Florence, Italy 12-16 July 2015
ICSV22
The 22nd International Congress on Sound and Vibration

# SOURCE, SENSOR AND REFLECTOR POSITION ESTIMATION FROM ACOUSTICAL ROOM IMPULSE RESPONSES

Luca Remaggi, Philip J. B. Jackson, Philip Coleman
*Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey GU2 7XH, UK.*
*email: {l.remaggi, p.jackson, p.d.coleman}@surrey.ac.uk*

The acoustic environment affects the properties of the audio signals recorded. Generally, given room impulse responses (RIRs), three sets of parameters have to be extracted in order to create an acoustic model of the environment: sources, sensors and reflector positions. In this paper, the cross-correlation based iterative sensor position estimation (CISPE) algorithm is presented, a new method to estimate a microphone configuration, together with source and reflector position estimators. A rough measurement of the microphone positions initializes the process; then a recursive algorithm is applied to improve the estimates, exploiting a delay-and-sum beamformer. Knowing where the microphones lie in the space, the dynamic programming projected phase slope algorithm (DYPSA) extracts the times of arrival (TOAs) of the direct sounds from the RIRs, and multiple signal classification (MUSIC) extracts the directions of arrival (DOAs). A triangulation technique is then applied to estimate the source positions. Finally, exploiting properties of 3D quadratic surfaces (namely, ellipsoids), reflecting planes are localized via a technique ported from image processing, by random sample consensus (RANSAC). Simulation tests were performed on measured RIR datasets acquired from three different rooms located at the University of Surrey, using either a uniform circular array (UCA) or uniform rectangular array (URA) of microphones. Results showed small improvements with CISPE pre-processing in almost every case.

## 1. Introduction

An audio signal is affected by the environment characteristics. To identify the interaction between the environment and the signal, the sound received by the listener is defined as the convolution between the reproduced sound and a room impulse response (RIR) (plus additive Gaussian noise). The room geometry and the relative microphone and source positions define a specific RIR. Knowledge of the room shape can improve algorithms used in various applications such as source separation, speech recognition, media production and music transcription. This also offers a potential in different research areas, including localization mapping, spatial audio and audio forensics.

To determine the position of each microphone utilized, different algorithms are available. Under the assumption of knowing the position of more than two loudspeakers in the 3D space, in [1] the authors presented a method based on a cost function implementing the triangulation technique. This was possible having calculated the distances between each sensor and source from the TOAs obtained producing chirp signals from every loudspeaker. The same technique to extract TOAs was used in [2] where two different methods exploiting TOAs and time differences of arrival (TDOAs) were presented. The maximum likelihood (ML) technique was exploited to estimate microphone and speaker

locations. Assuming TOAs known a-priori, microphone and source positions were estimated in [3] for either far- or near-field cases. Receivers and transmitters were localized in [4] applying minima solvers to a matrix containing the distances between the microphones and sources. A different approach was proposed in [5] where the classical multidimensional scaling (MDS) was extended into the so-called basis-point classical MDS (BCMDS). By measuring (with a tape) only the distances between each microphone and a small number of basis points, the entire squared-distance matrix, containing the squares of all the interpoint distances, was constructed and decomposed to find the wanted positions. In [6] an energy based approach was presented. Assuming the microphones and loudspeakers were lying on the same 2D plane, the energy of the audio segments was exploited to use the ML estimation for sensor and source positions. Making the assumption of having knowledge of the source position and the six reflectors position of a parallelepipedal room, using the image sources for the first and second order reflections, the microphone position was estimated in [7].

The reflector position can be estimated by exploiting the knowledge of a single RIR [8, 9] or using multiple channel systems [10, 11]. In [8], the authors estimated the geometry of the room by calculating the positions of the image sources based on the TOAs and TDOAs between high-order reflections. However, TOAs of second-order reflections are necessary, and with real RIRs it is not always possible to detect them reliably. In [12], the reflector position was estimated exploiting the inverse mapping of the acoustic multipath propagation problem in 2D. The strength was that they did not assume knowledge of RIRs directly, however, localization failed at low signal-to-noise ratios. Another way to find the reflector positions was proposed in [13], where the authors generated constraints from direct sound and first reflection DOAs in a 2D geometry. Exploiting image source theory, it is also possible to define the shape of a room considering the uniqueness between it and a single RIR, in case of polygonal geometries [9] and L-shaped rooms [14]. However, this algorithm is not robust to noise and cannot be applied to measured RIRs. The method in [15] iteratively searches for planes exploiting the image-source locations estimated through a maximum-likelihood 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 [16]. In [10], the authors presented a method to estimate the position of the walls using TOAs to generate ellipses tangent to them. 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. An extension of this method was presented in [17], considering floor and ceiling reflective and exploiting the projection of a 3D space into a 2D one. A full 3D model was then presented [18], localizing the reflector directly in a 3D space using ellipsoids instead of ellipses.

In this paper, a new method to estimate the localization of the microphones is presented. It is an iterative algorithm based on a rough initial position estimate. This is then applied to a source and reflector estimation model [17, 18] to create a complete room geometry estimation model. In Section 2 the estimation of sensors, sources and reflector position is presented. Section 3 shows simulations performed and results. Finally Section 4 draws the conclusion.

## 2. Room geometry estimation

In our previous work [18], we presented a method to estimate a reflector position having RIRs and microphone positions available. Despite the good performance, we identified the approximations made during each microphone position measurements as a cause of errors. In the following subsections, a new iterative approach to refine the microphones position using a uniform circular array (UCA) or uniform rectangular array (URA) will be presented. Then, the source position and reflector localization models will be reported.

**Algorithm 1** The cross-correlation based iterative sensor position estimation (CISPE) algorithm

**procedure** POSITION ESTIMATION
  $r^S_{i,j} \leftarrow$ Direct sound segmented $r_{i,j}$
  $h_{i,j}(n) \leftarrow$ Bandpass filtering $r^S_{i,j}$
  $X^{init}_{r,i} \leftarrow$ Position initialization
  **for** $i \leftarrow 1, M$ **do**
    **for** $j \leftarrow 1 : L$ **do**
      $h^B_j(n) \leftarrow$ Beamformed $h_{i,j}(n)$
      **while** $\overline{R_{j,\delta}} - \overline{R_j} \geq 0$ **do**
        **for** $\delta \leftarrow -\Delta : \Delta$ **do**
          $h^{BD}_j(n) \leftarrow$ Beamformed delayed $h_{i,j}(n)$
          $\overline{R_{j,\delta}} \leftarrow$ Max of the cross-correlation using $h^{BD}_j(n)$
          $h^B_j(n) \leftarrow$ Update the beamformed signal variable with $h^{BD}_j(n)$
      $E_{i,j} \leftarrow (c \cdot \operatorname{argmax} \overline{R_{j,\delta}})/F_s$
      $X^{part}_{r,i,j} \leftarrow$ Estimated position for the $j$-th source
    $X_{r,i} \leftarrow$ Mean of the estimated positions for every source

## 2.1 Sensor positions - The CISPE algorithm

Algorithms that assume knowledge of the microphone positions are affected by imprecise measurements. The cross-correlation based iterative sensor position estimation (CISPE) algorithm will be presented, based on the cross-correlation between the recorded RIRs and the beamformed signal. Providing performance good enough for our purposes and being simple, the delay-and-sum beamformer (DSB) [19] was used. The position of each microphone is updated every cycle. This procedure is applied to all the $M$ microphones used.

**Preprocessing and initialization.** A set of RIRs recorded through an UCA or URA of $M$ microphones and $L$ loudspeakers is available. Naming the RIR recorded between the $i$-th sensor and $j$-th source as $r_{i,j}(n)$, a preprocessing is performed over the data. Most of the RIR energy is concentrated on the direct sound, therefore, to avoid noise during the calculations, Hamming windows of $B = 101$ samples are applied to the RIRs to select the direct sounds only. A method for selecting the peaks of signals has been developed based on the dynamic programming projected phase slope algorithm (DYPSA) [20]. This was designed to estimate glottal closure instances from speech signals, and has been modified to make it applicable RIRs [17]. At this point, defining the RIR recorded between the $i$-th microphone and $j$-th loudspeaker and segmented using DYPSA as $r^S_{i,j}(n)$, where $n$ is the discrete time variable, they are filtered through bandpass filters $z_{i,j}(n)$:

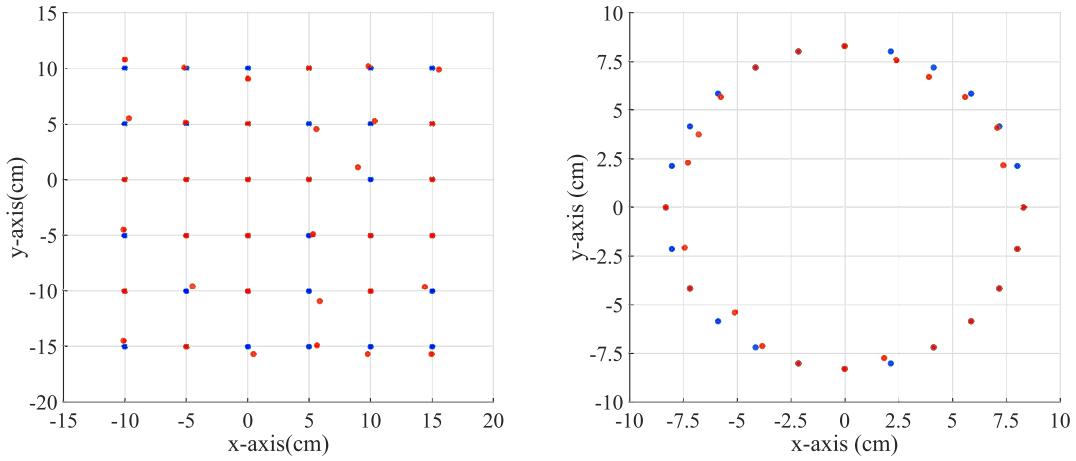
$$h_{i,j}(n) = z_{i,j}(n) * r^S_{i,j}(n), \tag{1}$$

where "*" stands for convolution. $z_{i,j}(n)$ is calculated heuristically, observing the RIRs' frequency content. The subband which results to have the flatter frequency content is selected for each RIR.

Since the presented algorithm is an iterative one, it must be initialized. A rough estimation of the $M$ microphone positions is used for this purpose. In the Cartesian coordinate system, these positions can be seen as points, and defined as $X^{init}_{r,i} = (x^{init}_{r,i}; y^{init}_{r,i})$.

**Iterative core.** The filtered RIRs $h_{i,j}(n)$ defined in Equation 1 are used as input of the DSB. Considering each single loudspeaker, the beamformed signals $h^B_j(n)$ are obtained, where $j$ is the loudspeaker index. The cross-correlation between the beamformed signal and the $M$ recorded RIRs is calculated, and the maximum values averaged over the $M$ microphones:

$$R_{i,j} = \sum_{q=0}^{Q-n-1} h_{i,j}(q+n) h^B_j(q); \qquad \overline{R_j} = \frac{1}{M} \sum_{i=1}^{M} \max_n R_{i,j}, \tag{2}$$




(a) (b)

Figure 1: Microphone positions in (a) 36-microphone URA and (b) 24-microphone UCA, initialized (blue) and estimated (red). Displacements are 3-times magnified for clarity.

where $q$ is the delay applied to the RIRs and $Q$ the length of either $h_{i,j}(n)$ or $h_j^B(n)$. The second step is to apply a time sample delay $\delta = \{-\Delta, -\Delta+1, ..., \Delta-1, \Delta\}$, where $\Delta \in \mathbb{N}$, to one microphone at a time checking the new $\overline{R_j}$. In other words, defining the delayed RIR as $h_{i,j}^D(n) = h_{i,j}(n-\delta)$ the new beamformed signal $h_j^{BD}(n)$ is calculated exploiting it, and the new cross-correlation average $\overline{R_{j,\delta}}$ is given by the Equation 2 substituting $h_{i,j}(q+n)$ with $h_{i,j}^D(q+n)$ and $h_j^B(q)$ with $h_j^{BD}(q)$. The $\delta$ value that gives the highest $\overline{R_{j,\delta}}$ provides the positional adjustment:

$$E_{i,j} = \frac{c \cdot \text{argmax}_\delta \, \overline{R_{j,\delta}}}{F_s}, \tag{3}$$

where $c$ is the sound speed and $F_s$ the sampling frequency. In the Cartesian coordinate system, the new position estimated through this adjustment is $X_{r,i,j}^{part} = (x_{r,i,j}^{part}; y_{r,i,j}^{part})$, where:

$$x_{r,i,j}^{part} = E_{i,j} \cdot \cos(\Phi_j) + x_{r,i}^{init} \quad \text{and} \quad y_{r,i,j}^{part} = E_{i,j} \cdot \sin(\Phi_j) + y_{r,i}^{init} \tag{4}$$

and $\Phi_j$ is the DOA relative to the $j$-th loudspeaker. Calculating $X_{r,i,j}^{part}$ for every $L$ source, the final estimated position of the $i$-th microphone $X_{r,i}^{est} = (x_{r,i}^{est}; y_{r,i}^{est})$ is given by:

$$x_{r,i} = \frac{1}{L}\sum_{j=1}^{L} x_{r,i,j}^{part} \quad \text{and} \quad y_{r,i} = \frac{1}{L}\sum_{j=1}^{L} y_{r,i,j}^{part}. \tag{5}$$

CISPE is recursive since it is repeated until the calculated $\overline{R_j}$ does not increase from the previous one, and it is applied to every microphone. The pseudo-code is listed in Algorithm 1.

## 2.2 Source and reflector positions

**Triangulation technique.** Knowing the microphone positions, the following method needs the distances from the microphones to the source and the DOA to estimate the source position. Distances are obtained using the TOAs of the direct sounds, extracted from RIRs using the DYPSA algorithm introduced in Section 2.1. Since the output is a sequence of non-zero values placed on the time samples corresponding to the RIR peaks, TOAs for direct sound and first order reflections can be calculated $\tau_{i,k} = s_{i,k}/F_s$, where $s_{i,k}$ is the time sample relative to the $k$-th reflector (i.e. $k = 0$ defines the direct sound), $i$ indicates the $i$-th microphone. Distances from the source are then obtained $d_{i,0} = \tau_{i,0} \cdot c$ [18].

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 ESPRIT [19]. The MUSIC algorithm [19] was chosen for the present study, since it can estimate DOAs relative to sources and image sources with the best accuracy and stability [17]. Beyond this, the fundamental requirement for using MUSIC, i.e. knowledge of the steering vector, is observed, since the microphone positions are estimated through the algorithm shown in Section 2.1. To implement MUSIC, either an URA or an UCA of microphones is used. The microphone array shapes enable the estimation of the azimuth $\Phi$. The radial distance $\rho$ is the distance given by DYPSA. For this reason, given $\rho = d_{i,0}$ and $\Phi$, and placing the $i$-th microphone on the point with coordinates given by Equation 5, the source position coordinates are found $x_s = x_{r,i} + d_{i,0}\cos(\Phi)$, $y_s = y_{r,i} + d_{i,0}\sin(\Phi)$ [17].

**Ellipsoid generation**. With knowledge of the microphone and source positions, the reflector position can be estimated in a two-stage approach. Firstly, ellipsoids are generated, then the reflector can be sought using two different ways: either the 3D common tangent algorithm (3D-COTA) together with the refinement through the 3D Hough transform to obtain more accuracy on the solutions, or RANSAC for fast processing [18].

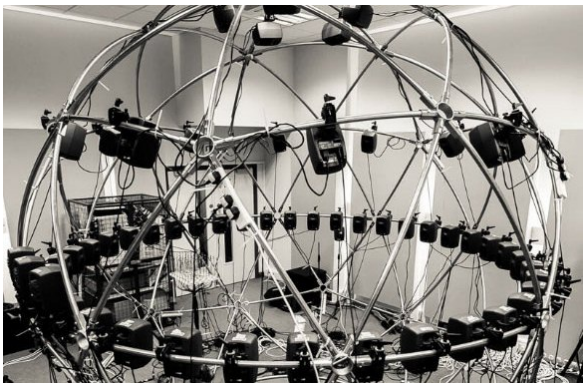
The idea is to construct an ellipsoid with its major axis equal to the first order reflection path and foci on the microphone and source positions, creating an ellipsoidal set of possible points where the reflector is tangent [18]. The general equation characterizing a quadratic surface in the 3D continuous space includes 10 parameters: $\{a, b, c, d, e, f, g, h, i \text{ and } j\}$. They can be placed in a $4 \times 4$ symmetric matrix $\mathbf{E}$ to create a model in homogeneous coordinates. A unitary sphere centred 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 centre position, axes directions and lengths. Therefore, the matrix defining the ellipsoid relative to the $i$-th microphone and the $k$-th reflector is:

$$\mathbf{E}_{i,k} = \mathbf{T}_i^{-T}\mathbf{R}_i^{-T}\mathbf{S}_{i,k}^{-T}\mathbf{E}_I\mathbf{S}_{i,k}^{-1}\mathbf{R}_i^{-1}\mathbf{T}_i^{-1}. \tag{6}$$

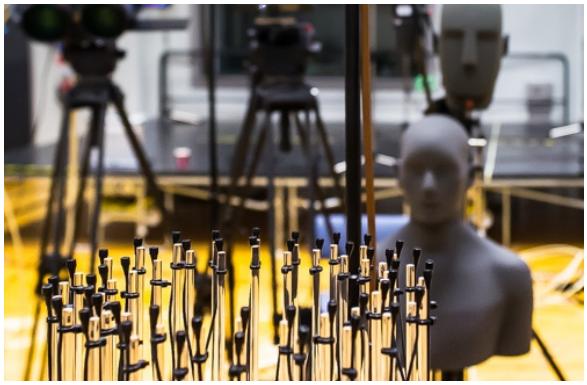
Considering the source position $(x_s; y_s; z_s)$ and the $i$-th microphone lying on the point $(x_{r,i}; y_{r,i}; z_{r,i})$, the sphere centre position is calculated as the midpoint between the two foci. The scaling matrix $\mathbf{S}_{i,k}$ enlarges (or shrinks) the sphere to have the major axis defined as $Q_{i,k}^{maj} \equiv d_{i,k}$, whereas the two minor axes are identical and coincide with $Q_{i,k}^{min} \equiv \sqrt{d_{i,k}^2 - d_{i,0}^2}$. Finally, a rotation transformation is applied to each axis, and the three rotation matrices are combined as $\mathbf{R}_i = \mathbf{R}_{x,i}\mathbf{R}_{y,i}\mathbf{R}_{z,i}$ [18].

**Reflector search**. The required plane is the one which is tangent to every ellipsoid. A plane can be defined in homogeneous coordinates and written as an array $\mathbf{p} = [p_1\ p_2\ p_3\ p_4]^T$, 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 $\mathbf{E}$. The 3D-COTA [18] is then defined as the algorithm that finds the plane which minimizes the cost function $J(\mathbf{p}) = \sum_{r=1}^{M} |\mathbf{p}^T\mathbf{E}_r^*\mathbf{p}|^2$, where $M$ is the number of microphones. However, every combination of the $M$ plane parameters has to be tested, highly increasing the run time. To refine the result, the 3D-Hough transform is applied to the COTA output. For a more in-depth explanation refer to [18].

Due to the high run time of the 3D-COTA, a reflector position search method based on RANSAC is also used [18]. The idea is to randomly select points on the ellipsoid surface and verify, by setting a threshold, which corresponding plane achieves the greatest consensus. A point $\mathbf{c}_l = [x_{c_l}\ y_{c_l}\ z_{c_l}]^T$ lying on one of the ellipsoids is randomly selected, a single sample, and the normal vector $\mathbf{n}_l$ is calculated. The $l$-th plane hypothesis is calculated $\mathbf{p}_l = \mathbf{n}_l^T(\mathbf{x} - \mathbf{c}_l)$, where $\mathbf{x} = [x\ y\ z]^T$. To count how many of the $N = M \cdot L$ ellipsoids are tangent to the plane, where $M$ is the number of microphones and $L$ the number of sources, a distance measure $|\mathbf{p}_l^T\mathbf{E}_m^*\mathbf{p}_l| = t$ is calculated for each of them, where $m$ refers to the $m$-th ellipsoid. A threshold $T$ is set and, when $t \leq T$, the ellipsoid is considered tangent. The plane that has the most ellipsoid support across test points is selected to represent the reflector position.



(a)



(b)

Figure 2: Photographs of (a) the "Surrey Sound Sphere" in the "Vislab" with the microphone URA, and (b) measurement setup in "Studio1" with close-up of the double UCA of microphones.

## 3. Simulations

The algorithms described above have been implemented in Matlab and several simulations have been performed. Exploiting measured RIRs, the performance of the reflector position algorithm, based on RANSAC [18], has been observed applying the CISPE algorithm as preprocessing. RIR measurements from three different laboratories at the University of Surrey have been used. One dataset was recorded exploiting a double concentric UCA, whereas the other two used a URA.

### 3.1 Recording setup

**UCA recordings** RIRs were recorded in a large recording studio called "Studio1" with dimensions $17.08 \times 14.55 \times 6.50\,\mathrm{m}^3$ and a RT60 of 1.1–1.5 s. 15 different loudspeaker positions were used and 3 of them were selected for the purposes of this article, named from "A" to "C" [21]. These 3 loudspeakers were positioned at a height of 1.5 m, lying on a circle around the UCA (24 microphones) with radius of 1.5 m. Defining the loudspeaker B as the one at $0°$, A was positioned at $-\frac{\pi}{4}$ and C at $\frac{\pi}{4}$ radians. The UCA is formed by a double concentric set of microphones with radius 8.3 cm and 10.4 cm respectively. For the aim of this paper we utilized the inner UCA only. The sample frequency used was $48\,\mathrm{kHz}$ and the swept-sine technique was used to measure RIRs.

**URA recordings** A reproduction and measurement system was mounted on a spherical structure, the "Surrey Sound Sphere" [22]. It was placed in two acoustically treated rooms. The first one is called "Studio2", with dimensions $6.55 \times 8.78 \times 4.02\,\mathrm{m}^3$ and RT60 235 ms averaged over the $0.5\,\mathrm{kHz}$, $1\,\mathrm{kHz}$ and $2\,\mathrm{kHz}$ octave bands. The second room is called "Vislab", with dimensions $7.90 \times 6.00 \times 3.98\,\mathrm{m}^3$, 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 $1.68\,\mathrm{m}$). $48$ microphones (Countryman B3 omni) were attached to a grid mounted on a microphone stand. The height of the equator and the microphones, was $1.62\,\mathrm{m}$. The sample frequency used was $48\,\mathrm{kHz}$. For this article, 8 sources lying on the equator with azimuth $0$, $\frac{\pi}{2}$, $\frac{2}{3}\pi$, $\frac{5}{6}\pi$, $\pi$, $\frac{3}{2}\pi$, $\frac{5}{3}\pi$ and $\frac{11}{6}\pi$ radians, and 36 microphones having a $6 \times 6$ square configuration with an inter-element spacing of $5\,\mathrm{cm}$, were used. Considering the centre of the sphere as the origin of the coordinate system, the central microphone of the URA was placed at $(0.0; 0.0; 1.62)\,\mathrm{m}$ for "Studio2", whereas for the "Vislab" dataset it was at $(0.675; 0.000; 1.620)\,\mathrm{m}$.

### 3.2 Reflector estimation

Reflector positions were estimated by the RANSAC-based algorithm (5000 test points). To assess the improvements introduced by CISPE algorithm, the RMSE was calculated considering the $z$-axis value ($z_v$) at $X = 5$ points, lying on the estimated plane, equally spaced between the sources and

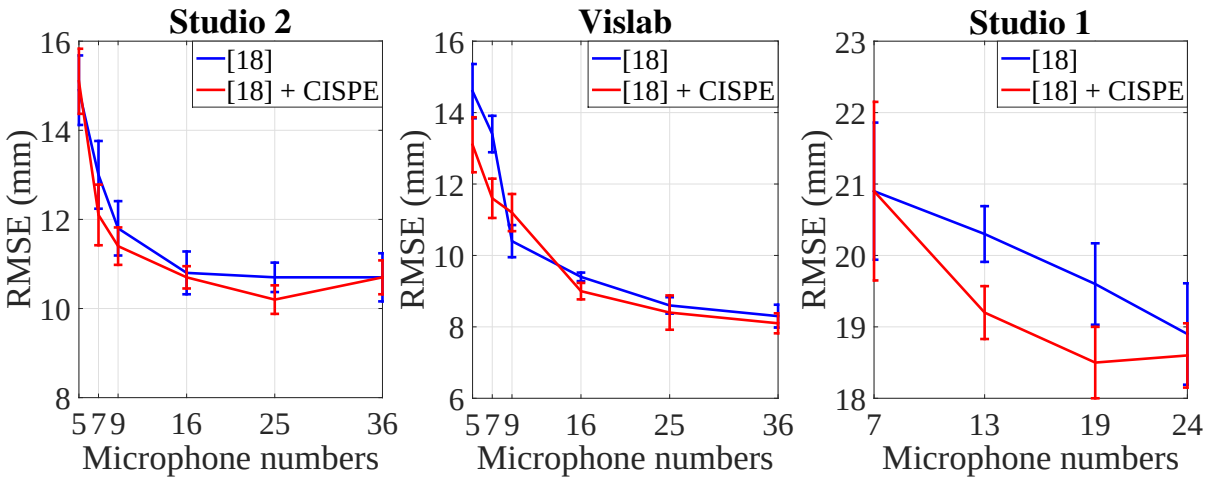



Figure 3: Average reflector position error and 95% confidence intervals versus microphone number, without (blue) and with (red) CIPSE: Studio2 (URA), Vislab (URA) and Studio1 (UCA) datasets.

microphones. From these values, the expected ones ($z_{ideal}$) were subtracted $e_r = z_v - z_{ideal}$, hence $\mathrm{RMSE} = \sqrt{\frac{1}{XN}\sum_{r=1}^{XN} e_r^2}$, where $N = ML$ ellipsoids, as defined in Section 2.2. The model was tested using different numbers of microphones, $M \in \{5, 7, 9, 16, 25, 36\}$ for the URA and $M \in \{7, 13, 19, 24\}$ for the UCA. $H = 100$ combinations of $L = 3$ loudspeakers (randomly taken over the 8 selected for "Studio2" and "Vislab" and 3 selected for "Studio1" were used for each different number of microphones. Given that CISPE is working, for now, with 2D space, three sources (A-C) lying on the same plane of the UCA were selected. The RMSE for each number of microphones used, averaged over $H$ trials, is reported in Figure 3. These results show the improvement given by the introduction of CISPE in every dataset. Placing too wide URAs inside the Surrey Sound Sphere, the far field assumption is not respected any more. Following the Fraunhofer rule [23], with the sphere radius 1.68 m, using 36 microphones, the signals are considered in the near field for frequencies over 3 kHz. For 25 microphones the critical frequency increases to ∼10 kHz. For this reason, we expect to observe a degradation in perfomance with over 25 microphones in the URA with fixed, 5-cm spacing and far-field beamforming, which may be recovered with more advanced beamformer designs.

## 4. Conclusion

CISPE, a new algorithm to estimate the microphone positions in a URA or UCA, has been presented, together with an already available source and reflector position estimator. Tests on real RIRs, recorded using the two different microphone array configurations, were performed and the reflector estimation evaluated. RMSEs showed improvements with the introduction of CISPE as preprocessing. Future work could investigate the effect of applying CISPE to the reflector position variant (COTA and Hough transform), and alternative beamforming techniques.

## 5. Acknowledgement

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

## References

1. Sachar, J. M., Silverman, H. F., and Patterson, W. R., "Microphone position and gain calibration for a large-aperture microphone array," *IEEE Transaction on Speech and Audio Processing* **13**(1), 42–52 (2005).

2. Raykar, V. C., Kozintsev, I. V., and Lienhart, R., "Position calibration of microphones and loudspeakers in distributed computing platforms," *IEEE Transaction on Speech and Audio Processing* **13**(1), 70–83 (2005).

3. Crocco, M., Del Bue, A., Bustreo, M., and Murino, V., "A closed form solution to the microphone position self-calibration problem," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Kyoto, Japan, 2012).

4. Kuang, Y., Burgess, S., Torstensson, A., and Åström, K., "A complete characterization and solution to the microphone position self-calibration problem," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Vancouver, Canada, 2013).

5. Birchfield, S. T. and Subramanya, A., "Microphone array position calibration by basis-point classical multidimensional scaling," *IEEE Transaction on Speech and Audio Processing* **13**(5), 1025–1034 (2005).

6. Chen, M., Liu, Z., He, L. W., Chou, P., and Zhang, Z., "Energy-based position estimation of microphones and speakers for ad hoc microphone arrays," in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, (New Paltz, USA, 2007).

7. Parhizkar, R., Dokmanić, I., and Vetterli, M., "Single-channel indoor microphone localization," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Florence, Italy, 2014).

8. Moore, A. H., Brookes, M., and Naylor, P. A., "Room geometry estimation from a single channel acoustic impulse response," in *Proc. European Signal Processing Conference (EUSIPCO)*, (Marrakech, Morocco, 2013).

9. Dokmanić, I., Lu, Y. M., and Vetterli, M., "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).

10. Antonacci, F., Filos, J., Thomas, M. R. P., Habets, E. A. P., Sarti, A., Naylor, P. A., and Tubaro, S., "Inference of room geometry from acoustic impulse responses," *IEEE Transaction on Audio, Speech and Language Processing* **20**(10), 2683–2695 (2012).

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

12. Tervo, S. and Korhonen, T., "Estimation of reflective surfaces from continuous signals," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Dallas, USA, 2010).

13. Canclini, A., Annibale, P., Antonacci, F., Sarti, A., Rabenstein, R., and Tubaro, S., "From direction of arrival estimates to localization of planar reflectors in a two dimensional geometry," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Prague, Czech Republic, 2011).

14. Marković, D., Antonacci, F., Sarti, A., and Tubaro, S., "Estimation of room dimensions from a single impulse response," in *Proc. of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, (New Paltz, USA, 2013).

15. Tervo, S. and Tossavainen, T., "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).

16. Sun, H., Mabande, E., Kowalczyk, K., and Kellermann, W., "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).

17. Remaggi, L., Jackson, P. J. B., Coleman, P., and Wang, W., "Room boundary estimation from acoustic room impulse responses," in *Proc. of the Sensor Signal Processing for Defence conference (SSPD)*, (Edinburgh, UK, 2014).

18. Remaggi, L., Jackson, P. J. B., Wang, W., and Chambers, J. A., "A 3D model for room boundary estimation," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Brisbane, Australia, 2015).

19. Van Trees, H. L., *Optimum Array Processing - Part IV of Detection, Estimation and Modulation Theory*, Wiley-Interscience (2002).

20. Naylor, P. A., Kounoudes, A., Gudnason, J., and Brookes, M., "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," *IEEE Transactions on Audio, Speech, and Language Processing* **15**(1), 34–43 (2007).

21. Remaggi, L., Jackson, P. J. B., and Coleman, P., "Estimation of room reflection parameters for a reverberant spatial audio object," in *Proc. of the 138th Audio Engineering Society Convention (AES)*, (Florence, Italy, 2015).

22. Coleman, P., Jackson, P. J. B., Olik, M., and Pedersen, J. A., "Personal audio with a planar bright zone," *J. Acoustic Society of America* **136**(4), 1725–1735 (2014).

23. Balanis, C. A., *Antenna theory: analysis and design - Third edition*, Wiley interscience (2005).