Room Boundary Estimation from Acoustic Room Impulse Responses

Luca Remaggi, Philip J. B. Jackson, Philip Coleman and Wenwu Wang

Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK

Email: l.remaggi@surrey.ac.uk

Abstract—Boundary estimation from an acoustic room impulse response (RIR), exploiting known sound propagation behavior, yields useful information for various applications: e.g., source separation, simultaneous localization and mapping, and spatial audio. The baseline method, an algorithm proposed by Antonacci et al., uses reflection times of arrival (TOAs) to hypothesize reflector ellipses. Here, we modify the algorithm for 3-D environments and for enhanced noise robustness: DYPSA and MUSIC for epoch detection and direction of arrival (DOA) respectively are combined for source localization, and numerical search is adopted for reflector estimation. Both methods, and other variants, are tested on measured RIR data; the proposed method performs best, reducing the estimation error by 30%.

I. Introduction

The acoustic environment can have a significant effect on the properties of acquired sound signals, which may degrade the performance of the related technologies used in applications such as speech recognition, speaker verification, source separation, music transcription and media production. Equally, the room effects provide an opportunity to extract information about the environment that may be useful in localization, mapping and spatial representation of the room, e.g., in spatial audio or audio forensics. Accordingly, the estimation of room geometry from acoustic room impulse responses (RIRs) is a topic that has been of growing interest in signal processing community.

The main purpose of the work presented in this paper is to estimate the position of reflecting surfaces in a room. The objective is to create a method that is able to do this for a 3-D environment, by utilizing measured RIRs. This work can be also observed as a supporting task for two main research areas: it can be used to create a convolutive source separation model useful for Defence aims, and to build surveillance systems, which will be able to constantly monitor static environments.

Considering a room as parallelepiped, six reflector positions are implicitly included in the first-order room reflections. First, it is necessary to extract the individual reflections. In [1], the authors presented a method to extract the direction of arrival (DOA) of the sound using spherical arrays of sources and microphones. In [2], two different ways to extract global room parameters from reverberant speech and music were compared.

The literature provides a few approaches to estimate the room geometry. Some used multiple channel systems [3] [4], whereas others exploit knowledge of a single RIR [5]. In [3], the authors presented a method to estimate the position of the walls in a room using times of arrival (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 2-D scenario they consider creates reflector hypothesis on the assumption of having a perfectly absorbent floor and ceiling. In [5] the authors estimated the geometry of the room by calculating the positions of the image sources based on the TOAs and time differences of arrival (TDOAs) between high-order reflections from the walls. The floor and ceiling were again considered to be completely absorbing. In addition, TOAs of second-order reflections are necessary, and with real RIRs it is not always possible to detect them reliably. Exploiting image source theory, it is also possible to define the shape of a polygonal room considering the uniqueness between it and a single RIR [6]. However, this algorithm is not robust to noise and for this reason cannot be applied to measured RIRs.

In this article, an algorithm that employs the TOAs for the direct sound and first-order reflections is proposed. The DOAs of the sources are used to estimate the source positions and hence those of the reflectors, using a 2-D set of microphones configured as a uniform rectangular array (URA) [7]. The method is based on [3], by adding new ways to extract peaks, estimate the source positions and calculate the reflectors. These improvements to the algorithm extend its utility for 3-D environments using real recorded RIRs.

The paper is organized as follows: Antonacci's method [3] is briefly presented in Section II. The proposed algorithms are then discussed in Section III. Simulations from measured data are provided and the output of the general baseline is compared to the new methods in Section IV using the root mean square error (RMSE). Finally Section V summarizes the paper and draws overall conclusions.

II. Overview of the baseline method

The main purpose of this work is to improve the method in [3], allowing its use on 3-D environments considering recorded RIRs. Different approaches to the sub-parts are investigated. The general overview of all the components in the proposed methods is shown in Figure 1. It is composed of two main parts, the source localization and the reflector estimation. Each of these has three different components, giving a total of six processing stages.

The first block is the epoch detection, where the peaks present in the RIR are selected. The positions of these im-
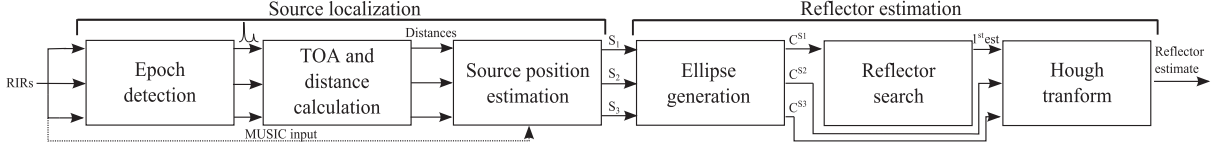
Fig. 1. System diagram overview: Source estimation (left) and reflector estimation (right). S_1, S_2 and S_3 are the three source positions; C^{S1}, C^{S2} and C^{S3} are the ellipses generated for S_1, S_2 and S_3 respectively.

pulses, which correspond to the direct sound and the reflections, provide information about the signal's TOAs. Given a RIR, the output of the epoch detection block is a sequence of non-zero values placed in the time samples corresponding to the peaks of the signal. In the baseline method by Antonacci et al. [3], the "find peak" algorithm, a method based on finding local centers of energy, included in the VOICEBOX tool [8] was proposed. Hence, the distances are simply calculated in the second block of the method following the equation $d_{i,0} = \tau_{i,0} c$, where $d_{i,0}$ and $\tau_{i,0}$ represent the length and TOA for the direct path considering the i-th microphone, and c is the sound speed. In the simulations described in Section IV, the sound speed was assumed to be $343.1\,\mathrm{m/s}$, which is the standard value in air considering a temperature of $20°\mathrm{C}$. The third block is the source position estimation, for which a Least-Squares (LS) based technique was adopted in [3].

Concerning the reflector estimation, the first block is the ellipse generation. Constructing an ellipse with its major axis equal to a first order reflection path means generating an elliptical path, formed by possible points where the reflector is tangent. The generation of an ellipse in the plane is based on making a valid correspondence between its characteristics and the six parameters included in the general 2-D conic equation. As in [3], these parameters are defined as a, b, c, d, e and f. The points are represented using homogeneous coordinates, setting $\mathbf{x} = [x\ y\ 1]^T$. Therefore the ellipse equation can be written as $\mathbf{x}^T \mathbf{C} \mathbf{x} = 0$, where $\mathbf{C}$ is a 3×3 matrix containing the six ellipse coefficients (Equation 1). The ellipse coefficients for each microphone-reflector combination can be obtained by starting from the basis matrix $\mathbf{C}_I$:

$$\mathbf{C} = \begin{bmatrix} a & b & d \\ b & c & e \\ d & e & f \end{bmatrix}; \quad \mathbf{C}_I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \quad (1)$$

The following operations are applied to $\mathbf{C}_I$:

$$\mathbf{C}_{i,k} = \mathbf{T}_i^{-T} \mathbf{R}_i^{-T} \mathbf{S}_{i,k}^{-T} \mathbf{C}_I \mathbf{S}_{i,k}^{-1} \mathbf{R}_i^{-1} \mathbf{T}_i^{-1}, \quad (2)$$

which refers to the i-th microphone in the array and the k-th reflector (with $i \in \{1, ..., M\}$ and $k \in \{1, ..., N\}$), where M is the number of microphones used and N is the number of reflectors in the room. $\mathbf{T}_i$, $\mathbf{R}_i$ and $\mathbf{S}_{i,k}$ are matrices of translation, rotation and scaling respectively [9]. Regarding the estimation of the reflector, a method was used to minimize the gradient of $J(\mathbf{l})$, a cost function which acts as a system of equations created to find points of contact between the ellipses and the line (reflector), where $\mathbf{l}$ denotes the line parameters. The origin of $J(\mathbf{l})$ will be explained more in detail in Section

III. Finally, the Hough transform is exploited to refine the result. The Hough transform converts a point on the Cartesian plane to a sinusoid and conversely a point in the Hough domain corresponds to a line in the Cartesian one. The objective is to find a number of ellipse points generated using multiple other source positions that are geometrically related to the first estimation of the line. These points are then Hough transformed and the one common to most of the sinusoids is inversely transformed in order to have the final estimation of the reflector. For further implementation explanation refer to [10].

III. PROPOSED METHOD

To improve the performance and allow its use for 3-D environments, three blocks are modified: the epoch detection, the source position estimation and the reflector search. The general idea is to exploit two algorithms like Dynamic Programming Projected Phase Slope Algorithm (DYPSA) [11] and Multiple Signal Classification (MUSIC) [12] in order to extract relevant information from RIRs. This information of TOAs and DOAs, permits the source position estimation which enables the generation of ellipses. In addition, a numerical search for the reflector is applied instead of the gradient-based one.

A. Epoch detection

To calculate the distances for the direct sound and reflection paths, it is necessary to extract the TOAs from the RIRs. A method for selecting the peaks of the signal has been developed based on DYPSA [11], which was designed to estimate glottal closure instances (GCIs) from speech signals. Some modifications have been made. The general idea is to take advantage of the properties of the group delay function to estimate the signal centers of energy [13].

The average of the slope function $S(\omega)$ is observed as the group delay function $G(\omega)$ with the opposite sign $S(\omega) = -G(\omega) = \mathrm{d}\Phi(\omega)/\mathrm{d}\omega$, where $\Phi(\omega)$ is the phase shift. To select the values corresponding to the instants where the phase slope function has a positive zero-crossing, this function is smoothed using a Hann window. Finally, two main processes are applied, the first is to compare an ideal slope function creating a level of confidence and the second is to calculate the weighted gain of each peak considering its importance on the original signal.

To adapt the algorithm to the purposes of this article, a threshold is defined on $S(\omega)$ in order to take only the most significant peaks. In fact, parts of the input signal $r(t)$ with significant peaks correspond to large slope values in $S(\omega)$.
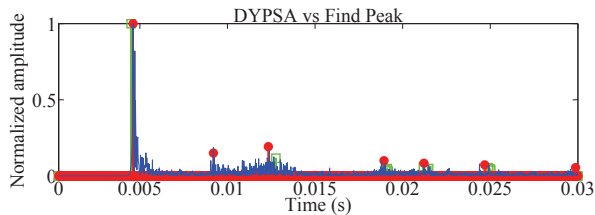
Fig. 2. DYPSA algorithm output (red) for a RIR with $RT_{25} = 0.04\,$s (blue) and "Find Peaks" output (green).

Fig. 3. MUSIC algorithm output, spectral power $P(\theta)$ for a source lying at $288°$ with respect to the microphone array.

The slope threshold is set to 0.2. Another threshold is applied on the time domain amplitude, to eliminate the peaks that are more than $25\,$dB below the main one. To have the same amplitude decay as the RIR, the signal is windowed in the neighborhood of the detected peaks. The energy of the RIR is calculated and used to obtain the amplitude of the output peaks. Figure 2 shows the output of the DYPSA algorithm for a measured RIR.

B. Source position estimation

To estimate the source position from a set of RIRs, both the distances from the microphones to the source, and the DOA are needed. To calculate DOAs for signals received by a microphone array composed of M elements, several classical methods can be adopted such as Bartlett, Capon, or the Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) [12]. The MUSIC algorithm has been chosen for the present study, since it can estimate DOAs relative to D sources present in the mixture signal (of sources and image sources), where $D \geq 1$, with the best accuracy and stability [14]. Beyond to this, assuming knowledge of the microphones position, the steering vector is consequently known, a fundamental requirement for using MUSIC.

This technique is based on the additive noise signal representation and the steering vector matrix $\mathbf{A}$ has a key role. Each element depends on θ, the angle formed among the source, the sensors and the x-axis considering that the central microphone in the array lies on the origin of the coordinate system. The implementation adopted is the 2-D MUSIC algorithm, and for this reason, an URA of microphones is used. The steering vector is formed of exponential elements which have the functionality of transfer functions. In narrowband signals it corresponds to exponentials defined on the wave-number of the signal sent, the distance between the sensors in the array and θ. The steering vector elements are $a_i = \exp(j\beta_i)$, where β_i is the phase angle of the i-th sensor:

$$\beta_i = \frac{2\pi}{\lambda}(\mathrm{x}_i \cos(\theta) + \mathrm{y}_i \sin(\theta)), \tag{3}$$

where $(\mathrm{x}_i, \mathrm{y}_i)$ are the coordinates of the i-th element of the array and λ the wavelength of the signal.

The spectral matrix $\mathbf{S}$ can be calculated by multiplying the received signal $\mathbf{X}$ with its complex conjugate transpose, $\mathbf{S} = \mathbf{X}\mathbf{X}^*$. Considering the eigenvalues placed in decreasing order, it is possible to divide the noise subspace $\mathbf{U}_n$ and the
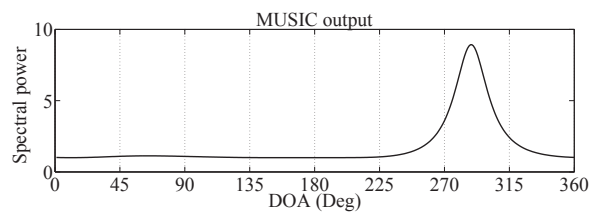
signal subspace $\mathbf{U}_s$, which corresponds to the first D eigenvectors. Thus, the noise subspace is taken using the eigenvector matrix $\mathbf{U}_n$. Through the reciprocal squared Euclidean distance between the curve representing $\mathbf{A}$ in the M-D space and the $M - D$ noise subspace, the spatial spectrum $\mathbf{P}(\theta)$ is created:

$$\mathbf{P}(\theta) = \frac{1}{\mathbf{A}^*(\theta)\mathbf{U}_n\mathbf{U}_n^*\mathbf{A}(\theta)}. \tag{4}$$

The DOAs of the signals correspond to the θ values that generate local maxima. The output of the MUSIC algorithm for a measured RIR is plotted in Figure 3.

The length of the path between the loudspeaker and the receiver obtained from DYPSA creates a circle in the space. The DOA supplied by MUSIC generates a half line with the angular coefficient corresponding to the angle of arrival. The combination of these parameters determines the source position, as drawn in Figure 4.
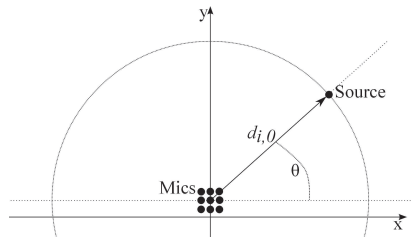

Fig. 4. Range from DYPSA and θ angle from MUSIC intersect to give the source localization.

Given the distance $d_{i,0}$ between the i-th microphone and the source and the DOA θ, and supposing the i-th microphone lies on a plane with coordinates (x_i, y_i), the source position can be found through the equations:

$$x_s = x_i + d_{i,0} \cos(\theta); \qquad y_s = y_i + d_{i,0} \sin(\theta). \tag{5}$$

C. Numerical search for a reflector

The method introduced in [3] makes the assumption that source and receiver lie in the same plane. In contrast, our work here considers a 3-D environment. For this reason, starting from a point position in 3-D space at the coordinates (x_0, y_0, z_0), a slice of that space is considered, by projecting the points onto the plane selected.

The reflector is the line in 2-D space that is tangential to the generated ellipse. Consider the definition of a line in homogeneous coordinates $\mathbf{l}^T\mathbf{x} = 0$, and set the line parameters

Fig. 5. Photograph of the experimental system showing the circular loud-speaker array and microphone grid (circled).

Fig. 6. The estimated floor position (red line), the ground-truth (black line) and the ellipses (different colors are used for different sources).

in a vector $\mathbf{l} = [l_1 \ l_2 \ l_3]^T$. The non-linear equation $\mathbf{l}^T\mathbf{C}^*\mathbf{l} = 0$ needs to be solved to calculate the tangent line to an ellipse, where $\mathbf{C}^*$ is the adjoint of the conic matrix $\mathbf{C}$.

Having available an array of M microphones, it is possible to create M ellipses and calculate the common tangent line corresponding to the wanted reflector. Since there are three line parameters, the minimum number of ellipses required is also three. Therefore, the new problem is to solve a system of non-linear equations. The Common Tangent Algorithm (COTA) [15] was used. It is based on minimization of the cost function:

$$J(\mathbf{l}) = \sum_{i=1}^{M} \left\| \mathbf{l}^T \mathbf{C}_i^* \mathbf{l} \right\|^2 . \tag{6}$$

To achieve the minimization, every combination of the line co-efficients has to be tried. To simplify the algorithm, the number of variables is reduced from three (l_1, l_2 and l_3) to two, and l_1 and l_2 are imposed to lie on a unitary circle. In this way these two coefficients can be rewritten as $l_1 = \cos(\alpha)$; $l_2 = \sin(\alpha)$. With this simplification, the variables to seek for minimizing $J(\mathbf{l})$ are α and l_3 only.

IV. SIMULATIONS

In this section, the performance of the proposed method are evaluated and compared with the baseline algorithm [3]. A subset of the 51840 RIRs recorded in the University of Surrey's acoustic laboratory was taken [16]. Those recordings were given as input to five different methods to simulate the results and compare them. The first tried is the baseline. The last was the one proposed where all the modifications were applied. The other three were hybrids between the old method and the new one, each using only one of the previous algorithms. The first hybrid employs the "Find peak" algorithm, MUSIC and the numerical search for reflector; the second is created using DYPSA, MUSIC and the search of the reflector exploiting the gradient of $J(\mathbf{l})$; the third hybrid selects DYPSA, LS and the numerical search; the last method is the proposed method, with DYPSA, MUSIC and numerical calculation for the minimization of the cost function $J(\mathbf{l})$. The estimation of the floor is performed, by extending the previous restrictions which considered ground and ceiling as completely absorbing.
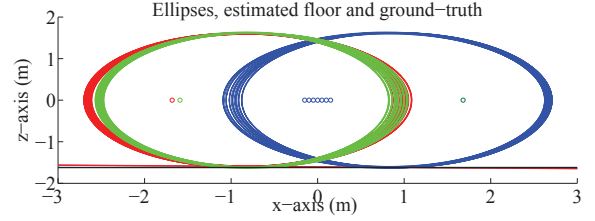
A. RIRs reproduction system

A reproduction and measurement system was designed and mounted on a bespoke spherical structure, the "Surrey Sound Sphere", placed in an acoustically treated room of dimensions $6.55 \times 8.78 \times 4.02$ m (RT60 235 ms averaged over 0.5 kHz, 1 kHz and 2 kHz octave bands) [17]. Loudspeakers (Genelec 8020b) were clamped to the equator of the sphere to form a 60 channel circular array (radius of 1.68 m), and 48 microphones (Countryman B3 omni) were attached to a grid mounted on a microphone stand. Since the sphere has a cut portion on the bottom, allowing it to be stably fixed on the ground, the height of the equator, and so of the microphones and loudspeakers, is 1.62 m. A computer running Matlab was used to play and record the signals from the microphones, via the 'playrec' utility at 48 kHz. A 72 channel MOTU PCIe 424 sound card was used for the analogue to digital interface, with the microphone inputs pre-amplified by a PreSonus Digimax D8. Level differences between the input and output signal channels were compensated through calibration. RIRs between each microphone and each loudspeaker were measured using the maximum length sequence (MLS) approach (15-th order).

For the purpose of this paper, 336 RIRs were taken, for 42 microphones and 8 loudspeakers. The sources selected lie on the equator with azimuth angles of $0°$, $90°$, $120°$, $150°$, $180°$, $270°$, $300°$ and $330°$. Figure 5 shows the sphere with the microphone array. The 42 microphones used have a 7×6 rectangular configuration, at the center of the sphere, with an inter-element spacing of 5 cm.

B. Results and comparisons

The methods were implemented in Matlab. Figure 6 shows the line estimated via the proposed method compared with the ground-truth line.

To compare the results, the RMSE of the reflector position was calculated. It was obtained considering the z-axis value (z_i) of the estimated line in $X = 5$ points along the x-axis: the five points equally spaced between the source and the receiver were chosen for the simulation. From these values the expected ones (z_{ideal}) were subtracted to find the errors $e_i = z_i - z_{ideal}$. This procedure was applied to every ellipse generated. For this reason, considering N ellipses, the general equation to calculate the RMSE is:

$$RMSE = \sqrt{\frac{1}{XN} \sum_{i=1}^{XN} e_i^2} \tag{7}$$

To generalize the results, an iterative test method was employed. Having 42 microphones and 8 sources, starting with 3 receivers and incrementing up to all 42, 50 combinations of 3 loudspeakers were used for each different number of microphones. Random combinations of microphone and source were tested in this way ($N = 1950$). The RMSE calculated for each combination and averaged over 10000 trials is reported in Table III, which demonstrates the superior performance of the proposed method.

Finally, tests were performed to analyze the new blocks sensitivity. The ranges calculated through the "Find peak" algorithm and DYPSA were compared with the expected values relative to the direct sound and the first 2 reflections. The results are reported in Table I . The MUSIC algorithm has been compared with other two classical methods for DOAs estimation, the Barlett and Capon algorithms. The MUSIC algorithm performs better than the others as can be seen in Table II, where the deviations from the ground-truth, averaged over 8 different sources tested, are reported. Regarding the reflector search block, it introduces improvements, which can be observed from the higher RMSE value of the second hybrid approach with respect to the proposed method (Table III).

TABLE I
ERRORS (MM (MS)) CALCULATED FOR THE "FIND PEAK" AND DYPSA.

	Direct sound	1st refl.	2nd refl.
DYPSA	**6.8 (0.020)**	**2.8 (0.008)**	**148.5 (0.433)**
Find peak	10.9 (0.032)	19.6 (0.057)	691.3 (2.000)

TABLE II
ERRORS (DEG) CALCULATED USING MUSIC, CAPON AND BARTLETT.

	MUSIC algorithm	Bartlett algorithm	Capon algorithm
Errors	**1.3**	2.6	3.0

TABLE III
RMSEs (MM) FOR THE SIMULATED CONFIGURATIONS.

	Baseline [3]	Hybrid 1	Hybrid 2	Hybrid 3	Proposed
RMSE	117.0	113.2	67.4	59.9	**18.9**

V. CONCLUSION

Four improved versions of the Antonacci et al. baseline method for the geometric estimation of a room given acoustic RIRs have been presented. Simulations using measured RIRs have been performed for these methods to compare their performance with the baseline. The RMSEs have shown that the method based on the combination of the DYPSA and MUSIC algorithms for the estimation of the source position, and the exploiting of a numerical search for the reflector based on a cost function, improves the results. Therefore, the new method can be considered a useful algorithm for estimating the geometry of a 3-D environment based on a set of real RIRs. Future work will consider a full 3-D implementation and expansion of the RIR dataset to validate these preliminary performance gains.

REFERENCES

[1] 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 2013)*, vol. 19, pp. 1–7, Montréal, Canada, 2013.

[2] P. Kendrick, T. J. Cox, F. F. Li, Y. Zang, and J. A. Chambers, "Monaural room acoustic parameters from music and speech," *J. Acoustical Society of America*, pp. 278–287, July 2008.

[3] 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.

[4] J. Filos, A. Canclini, F. Antonacci, A. Sarti, and P. A. Naylor, "Localization of planar acoustic reflectors from the combination of linear estimates," in *Proc. European Signal Processing Conference (EUSIPCO 2012)*, pp. 1019–1023, Bucharest, Romania, 2012.

[5] 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 2013)*, Marrakech, Morocco, 2013.

[6] 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)*, pp. 321–324, Prague, Czech Republic, 2011.

[7] J. D. Flanagan, J. D. Johnson, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoustical Society of America*, vol. 78, no. 5, pp. 1508–1518, 1985.

[8] M. Brookes, "Voicebox: a speeching processing toolbox for MATLAB." http://www.ee.imperial.ac.uk/hp/staff/dmb/voicebox/voicebox.html, 1997.

[9] J. Filos, A. P. Habets, and P. A. Naylor, "A two-step approach to blindly infer room geometries," in *Proc. of the International Workshop on Acoustic Echo Noise Control (IWANEC)*, Tel Aviv, Israel, 2010.

[10] 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. of the European Signal Processing Conference (EUSIPCO)*, pp. 161–165, Barcelona, Spain, 2011.

[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*, ch. 9. Wiley-Interscience, 2002.

[13] R. Smits and B. Yegnanarayana, "Determination of instance of significant excitation in speech using group delay function," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 5, pp. 325–333, 1995.

[14] T. B. Lavate, V. K. Kokate, and A. M. Sapkal, "Performance analysis of MUSIC and ESPRIT DOA estimation algorithms for adaptive array smart antenna in mobile communication," in *Proc. of the Second International Conference on Computer and Network Technology (ICCNT)*, pp. 308–311, Bangkok, Thailand, 2010.

[15] 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)*, pp. 2822–2825, Dallas, USA, 2010.

[16] P. Coleman, *Loudspeaker array processing for personal sound zone reproduction*. PhD thesis, University of Surrey, Guildford, UK, 2014.

[17] P. Coleman, P. J. B. Jackson, and M. Olik, "Stereophonic personal audio reproduction using planarity control optimization," in *Proc. of the 21st International Congress on Sound and Vibration (ICSV 21)*, Beijing, China, 2014.