: antonacc@elet.polimi.it; augusto.sarti@polimi.it).

J. Filos and P. A. Naylor are with the Electrical and Electronic Engineering Department, Imperial College London, London SW7 2AZ, U.K. (e-mail: jason.filos03@imperial.ac.uk; p.naylor@imperial.ac.uk).

M. R. P. Thomas was with the Electrical and Electronic Engineering Department, Imperial College London, London SW7 2AZ, U.K.. He is with now Microsoft Research, Redmond, WA 98052 USA (e-mail: markth@microsoft.com).

E. A. P. Habets is with the International Audio Laboratories Erlangen, University of Erlangen-Nuremberg, 91054 Erlangen, Germany (e-mail: emanuel.habets@audiolabs-erlangen.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2012.2210877

structed. Inferring room geometries can be advantageous for applications such as acoustic source localization [4], [5] and wavefield rendering [6]. In [7] the authors show that an accurate modelling of the acoustic propagation enables localization algorithms to improve their accuracy of range and elevation discrimination. In order to model the acoustic propagation, the authors infer the geometry of the environment through the methodology described in [8]. In [6] the authors compensate for the effect of reverberation in wavefield rendering by modelling the room transfer function in the rendering process. Additionally, it has been shown in [9] that the attenuation of interfering signals can be increased when the direction of arrival of the interfering signal and its dominant reflections are available.

One way of locating reflective boundaries employs an inverse mapping of the acoustic multi-path propagation problem in a 2-D geometry representing the reflective surface location [10]. The interesting aspect of this work is that the estimation is accomplished starting from continuous signals. However, typical for a system that uses cross-correlation as the primary source of information, the localization fails at low signal-to-noise ratios. A different approach is adopted by the authors in [11] where a solution for the localization of 2-D reflectors is envisioned by using a single room impulse response (RIR), and proving the uniqueness of the solution. This methodology, however, assumes that all the first-order and second-order reflections generate unique impulsive features in the impulse response, which is not always the case for environments different from the shoe-box shaped room. Moreover, the algorithm must first identify first and second-order reflections. Though it solves the problem of inference of the room geometry in an elegant way, this approach is not always a viable solution in real scenarios, due to the strong assumptions. In [8] a modeling algorithm is proposed that uses a constrained room model and L_1-regularized least-squares method to infer the room geometry. This algorithm is evaluated with a compact microphone array that also houses a speaker emitting a test signal. However, due to the limitations on the position of the sound emitter this approach only correctly classifies a limited number of reflectors in real acoustic environments. Recently, in [12] the authors have extended the methodology to 3D environments and to real world signals.

A geometric approach to localize reflectors in a room was first proposed in [1]. In this work, one microphone and one moving sound source were used to obtain the time of arrivals (TOAs) of the reflections assuming that the source and receiver were synchronized. The TOA information was used to form a set of elliptical constraints on the possible locations of the reflector. The

common tangent of these constraints was shown to correspond to the reflector location that can be found by minimizing a specific cost function in the least-squares sense. In [2] the authors considered the case when measurements were unsynchronized. In this case, the TOAs cannot be found directly, and an additional step was proposed that estimates TOAs from the TDOA estimates. A technique for the estimation of multiple reflectors from a single set of measurements was also proposed, which iteratively minimizes a global solution space. More recently, the authors proposed a robust inference method [3] which utilizes a closed-form solution to minimize the cost function. In addition, a parametrization based on the Hough transform was introduced which could increase the robustness to errors in the TOA estimation.

In this paper, we present a general geometric approach for localizing reflectors that is practical in multipath environments and robust to measurement errors. Three different scenarios are considered: i) the source and receiver signals are synchronized and the source signal is known, ii) the source and receiver signals are unsynchronized and the source signal is known and iii) the source and receiver signals are unsynchronized and the source signal consists of an unknown impulse-like sound like finger-snaps or hand-claps. When impulse-like signals are used (scenario iii), the estimation of TOAs and TDOAs could be affected by errors, due to the non-impulsive nature of the probing signal. TOAs or TDOAs could be affected by errors even in scenario i) and ii), due to non-ideal emission and acquisition systems, thus degrading the accuracy of the localization. In this work we reduce the impact of such errors using a template matching filter technique, which partially compensates for the non-impulsive nature of the probing signal. The iterative localization method proposed in [2] works well on simulated room impulse responses (RIRs). However, since it considers only a single stationary source, it is often impossible to obtain a complete set of TOAs, i.e., a set that will contain TOAs of all reflectors, in real acoustic environments. An alternate approach, proposed in this paper, considers a space parametrization based on the Hough transform that is practical for real environments in which higher-order reflections can arrive at the microphones before the first-order reflections (e.g., corridor-style rooms) or when first-order reflections coming from different walls are hardly distinguishable in the impulse response (e.g., square rooms). By considering acquisitions from multiple source positions, the benefit of this method is studied in detail in terms of improved localization accuracy for a single reflector and is then extended to the case of multiple reflectors. Finally, experimental results using simulated and measured data, obtained using different microphone arrangements, are presented.

The main achievements of this article with respect to the state of the art are:

- the use of the Hough transform (not adopted in [1], [2]) for a refined localization of the reflectors even in adverse environments;
- the removal of the need of distinguishing among first- and second-order reflections, which is a necessary condition in [11];
- with respect to [13], [3], the use of template matching, which allows to estimate TOAs of reflective paths in

a more accurate way, thus enabling the use of imperfect impulsive signals, such as hand clapping or finger snapping.

The remainder of the paper is organized as follows: Section II describes the theory to formulate the problem of reflector localization and provides an overview of the proposed method. Section III discusses the measurement and estimation of TOAs from AIRs. Section IV considers the localization of a single reflector. We present an extension to the single reflector case in Section V that aims to improve the robustness to noise. Section V.E generalizes the reflector estimation to the case of multiple reflectors. Section VI presents experimental results to demonstrate the feasibility of the technique. Finally, Section VII summarizes the paper and suggests directions for future work.

II. PROBLEM FORMULATION AND METHOD OVERVIEW

A. Notation and Problem Formulation

With reference to Fig. 1(a), a sound source located at $\mathbf{r}_{\mathrm{s}} \triangleq [x_{\mathrm{s}} \, y_{\mathrm{s}}]^T$ emits the signal $s(t)$. The signals $x_i(t)$ are observed by M microphones at positions

$$\mathbf{r}_i \triangleq [x_i \, y_i]^T, \quad i = 0, \dots, M-1. \quad (1)$$

The observed signals are given by the convolution of the source $s(t)$ with the corresponding acoustic room impulse responses $h_i(t)$, $i = 0, \dots, M-1$:

$$x_i(t) = \int_0^\infty h_i(t')s(t - t')\,dt' + n_i(t),$$
$$i = 0, \dots, M-1, \quad (2)$$

where $n_i(t)$ is additive environmental noise. Under the hypothesis of ideal reflections AIRs are given by

$$h_i(t) = \sum_{q=0}^{Q} \alpha_{i,q}\delta(t - \tau_{i,q}), \quad (3)$$

where Q is the total number of reflections of all orders, $\alpha_{i,q}$ is an attenuation term and $\tau_{i,q}$ is defined as the TOA associated with the ith microphone and the qth reflection. Note that the TOA of the direct-path is defined with respect to the *null* reflector, i.e., $q = 0$. When sampling AIRs and for subsequent discrete-time processing, delays corresponding to fractions of the sampling period can be taken into account using, for example, [14]. Let $s(n)$, $h_i(n)$, $x_i(n)$ represent sampled versions of the source, channel and observation at microphone i.

For the remainder of this paper, we will consider only TOAs that are related to M direct-paths and N first-order reflections. With reference to Fig. 1(b) we can obtain estimates of the TOAs related to the direct path and first-order reflections by analyzing $h_i(t)$. For this we note that the first peak in $h_i(t)$ is related to the time of propagation of the direct-path from $\mathbf{r}_{\mathrm{s}}$ to $\mathbf{r}_i$ given by $\tau_{i,0}$. Any subsequent peak in $h_i(t)$ is related to the composite time of propagation of the sound due to reflection. By defining $\mathbf{r}_{\mathrm{p},i}$ as the reflection point on any reflector, we obtain $\tau_{i,k}$, $k = 1, \dots, N$ as the sum of the propagation times from $\mathbf{r}_{\mathrm{s}}$ to $\mathbf{r}_{\mathrm{p},i}$, and then from $\mathbf{r}_{\mathrm{p},i}$ to $\mathbf{r}_i$ for any N reflectors present in
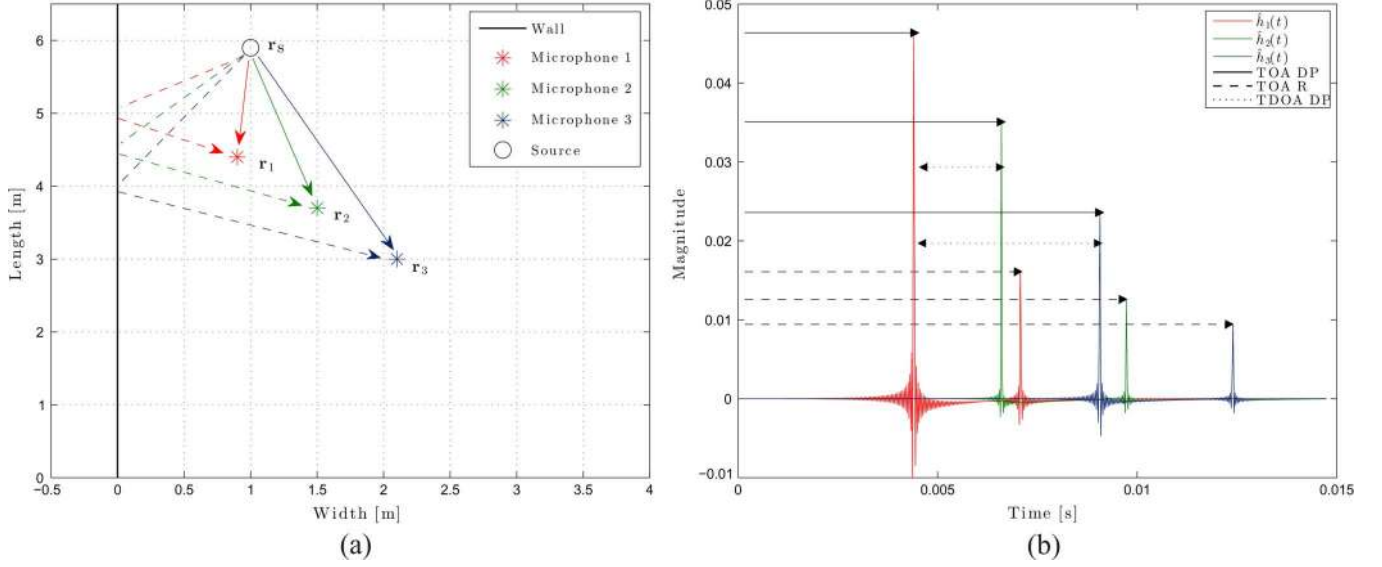
Fig. 1. Problem geometry and associated AIRs. (a) Geometry of the problem: A single reflector in an anechoic space. Three microphones pick up the direct-path (solid line) along with the reflection (dashed line). (b) AIR of three microphone estimates. The solid line TOA DP refers to the time of flight of the direct-path, the dashed line TOA R refers to the composite time of flight from source to reflector and then from reflector to microphone. TDOA DP refers to the time difference of arrival between the TOAs of the direct-path between microphone 1 and the other two.
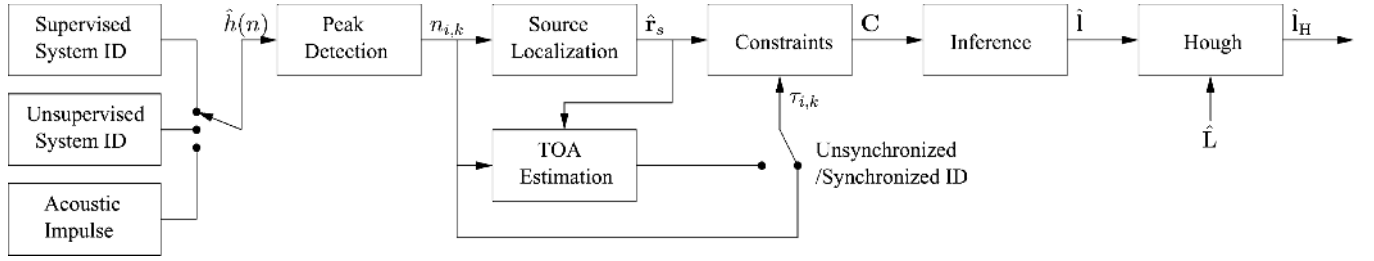

Fig. 2. System diagram. An acoustic impulse response is generated, containing peaks corresponding to sound from reflected boundaries. Peak picking from either the true or estimated system yields TOAs if the measurements are synchronized, else TOAs are estimated with the aid of source localization. TOAs, combined with an estimate of the source and knowledge of the geometry of the receiver array, are used to parameterize a set of ellipses. Geometric inference is performed with the ellipses by finding lines of common tangency that correspond to the reflector locations.

our environment. Additionally we define $\Delta_{i,j}$ as the TDOA of the direct-path between the ith and jth microphone:

$$\Delta_{i,j} = |\tau_{i,0} - \tau_{j,0}|. \tag{4}$$

The geometric inference problem can be summarized as follows: locate the reflective boundaries of an acoustic environment based upon TOAs estimated from acoustic measurements.

B. Method Overview

The proposed approach is outlined in Fig. 2. A microphone array of M microphones, whose relative geometry is assumed known, is placed in a reverberant environment. The AIRs $\hat{h}_i(n)$, from a single source to the array of receivers $i = \{0, \ldots M - 1\}$, are estimated using either supervised or unsupervised techniques. Peaks within the AIR correspond to the direct-path from source to receiver and the summed reflective paths from source to reflector and reflector to receiver. The temporal location of these peaks is found in discrete time as $n_{i,k}$. In the case where source and receiver are synchronized, these time instants correspond to the TOAs. However, in many practical cases the synchronization cannot be achieved; therefore the TDOAs, which are preserved, are used to localize the source relative to the mi-

crophone array from which TOAs are estimated. If any one of the TOAs, TDOAs or source location are unavailable then it can always be estimated from the remaining two.

The TOAs and the location of the source relative to the microphone array are used to form a set of elliptical constraints on the possible locations of the reflector, where the common tangent of the ellipses corresponds to the location of the reflector.

III. MEASUREMENT AND ESTIMATION OF TOAs

In this Section we illustrate the steps that, starting from the AIRs, lead to the estimation of TOAs. In particular, we consider both the cases of synchronized and unsynchronized AIRs. In the former case the TOAs are directly extracted from the impulse response. In the latter situation (Section III.B), instead, TOAs can be estimated once the source has been localized. In order to make the peak detection algorithm robust against non-ideal acquisition and emission systems, we propose a template matching procedure in Section III.A to improve the relevant temporal characteristics of the received source signal.

A. Peak Detection From Acoustic Impulse Responses

Fractional delays result from path lengths that are not multiples of the distance propagated by sound in one sample pe-
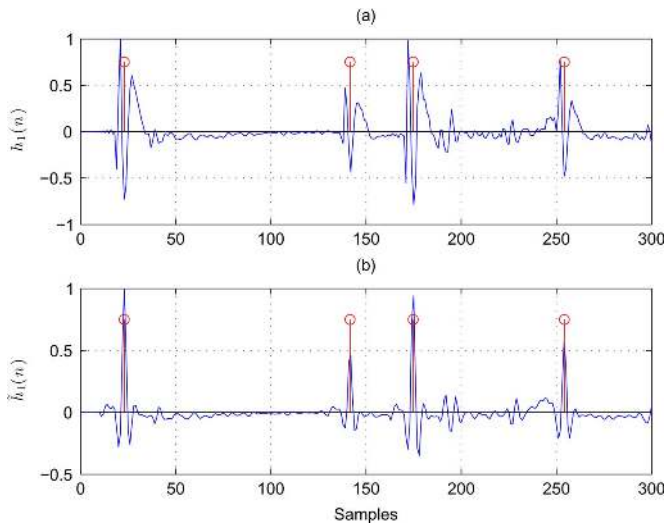
Fig. 3. Direct-path and three first-order reflections for (a) measured impulse response, (b) modified impulse response according to (7). Red 'o' mark the estimated peak locations.

riod. Detection of impulsive events can be achieved to within one sample by considering local centres of energy with algorithms such as the sliding group delay function [15] and the *find-peaks* function [16]. In this work we detect the $N+1$ most relevant peaks in the AIRs. More specifically, the Hough transform-based localization technique initially requires only the knowledge of the propagation time of the direct path and the echo related to the closest reflector for the prescribed source location. In a later stage, the remaining $N-1$ impulses are used.

AIRs measured in real acoustic environments present another challenging problem as the source impulse-like emission $h^{\mathrm{s}}(n)$ is convolved with the AIR $h_i(n)$. Assuming supervised identification with which estimation error can be ignored, the measured AIR is

$$\hat{h}_i(n) = h^{\mathrm{s}}(n) * h_i(n). \tag{5}$$

An example impulse response for a measured system is seen in Fig. 3(a), showing respectively the direct-path and three first-order reflections for a single channel. The centres of each event are marked by 'o', each of which are surrounded by nearby ripples caused by $h^{\mathrm{s}}(n)$. The ripples cause ambiguity in determining the exact time corresponding to the peak and therefore a matched filter is proposed to alleviate this problem.

The length of $h^{\mathrm{s}}(n)$ is usually sufficiently short that it has decayed before the arrival of the first-order reflections [17], as in the example. Therefore, $h^{\mathrm{s}}(n)$ can be observed from the first few nonzero taps in $\hat{h}_i(n)$. Let n_i^{DP} be the propagation time of the direct-path signal from the source to microphone i and N^{s} be the approximate length of the loudspeaker impulse response. The impulsive source emission can be estimated by $\hat{h}_i^{\mathrm{s}}(n) = \hat{h}_i(n + n_i^{\mathrm{DP}})w_i(n)$, where

$$w_i(n) = \begin{cases} 1 & \text{if } 0 \le n < N^{\mathrm{s}} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The filter $\hat{h}^{\mathrm{s}}(n)$ is equalized through the sliding correlation or matched filter [18],

$$\tilde{h}_i(n) = \sum_{j=0}^{N^{\mathrm{s}}-1} \hat{h}_i^{\mathrm{s}}(j) h_i(n+j), \tag{7}$$

that equalizes $\hat{h}_i^{\mathrm{s}}$ to a single peak as demonstrated in Fig. 3 for a measured AIR. In (b) the mean group delay of $\hat{h}_i^{\mathrm{s}}$ has been compensated. The detected peaks are denoted by $n_{i,k}$ where i and k are the microphone and reflector index respectively. In the case of synchronized measurements, $n_{i,k} = \tau_{i,k} f_{\mathrm{s}}$.

B. Estimation of TOAs From Unsynchronized Airs

In order to estimate TOAs from unsynchronized AIRs, the TDOAs of the direct-paths are used to localize the acoustic source and consequently estimate the propagation time of the direct sound from the source to a reference microphone. Subsequently, the propagation times of all the other arrivals can be inferred.

The reference microphone $(i = 0)$ is placed at the origin of the coordinate system $\mathbf{r}_0 = [0\,0]^T$. The distances from the origin to the ith microphone and the source are denoted by R_i and R_{s}, respectively, where

$$R_i \triangleq \|\mathbf{r}_i\| = \sqrt{x_i^2 + y_i^2}, \quad i = 1, \ldots, M-1$$
$$R_{\mathrm{s}} \triangleq \|\mathbf{r}_{\mathrm{s}}\| = \sqrt{x_{\mathrm{s}}^2 + y_{\mathrm{s}}^2}.$$

The difference in the distances of microphones i and j from the source is the *range difference*, $d_{i,j}$, and is proportional to the TDOA of the direct-path between the ith and jth microphone, $\Delta_{i,j}$. If the speed of sound is η, then

$$d_{i,j} = \eta \cdot \Delta_{i,j} \tag{8}$$

and the propagation time $\tau_{0,0}$ of the direct path from $\mathbf{r}_{\mathrm{s}}$ to $\mathbf{r}_0$ is given by

$$\tau_{0,0} = \frac{R_{\mathrm{s}}}{\eta}. \tag{9}$$

Recall that the subscript 0 denotes the null reflector. In order to estimate R_{s} we employ the passive source localization algorithm [4] that is based on a least squares (LS) estimator employing a spherical least squares error criterion defined in 3D space. For our purposes this algorithm is modified for a 2D space so that the spherical LS error function is modified to a circular LS error criterion. From [4] we establish the distance D_i from the ith microphone to the source

$$\hat{D}_i = R_{\mathrm{s}} + \hat{d}_{i,0}. \tag{10}$$

The error function is then defined as the difference between the measured and true values,

$$\mathbf{e}(\mathbf{r}_{\mathrm{s}}) = \mathbf{A}\boldsymbol{\theta} - \mathbf{b} \tag{11}$$

where

$$\mathbf{A} \triangleq [\mathbf{S} \mid \hat{\mathbf{d}}], \quad \mathbf{S} \triangleq \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_{M-1} & y_{M-1} \end{bmatrix}$$

$$\boldsymbol{\theta} \triangleq \begin{bmatrix} x_{\mathrm{s}} \\ y_{\mathrm{s}} \\ R_{\mathrm{s}} \end{bmatrix}, \quad \mathbf{b} \triangleq \frac{1}{2} \begin{bmatrix} R_1^2 - \hat{d}_{1,0}^2 \\ R_2^2 - \hat{d}_{2,0}^2 \\ \vdots \\ R_{M-1}^2 - \hat{d}_{M-1,0}^2 \end{bmatrix}$$

and $\mathbf{S} \mid \hat{\mathbf{d}}$ indicates that $\mathbf{S}$ and $\hat{\mathbf{d}}$ are stacked side-by-side with $\hat{\mathbf{d}} = [\hat{d}_{1,0}, \hat{d}_{2,0}, \ldots, \hat{d}_{M-1,0}]^T$. The corresponding LS criterion is then given by

$$J = \mathbf{e}^T \mathbf{e} = [\mathbf{A}\boldsymbol{\theta} - \mathbf{b}]^T [\mathbf{A}\boldsymbol{\theta} - \mathbf{b}]. \qquad (12)$$

The solution for $\boldsymbol{\theta}$ is given by [4]

$$\hat{\boldsymbol{\theta}}_1 = \mathbf{A}^\dagger \mathbf{b}, \qquad (13)$$

where $(\cdot)^\dagger$ defines the pseudo-inverse. We can now use the estimate of the distance from the ith microphone to the source in order to estimate the TOAs of the direct-path for each of these microphones (note that $D_0 \triangleq R_s$),

$$\hat{\tau}_{i,0} = \frac{\hat{D}_i}{\eta}, \quad i = 0, \ldots, M-1. \qquad (14)$$

The TOAs of the reflective paths are straightforwardly obtained since both $\hat{\tau}_{i,0}$ and the TDOAs between the direct-paths and the reflective paths are known from inspection of $h_i(t)$, even if the source and microphone signals are not synchronized. In the next section the localization of a single reflector based on knowledge of the TOAs is formulated.

IV. LOCALIZATION OF A SINGLE REFLECTOR FROM TOAS

With reference to the system diagram in Fig. 2, our aim is to use the TOAs and positions of multiple microphones and a single source to obtain a geometrical interpretation of the boundaries of the acoustic enclosure. We consider initially an acoustic scene in which a single planar reflector is present. We then introduce a cost function whose minimum in an error-free scenario corresponds to the line parameters of the reflector. The same cost function will be extended in Section V.E to the case of a more complex acoustic scene in which multiple reflectors are present.

A. Geometric Constraint

We make the fundamental assumption that source and receivers lie on the same plane and the lying plane of the reflector is orthogonal to this plane. In this scenario, the geometry of the acoustic scene is described by the plane in which sources and receivers lie. In a two-dimensional geometry a line is the collection of points $[x \ y]^T$ such that

$$l = \{(x, y) \in \mathbb{R}^2 | l_1 x + l_2 y + l_3 = 0\}, \qquad (15)$$

which after setting the line parameter $\mathbf{l} = [l_1 \ l_2 \ l_3]^T$ can be written as

$$\mathbf{l}^T [x \ y \ 1] = 0. \qquad (16)$$

With reference to the notation introduced in Section II and to Fig. 4, we observe that $\tau_{i,k}$ is the sum of two terms: the propagation time from the source $\mathbf{r}_s$ to the unknown reflection point $\mathbf{r}_{p,i}$ on the reflector $\mathbf{l}$ and the propagation time $\mathbf{r}_{p,i}$ to the microphone position $\mathbf{r}_i$. The knowledge of $\tau_{i,k}$, $\mathbf{r}_s$ and $\mathbf{r}_i$, therefore, bounds the reflection point $\mathbf{r}_{p,i}$ to lie on an ellipse with foci in $\mathbf{r}_s$ and $\mathbf{r}_i$ and whose major diameter is $Q_{i,k}^{\{maj\}} = \eta \tau_{i,k}$, η being the sound speed. We assume, moreover, that the reflection undergoes Snell's law, therefore the line perpendicular to $\mathbf{l}$
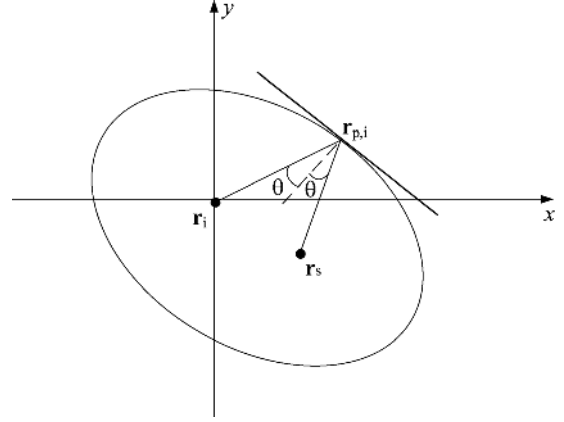

Fig. 4. The TOA of the reflective path is constituted by the time of propagation from $\mathbf{r}_s$ to $\mathbf{r}_{p,i}$ and from $\mathbf{r}_{p,i}$ to $\mathbf{r}_i$. Possible reflection points lie on an ellipse.

is also the bisector of the angle $\mathbf{r}_s \widehat{\mathbf{r}_{p,i}} \mathbf{r}_i$. According to the properties of ellipses, this means that $\mathbf{l}$ is tangential to the ellipse. We notice that the reflection point $\mathbf{r}_{p,i}$ depends on the positions $\mathbf{r}_i$ and $\mathbf{r}_s$, therefore if we consider another microphone in the set $\mathbf{r}_i, i = 0, \ldots, M-1$, the reflection point on the ellipse changes. However, what remains unchanged for all the ellipses is that they are all tangential to the reflector. The common tangent estimation takes inspiration from the observation: the reflector line is found as the line that is tangential to all M ellipses. In order to accomplish this task, however, we need to parameterize the above tangential constraint in such a way that the tangent lines directly appear in the ellipse equation. This is where the projective geometry provides a convenient formulation.

B. Parametrization of the Ellipse

The aim of this section is to find the parameters of the ellipse given the foci $\mathbf{r}_i = [x_i \ y_i]^T$ and $\mathbf{r}_s = [x_s \ y_s]^T$ and the major axis $Q_{i,k}^{\{maj\}}$. Using the parameters $\{a, b, c, d, e, f\}$ the conic can be expressed as [19]

$$C = \{(x, y) \in \mathbb{R}^2 | ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0\}. \qquad (17)$$

A parametrization that is convenient for our purposes is based on the representation of points using homogeneous coordinates. The homogeneous coordinates for the point $[x \ y]^T$ are $[\lambda x \ \lambda y \ \lambda]^T$, λ being a scalar different from zero. In such a representation the point $[x \ y]^T$ in the Euclidean space is mapped into a three-dimensional space and all points aligned on the direction $[\lambda x \ \lambda y \ \lambda]^T$ correspond to the same point in the Euclidean space, thus defining an equivalence class between homogeneous and Euclidean coordinates.

In homogeneous coordinates, the conic in (17) becomes

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = 0, \qquad (18)$$

where $\mathbf{x} = [\lambda x \ \lambda y \ \lambda]^T$ and $\mathbf{C}$ is the conic matrix, given by

$$\mathbf{C} = \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & f \end{bmatrix}. \qquad (19)$$

This defines an ellipse after constraining

$$\det(\mathbf{C}) \neq 0, \quad \begin{vmatrix} a & b \\ b & c \end{vmatrix} > 0, \quad \det(\mathbf{C})/(a+c) < 0. \quad (20)$$

The implicit equation of an ellipse with foci in (x_s, y_s) and (x_i, y_i) and with major diameter $Q_{i,k}^{\mathrm{maj}} = c\tau_{i,k}$ is

$$\sqrt{(x-x_i)^2 + (y-y_i)^2} + \sqrt{(x-x_s)^2 + (y-y_s)^2} = Q_{i,k}^{\mathrm{maj}}. \quad (21)$$

By expanding (21) and comparing it term by term with a conic with parameters $(a_{i,k}, b_{i,k}, c_{i,k}, d_{i,k}, e_{i,k}, f_{i,k})$ we obtain

$$a_{i,k} = Q_{i,k}^{\{\mathrm{maj}\}^2} - (x_s - x_i)^2,$$
$$b_{i,k} = -2(x_s - x_i)(y_s - y_i),$$
$$c_{i,k} = Q_{i,k}^{\{\mathrm{maj}\}^2} - (y_s - y_i)^2,$$
$$d_{i,k} = (x_i - x_s)\left(x_i^2 + y_i^2 - x_s^2 - y_s^2 - Q_{i,k}^{\{\mathrm{maj}\}^2}\right) - x_s Q_{i,k}^{\{\mathrm{maj}\}^2},$$
$$e_{i,k} = (y_i - y_s)\left(x_i^2 + y_i^2 - x_s^2 - y_s^2 - Q_{i,k}^{\{\mathrm{maj}\}^2}\right) - y_s Q_{i,k}^{\{\mathrm{maj}\}^2},$$
$$f_{i,k} = \left(x_s^2 Q_{i,k}^{\{\mathrm{maj}\}^2} + y_s^2 Q_{i,k}^{\{\mathrm{maj}\}^2} \frac{y_i^2 - y_s^2 + x_i^2 - x_s^2 - Q_{i,k}^{\{\mathrm{maj}\}^2}}{4}\right).$$

In the remainder of this section we consider a scenario in which we only have a single reflector, therefore $\mathbf{C}_{i,k}$ reduces to $\mathbf{C}_i$. More meaningful for our purposes in the definition of the *line conic* associated with the point conic defined in (18) and (19). The line $\mathbf{l}$ passes through the point $\mathbf{x}$ *iff*

$$\mathbf{l}^T \mathbf{x} = 0$$

and is tangential to the point ellipse $\mathbf{C}_i$ iff

$$\mathbf{l}^T \mathbf{C}_i^* \mathbf{l} = 0, \quad (22)$$

where $\mathbf{C}_i^* = \det(\mathbf{C}_i)\mathbf{C}_i^{-1}$ is the adjoint of the conic matrix $\mathbf{C}_i$.

C. Common Tangent Estimation Algorithm

As mentioned previously, if we acquire multiple impulse responses from $\mathbf{r}_s$ to $\mathbf{r}_i$, $i = 0, \ldots, M-1$, the line $\mathbf{l}$ is tangent to all the ellipses $\mathbf{C}_i$, $i = 0, \ldots, M-1$ at points $\mathbf{r}_{\mathrm{p},i}$, $i = 0, \ldots, M-1$. Fig. 5 shows an example where $M = 3$. Combining the constraints in (22), the reflector line is the simultaneous solution of [1]

$$\begin{cases} \mathbf{l}^T \mathbf{C}_0^* \mathbf{l} & = 0 \\ \mathbf{l}^T \mathbf{C}_1^* \mathbf{l} & = 0 \\ \ldots \\ \mathbf{l}^T \mathbf{C}_{M-1}^* \mathbf{l} & = 0 \end{cases} \quad (23)$$

Since we have three unknowns (the parameters l_1, l_2, l_3) we need at least $M = 3$.
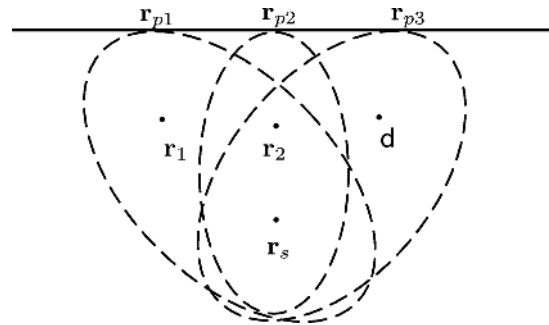

Fig. 5. The reflector line is the common tangent to the ellipses traced for $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$.

From a geometrical standpoint, solving (23) corresponds to finding the line $\mathbf{l}$, in the line parameter space, that lies on all the manifolds representing constraints in (23).

The solution of a nonlinear system as in (23) is non-trivial when the measures of $\tau_{i,k}$ are affected by measurement errors and the positions $\mathbf{r}_s$ and $\mathbf{r}_i$ are known up to some uncertainty. We shall combine the individual equations in (23) into the cost function [1]

$$J(\mathbf{l}) = \sum_{i=0}^{M-1} \|\mathbf{l}^T \mathbf{C}_i^* \mathbf{l}\|^2, \quad (24)$$

which is a multivariate fourth-order polynomial in l_1, l_2, l_3. We notice that the cost function admits the trivial solution $\mathbf{l} = \mathbf{0}$. In order to find the global minimum we resort to an analytical minimization technique [3] by slicing the homogeneous coordinates space (l_1, l_2, l_3) with the two planes $l_1 = 1$ and $l_3 = 1$. On these two planes the cost function $J(\mathbf{l})$ is not homogeneous and the set of local minima can be found in an analytical way. By merging the minima found on the two planes we obtain the global solution.

If we consider the case in which we iteratively estimate the line reflectors one at a time then we can denote the coefficients of the adjoint conic associated to the ith ellipse with the matrix

$$\mathbf{C}_i^* = \begin{bmatrix} \alpha_i & \beta_i/2 & \delta_i/2 \\ \beta_i/2 & \gamma_i & \varepsilon_i/2 \\ \delta_i/2 & \varepsilon_i/2 & \zeta_i \end{bmatrix}.$$

Using this notation the cost function can be expanded as

$$\begin{aligned} J(\mathbf{l}) = \sum_{i=0}^{M-1} \big[& \alpha_i^2 l_1^4 + \gamma_i^2 l_2^4 + \zeta_i^4 l_3^4 + 2\alpha_i\beta_i l_1^3 l_2 \\ & + 2\alpha_i\delta_i l_1^3 l_3 + 2\beta_i\gamma_i l_1 l_2^3 + 2\gamma_i\varepsilon_i l_2^3 l_3 + 2\delta_i\zeta_i l_1 l_3^3 \\ & + 2\varepsilon_i\zeta_i l_2 l_3^3 + \left(2\alpha_i\gamma_i + \beta_i^2\right) l_1^2 l_2^2 + \left(2\alpha_i\zeta_i + \delta_i^2\right) l_1^2 l_3^2 \\ & + \left(2\gamma_i\zeta_i + \varepsilon_i^2\right) l_2^2 l_3^2 + 2(\alpha_i\varepsilon_i + \beta_i\delta_i) l_1^2 l_2 l_3 \\ & + 2(\beta_n\varepsilon_n + \gamma_n\delta_n) l_1 l_2^2 l_3 + 2(\beta_n\zeta_n + \delta_n\varepsilon_n) l_1 l_2 l_3^2 \big]. \end{aligned} \quad (25)$$

Slicing $J(\mathbf{l})$ with the planes $l_1 = 1$ and $l_2 = 1$ means computing $J(\mathbf{l})|_{l_1=1}$ and $J(\mathbf{l})|_{l_3=1}$, respectively.
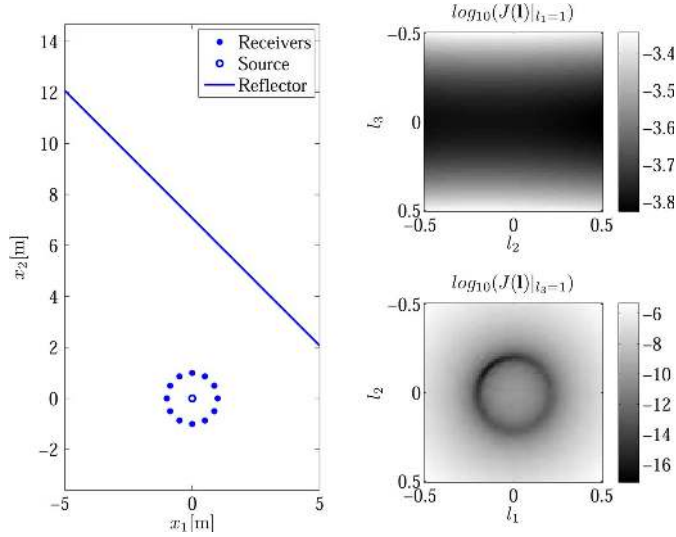
Fig. 6. Example of cost functions $J(\mathbf{l})|_{l_1=1}$ and $J(\mathbf{l})|_{l_3=1}$ for a specific configuration of microphones and sources.

We proceed by finding the zeros of the gradient of $J(\mathbf{l})|_{l_1=1}$ and $J(\mathbf{l})|_{l_3=1}$, so that we obtain the sets

$$L_1 = \left\{ \mathbf{l}: \left.\frac{\partial J(\mathbf{l})}{\partial l_2}\right|_{l_1=1} = 0 \;\wedge\; \left.\frac{\partial J(\mathbf{l})}{\partial l_3}\right|_{l_1=1} = 0 \right\}, \quad (26)$$

and

$$L_3 = \left\{ \mathbf{l}: \left.\frac{\partial J(\mathbf{l})}{\partial l_1}\right|_{l_3=1} = 0 \;\wedge\; \left.\frac{\partial J(\mathbf{l})}{\partial l_2}\right|_{l_3=1} = 0 \right\}. \quad (27)$$

Notice that the partial derivatives of the slices $J(\mathbf{l})_{l_1=1}$ and $J(\mathbf{l})_{l_3=1}$ are polynomials of order 3, and therefore L_1 and L_3 contain 9 solutions each. Some of them are in the complex domain and do not admit a solution. We denote with $\bar{L}_1$ and $\bar{L}_3$ the subsets of purely real solutions of L_1 and L_3, respectively. We then define

$$\bar{L} = \bar{L}_1 \cup \bar{L}_3 = \{\mathbf{l}_1 \ldots \mathbf{l}_K\}, \quad (28)$$

which contains $K \leq 18$ candidate solutions. The global minimum of $J(\mathbf{l})$ is selected as

$$\hat{\mathbf{l}} = \arg \min_{\mathbf{l}_m} J(\mathbf{l}_m), \mathbf{l}_m \in \bar{L}. \quad (29)$$

Notice that the trivial solution $\mathbf{l} = 0$ is inherently avoided by cutting the line parameter space with the planes $l_1 = 1$ and $l_3 = 1$. Note also that $J(\mathbf{l})|_{l_1=1}$ and $J(\mathbf{l})|_{l_3=1}$ are no longer homogeneous.

Finally, Fig. 6 shows an example of slices $J(\mathbf{l})|_{l_1=1}$ and $J(\mathbf{l})|_{l_3=1}$ (right-hand side), for the configuration of microphones and sources on the left-hand side. The correct line parameters for the configuration under analysis are $l_1 = -0.14, l_2 = -0.14, l_3 = 1$. We also notice that the minimum of $J(\mathbf{l})|_{l_3=1}$ has an asymmetric shape, and in particular is sharper along the radius of the circumference centered in $l_1 = 0, l_2 = 0$. As a consequence, the distance of the reflector will be identified better than its orientation. This fact depends on the mutual configuration of the source and the microphones and cannot be attributed to the proposed methodology.

V. LOCALIZATION USING THE HOUGH TRANSFORM

In a noise-free scenario, and neglecting the effects of machine precision, the global minimum of (29) is also the true solution, so that all ellipses are perfectly aligned and yield a single solution that is the common tangent to all ellipses considered. However, due to errors in the TOA information, the ellipses are prone to mismatch and (29) is not guaranteed to correspond to the true reflector. In this Section we present an ad-hoc method to robustly estimate the line parameters of a reflector. TOA information, along with their geometrical representation as ellipses, can be parameterized to points in the Hough parameter space. Peaks in the parameter space correspond to reflectors. By considering multiple TOA estimates, achieved by moving the sound source in the acoustic scene, the accuracy and robustness of the reflector localization can be improved. The same approach is also used in the following section to handle the case of multiple reflectors present in the acoustic scene.

A. Geometrical Relation Between Line Estimates and Ellipses

There are different ways to convert TOA information to points in the Hough space. Rather than establishing a direct parametrization of the TOAs we adhere to the geometric framework of the previous Section. In order to initialize the localization algorithm based on the Hough transform, we need an initial estimate $\mathbf{l}$ of the reflector line using the minimization in (29). This is accomplished by placing the source close to the reflector of interest, so that the first echo after the direct path in the impulse response comes from the reflector under analysis. In this way, we obtain a set of M ellipses coming from the same reflector that are used for the initial estimate $\mathbf{l}$. In the following stage we use the information coming from the remaining $N - 1$ source positions. Given M microphones and N source positions, the aim is to define a set of candidate points to be used for the refinement of the first estimate. These points are defined as

$$\mathbf{p}_j \triangleq [x_j \, y_j]^T, \quad j = 0, \ldots, P, \quad (30)$$

where $MN - 1 \leq P \leq 2MN - 1$. The elements in $\mathbf{P} = [\mathbf{p}_0 \, \mathbf{p}_1 \, \ldots \, \mathbf{p}_P]$ are either points of intersection, points of tangency, or closest coordinate points of an ellipse and the initial reflector line estimate $\mathbf{l}$. Consequently, for every ellipse $\mathbf{C}$ and reflector line $\mathbf{l}$ the following hold:

- If $\mathbf{l}$ goes through $\mathbf{C}$ then we obtain two points of intersection.
- If $\mathbf{l}$ touches $\mathbf{C}$ at one point, or in other words if $\mathbf{l}$ is tangent to $\mathbf{C}$, then we obtain one point of tangency.
- If $\mathbf{l}$ does not go through $\mathbf{C}$ then we need to calculate the closest point on the line with respect to the conic.

B. Analytical Framework

In homogenous coordinates, any line $\mathbf{l}$ cutting through the ellipse $\mathbf{C}$ intersects the ellipse at the two points of intersection $\mathbf{p}_\alpha \triangleq [x_\alpha \, y_\alpha \, 1]^T$ and $\mathbf{p}_\beta \triangleq [x_\beta \, y_\beta \, 1]^T$. Furthermore, there exist two lines parallel to $\mathbf{l}$, i.e., with slope $m = -(l_1)/(l_2)$, that touch the ellipse at the tangential points $\mathbf{p}_{\bar{\alpha}} \triangleq [x_{\bar{\alpha}} \, y_{\bar{\alpha}} \, 1]^T$ and $\mathbf{p}_{\bar{\beta}} \triangleq [x_{\bar{\beta}} \, y_{\bar{\beta}} \, 1]^T$. Calculating these points is outlined in detail

in [3]. Any line $\mathbf{l}$ can be classified in three ways based upon its intersection with an ellipse:

1) If $\mathbf{l}$ cuts through the ellipse, then one of its parallel lines, touching the ellipse at point $\mathbf{p}_{\bar{\alpha}}$, will be either to the left or right, above or below $\mathbf{l}$. Substitution into (16) yields $\mathbf{l}^T \mathbf{p}_{\bar{\alpha}} \neq 0$. If $\mathbf{l}^T \mathbf{p}_{\bar{\alpha}} > 0$ and $\mathbf{l}$ passes through the ellipse, then by definition $\mathbf{l}^T \mathbf{p}_{\bar{\beta}} < 0$. Consequently if $\mathbf{l}^T \mathbf{p}_{\bar{\alpha}} < 0$ then $\mathbf{l}^T \mathbf{p}_{\bar{\beta}} > 0$.

2) If $\mathbf{l}$ is tangential to the ellipse, then $\mathbf{l}^T \mathbf{p}_{\bar{\alpha}}$ is either equal to zero, or not equal to zero. Therefore if $\mathbf{l}^T \mathbf{p}_{\bar{\alpha}} = 0$, then $\mathbf{l}^T \mathbf{p}_{\bar{\beta}} \neq 0$. A similar argument holds for the case when $\mathbf{l}^T \mathbf{p}_{\bar{\beta}} = 0$.

3) If $\mathbf{l}$ neither intersects or is tangential to the ellipse, then the two parallel lines touching the ellipse at points $\mathbf{p}_{\bar{\alpha}}$ and $\mathbf{p}_{\bar{\beta}}$ are either both below, above, left or right of $\mathbf{l}$. Therefore if $\mathbf{l}^T \mathbf{p}_{\bar{\alpha}} > 0$ then $\mathbf{l}^T \mathbf{p}_{\bar{\beta}} > 0$. If $\mathbf{l}^T \mathbf{p}_{\bar{\alpha}} < 0$ then $\mathbf{l}^T \mathbf{p}_{\bar{\beta}} < 0$.

Consequently, in order to determine the relationship between $\mathbf{l}$ and the ellipse, it is sufficient to compute

$$\Phi = |\mathrm{sgn}(\mathbf{l}^T \mathbf{p}_{\bar{\alpha}}) + \mathrm{sgn}(\mathbf{l}^T \mathbf{p}_{\bar{\beta}})|, \tag{31}$$

where $\mathrm{sgn}(\cdot)$ is defined as

$$\mathrm{sgn}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

If $\Phi < 1$, then $\mathbf{l}$ goes through the ellipse. If $\Phi = 1$, then $\mathbf{l}$ is tangential to the ellipse. Finally, if $\Phi > 1$, then $\mathbf{l}$ does not intersect the ellipse.

C. Obtaining Candidate Points

New candidate points are appended to $\mathbf{P}$ in the following way. First (31) is used to classify the line into one of the three classes. In the first case it is sufficient to calculate the two points of intersection and add the resulting points to $\mathbf{P}$. In the second case the single point of intersection is calculated to obtain one point of tangency and the result stored in $\mathbf{P}$. In the final case the two tangential points $\mathbf{p}_{\bar{\alpha}}$ and $\mathbf{p}_{\bar{\beta}}$ are used since one of them will be the closest point on the line to the ellipse and the other the furthest. Since we are only interested in the closest point, it is sufficient to compute the distance of points $\mathbf{p}_{\bar{\alpha}}$ and $\mathbf{p}_{\bar{\beta}}$ and the line, by projecting them both onto the line and selecting the shortest distance from

$$\min\left\{ \frac{|l_1 x_{\bar{\alpha}} + l_2 y_{\bar{\alpha}} + l_3|}{\sqrt{l_1^2 + l_2^2}}, \frac{|l_1 x_{\bar{\beta}} + l_2 y_{\bar{\beta}} + l_3|}{\sqrt{l_1^2 + l_2^2}} \right\}, \tag{32}$$

and adding the corresponding coordinate point to $\mathbf{P}$.

D. Reflector Localization Using the Hough Transform

In this paragraph we process the set of candidate points $\mathbf{P} = [\mathbf{p}_0 \, \mathbf{p}_1 \, \cdots \, \mathbf{p}_P]$ through the Hough transform in order to refine the initial estimate $\mathbf{l}$ of the reflector. The Hough transform can be used for estimating the parameters of a shape from its boundary points [20]. It considers the following normal parametrization

$$\rho = x \cos \theta + y \sin \theta, \tag{33}$$

which specifies a straight line by the angle θ of its normal and its algebraic distance ρ from the origin. A point in the cartesian

space maps to a sinusoid in the Hough parameter space that corresponds to all the lines passing through it. Conversely, points in the parameter space are transformed into lines in the Cartesian coordinate space. Given two points lying on a line with parameters (ρ, θ), in the Hough parameter space the sinusoids corresponding to these two points intersect at ρ, θ. Therefore, given the points $\mathbf{p}_j$ in the coordinate space, the parameters of a line corresponding to the best-fit of $\mathbf{P}$ can be found. Let $\rho \in \mathbb{R}$ and $\theta \in [0, \pi]$. For each point $[x_j \, y_j]^T$ we calculate

$$\hat{\rho} = x_j \cos \hat{\theta} + y_j \sin \hat{\theta}. \tag{34}$$

The results are stored in an accumulator $\mathcal{A}$, initially set to zero, which is incremented at every step such that:

$$\mathcal{A}(\hat{\rho}, \hat{\theta}) = \mathcal{A}(\hat{\rho}, \hat{\theta}) + 1. \tag{35}$$

The position of the largest maximum of the accumulator given by

$$[\hat{\theta}_{\max}, \hat{\rho}_{\max}] = \arg\max\{\mathcal{A}(\rho, \theta)\}, \tag{36}$$

is then picked, which finally leads to the line parameters of the best-fit:

$$\hat{\mathbf{l}}_{\mathrm{H}} = [\cos(\hat{\theta}_{\max}) \, \sin(\hat{\theta}_{\max}) \, (-\hat{\rho}_{\max})]^T. \tag{37}$$

By taking repeated measurements of TOAs, using a source that is placed at different locations in the acoustic scene, it is possible to append additional data points to $\mathbf{P}$ for a single reflector. True solutions will cluster around the same point in the Hough space, while outliers will receive fewer votes in the accumulator space. There are many robust evaluators available that dynamically remove contributions of backgrounds and analyze voting patterns around peaks in the accumulator space [21]. However, when considering a single reflector in the Hough space, it is often sufficient in practice to estimate the single most voted bin to obtain $\hat{\mathbf{l}}_{\mathrm{H}}$. By computing local centers of energy and discarding outliers in the Hough space, more accurate reflector results can be obtained even when TOA measurements are affected by noise.

E. Extension to Multiple Reflectors

There are many additional challenges when considering multiple reflectors. One of the challenges is accurately obtaining TOA estimates for all reflectors present in the acoustic environment. Another challenge is matching the TOAs to the correct reflectors. The latter problem is known in the literature as measurement disambiguation. One of the biggest challenges when analyzing RIRs is distinguishing between direct-path and echo-path peaks and between peaks related to different reflectors (virtual sources). Furthermore, it is necessary to find the set of corresponding TOAs for different sensor groups. Recently, [22] presented an approach for disambiguation of TDOA estimates for multiple sources. This approach is based on graph theory, making use of an efficient algorithm suitable for real-time implementation. The authors observed that by exploiting two TDOA constraints they could match peaks correctly to different sources. Although not mentioned in the paper, this also holds for reflections, which can be seen as virtual sources.

In contrast to [22], where disambiguation of TOA information is performed directly after capturing the TDOA estimates, we propose an alternative approaches in this article based on the Hough transform. In a multiple source scenario, we assume that for each position of the source, the first dominant peak after the direct one is related for all the M RIRs captured at the microphones to the same reflector (i.e., the dominant one for the prescribed source position). Therefore, by combining measurements related to the same reflector, we obtain the initial estimate $\mathbf{l}$ to be used for the initialization of the Hough transform. By using the Hough transform, the initial estimate is refined to provide the estimate $\mathbf{l}_H$. In particular, we consider the case in which repeated acquisitions of TOA information is performed using varying source positions. Closely following the methodology presented in Section V, given M microphones and N source positions, we first populate $\mathbf{P}$. If enough measurements are made at varying source positions, then data points in the Hough space related to all reflectors in the acoustic environment should exist. Correctly identifying maxima in the accumulator space and matching them to real reflectors is a challenging problem, if prior information about their number is unavailable. There are many clustering algorithms that iteratively partition the space into clusters without *a priori* information [23]. Addressing this issue is, however, beyond the scope of this paper and for the sake of simplicity we only consider a rectangular room in the experimental verification using real-world data. Consequently the four greatest bins in the Hough space are identified. Using (37), the reflectors are found in the line parameter space yielding

$$\hat{\mathbf{l}}_{\mathrm{H}_k}, \, k \in \{1, \ldots, 4\}. \tag{38}$$

An additional problem lies in the quantization of the Hough space [23]. These effects can be reduced, but not completely eliminated through dynamic quantization [24]. For our purposes it has been found that the empirical parameters $d_\rho = 0.5$ mm and $d_\theta = 0.25°$, with a range from 0 to 6 meters, provide good results. Finally, we notice that finite-length reflectors can be localized with the proposed methodology, under the assumption that, for prescribed source locations, first-order echoes coming from these reflectors are observed in the impulse responses.

VI. EXPERIMENTAL RESULTS

The performance of the proposed inference algorithm is evaluated with three experiments:

1) Multiple source positions are employed to obtain simulated AIRs and perform geometric inference.
2) The localization robustness is examined by adding temporal error (noise) to the TOA estimates and comparing the averaged individual results of the COTA method at multiple source positions with the results of the source positions combined using the Hough transform correction.
3) Geometric inference is performed using measurements made in a real conference room.

A. Evaluation Criteria

Given reference source location $\mathbf{r}_s$ and estimated source location $\hat{\mathbf{r}}_s$, the source localization error is given by the Euclidian distance $\epsilon_s = \|\hat{\mathbf{r}}_s - \mathbf{r}_s\|$. Let $\mathbf{l}$ and $\hat{\mathbf{l}}$ be the true and estimated
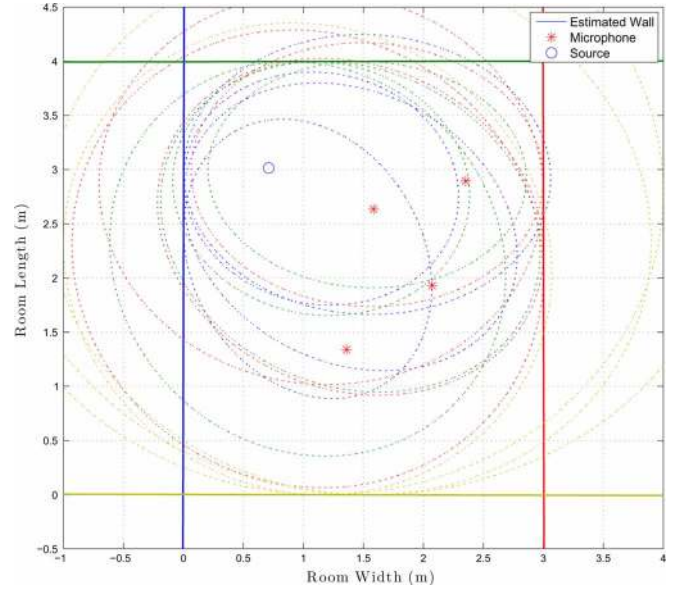

Fig. 7. Example inference result for a rectangular room measuring 3×4 m. Reflectors and their corresponding ellipses are drawn in the same color.

reflector lines, respectively. From these we can evaluate the distance κ from $\mathbf{r}_0$ to a point on each line and the orientation α. The distance can be evaluated by projecting $\mathbf{r}_0$ onto the line such that

$$\kappa = \frac{|l_1 x_0 + l_2 y_0 + l_3|}{\sqrt{l_1^2 + l_2^2}}, \tag{39}$$

and the orientation from

$$\alpha = \arctan \frac{l_2}{l_1}. \tag{40}$$

The accuracy of the reflector localization is measured using:
- distance error $\epsilon_{\mathrm{d}} = |\kappa - \hat{\kappa}|$;
- angular error $\epsilon_{\mathrm{a}} = |\alpha - \hat{\alpha}|$;

B. Simulated Experiment

Simulated AIRs were obtained with the source-image method [25], [14], taking fractional delays into account, for random source and receiver placement in a rectangular room of random dimensions of width and height $(X \times Y)$, with $X \in [3, 5]$ m and $Y \in [4, 6]$ m. The floors and ceilings were perfectly absorbing. An example simulation is depicted in Fig. 7, showing the source, microphones, ellipses and estimated reflectors.

The performance was assessed by averaging the results of 100 Monte Carlo runs. The mean and variance of ϵ_s, ϵ_{d} and ϵ_{a} were calculated considering all located reflectors and individual reflectors ranked in order of error. In some cases not all reflectors are identified with the same degree of accuracy; ranking the error in this way provides insight into the distribution of errors as a function of the number of identified reflectors.

For each run the sound source $(\mathbf{r}_s)$ was placed at four positions inside the room, close to one of the reflectors each time. The microphone positions $(\mathbf{r}_i)$ were picked from a uniform distribution inside the room, constraining the positions to be at a distance of at least 0.5 m from each wall and with each microphone being kept at a minimum distance of 0.5 m from the source positions. We exclude those cases in which the inference

TABLE I
DISTANCE AND ANGULAR ERROR RESULTS FOR SIMULATED DATA

Walls	$\mu(\epsilon_d)$ [cm]	$\sigma(\epsilon_d)$ [cm]	$\mu(\epsilon_a)$ [°]	$\sigma(\epsilon_a)$ [°]
All	**0.926**	**1.169**	**0.175**	**0.351**
Best	0.206	0.210	0.036	0.033
2nd best	0.505	0.295	0.079	0.049
2nd worst	0.884	0.421	0.141	0.080
Worst	2.109	1.756	0.442	0.622

algorithm fails due either to the inseparability of neighboring peaks in the AIR, if a source position does not uniquely identify one of the four reflectors or if the matrices involved in the source localization, particularly in (13), are rank deficient. In the latter case when the microphones are arranged as a linear array it might not be possible to estimate the source location, because of the front-back ambiguity [26]. Additionally the simulation was limited to include only first-order reflections since for each reflector only a single source position is used. When TOAs from higher-order reflections arrive before the first-order reflections of the dominant reflector then their corresponding points in the Hough space do not cluster around the true peaks and are therefore considered as outliers. The sampling frequency is 44.1 kHz using $M = 4$ microphones. We consider unsynchronized AIRs. Source localization was applied as described in Section III.B to estimate TOAs from the TDOAs. Inference was performed using the approach outlined in Section V.E.

The results of the source localization are $\mu(\epsilon_s) = 0.92$ cm and $\sigma(\epsilon_s) = 1.62$ cm. The line parameters of the four reflectors were obtained from (38). The distance and angular error for the reflector inference are given in Table I. Averaged across all walls our approach achieves a $\mu(\epsilon_d)$ and $\mu(\epsilon_a)$ of around one cm and less than half degree respectively.

C. Robustness Analysis

In order to study the robustness of our method with respect to noise, additional white noise was added to the TOA estimates of a single reflector:

$$n_{i,1}^\star = n_{i,1} + \xi, \qquad (41)$$

where ξ is zero-mean gaussian noise with standard deviation variable between 0 and 5 samples.

Two arrangements of source positions were considered: a linear and circular arrangement. In the first case the source was placed 0.5 meters behind the centre of the microphone array (with respect to the wall) and moved at five equidistant intervals between $[-10, 10]$ cm along the length of the room. In the second case the source was moved on a half circle of diameter 1 m from the centre of the array at five equiangle positions between $[-270°, 90°]$.

The performance was assessed by averaging the results of 50 Monte Carlo runs. In each run five source positions were used. For each source position the line parameters of the reflector were calculated using the initial estimate (29) and the Hough data points. At the end of each run the average error of the COTA method was computed along with the best fit, obtained from the analysis of the Hough parameter space, of the five repetitions combined.
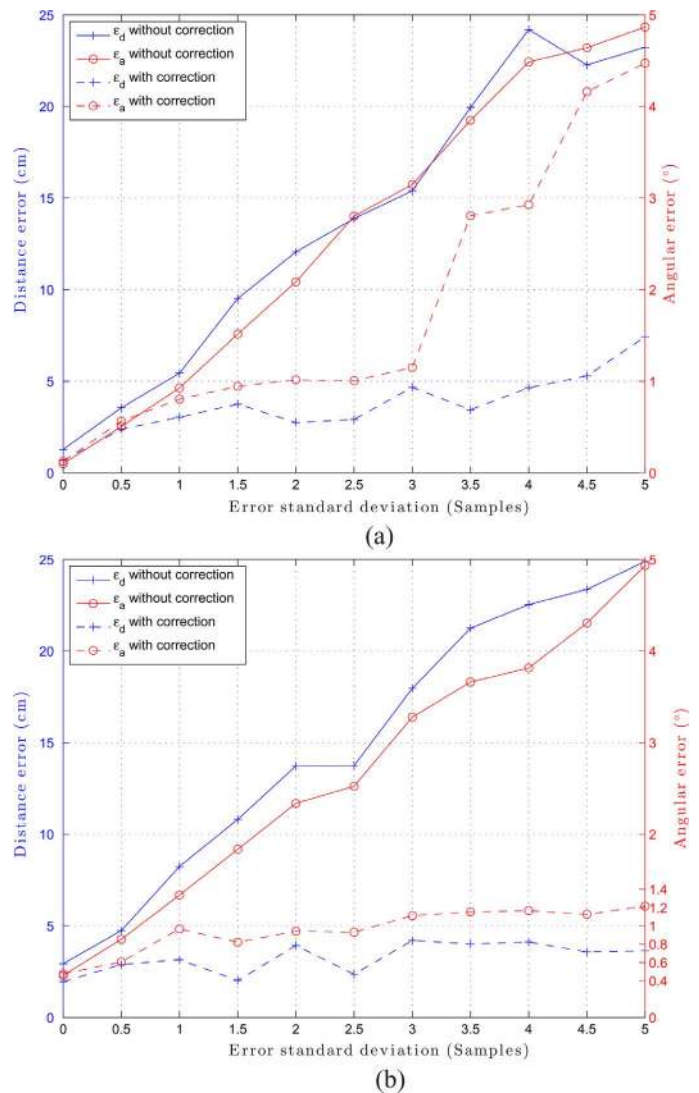

Fig. 8. Average distance, shown as the left y-axis, and angular reflector localization error, shown as the right y-axis, for a single reflector using five linearly (a) and circularly (b) arranged source positions, as a function of additive noise to the TOA estimates for the estimate without the Hough transform (without correction) and with the Hough transform (with correction). (a) Linear source arrangement. (b) Circular source arrangement.

The results for both arrangements are shown in Figs. 8(a)–(b) respectively. In both cases it is observed that the robustness to errors in the TOAs is improved by clustering multiple measurements. When the source positions are close to each other, i.e., yielding a low spatial variation, such as in Fig. 8(a), we see that the angular error (and also to some extent the distance error) increases rapidly with added errors to the TOA information. Consequently, enough spatial variation should be introduced in the positioning of the source if high accuracy is desired, especially in adverse conditions, i.e., when TOA estimates are noisy.

D. Real Experiment

The simulated experiments represent idealized environments in which the transfer function of the measurement apparatus is negligible and the floor and ceiling are perfectly absorbing. In the case of real-world measurements, geometric inference is a much more challenging problem. An experiment was devised in

TABLE II
REFLECTOR LOCALIZATION RESULTS WITH REAL-WORLD DATA

Wall	$N = 4$		$N = 8$		$N = 12$		$N = 16$	
	ϵ_d [cm]	ϵ_a [°]	ϵ_d [cm]	ϵ_a [°]	ϵ_d [cm]	ϵ_a [°]	ϵ_d [cm]	ϵ_a [°]
All	**2.358**	**0.812**	**1.580**	**0.603**	**1.025**	**0.309**	**0.853**	**0.205**
North	2.240	0.751	2.240	0.751	0.700	0.253	0.700	0.253
East	1.780	0.720	0.850	0.751	0.610	0.504	0.500	0.252
South	2.960	0.877	1.780	0.419	1.680	0.226	1.560	0.126
West	2.450	0.899	1.450	0.490	1.111	0.252	0.650	0.188

a small conference a room measuring $3.31 \times 3.58 \times 3.00$ m, with concrete walls and two flush-mounted wooden doors in the south and east walls. A microphone array consisting of four microphones spaced by 0.5 m in a '+' configuration and a fifth placed in the centre was positioned at $(1.75, 1.5)$ m from the south-west corner. A Genelec 8030A loudspeaker was positioned around the array in 16 equiangle positions at a range of 1 m from the array centre, ensuring that the loudspeaker was always faced towards the array. The loudspeaker positions used in this experiment are similar to those used in a 2-D wavefield synthesis array. The microphone signals were sampled at 96 kHz. At each position, the acoustic impulse response between the source and microphone array was estimated using the MLS method [27], followed by localization of the source and a single dominant reflector using the techniques described in Sections IV and V. No effort was made to synchronize the recorded signals with the input stimulus. The line estimates were combined using the Hough transform and the parameters corresponding to the top four bins were used to estimate the bounding line reflectors.

We proceeded to evaluate the improvement in localization accuracy when an increasing number of source positions is employed. First, the sound source was arranged in a '+' configuration, i.e., displaced on the north, west, south and east directions with respect to the array. As a next step, four further source positions are considered at a rotation of $45°$, i.e., including measurements coming from the north-west, south-west, south-east, north-east. Finally, two further rotations of $+22.5°$ and $-22.5°$ yield results for 12 and 16 source positions. Table II shows the localization accuracy for each of the walls along with the average accuracy for all the four configurations described above. Notice that the accuracy improves as the number of sources increases. Even for the case of 4 source positions an error of only a few centimeters is observed, which is suitable for many application scenarios. Using 16 source positions, effectively mimicking a wavefield synthesis array, the localization accuracy approaches the limits of the hand-measured ground truths. Localization results for the 16 sources case are shown in Fig. 9. The error between the intended and estimated positions is due, in part, to the manual positioning of the loudspeaker. This is not problematic as the system makes no prior assumptions about the source location. The Hough data points, marked as '+', lie very close to the room boundaries and are well-fitted by the estimated line reflectors. Some erroneous data points are due to the source positions near multiples of $45°$ in which no single reflector is dominant; they are however treated as outliers by the algorithm and do not affect the estimated reflectors. Reflections from the walls were always dominant over those arising from the floor and ceiling as they are less reflective than the walls.
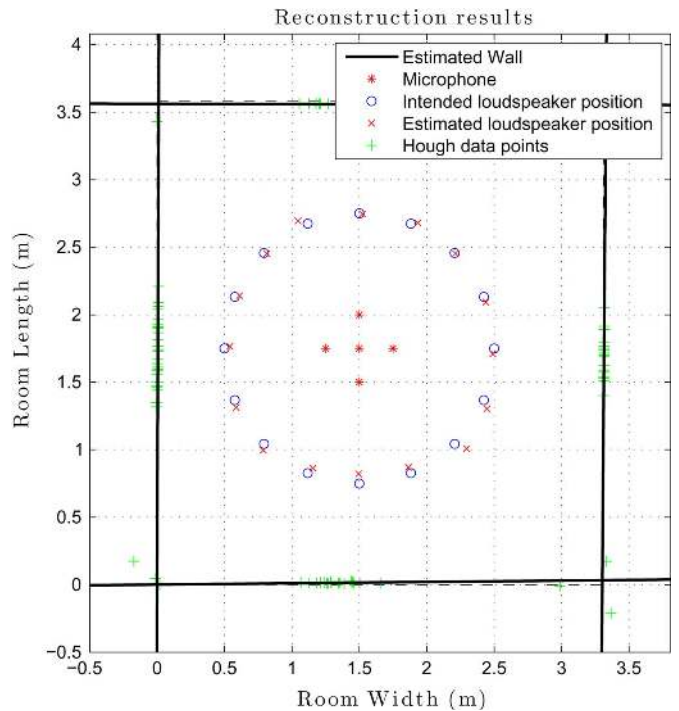

Fig. 9. Room inference results using a microphone array, placed centrally in a small conference room, capturing a MLS sequence from 16 source positions in turn.

VII. CONCLUSION

Inference of the geometry of an acoustic environment in 2-D from AIR estimates has been considered in this paper. Peaks in the AIRs correspond to the TDOAs related to the dominant reflections in a room, from which the location of the source relative to the receivers can be estimated. The TOA corresponding to each peak can then be used in conjunction with the relative source and receiver locations to parameterize an ellipse that describes the locus of possible reflector locations. The common tangent between multiple ellipses corresponds to the location of a particular reflector. An algorithm has been proposed that automatically locates multiple line reflectors in a 2-D plane from estimates of the AIR. Monte Carlo simulations reveal that the proposed method works reliably even when the source location is unknown and the AIRs are unsynchronized. Further simulations show that, by using the Hough transform and taking repeated measurements at different source positions, the robustness to noise in the TOA information can be improved. Real-world measurements show that the proposed technique provides reliable results in a practical setting.

REFERENCES

[1] F. Antonacci, A. Sarti, and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 2822–2825.

[2] J. Filos, E. A. P. Habets, and P. A. Naylor, "A two-step approach to blindly infer room geometries," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel Aviv, Israel, Sep. 2010.

[3] J. Filos, A. Canclini, M. R. P. Thomas, F. Antonacci, A. Sarti, and P. A. Naylor, "Robust inference of room geometry from acoustic measurements using the Hough transform," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Barcelona, Spain, Aug. 2011, pp. 161–165.

[4] Y. A. Huang, J. Benesty, G. W. Elko, and R. M. Mersereati, "Real-time passive source localization: A practical linear-correction least-squares approach," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 8, pp. 943–956, Nov. 2001.

[5] F. Ribeiro, D. Ba, C. Zhang, and D. Florêncio, "Turning enemies into friends: Using reflections to improve sound source localization," in *Proc. Int. Conf. Multimedia and Expo (ICME)*, Singapore, Jul. 2010, pp. 731–736.

[6] T. Betlehem and T. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *J. Acoust. Soc. Amer.*, vol. 117, pp. 2100–2111, 2005.

[7] F. Ribeiro, C. Zhang, D. Florêncio, and D. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Trans. Speech Audio Process.*, vol. 18, no. 7, pp. 1781–1792, Sep. 2010.

[8] D. Ba, F. Ribeiro, C. Zhang, and D. Florêncio, "L1 regularized room modeling with compact microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 157–160.

[9] F. Antonacci, D. Riva, G. Prandi, M. Tagliasacchi, A. Sarti, and S. Tubaro, "Efficient interferer cancellation based on geometrical information of the reverberant environment," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.

[10] S. Tervo and T. Korhonen, "Estimation of reflective surfaces from continuous signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, Mar. 2010, pp. 153–156.

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

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

[13] A. Canclini, F. Antonacci, M. R. P. Thomas, J. Filos, A. Sarti, P. A. Naylor, and S. Tubaro, "Exact localization of acoustic reflectors from quadratic constraints," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, 2011, pp. 17–20.

[14] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, vol. 80, no. 5, pp. 1527–1529, Nov. 1986.

[15] M. Brookes, P. A. Naylor, and J. Gudnason, "A quantitative assessment of group delay methods for identifying glottal closures in voiced speech," *IEEE Trans. Speech Audio Process.*, vol. 14, no. 2, pp. 456–466, Mar. 2006.

[16] D. M. Brookes, VOICEBOX: A speech processing toolbox for MATLAB,, 1997 [Online]. Available: http://www.ee.imperial.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[17] L. R. Fincham, "Refinements in the impulse testing of loudspeakers," *J. Audio Eng. Soc.*, vol. 33, no. 3, pp. 133–140, Mar. 1985.

[18] G. Turin, "An introduction to matched filters," *IRE Trans. Inf. Theory*, vol. 6, no. 3, pp. 311–329, Jun. 1960.

[19] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[20] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Commun. ACM*, vol. 15, pp. 11–15, Jan. 1972.

[21] Y. Furukawa and Y. Shinagawa, "Accurate and robust line segment extraction by analyzing distribution around peaks in Hough spaces," *Comput. Vis. Image Underst.*, vol. 92, pp. 1–25, Oct. 2003.

[22] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Trans. Speech Audio Process.*, vol. 16, no. 8, pp. 1479–1489, Nov. 2008.

[23] J. M. Jolion, P. Meer, and S. Bataouche, "Robust clustering with applications in computer vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 8, pp. 791–802, Aug. 1991.

[24] J. O'Rourke and K. R. Sloan, "Dynamic quantization: Two adaptive data structures for multidimensional spaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 3, pp. 266–280, May 1984.

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

[26] H. Kuttruff, *Room Acoustics*, 4th ed. New York: Taylor & Frances, 2000.

[27] J. Vanderkooy, "Aspects of MLS measuring systems," *J. Audio Eng. Soc.*, vol. 42, pp. 219–231, 1994.

Fabio Antonacci was born in Bari (Italy) on July 26, 1979. He received the Laurea degree in 2004 in telecommunication engineering and the Ph.D. in information engineering in 2008, both at Politecnico di Milano, Italy.

He is currently a post-doc researcher at Image and Sound Processing Group in Dipartimento di Elettronica ed Informazione at Politecnico di Milano, Italy. His research focuses on space-time processing of audio signals, for both speaker and microphone arrays (source localization, acoustic scene analysis, rendering of spatial sound) and on modeling of acoustic propagation (visibility-based beam tracing).

He co-authored approximately 40 articles in proceedings of international conferences and on peer-reviewed journals.

Jason Filos (M'12) received the M.Eng. degree in electrical and electronic engineering and the DIC degree in imaging physics from Imperial College London, London, U.K., in 2007 and 2008, respectively.

His research interests are in the area of acoustic signal processing and he has worked in particular on SIMO/MIMO systems, geometric inference, and separable convex and non-convex optimization. He has industrial experience with audio and video in the field of high-definition TV with the IBM Microprocessor Hardware Development department where he worked as a Speed Team Developer.

Mark Thomas (S'06–M'09) received the M.Eng. degree in electrical and electronic engineering and the Ph.D. degree from Imperial College London, London, U.K., in 2006 and 2010, respectively. His research interests are in the areas of speech, audio and acoustic signal processing. He has worked on glottal-synchronous speech processing with emphasis on speech production modelling, single and multichannel speech enhancement, blind multichannel system identification and equalization, and robust optimal beamforming. He has industrial experience with audio, video, and RF in the field of broadcast engineering. From 2010/2011 he was a Postdoctoral Research Associate at Imperial College London where his research focussed on multichannel space-time acoustic signal processing. He is currently a Postdoctoral Researcher with the Speech Research Group at Microsoft Research, Redmond, USA. Dr. Thomas has been a member of the IEEE Signal Processing Society since 2006.

Emanuël A. P. Habets (S'02-M'07-SM'11) received his B.Sc. degree in electrical engineering from the Hogeschool Limburg, The Netherlands, in 1999, and his M.Sc. and Ph.D. degrees in electrical engineering from the Technische Universiteit Eindhoven, The Netherlands, in 2002 and 2007, respectively. From March 2007 until February 2009, he was a Postdoctoral Fellow at the Technion—Israel Institute of Technology and at the Bar-Ilan University in Ramat-Gan, Israel. From February 2009 until November 2010, he was a Research Fellow in the Communication and Signal Processing group at Imperial College London, United Kingdom. Since November 2010, he is an Associate Professor at the International Audio Laboratories Erlangen (a joint institution of the University of Erlangen and Fraunhofer IIS) and a Chief Scientist at Fraunhofer IIS, Germany.

His research interests are in signal processing for acoustical applications, more specifically microphone array processing for noise reduction, dereverberation and source localization.

Dr. Habets was a member of the organization committee of the 9th International Workshop on Acoustic Echo and Noise Control (IWAENC) in Eindhoven, The Netherlands, 2005. He is a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing (2011–2013).

Augusto Sarti (M'04) received the M.S. and the Ph.D. degrees in electronic engineering, both from the University of Padua, Italy, in 1988 and 1993, respectively, with research on nonlinear communication systems. His graduate studies included a joint graduate program with the University of California at Berkeley, where he worked on nonlinear system theory.

In 1993, he joined the Dipartimento di Elettronica e Informazione of the Politecnico di Milano, Milan, Italy, where he is currently an Associate Professor. His research interests are in the area of digital signal processing, with particular focus on sound analysis, processing and synthesis; space-time audio processing; geometrical acoustics; music information retrieval. He also worked on problems of multidimensional signal processing, vision-based 3D scene reconstruction; camera calibration; image analysis; motion planning and nonlinear system theory.

He coauthored over 180 scientific publications on international journals and congresses as well as numerous patents in the multimedia signal processing area. He coordinates the Sound and Music Computing Lab of the Image and Sound Processing Group of the Politecnico di Milano. He promoted and coordinated or contributed to numerous (20+) EC-funded project. He is a member of the IEEE Technical Committee on Audio and Acoustics Signal Processing, and Associate Editor of IEEE Signal Processing Letters. He has served as guest editor for numerous special issues of international journals. He was co-chairman of the 2005 Edition of the IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS); Chairman of 2009 edition of the Digital Audio Effects conference, (DAFx); and in the organizing committees of numerous other conferences in the area of signal processing.

Patrick Naylor (M'89, SM'07) received his B.Eng. degree in electronic and electrical engineering from the University of Sheffield, U.K., in 1986 and the Ph.D. degree from Imperial College, London, U.K., in 1990. Since 1990 he has been a member of academic staff in the Department of Electrical and Electronic Engineering at Imperial College London where he is also Director of Postgraduate Studies. His research interests are in the areas of speech, audio and acoustic signal processing. He has worked in particular on adaptive signal processing for dereverberation, blind multichannel system identification and equalization, acoustic echo control, speaker identification, single and multi-channel speech enhancement and speech production modelling with particular focus on the analysis of the voice source signal. In addition to his academic research, he enjoys several fruitful links with industry in the UK, USA and in mainland Europe. He is an associate editor of IEEE Transactions on Audio Speech and Language Processing and an associate member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing.

Stefano Tubaro (M'01) was born in Novara, Italy, in 1957. He received his Electronic Engineering degree at the Politecnico di Milano, Milano, Italy, in 1982. He then joined the Dipartimento di Elettronica e Informazione of the Politecnico di Milano, first as a researcher of the National Research Council; then as an Associate Professor (1991) and finally as a Full Professor (2004).

He initially worked on problems related to speech analysis; motion estimation/compensation for video analysis/coding; and vector quantization applied to hybrid video coding. In the past few years, his research interests have focused on image and video analysis for the geometric and radiometric modeling of 3-D scenes; and advanced algorithms for video coding and sound processing. He has authored over 150 scientific publications on international journals and congresses. He co-authored two books on digital processing of video sequences. He also co-authored several patents on image processing techniques.

He coordinates the research activities of the Image and Sound Processing Group (ISPG) at the Dipartimento di Elettronica e Informazione of the Politecnico di Milano; which is involved in several research programs funded by industrial partners, the Italian Government, and by the European Commission. He has been involved in the IEEE Technical Committee of Multimedia Signal Processing (2005–2009), and is currently involved in that of Image Video and Multidimensional Signal Processing.