

# A 3D MODEL FOR ROOM BOUNDARY ESTIMATION

*Luca Remaggi, Philip J. B. Jackson, Wenwu Wang and Jonathon A. Chambers*

Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK  
[l.remaggi, p.jackson, w.wang, j.a.chambers]@surrey.ac.uk

## ABSTRACT

Estimating the geometric properties of an indoor environment through acoustic room impulse responses (RIRs) is useful in various applications, e.g., source separation, simultaneous localization and mapping, and spatial audio. Previously, we developed an algorithm to estimate the reflector's position by exploiting ellipses as projection of 3D spaces. In this article, we present a model for full 3D reconstruction of environments. More specifically, the three components of the previous method, respectively, MUSIC for direction of arrival (DOA) estimation, numerical search adopted for reflector estimation and the Hough transform to refine the results, are extended for 3D spaces. A variation is also proposed using RANSAC instead of the numerical search and the Hough transform which significantly reduces the run time. Both methods are tested on simulated and measured RIR data. The proposed methods perform better than the baseline, reducing the estimation error.

**Index Terms**— Room geometry estimation, room impulse response, reflection, 3D model

## 1. INTRODUCTION

The shape of a room affects the properties of the acoustic signals acquired. Estimation of room geometry from given room impulse responses (RIRs) would have impacts on various application such as source separation, speech recognition, media production and music transcription. Knowing the positions of reflectors facilitates the extraction of parameters which characterize the specific room. This offers a potential for improving models applied in different research areas, including localization mapping, spatial audio or audio forensics.

Existing models for room geometry estimation are often divided in two steps, source localization and reflector estimation. To localize the source, two different parameters can be used: the sound time of arrivals (TOAs) (or similarly time difference of arrivals (TDOAs) between two different microphones) which can be estimated by the generalized cross-correlation with phase transform (GCC-PHAT) method [1]

[2], and the direction of arrivals (DOAs) through the spatial method proposed in [3]. The reflector position can be estimated by exploiting the knowledge of a single RIR [4, 5] or using multiple channel systems [6]. In [6], the authors presented a method to estimate the position of the walls in a room using TOAs to generate ellipses tangential to the reflectors. This algorithm relates distances calculated directly from RIRs with the ellipse's property that the sum of the distances from the two foci to any point on the ellipse is a constant. However, the 2D scenario they have considered assumes that a perfectly absorbent floor and ceiling exist. To improve the estimation however, it is necessary to create a 3D model, which is able to estimate the floor and ceiling positions too. For example, the work in [7] considers the ellipses as 2D projections of 3D ellipsoids, while the geometry is estimated through visual data processing in [8]. The method in [2] iteratively searches the planes exploiting the image-source locations estimated through a maximum likelihood based algorithm. Another approach is to estimate DOAs relative to all the reflections, direct sound and interference using a spherical harmonics domain minimum variance distortionless response (MVDR) beamformer, and then to extract the TDOAs of the direct sound and reflections through a cross-correlation method [9]. In [10] the reflector's position is estimated by observing the room response in its frequency domain.

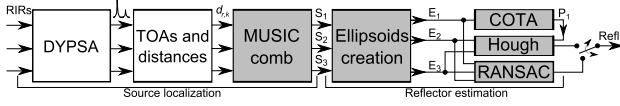
In this article, a model is proposed to estimate the reflector position by combining TOAs and DOAs. The DOAs of the sources are used to estimate the source positions and hence those of the reflectors, using a uniform perpendicular rectangular array (UPRA) of microphones. The method is an improved version of [7], transforming it into a full 3D model.

The paper is organized as follows: Sec. 2 discusses our previous 2D method [7] and presents the proposed model; Sec. 3 shows results achieved with simulated and measured RIRs; Sec. 4 reports the relation of this paper to the state of the art; Sec. 5 draws overall conclusions.

## 2. MODELING THE 3D ROOM

Two main tasks form the proposed method, the source localization and reflector estimation. The overview of the system is given in Fig. 1, showing every algorithm composing it: the dynamic programming projected phase slope al-

This work was supported in the UK by the Engineering and Physical Sciences Research Council (EPSRC) Grants EP/K014307/1 and EP/L000539/1, and the MOD University Defence Research Collaboration in Signal Processing.



**Fig. 1:** System overview:  $d_{r,k}$  is the distance estimated for the  $r$ -th microphone and  $k$ -th reflector;  $S_1$ ,  $S_2$  and  $S_3$  are the three source positions;  $E_1$ ,  $E_2$  and  $E_3$  are the ellipsoids generated;  $P_1$  is the first plane estimation. The grey blocks are the algorithms modified (or added) from [7].

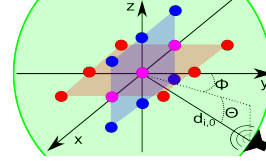
gorithm (DYPsA) [11] for epoch detection, the TOAs and distances calculation from DYPsA output, the multiple signal classification (MUSIC) [12] combined with previous outputs to achieve the source localization, the ellipsoid generation, the common tangent algorithm (COTA) [13] together with the Hough transform [6] to refine the results. The random sample consensus algorithm (RANSAC) [14] is also proposed as an alternative to the COTA and Hough transform. The focus of our work has been on modifying the reflector estimation task of our 2D version [7], creating ellipsoids instead of ellipses and finding the common tangent plane rather than a line. In addition, MUSIC is now available in its 3D version.

## 2.1. Source localization

Assuming knowledge of the microphone positions and RIRs, the following methods are used to estimate the source position. First, exploiting RIRs, the distances between each microphone and source are calculated using DYPsA. Then, the DOAs of the sources are estimated using MUSIC. These two outputs are combined to obtain the 3D positions of the sources [7].

**Epoch detection and distance calculation.** DYPsA was originally designed to estimate glottal closure instances in speech signals, thus some modifications have been made [7]. The DYPsA output is a sequence of non-zero values placed on the time samples corresponding to the RIR peaks. Computing the direct path TOAs  $\tau_{r,0} = \sigma_{r,0}/f_s$ , where  $\sigma_{r,0}$  is the first peak time sample relative to the  $r$ -th microphone and  $f_s$  the sampling frequency, distances from the source are calculated:  $d_{r,0} = \tau_{r,0}c$ , where  $c$  is the sound speed. For compact UPRA, we assume correspondence of the  $k$ -th reflector across microphones.

**Source position estimation.** To estimate the source position, DOAs are also needed. To calculate them from the RIRs received by an  $M$ -element uniform rectangular array (URA) of microphones, several techniques can be adopted [12, 15]. MUSIC has been chosen since it estimates DOAs with a lower deviation from the ground-truth [7]. To implement the 3D MUSIC, a UPRA of microphones is used, i.e. a set of 15 receivers, 9 lying on the  $x$ - $z$  plane and 9 on the  $x$ - $y$  (Fig. 2). This configuration enables the estimation of two spherical coordinates for defining a point in 3D, i.e. the azimuth  $\Phi$  and the elevation  $\Theta$ . The third coordinate, the ra-



**Fig. 2:** The red microphones are lying on the  $x$ - $y$  plane, the blue ones on the  $x$ - $z$  plane, the three purple ones are on the intersection line.  $d_{i,0}$  is the range obtained from DYPsA, used together with the azimuth  $\Phi$  and elevation  $\Theta$  to localize the source drawn at the bottom-right.

dial distance  $\rho$ , is given by DYPsA [7]. Given  $\rho = d_{r,0}$ ,  $\Phi$  and  $\Theta$ , and supposing the  $r$ -th microphone lies on a point with coordinates  $(x_r, y_r, z_r)$ , the source position is found as:  $x_s = x_r + d_{r,0} \cos(\Phi) \cos(\Theta)$ ,  $y_s = y_r + d_{r,0} \sin(\Phi) \cos(\Theta)$ ,  $z_s = z_r + d_{r,0} \sin(\Theta)$ .

## 2.2. Reflector estimation

Once the source is localized, the reflector position can be estimated with a two-stage method: firstly, ellipsoids are generated, then the reflector is searched by COTA together with the refinement through the Hough transform. For the proposed method, a RANSAC based algorithm is also implemented.

**Ellipsoid generation.** For a 2D reflector search, an ellipse is constructed with its major axis equal to the first-order reflection path and foci on the microphone and source positions, creating an elliptical set of possible points where the reflector is tangent [7]. This is done by fitting its characteristics to the general conic equation parameters [6].

Extending the model to 3D space, the idea is the same, but ellipsoids are generated instead [16]. The general equation for a quadratic surface in the 3D space is defined in [17], and the parameters included are:  $a, b, c, d, e, f, g, h, i$  and  $j$ . They can be placed in a  $4 \times 4$  symmetric matrix  $\mathbf{E}$  to create a model in homogeneous coordinates by setting  $\mathbf{x} = [x \ y \ z \ 1]^T$ :

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_g & \mathbf{q} \\ \mathbf{q}^T & j \end{bmatrix}, \quad \mathbf{E}_g = \begin{bmatrix} a & d & f \\ d & b & e \\ f & e & c \end{bmatrix}, \quad (1)$$

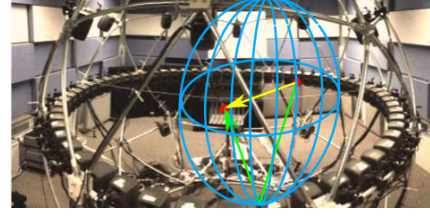
where  $\mathbf{q} = [g \ h \ i]^T$  and  $[\cdot]^T$  denotes the vector transpose. The equation of a quadratic surface can be now written as  $\mathbf{x}^T \mathbf{E} \mathbf{x} = 0$ , representing an ellipsoid only when the following constraints are satisfied:  $\det(\mathbf{E}) \neq 0$ ,  $\mathbf{E}_g > 0$  and  $\det(\mathbf{E})/(a + b + c) < 0$ . A unitary sphere centered on the origin of the system is defined as  $\mathbf{E}_I = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & -1 \end{bmatrix}$ , where  $\mathbf{I}$  is a  $3 \times 3$  identity matrix. Transformations of translation, rotation and scaling are applied to model the ellipsoid with the required center position, axes directions and lengths. Therefore, the matrix defining the ellipsoid relative to the  $r$ -th microphone and the  $k$ -th reflector is written as:

$$\mathbf{E}_{r,k} = \mathbf{T}_r^{-T} \mathbf{R}_r^{-T} \mathbf{S}_{r,k}^{-T} \mathbf{E}_I \mathbf{S}_{r,k}^{-1} \mathbf{R}_r^{-1} \mathbf{T}_r^{-1}. \quad (2)$$

The sphere center is translated to the point  $(\Delta x_r; \Delta y_r; \Delta z_r)$  through the translation matrix  $\mathbf{T}_r$  [18]. Considering the source position  $(x_s; y_s; z_s)$  and the  $r$ -th microphone lying on the point  $(x_r; y_r; z_r)$ , the center point is calculated as the midpoint between the two foci:  $\Delta x_r \equiv \frac{x_s + x_r}{2}$ ,  $\Delta y_r \equiv \frac{y_s + y_r}{2}$ ,  $\Delta z_r \equiv \frac{z_s + z_r}{2}$ . The scaling matrix  $\mathbf{S}_{r,k}$  [18] enlarges (or shrinks) the sphere: the major axis is defined as  $Q_{r,k}^{maj} \equiv d_{r,k}$ , whereas the two minor axes are identical and coincide with  $Q_{r,k}^{min} \equiv \sqrt{d_{r,k}^2 - d_{r,0}^2}$ . Finally, a rotation transformation is applied to each axis, and the three rotation matrices [18] are inserted into Equation (2) as  $\mathbf{R}_r = \mathbf{R}_{x,r} \mathbf{R}_{y,r} \mathbf{R}_{z,r}$ . The angles of rotation  $\alpha_r$ ,  $\beta_r$  and  $\gamma_r$  are obtained by following the equations  $\alpha_r \equiv \text{atan}\left(\frac{z_s - z_r}{y_s - y_r}\right)$ ,  $\beta_r \equiv \text{atan}\left(\frac{x_s - x_r}{z_s - z_r}\right)$ ,  $\gamma_r \equiv \text{atan}\left(\frac{y_s - y_r}{x_s - x_r}\right)$ .

**Reflector search and 3D Hough transform.** Having now a propagation model, the reflector can be estimated. In a similar way to the line in 2D space [7], the required plane is the one which is tangent to every ellipsoid. A plane in the 3D space [17] can be defined in homogeneous coordinates, and its characteristic equation  $\mathbf{p}^T \mathbf{x} = 0$  expresses that the point  $\mathbf{x}$  lies on the plane  $\mathbf{p} = [p_1 \ p_2 \ p_3 \ p_4]^T$ , which is tangent to  $\mathbf{E}$  if it satisfies the equation  $\mathbf{p}^T \mathbf{E}^* \mathbf{p} = 0$ , where  $\mathbf{E}^*$  is the adjoint matrix of  $\mathbf{E}$ . COTA [7] can be modified for 3D spaces:  $J(\mathbf{p}) = \sum_{r=1}^M |\mathbf{p}^T \mathbf{E}_r^* \mathbf{p}|^2$ , where  $M$  is the number of microphones. However, every combination of the  $M$  plane parameters has to be tested, increasing the computational cost compared to the 2D version. Therefore, a new algorithm has been also developed, based on RANSAC [14], as shown later.

To refine the result, the Hough transform is applied to the COTA output. The extension of this algorithm to 3D spaces allows the detection of planes [19]. The idea is to apply the Hough transform to the points  $T_a = (x_a; y_a; z_a)$  and  $T_b = (x_b; y_b; z_b)$  lying on the ellipsoid's surface.  $T_a$  (and  $T_b$ ) has the same (and opposite) normal vector of the initially estimated plane. The normal vector of a point placed on the ellipsoid surface is defined as  $\mathbf{n}_e = \frac{\nabla E}{\|\nabla E\|} = [n_1 \ n_2 \ n_3]^T$ , where  $\nabla E$  is the gradient of the ellipsoid parametric equation  $E$  defined in [17] and  $\|\cdot\|$  denotes Euclidean norm. Equalizing each component to the normal vector coordinates of the first estimated plane  $\mathbf{n}_p = [p_1 \ p_2 \ p_3]^T$ , matrix products are exploited to obtain the coordinates of the two points:  $\mathbf{X}_{T_a} = \mathbf{E}_g^{-1} \mathbf{K}_{T_a}$  and  $\mathbf{X}_{T_b} = \mathbf{E}_g^{-1} \mathbf{K}_{T_b}$ , where  $\mathbf{E}_g$  is defined in Equation (1),  $\mathbf{X}_{T_a, T_b} = [x_{a,b} \ y_{a,b} \ z_{a,b}]^T$  and  $\mathbf{K}_{T_a, T_b} = \frac{1}{2} \mathbf{n}_p \mp \frac{1}{\|\nabla E\|} \mathbf{q}$ . The selection of the relevant points from the set available is based on the position of the plane with respect to the ellipsoid as in [7], but extended by adding the  $z$  coordinate. The 3D Hough transform is then applied to the subset of points obtained,  $\rho = x \cos(\theta) \sin(\phi) + y \sin(\theta) \sin(\phi) + z \cos(\phi)$ , where  $\rho$  is the distance from the center of the polar coordinate system,  $\theta$  is the azimuth and  $\phi$  the elevation [19]. Using an accumulator, the coefficients of the best fitting plane are given by  $p_1 = \cos(\theta_{max}) \sin(\phi_{max})$ ,  $p_2 =$



**Fig. 3:** The experimental system showing the sphere, a blue ellipsoid hypothesized using the direct sound (yellow arrow) and the first order reflection (green arrow) TOAs, for the selected loudspeaker and microphone (red circles).

$\sin(\theta_{max}) \sin(\phi_{max})$ ,  $p_3 = \cos(\phi_{max})$  and  $p_4 = \rho_{max}$ .

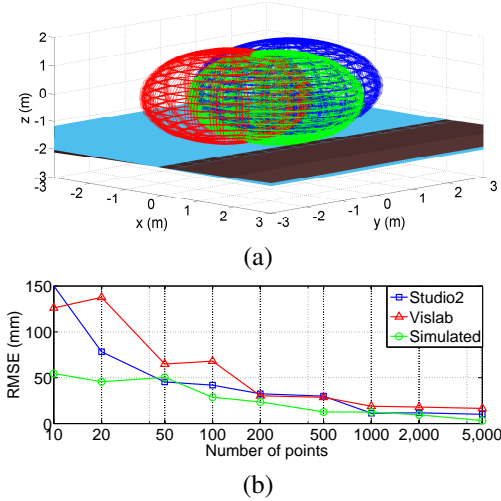
**RANSAC based algorithm.** The idea is to randomly select a certain number of points [14] on the ellipsoids and verify, by setting a threshold, which subset generates the plane closest to the required one. A point  $\mathbf{c}_l = [x_{c_l} \ y_{c_l} \ z_{c_l}]^T$  lying on one of the ellipsoids is randomly selected and the normal vector  $\mathbf{n}_l$  is calculated; thus the sample number used is 1. The  $l$ -th plane tried during the algorithm is given by  $\mathbf{p}_l = \mathbf{n}_l^T (\mathbf{x} - \mathbf{c}_l)$ , where  $\mathbf{x} = [x \ y \ z]^T$ . To verify if the plane is tangent to all the  $N = M \cdot L$  ellipsoids, where  $M$  is the number of microphones and  $L$  the number of sources,  $|\mathbf{p}^T \mathbf{E}^* \mathbf{p}| = t$  is calculated for each of them. Since the plane is perfectly tangent if  $t = 0$ , a threshold  $T$  is set and, when  $t > T$ , the ellipsoid is considered non-tangent. The plane that has the most ellipsoid support is selected.

### 3. EXPERIMENTAL EVALUATION

The proposed method is evaluated and compared with [7]. Simulations were performed using three different datasets. Two of them are real RIRs recorded inside two laboratories at the University of Surrey. The third one is simulated [20]. The threshold used by DYPSA on the slope function [7] and the RANSAC threshold  $T$  were changed to use different datasets.

#### 3.1. RIRs reproduction system and simulated data

A reproduction and measurement system was mounted on a spherical structure, the ‘‘Surrey Sound Sphere’’ [21] (Fig. 3). It was placed in two acoustically treated rooms. The first one is called ‘‘Studio2’’, with dimensions  $6.55 \times 8.78 \times 4.02 \text{ m}^3$  and RT60 235 ms averaged over the 0.5 kHz, 1 kHz and 2 kHz octave bands. This dataset is available online [22]. The second room is called ‘‘Vislab’’, with dimensions  $7.90 \times 6.00 \times 3.98 \text{ m}^3$ , and RT60 of 215 ms averaged as for ‘‘Studio2’’. 60 Loudspeakers (Genelec 8020b) were clamped to the equator to form a circular array (radius of 1.68 m). 48 microphones (Countryman B3 omni) were attached to a grid mounted on a microphone stand. The height of the equator and the microphones, is 1.62 m. The sample frequency used is 48 kHz. For this article, 8 sources lying on the equator with azimuth 0,



**Fig. 4:** (a) The estimated floor position (brown plane), the ground-truth (blue plane) and the ellipsoids (different colors are used for different sources). (b) RMSEs calculated using different number of points (RANSAC) for the three datasets.

$\frac{\pi}{2}$ ,  $\frac{2}{3}\pi$ ,  $\frac{5}{6}\pi$ ,  $\pi$ ,  $\frac{3}{2}\pi$ ,  $\frac{5}{3}\pi$  and  $\frac{11}{6}\pi$  radians, and 48 microphones having a  $8 \times 6$  rectangular configuration with an inter-element spacing of 5 cm, are used. The third dataset was generated [20] and simulates a room with dimensions  $14.55 \times 17.08 \times 6.05 \text{ m}^3$  and an RT60 of 820 ms. RIRs have been produced, either simulating 3 sources placed in the 3D space and 15 microphones (Fig. 2) or the same configuration as the real ones.

### 3.2. Results and comparisons

A comparison between the proposed method implemented in Matlab and the ground-truth is shown in Fig. 4 (a). The RMSE of the reflector position was calculated considering the  $z$ -axis value ( $z_v$ ) at  $X = 5$  points, lying on the estimated plane, equally spaced between the sources and microphones. From these values, the expected ones ( $z_{ideal}$ ) were subtracted  $e_r = z_v - z_{ideal}$ . Considering  $N$  ellipsoids, where  $N$  is defined in Section 2.2,  $RMSE = \sqrt{\frac{1}{XN} \sum_{r=1}^{XN} e_r^2}$ . The model was tested using different number of microphones  $M \in \{5, 7, 9, 16, 25, 36, 42\}$ .  $H = 50$  combinations of  $L = 3$  loudspeakers, randomly selected over the 8 available, were used for each different number of microphones. The RMSE for each microphones number used, averaged over  $H$  trials, is reported in Table 1. These results show the improvement given by the two variants of the 3D model with respect to [7], considering every dataset tested. Beyond this, generally the combination of COTA and Hough transform yielded better performance. However using the “Vislab” dataset, RANSAC generated lower RMSEs, due to computational limitation given by COTA, which did not allow tests with better resolution here. Depending on the aim of the application, if more processing time is available, the

**Table 1:** RMSEs (mm) for different microphones number and 5000 points randomly chosen for RANSAC. “St.” stands for “Studio2”, “Vi.” for “Vislab” and “Si.” for “Simulated”.

Mics	5	7	9	16	25	36	42
Model [7]-St.	19.3	19.5	18.9	17.9	18.4	17.5	17.3
COTA-St.	14.4	<b>13.9</b>	<b>8.6</b>	<b>6.6</b>	<b>6.5</b>	<b>6.6</b>	<b>6.8</b>
RANSAC-St.	<b>13.9</b>	14.1	14.2	13.6	10.8	9.5	9.6
Model [7]-Vi.	30.2	31.5	30.0	28.3	28.0	27.9	27.8
COTA-Vi.	26.3	20.8	17.9	18.5	17.0	17.1	16.0
RANSAC-Vi.	<b>17.9</b>	<b>17.5</b>	<b>17.1</b>	<b>17.4</b>	<b>16.5</b>	<b>14.2</b>	<b>13.5</b>
Model [7]-Si.	28.9	28.7	26.7	17.3	14.9	13.9	14.4
COTA-Si.	<b>3.8</b>	<b>3.9</b>	<b>3.5</b>	<b>3.5</b>	<b>1.3</b>	<b>1.9</b>	<b>1.2</b>
RANSAC-Si.	10.0	11.6	7.8	5.7	6.3	4.5	4.1

COTA-Hough combination is suggested; if a quick estimation is the priority, RANSAC is recommended, reducing the model run time from 3854 s to 26 s (averaged over all the tests performed). Specific tests were also performed over the proposed RANSAC. RMSEs, trying different numbers of points (i.e. planes) randomly chosen on the ellipsoid surface, 3 sources and 25 microphones, are presented in Fig. 4 (b), showing that the best performances are given testing more than 1000 points.

## 4. RELATION TO PRIOR WORK

Different approaches can be followed to estimate the room geometry given acoustic RIRs. In [6], the state of the art was extended by adding the Hough transform to refine the localization technique proposed in [13], and creating a model that does not use the second order reflections as it was in [5]. However, this method presented some weaknesses: it was a 2D model and either the floor or the ceiling was considered perfectly absorbing; thus we presented a modified version for 3D environments in [7] that exploited the projections of 3D to 2D. Here, a full 3D model which improves the estimation performances is proposed. In addition to this, the RANSAC algorithm is introduced as an alternative to the COTA and Hough transform, reducing the overall algorithm run time.

## 5. CONCLUSIONS

Two versions of a model to estimate the room geometry have been presented, by extending a previous method and creating 3D techniques that are able to estimate directly indoor environments. Simulations for both simulated and real RIRs have been performed, and the results compared calculating RMSEs. The new 3D model showed a significant improvement over the 2D one. The use of RANSAC instead of COTA and Hough reduces considerably the run time of the algorithm, thus it is recommended for fast processing. Future work will investigate datasets with different heights of microphones and sources, and extend analysis beyond the first reflection.

## 6. REFERENCES

- [1] C. Blandin, A. Ozerov, and E. Vincent, "Multi-source TDOA estimation in reverberant audio using angular spectra and clustering," *Signal Processing*, vol. 92, no. 8, pp. 1950–1960, 2012.
- [2] S. Tervo and T. Tossavainen, "3D room geometry estimation from measured impulse responses," in *Proc. of the IEEE international Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, 2012.
- [3] H. Morgenstern and B. Rafaely, "Enhanced spatial analysis of room acoustics using acoustic multiple-input multiple-output (MIMO) systems," in *Proc. ASA Meeting on Acoustics (ICA)*, Montréal, Canada, 2013.
- [4] A. H. Moore, M. Brookes, and P. A. Naylor, "Room geometry estimation from a single channel acoustic impulse response," in *Proc. European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, 2013.
- [5] I. Dokmanić, Y. M. Lu, and M. Vetterli, "Can one hear the shape of a room: the 2-D polygonal case," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011.
- [6] 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 Transaction on Audio, Speech and Language Processing*, vol. 20, no. 10, pp. 2683–2695, 2012.
- [7] L. Remaggi, P. J. B. Jackson, P. Coleman, and W. Wang, "Room boundary estimation from acoustic room impulse responses," in *Proc. of the Sensor Signal Processing for Defence conference (SSPD)*, Edinburgh, UK, 2014.
- [8] W. Hussain, J. Civera, and L. Montano, "Grounding acoustic echoes in single view geometry estimation," in *Proc. of the AAAI conference on artificial intelligence*, Québec City, Canada, 2014.
- [9] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Joint DOA and TDOA estimation for 3D localization of reflective surfaces using eigenbeam MVDR and spherical microphone arrays," in *Proc. of the IEEE international Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011.
- [10] L. Zamaninezhad, P. Annibale, and R. Rabenstein, "Localization of environmental reflectors from a single measured transfer function," in *Proc. of the IEEE International Symposium on Communications, Control and Signal Processing (ISCCSP)*, Athens, Greece, 2014.
- [11] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 34–43, 2007.
- [12] H. L. Van Trees, *Optimum Array Processing - Part IV of Detection, Estimation and Modulation Theory*, chapter 9, Wiley-Interscience, 2002.
- [13] F. Antonacci, A. Sarti, and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *Proc. of the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Dallas, USA, 2010.
- [14] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [15] B. D. VanVeen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE Acoustic, Speech and Signal Processing Magazine*, vol. 5, no. 2, pp. 4–24, 1988.
- [16] E. Nastasia, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of planar acoustic reflectors through emission of controlled stimuli," in *Proc. of the European Signal Processing Conference (EUSIPCO)*, Barcelona, Spain, 2011.
- [17] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision - Second edition*, Cambridge University Press, 2004.
- [18] T. Akenine-Moller, E. Haines, and N. Hoffman, *Real-Time Rendering - Third Edition*, chapter 4, A. K. Peters Ltd., 2008.
- [19] D. Borrmann, J. Elseberg, K. Lingermann, and A. Nüchter, "The 3D Hough transform for plane detection in point clouds: a review and a new accumulator design," *3D Research, Springer-Verlag New York*, vol. 2, no. 2, 2011.
- [20] E. A. P. Habets, *Room Impulse Response Generator*, Eindhoven University of Technology, 2006.
- [21] P. Coleman, P. J. B. Jackson, M. Olik, and J. A. Pedersen, "Personal audio with a planar bright zone," *J. Acoustic Society of America*, vol. 136, no. 4, pp. 1725–1735, 2014.
- [22] P. Coleman, P. J. B. Jackson, A. Duque, J. Francombe, and S. Bech, "Surrey Studio-2 RIR Dataset (SS2RD)," 2014, <http://www.cvssp.org/soundzone/resource/Studio2RIR/>.